2013

Bayesian Stochastic Differential Equation Modelling with Application to Finance

Al-Saadony, Muhannad

Bayesian Stochastic Differential Equation Modelling with Application to Finance

by

Muhannad Al-Saadony

A thesis submitted to Plymouth University in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

School of Computing and Mathematics

Faculty of Science and Technology

Plymouth University

February 2013

Bayesian Stochastic Differential Equation Modelling with Application to Finance

Muhannad Al-Saadony

Abstract

In this thesis, we consider some popular stochastic differential equation models used in finance, such as the Vasicek Interest Rate model, the Heston model and a new fractional Heston model. We discuss how to perform inference about unknown quantities associated with these models in the Bayesian framework.

We describe sequential importance sampling, the particle filter and the auxiliary particle filter. We apply these inference methods to the Vasicek Interest Rate model and the standard stochastic volatility model, both to sample from the posterior distribution of the underlying processes and to update the posterior distribution of the parameters sequentially, as data arrive over time. We discuss the sensitivity of our results to prior assumptions.

We then consider the use of Markov chain Monte Carlo (MCMC) methodology to sample from the posterior distribution of the underlying volatility process and of the unknown model parameters in the Heston model. The particle filter and the auxiliary particle filter are also employed to perform sequential inference. Next we extend the Heston model to the fractional Heston model, by replacing the Brownian motions that drive the underlying stochastic differential equations by fractional Brownian motions, so allowing a richer dependence structure across time. Again, we use a variety of methods to perform inference. We apply our methodology to simulated and real financial data with success. We then discuss how to make forecasts using both the Heston and the fractional Heston model. We make comparisons between the models and show that using our new fractional Heston model can lead to improve forecasts for real financial data.

Contents

List of Figures

List of Tables

Acknowledgements

Completing a PhD is truly a marathon event, and I would not have been able to complete this journey without the aid and support of countless people over the last three years. I must first express my gratitude towards Dr Julian Stander, my Director of Studies, for his excellent supervision. His guidance has been invaluable. I have greatly enjoyed all of our discussion, for which he has always been available. I would like to thank him for letting me be free to develop and pursue my own ideas and for proof–reading so carefully several drafts of this thesis. On a personal level I would like to thanks Dr Julian for his constant support and sensitivity in helping me during my the last three years.

I would also like to thank Dr Paul Hewson for his unselfish and unfailing support as my second supervisor. I would like to extend my thanks to my friends on Plymouth PhD programmes for their help and encouragement.

Special thanks to my parents and my family for all their love and inspirations.

Author's declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other university award.
Relevant scientific seminars and conferences were regularly attended at which work was often presented. One paper was prepared for publication.

Scientific papers :

Inference for the Fractional Heston Model using the Auxiliary Particle Filter.
Muhannad Al-Saadony, Julian Stander and Paul Hewson.
(To be submitted)

Presentation given :

2012 : Inference for the Fractional Heston Model using the Auxiliary Particle Filter.
SIAM Conference on Financial Mathematics and Engineering (FM12), Minneapolis, Minnesota, USA.
Royal Statistical Society 2012, International Conference, Telford, UK.

2011 : Application of the Auxiliary Particle Filter to Financial Models.
Research Students' Conference, University of Cambridge, Cambridge, UK.

2010 : MCMC Algorithms for the Vasicek Interest Rate Model.
Plymouth University, Plymouth, UK.

Conferences and Courses attended :

2010 : Research Students' Conference, University of Warwick, Warwick, UK.

- Academy for PhD training in statistics

  . Statistical Computing and Statistical Inference, University of Cambridge, Cambridge, UK.

. Statistical Modelling and Statistical Asymptotics, University of Nottingham, Nottingham, UK.

. Applied Stochastic Process and Computer Intensive Statistics, University of Bristol, Bristol, UK.

Interest Rate Theory and Practice, OptimSystem Organization, London, UK.

Word count for the main body of this thesis: 24407

Signed: _____

Date: _____

Chapter 1

Introduction

In this thesis we extend and develop models and inference techniques for financial data. In particular, we extend the popular Heston model for financial modelling to a more general fractional Heston model and perform inference about the unknown quantities of this model from financial data using various sequential Monte Carlo techniques. In Section 1.1, we present some background to fractional stochastic differential equation modelling and to associated Bayesian inference. In Section 1.2 we outline the aim and structure of the thesis.

1.1 Background

A large number of statistical models used in financial, including the Heston model, are defined by means of stochastic differential equations, a major topic in measure theory. This is because stochastic differential equations are effectively ordinary differential equations extended by the addition of a term defined using a random measure. Brownian motion plays a central role in stochastic differential equations. Brownian motion is a stochastic Gaussian process with mean zero, variance that increases with time, and independent increments; for a further discussion, see Mikosch (1998), for example.

Stochastic differential equations have been developed in many different areas such as finance, biology, engineering and physics. Klebaner (2012) presents a thorough introduction to stochastic calculus and discusses a range of applications from finance (including currency options), biology (including population growth, genetric evolution and cell biology) and

physics (including signal processing and random oscillators). Nielsen and Shiryaev (2010) work on some importance features related to stochastic processes such as random change of time and the modelling and interpretation of empirically observed dynamic processes. Many authors discuss both stochastic models for financial applications as well as Monte Carlo methods to make inference about unknown model parameters. Korn et al. (2010), for example, present Monte Carlo methods and algorithms, including multilevel Monte Carlo methods, for financial and actuarial models using simulation algorithms.

In recent years, some statistical methods have been developed specifically for stochastic differential equation. As an example, Kessler et al. (2012) present a summary of current research trends and recent developments in statistical methods for stochastic differential equations. They discuss a spectrum of estimation methods, including non-parametric as well as parametric estimation based on likelihood methods, estimating functions, and simulations based techniques for high-frequency data.

A good example of a stochastic differential equation is the Vasicek Interest Rate model. Vasicek (1977) describes the evolution of interest rate as a factor of market risk. We will discuss the Vasicek Interest Rate model in Chapter 2.

Much research follows from the model for European call and put options presented by Fisher and Scholes (1973). One recent study (Buchen, 2012) explains some types of options in assets markets and illustrates how to use straightforward techniques to price a wide range of exotic options within the Black & Scholes framework (Fisher and Scholes, 1973). Such options include barrier options (with payoffs which are related to some aspect of the actual asset path such as whether a given barrier level has been crossed), lookback options (with payoffs which depend on the maximum or minimum asset price over some monitoring windows), Asian options (with payoffs which depend in various ways on the average of the asset prices over some specified time windows) and exotic multi-options (multi-time periods and multi-asset options). From a practical point of view, Iacus (2010) presents some elementary and advanced topics on modern option pricing, from the basic models of the Black & Scholes theory to more sophisticated approaches based on Lévy and other jump processes. Iacus (2007) discusses the use of appropriate statistical techniques, with

the choice of particular financial models starting from real financial data.

The Heston (1993) model is one of the most popular stochastic volatility models in the statistical literature. Aït-Sahalia and Kimmel (2007) and Atiya and Wall (2009) studied maximum likelihood parameter estimation for the Heston model. Benhamou et al. (2010) derive an accurate analytical formula for the price of vanilla options for any time dependent Heston model and Drăgulescu and Yakovenko (2002) present an analytic formula for the time-dependent probability distribution of stock price changes. Based on the setting of the Black & Scholes model, if the underlying volatility process is estimated using sequential price process data, then it is called historical volatility, otherwise it is called implied volatility, which is shown in Miyahara (2012) to depend on the strike price, which is the specified price at which an option contract can be exercised. In more applied work, Bauwens et al. (2012) explore key concepts and topics essential for modelling the volatility of financial time series, both univariate and multivariate, and high- and low-frequency data, using parametric and non-parametric techniques.

A fractional stochastic differential equation is a generalization case of a stochastic differential equation. The general idea of a fractional stochastic differential equation is to replace Brownian motion by the more general fractional Brownian motion. Biagini et al. (2010) present fractional Brownian motion models which have been used for a number of phenomena in different fields in finance. As fractional Brownian motion represents a natural one-parameter extension of Brownian motion, with the parameter being the Hurst index, it is logical to ask if a stochastic calculus for fractional Brownian motion can be developed by generalizing the Itô calculus for processes based on Brownian motion as presented in Iacus (2007), for example. Chapter 1 of Tsoi et al. (2011) presents a multidimensional Wick-Itô formula, which is a very important tool for analysing fractional Brownian motion, fractional Gaussian processes and fractional white noise. More generally, multivariate fractional Brownian motion has been studied recently by Jiang and Zhou (2011), Amblard and Coeurjolly (2011), Lavancier et al. (2009) and Coeurjolly et al. (2010), who consider a special case, bivariate fractional Brownian motion and the associated cross-correlation. Coeurjolly (2000) and Dieker (2004) study the simulation of fractional Brownian motion and present

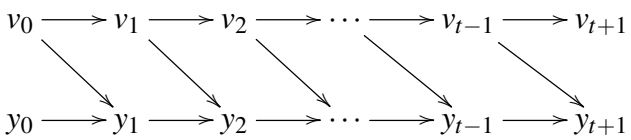
associated algorithms. An interesting role of fractional Brownian motion is making inference for long memory processes. Some basic asymptotic analysis for this is given in Giratis et al. (2012). The problem of estimating the Hurst index is discussed in Gloter and Hoffmann (2004), Chronopoulou and Viens (2011), Chang and Chang (2002) and Gloter and Hoffmann (2007), using both non-parametric and parametric methods. Prakasa Rao (2010) presents statistical inference methodology for fractional diffusion processes and considers parametric and non-parametric inference problems for fractional processes when a complete path of the process over a finite interval is observable.

The Bayesian approach to statistical inference has been increasingly applied in the area of finance in recent years. In the Bayesian approach, prior knowledge about unknown parameters is quantified by means of a probability density function. This probability density function, know as the prior density is updated when data arrive using Bayes' Theorem to give the posterior density function, upon which inference about the unknown quantities is based. Markov chain Monte Carlo (MCMC) methodology provides popular algorithms for sampling from the posterior distribution. Some important foundations and algorithms for the Bayesian approach are presented by Brooks et al. (2011) and Lunn et al. (2013). One recently developed method is the particle filter, the basic aim of which is to update the posterior distribution of the unknown quantities of an underlying state space model. A state space model is a mathematical model which consists of observed and hidden processes. The simplest diagram of a state space model (Petris et al., 2009) takes the form:

$$\begin{array}{ccccccccccc} x_0 & \longrightarrow & x_1 & \longrightarrow & x_2 & \longrightarrow \cdots \longrightarrow & x_{t-1} & \longrightarrow & x_t & \longrightarrow & x_{t+1} \longrightarrow \cdots \\ & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\ & & y_1 & & y_2 & \cdots & y_{t-1} & & y_t & & y_{t+1} \end{array} \tag{1.1}$$

In this case, the aim is to estimate the hidden process $x_1, x_2, \ldots$ and the model parameters sequentially as data $y_1, y_2, \ldots$ arrive. This approach fits nicely into the Bayesian framework, as discussed in Chapter 2, because it naturally models the observed data and the hidden

processes. The diagram for the model that we will study in this thesis is as follows:



(1.2)

In this case, our aim will be to estimate the hidden volatility process $v_1, v_2, \ldots$ and the model parameters sequentially as observed data $y_1, y_2, \ldots$ arrive, assuming that the initial values (y_0,v_0) are known. After defining our state space model, the particle filter will be used to estimate the hidden process assuming that the other parameters are known. In fact, the particle filter consists of two steps, filtering and smoothing. The aim of filtering is to update a target distribution at every iteration when the distribution is computed approximately. The aim of smoothing is to sample from a joint distribution and then to use this to approximated an associated marginal. We will extend the method to estimate the parameters using the auxiliary particle filter. Durbin and Koopman (2012) present a full theoretical treatment of linear, non-linear, non-Gaussian and multivariate time series models using the particle filter, the extended Kalman filter, and unscented filtering and smoothing based on approximation and simulation methods.

Stochastic filtering has been applied to finance as discussed by Date and Ponomareva (2011) and Javaheri et al. (2002). A review of linear and non-linear time series filtering to applications in mathematical finance and in particular to the Heston model, and associated filters is presented in Chapter 9 of Bhar (2010). These authors work in the continuous time framework, consider a general state space model and filter, linearize it and apply the Kalman Filter. They then extend their approach to non-linear state space model. White (2006) describes some non-linear filtering algorithms for stochastic volatility models. Jianhui (2008) presents a novel particle filter algorithm for the fractional Ornstein-Uhlenbeck model applied to finance. The fractional Ornstein-Uhlenbeck process is a generalization of the Ornstein-Uhlenbeck process, which is the solution of a one-dimensional homogeneous

linear stochastic differential equations, driven by fractional Brownian motion, as discussed Prakasa Rao (2010).

The aim of forecasting is to predict a future value of an underlying process from available data. The most popular forecasting methods based on stochastic differential equations are discussed in Hyndman et al. (2008). These authors perform inference in the Bayesian framework and discuss applications to business and industry.

All the computational for thesis have been performed using `R` (R Core Team, 2012), the software environment for statistical computing and graphical, and some of `R`'s contributed packages.

1.2 Aim and Outline of the Thesis

In this thesis we aim to extend the popular Heston model for financial modelling to a more general fractional Heston model and to perform inference about the unknown quantities of this model from financial data using various sequential Monte Carlo techniques.

More precisely, our objective is to compare the performance of the fractional Heston model with that of the standard Heston model in the Bayesian framework. In particular, we will discuss the estimation of the Hurst indices of the fractional Brownian motion processes that define the stochastic differential equations for volatility and asset price using simulated and real data. Our work will be illustrated by application to real data from the Standard & Poor SP350 European Market Index. We will develop and discuss particle filter and auxiliary particle filter algorithms to achieve these objectives.

The outline of this thesis is as follows:

In Chapter 2 we present some methods for Bayesian Inference, including sequential importance sampling, the particle filter and the auxiliary particle filter algorithms. We apply these algorithms to two examples from finance, namely the Vasicek Interest Rate model, defined by a stochastic differential equation, and the stochastic volatility model, which models the underlying variability of a share price. We then explore the sensitivity

of our posterior sample to prior assumptions, and make some recommendations for the choice of the prior distribution.

In Chapter 3 we present the Heston model. We use Markov chain Monte Carlo algorithms (MCMC) to sample from the posterior distribution of the underlying volatility process and the parameters. Next we use particle filter based methodology both to sample the posterior distribution of underling volatility process and to update the posterior distribution of the parameters sequentially, as data arrive over time. As mentioned, we apply our approach to simulated and real data.

In Chapter 4 we extend the Heston model to the fractional Heston model by replacing Brownian motion by fractional Brownian motion in the equations that define the model. After explicitly defining the fractional Heston model and showing its representation as a state space model, we use particle filter based methodology both to sample the posterior distribution of the underlying volatility process and to update the posterior distribution of the parameters sequentially. We apply our approach to simulated and real data. Finally, we apply forecasting beyond and within the data and make comparisons between the forecasts obtained from the Heston and fractional Heston models.
In Chapter 5 we review the main conclusion from each chapter and make suggestion for future work.

Chapter 2

Particle Filter

In this chapter we discuss particle filter based methodologies and apply then to the Vasicek Interest Rate and Stochastic Volatility Models. These models define as stochastic differential equations. In Section 2.1 we introduce some basic ideas of the particle filter and write the Vasicek Interest Rate model as a Hidden Markov model. In Section 2.2 we present some important features about importance sampling.
In Sections 2.3, 2.4 and 2.5 we explain the particle filter, the auxiliary particle filter and the auxiliary particle filter with unknown parameters algorithms, while in Section 2.6 we present all our results including sensitivity studies concerning the choice of the prior distribution for some of the parameters of our models.

2.1 Introduction

The particle filter, which is example of a Sequential Monte Carlo (SMC) method, approximates complicated posterior distributions in on-line filtering applications. Moreover, the particle filter is a very successful numerical methods for solving many problems in non-linear non-Gaussian scenarios. The particle filter can estimate a sequence of hidden parameters $x_k, k = 0,1,2,...$ from a sequence of observed data $y_k, k = 1,2,\ldots$; see 1.1 on page 4.
Firstly, we will explain the general model to which the particle filter can be applied. This is the Hidden Markov model or State Space model, formulated as follows:-

- We have an underlying process $x_0, x_1, ...$, assumed to follow a first order Markov process, such that $x_k|x_{k-1} \sim \pi_{x_k|x_{k-1}}(x_k|x_{k-1})$, with initial distribution $\pi(x_0)$;

- we have observations $y_1, y_2, ...$, assumed to be conditionally independent provided that $x_0, x_1, ...$, are known, allowing us to write $y_k|x_k \sim \pi_{y_k|x_k}(y_k|x_k)$.

Therefore, we can write a State Space model as:-

$$x_k = f(x_{k-1}, w_k) \tag{2.1}$$

$$y_k = h(x_k, v_k), \qquad k = 1, 2, ..., \tag{2.2}$$

in which f and h are known function, and w_k and v_k are independent identically distributed sequences with known probability density functions. An important special case of (2.1) and (2.2) is

$$\begin{aligned} x_k &= g(x_{k-1}) + w_k \\ y_k &= x_k + v_k. \end{aligned}$$

Usually, $\mathrm{E}(w_k) = 0$ and $\mathrm{E}(v_k) = 0$.

One of our aims will be to use simulation to understand the posterior density of the underlying process x up to time t, $\pi(x_{0:t}|y_{1:t})$ in which $x_{0:t} = (x_0, x_1, ..., x_t)$ and $y_{1:t} = (y_1, y_2, ...y_t)$. In general, we write $v_{s:t} = (v_s, v_{s+1}, ..., v_t)$. We will assume that $\pi_t(x_{0:t}|y_{1:t})$ is known up to a normalizing constant as the product of a likelihood function and a prior density.

We will begin by applying the particle filter to the Vasicek Interest Rate model, and other models such as the stochastic volatility model. So, as an illustration, we reformulate the Vasicek model as a State Space model. The Vasiceck Interest Rate model is formulated as a Stochastic Differential Equation (SDE) as:-

$$dx_t = \theta_1(\theta_2 - x_t)dt + \theta_3 dB_t, \tag{2.3}$$

where θ_1, θ_2 and θ_3 are (possibly unknown) positive parameters and B_t is Brownian motion. Note that (2.3) can be written as

$$dx_t = (\theta_1 \times \theta_2 - \theta_1 x_t)dt + \theta_3 dB_t,$$

so that often the re-parametrization $\beta_1 = \theta_1 \times \theta_2, \theta_1, \theta_3$ is used. We assume that x_t is observed as y_t through an observation equation of the same form as (2.2):-

$$y_t = x_t + v_t; \tag{2.4}$$

y_t is a noisy version as x_t.

We now reformulate the Vasicek Interest Rate model for the underlying process in the form (2.1). First, remember that (2.3) has solution:-

$$x_t = x_0 \exp(-\theta_1 t) + \theta_2(1 - \exp(-\theta_1 t)) + \theta_3 \exp(-\theta_1 t) \int_0^t \exp(\theta_1 s) dB_s; \tag{2.5}$$

for the definition of the integral $\int_0^t \exp(\theta_1 s) dB_s$, see Section 2.1.2 of Mikosch (1998).

Let

$$I(0,t) = \int_0^t \exp(\theta_1 s) dB_s,$$

so that (2.5) becomes

$$x_t = \theta_2 + (x_0 - \theta_2)\exp(-\theta_1 t) + \theta_3 \exp(-\theta_1 t) I(0,t).$$

Then, the Vasicek Interest Rate at time $t + \delta t$ is

$$x_{t+\delta t} = \theta_2 + (x_0 - \theta_2)\exp(-\theta_1 t)\exp(-\theta_2 \delta t) + \theta_3 \exp(-\theta_1 t)\exp(-\theta_1 \delta t) I(0, t + \delta t)$$

By the additive property of integration, it follows easily that

$$I(0,t+\delta t) = I(0,t) + I(t,t+\delta t).$$

Hence,

$$\begin{aligned} x_{t+\delta t} &= \theta_2 + (x_0-\theta_2)\exp(-\theta_1 t)\exp(-\theta_1\delta t) + \\ &\quad \theta_3\exp(-\theta_1 t)\exp(-\theta_1\delta t)I(0,t) + \theta_3\exp(-\theta_1 t)\exp(-\theta_1\delta t)I(t,t+\delta t) \\ &= \theta_2 + \exp(-\theta_1\delta t)[(x_0-\theta_2)\exp(-\theta_1 t) + \theta_3\exp(-\theta_1 t)I(0,t)] + \\ &\quad \theta_3\exp(-\theta_1 t)\exp(-\theta_1\delta t)I(t,t+\delta t) \\ &= \theta_2 + (x_t-\theta_2)\exp(-\theta_1\delta t) + \theta_3\exp(-\theta_1 t)\exp(-\theta_1\delta t)I(t,t+\delta t) \end{aligned}$$

Then,

$$x_{t+\delta t} = \theta_2(1-\exp(-\theta_1\delta t)) + x_t\exp(-\theta_1\delta t) + \xi_{t+\delta t}, \tag{2.6}$$

where $\xi_{t+\delta t} = \theta_3\exp(-\theta_1 t)\exp(-\theta_1\delta t)I(t,t+\delta t)$. By definition of the integral $I(t,t+\delta t)$, $\xi_{t+\delta t}$ is Gaussian with mean zero (Mikosch (1998)). From (2.6), the conditional variance has the following form:-

$$\begin{aligned} \mathrm{var}[x_{t+\delta t}|x_t] &= \mathrm{var}[\theta_2(1-\exp(-\theta_1\delta t)) + x_t\exp(-\theta_1\delta t) + \xi_{t+\delta t}|x_t] \\ &= \mathrm{var}[\xi_{t+\delta t}|x_t]. \end{aligned}$$

Since by definition of $I(t,t+\delta t), \xi_{t+\delta t}$ and x_t are independent variables, it follows that

$$\mathrm{var}[x_{t+\delta t}|x_t] = \mathrm{var}[\xi_{t+\delta t}].$$

So, setting $t = k-1$ and $\delta t = 1$, we have

$$x_k = \theta_2(1-\exp(-\theta_1)) + x_{k-1}\exp(-\theta_1) + \xi_k \tag{2.7}$$

where $\mathrm{E}[\xi_k] = 0$ and $\mathrm{var}[\xi_k] = \mathrm{var}[x_k|x_{k-1}]$. Equation (2.7) therefore takes the same form as equation (2.1). Moreover, we rewrite the observation equation (2.4) in the same form as used by Faff et al. (2006) as

$$y_k = x_k + Hz_k, \tag{2.8}$$

where $z_k \sim N(0,1)$ and the standard deviation H of the noise process may be unknown. This form corresponds to equation (2.2). We refer to the resulting model as the noise Vasicek Interest Rate model. In Section 2.2 we discuss Importance Sampling and Sequential Importance Sampling. The basic particle filter is set out in Section 2.3, and this is extended to the Auxiliary Particle Filter in Section 2.4. Our treatment is based on the excellent book by Petris, Petrone and Campagnoli (2009) (Petris et al., 2009). In Section 2.5 we discuss an extension of the basic auxiliary particle filter to the estimation of unknown parameters. We present illustrative applications to the stochastic volatility model and to the noisy Vasicek Interest Rate model in Section 2.6.

2.2 Importance Sampling

The general idea of *Importance Sampling* is to estimate an expectation of interest under a distribution π from which it is difficult to sample by re-weighting realization drawn from a different distribution g. Mathematically, if we have a interesting function f, then its expected value under the probability density function π is:

$$E_\pi(f(X)) = \int f(x)\pi(x)dx. \tag{2.9}$$

Let us assume that we cannot sample realization from π and let g be a probability density function from which we can sample realization. We can rewrite (2.9) as

$$\begin{aligned} E_\pi(f(X)) &= \int f(x)\frac{\pi(x)}{g(x)}g(x)dx \\ &= E_g(f(X)w^*(X)) \end{aligned}$$

where

$w^*(x) = \pi(x)/g(x)$ is called an *importance function* or weight and the density g is called an importance density. Therefore, the expected value $\mathrm{E}_\pi(f(X))$ can be estimated or approximated as:-

$$E_\pi(f(X)) \approx \frac{1}{N}\sum_{i=1}^{N} f(x^{(i)})w^*(x^{(i)}), \tag{2.10}$$

where $x^{(1)},...,x^{(n)}$ is a sample of N realization from g. Often in Bayesian applications the target density π is a posterior density known only up to a normalization constant C as the product of a likelihood function and a prior density; this means $\pi = \frac{\pi^*}{C}$, where we know π^* but not C. If we set $f = C$ in (2.10) , we have that

$$\begin{aligned} C &= \mathrm{E}_\pi(C) \\ &\approx \frac{1}{N}\sum_{i=1}^{N} Cw^*(x^{(i)}), \end{aligned}$$

in which

$$Cw^*(x^{(i)}) = C\frac{\pi(x^{(i)})}{g(x^{(i)})} \tag{2.11}$$

$$= C\frac{\frac{\pi^*(x^{(i)})}{C}}{g(x^{(i)})} \tag{2.12}$$

$$= \frac{\pi^*(x^{(i)})}{g(x^{(i)})} \tag{2.13}$$

is known.

Let

$$\widetilde{w}^{(i)} = Cw^*(x^{(i)}) = \frac{\pi^*(x^{(i)})}{g(x^{(i)})},$$

so that

$$C \approx \frac{1}{N}\sum_{i=1}^{N} \widetilde{w}^{(i)},$$

in which $\widetilde{w}^{(i)}$ is known. We can now estimate $\mathrm{E}_\pi(f(X))$ for a general f as follows:-

$$\begin{aligned}
\mathrm{E}_\pi(f(x)) &\approx \frac{1}{N}\sum_{i=1}^{N} f(x^{(i)})w^*(x^{(i)}) \\
&= \frac{\frac{1}{N}\sum_{i=1}^{N} f(x^{(i)})\widetilde{w}^{(i)}}{C} \\
&\approx \frac{\frac{1}{N}\sum_{i=1}^{N} f(x^{(i)})\widetilde{w}^{(i)}}{\frac{1}{N}\sum_{i=1}^{N} \widetilde{w}^{(i)}} \\
&= \frac{\sum_{i=1}^{N} f(x^{(i)})\widetilde{w}^{(i)}}{\sum_{i=1}^{N} \widetilde{w}^{(i)}} \\
&= \sum_{i=1}^{N} f(x^{(i)})\frac{\widetilde{w}(x^{(i)})}{\sum_{j=1}^{N} \widetilde{w}^{(j)}} \\
&= \sum_{i=1}^{N} f(x^{(i)})w^{(i)},
\end{aligned}$$

where

$$w^{(i)} = \frac{\widetilde{w}^{(i)}}{\sum_{j=1}^{N} \widetilde{w}^{(j)}}$$

are known normalized weights so that $\sum_{i=1}^{N} w^{(i)} = 1$. In the above, we can view the sample $x^{(1)},...,x^{(N)}$ with associated weights $w^{(1)},...,w^{(i)}$ as a discrete approximation of the target π. So, using δ_x for the unit mass at x and setting $\hat{\pi} = \sum_{i=1}^{N} w^{(i)}\delta_{x^{(i)}}$, we have $\hat{\pi} \approx \pi$. The realizations $x^{(1)},x^{(2)},...,x^{(N)}$ are referred to as Particles.

2.2.1 Sequential Importance Sampling

We can extend the above *Importance Sampling* approach to sample sequentially from the posterior π_t : at time $t = 1$ we require a sample from $\pi_1(x_{0:1}|y_1)$, at time $t = 2$ we require a sample from $\pi_2(x_{0:2}|y_{1:2})$, and so on.

The algorithm for doing this is as follows:-

For $i = 1,2,...,N$, sample $x_0^{(i)} \sim \pi(x_0)$

Set $t = 1$

For $i = 1,...,N$, sample $x_1^{(i)} \sim g_1$, where g_1 is a probability density function chosen by the user.

Calculate

$$\widetilde{w}_1^{(i)} = \frac{L(y_1|x_1^{(i)})\,\pi(x_1^{(i)}|x_0^{(i-1)})}{g_1(x_1^{(i)})},$$

where L is the likelihood function.

Normalize the weights $\widetilde{w}_1^{(i)}$, $i = 1,2,...,N$, to get $w_1^{(i)}$,

Set $t = t+1$

For $i = 1,...,N$, sample $x_t^{(i)} \sim g_t$, where g_t is a probability density function chosen by the user.

Calculate

$$\widetilde{w}_t^{(i)} = \frac{L(y_t|x_t^{(i)})\pi_t(x_t^{(i)}|x_{t-1}^{(i)})}{g_t(x_t^{(i)})} w_{t-1}^{(i)}.$$

Normalize the weight $\widetilde{w}_t^{(i)}, i = 1,...,N$, to get $w_t^{(i)}$.

The derivation of $\widetilde{w}_t^{(i)}$ in terms of $w_{t-1}^{(i)}$ will be seen in Section 2.3. An alternative approach would be to re-sample $x_t^{(1)}, x_t^{(2)}, ..., x_t^{(N)}$ according to the weights $w_t^{(1)}, w_t^{(2)}, \ldots, w_t^{(N)}$, and then to set the weights to be constant at $w_t^{(i)} = \frac{1}{N}$. This means that they cancel out in the next iteration. The above procedure can be modified in an obvious way if x_0 is considered to be known.

2.2.2 Sequential Monte Carlo Bayesian Algorithm using Weight Updating

We now present a variation on the Sequential Importance Sampling method of Section 2.2.1 applied to the Vasicek Interest Rate model observed without noise. In particular, let us return to the Vasicek Interest Rate model defined in equation (2.7), and let us assume that $x_t, t = 1,...,T$, are observed sequentially, but without noise.

Let $\theta = (\theta_1, \theta_2, \theta_3)$ and let us assume that we wish to estimate the posterior $\pi(\theta|x_{0:t})$ of θ given $x_{0:t}, t = 1,...,T$, sequentially. Let $\pi(\theta)$ be the prior distribution of θ and let x_0 be known. By Bayes Theorem

$$\begin{aligned}\pi(\theta|x_{0:t}) &\propto \pi(x_{0:t}|\theta)\cdot\pi(\theta)\\ &= \pi(x_t|x_{0:t-1},\theta)\cdot\pi(x_{0:t-1}|\theta)\cdot\pi(\theta)\\ &= \pi(x_t|x_{t-1},\theta)\cdot\pi(x_{0:t-1}|\theta)\cdot\pi(\theta),\end{aligned}$$

where the transition density $\pi(x_t|x_{t-1},\theta)$ is given by (2.7). We can sample from $\pi(\theta|x_{0:t})$ sequentially as follows:

For $i = 1,...,N$, sample $\theta^{(i)} \sim \pi(\theta)$ and set $w_0^{(i)} = \frac{1}{N}$.

Set $t = 1$

For $i = 1,...,N$, modify the weights by calculating

$$\widetilde{w}_1^{(i)} = \pi(x_1|x_0,\theta^{(i)})w_0^{(i)}.$$

Normalize the weights $\widetilde{w}_1^{(i)}, i = 1,...,N$, to get $w_1^{(i)}$.

Set $t = t+1$

For $i = 1,...,N$, modify the weights by calculating

$$\widetilde{w}_t^{(i)} = \pi(x_t|x_{t-1},\theta^{(i)})\, w_{t-1}^{(i)}.$$

Normalize the weights $\widetilde{w}_t^{(i)}, i = 1,...,N$, to get $w_t^{(i)}$.

At any time t, the weighted particles $(\theta^{(1)},w_t^{(1)}),(\theta^{(2)},w_t^{(2)}),\ldots,(\theta^{(N)},w_t^{(N)})$ represent a sample from the posterior $\pi(\theta|x_{0:t})$.

2.3 The Basic Particle Filter

The general idea of the particle filter is an extension of *Importance Sampling*. In fact, the particle filter is an extension of Importance Sampling to Filtering and Smoothing. In

Filtering we are interested in the sequential approximation of the distributions $\pi_1(x_{0:1}|y_1)$ and $\pi_1(y_1)$ at the first time instance, $\pi_2(x_{0:2}|y_{1:2})$ and $\pi_2(y_{1:2})$ at the second time instance, and so on. Therefore, it is clear that the target distribution will change every time that a new observation y is made. Smoothing is concerned with sampling from the joint distribution $\pi_T(x_{1:T}|y_{1:T})$ and approximating the associated marginals $\pi_T(x_k|y_{1:T})$ where $k = 1,2,\ldots,T$.

In general, we consider the problem of updating a discrete approximation of $\pi_{t-1}(x_{0:t-1}|y_{1:t-1})$ when the observation y_t becomes available. Then, there are two steps for updating. The first step involves drawing an additional component $x_t^{(i)}$ to obtain $x_{0:t}^{(i)}$ and the second step involves updating the weights $w_{t-1}^{(i)}$ to $w_t^{(i)}$.

We will adopt the following form for our importance density:-

$$g_t(x_{0:t}|y_{1:t}) = g_{t|t-1}(x_t|x_{0:t-1}, y_{1:t}).g_{t-1}(x_{0:t-1}|y_{1:t-1})$$

where

$g_{t|t-1}$ is an importance transition density which can depend on $x_{0:t-1}$.

We shall update the weights as follows:-

$$\begin{aligned} w_t &\propto \frac{\pi(x_{0:t}|y_{1:t})}{g_t(x_{0:t}|y_{1:t})} \propto \frac{\pi(x_{0:t}, y_t|y_{1:t-1})}{g_t(x_{0:t}|y_{1:t})} \\ &\propto \frac{\pi(x_t, y_t|x_{0:t-1}, y_{1:t-1}) \cdot \pi(x_{0:t-1}|y_{1:t-1})}{g_{t|t-1}(x_t|x_{0:t-1}, y_{1:t}) \cdot g_{t-1}(x_{0:t-1}|y_{1:t-1})} \\ &\propto \frac{\pi(y_t|x_t) \cdot \pi(x_t|x_{t-1})}{g_{t|t-1}(x_t|x_{0:t-1}, y_{1:t})}.w_{t-1} \end{aligned}$$

In general, we can rewrite the last formula for the i^{th} unnormalized weight corresponding to the i^{th} particle $x_t^{(i)}$ as:-

$$\widetilde{w}_t^{(i)} = \frac{\pi(y_t|x_t^{(i)}) \cdot \pi(x_t^{(i)}|x_{t-1}^{(i)})}{g_{t|t-1}(x_t^{(i)}|x_{0:t-1}^{(i)}, y_{1:t})}.w_{t-1}^{(i)}$$

The final step of updating the weights then becomes simple normalization:-

$$w_t^{(i)} = \frac{\widetilde{w}_t^{(i)}}{\sum_{j=1}^{N} \widetilde{w}_t^{(j)}}$$

It turns out that some of the normalized weights $w_t^{(i)}$ are sometimes very close to either 1 or 0. This can be problematic as having a weight close to 1 means that only one particle is important. To overcome this we calculate the *effective sample size* that is defined as

$$N_{eff} = \left[\sum_{i=1}^{N} (w_t^{(i)})^2 \right]^{-1}.$$

If $w_t^{(i)} = \frac{1}{N}, i = 1,...,N$, then the particles have equal weight and $N_{eff} = N$. If $w_t^{(1)} = 1$, for example, and $w_t^{(i)} = 0, i = 2,...,N$, then $N_{eff} = 1$ as only one particle is important. If N_{eff} drops below a certain value N_0, then we re-sample the particle paths $x_{0:t}^{(i)}, i = 1,...,N$, so that each re-sampled particle path has a new weight $\frac{1}{N}$. This keeps the particle filter in control.

So here is a summary of the basic particle filtering algorithm:-

- At time $t = 1$
  - sample $x_1^{(i)} \sim g_1(x_1|y_1)$ and set $w_0^{(i)} = N^{-1}, \quad i = 1,2,...,N$
- $t = 2,...,T$:
  - For $i = 1,...,N$
  - Draw $x_t^{(i)}$ from $g_{t|t-1}(x_t|x_{0:t-1}^{(i)}, y_{1:t})$
  - set
$$\widetilde{w}_t^{(i)} = \frac{\pi(y_t|x_t^{(i)}) \cdot \pi(x_t^{(i)}|x_{t-1}^{(i)})}{g_{t|t-1}(x_t^{(i)}|x_{0:t-1}^{(i)}, y_{1:t})} . w_{t-1}^{(i)}.$$
  - Normalize the weights
$$w_t^{(i)} = \frac{\widetilde{w}_t^{(i)}}{\sum_{j=1}^{N} \widetilde{w}_t^{(j)}}$$
  - compute $N_{eff} = \left[\sum_{i=1}^{N} (w_t^{(i)})^2 \right]^{-1}$

- if $N_{eff} \leq N_0$, re-sample
    - we will draw a sample of size N from a discrete distribution $p(x_{0:t} = x_{0:t}^{(i)}) = w_t^{(i)},\ i = 1,2,...,N$, and relabel this sample $x_{0:t}^{(1)},...,x_{0:t}^{(N)}$
    - Reset the weights : $w_t^{(i)} = N^{-1}, i = 1,2,...,N$.
- Set $\hat{\pi} = \sum_{i=1}^{N} w_t^{(i)} \delta_{x_{0:t}^{(i)}}$. This is referred to as a mixture approximation to the required posterior $\pi(x_{0:t}|y_{1:t})$.

An advantage of the particle filter is that it can be applied to a variety of models defined through equations (2.1) and (2.2), including models that deal with non-Gaussian noise. The particle filter can also be applied to more general models, although it can become very complicated to implement and inefficient, focusing too much on more probable regions of the state space.

2.4 Auxiliary Particle Filter

The auxiliary particle filter was introduced by Pitt and Shephard in 1999. Essentially, it is an improvement on the particle filter that yields better simulations from the tails of the underlying density. Moreover, the Particle Filter has two points of weakness. The first weakness, already mentioned in Section 2.3, is related to the efficiency of the sample and is caused by the fact that some of the weights may take very large values, known as outliers, leading to degeneracy. The second weakness is related to the often poor quality of the particle filter approximation in the tails of the density of interest. The mixture approximation $\hat{\pi}$ is often poor when outlying particle are sampled.

The basic idea of the auxiliary particle filter is to extend the particle filter by adding an auxiliary variable. An example of an auxiliary variable is the variable I in the following argument, where we write a density $P(x)$ as a mixture of densities defined by I :-

$$\begin{aligned} P(x) &= \sum_I P(x, I = i) \\ &= \sum_I P(I = i) P(x|I = i). \end{aligned}$$

We assume that at time $t-1$ we have a discrete approximation $\hat{\pi}_{t-1} = \sum_{i=1}^{N} w_{t-1}^{(i)} \delta_{x_{0:t-1}^{(i)}}$ of the joint posterior distribution $\pi(x_{0:t-1}|y_{1:t-1})$. Our aim is to update this approximate joint distribution when a new data point y_t is observed. We will write:-

$$\begin{aligned}
\pi(x_{0:t}|y_{1:t}) &\propto \pi(x_{0:t}, y_t|y_{1:t-1}) \\
&= \pi(y_t|x_{0:t}, y_{1:t-1}) \cdot \pi(x_t|x_{0:t-1}, y_{1:t-1}) \cdot \pi(x_{0:t-1}|y_{1:t-1}) \\
&= \pi(y_t|x_t).\pi(x_t|x_{t-1}) \cdot \pi(x_{0:t-1}|y_{1:t-1}) \\
&\approx \pi(y_t|x_t).\pi(x_t|x_{t-1}) \cdot \hat{\pi}_{t-1}(x_{0:t-1}) \\
&= \sum_{i=1}^{N} w_{t-1}^{(i)} \pi(y_t|x_t) \pi(x_t|x_{t-1}^{(i)}) \delta_{x_{0:t-1}^{(i)}} \qquad (2.14)
\end{aligned}$$

It is clear that the above equation is an un-normalized distribution for $x_{0:t}$ in which $x_{0:t-1}$ is a discrete component and x_t is continuous. Therefore, the target is approximated by a mixture distribution, that can be by expressed using an auxiliary variable as follows:-

$$\begin{aligned}
P(I=i) &= w_{t-1}^{(i)} \\
x_{0:t}|I=i &\sim C\pi(y_t|x_t)\pi(x_t|x_{t-1}^{(i)})\delta_{x_{0:t-1}^{(i)}}.
\end{aligned}$$

So, for a given i, the original target $\pi(x_{0:t}|y_{1:t})$ can be extended to become:-

$$\pi^{aux}(x_{0:t}, i|y_{1:t}) \propto w_{t-1}^{(i)} \pi(y_t|x_t) \pi(x_t|x_{t-1}^{(i)}) \delta_{x_{0:t-1}^{(i)}}.$$

Pitt and Shephard (1999) suggested the following formula for the *importance density*

$$g_t(x_{0:t}, i|y_{1:t}) \propto w_{t-1}^{(i)} \pi(y_t|\hat{x}_t^{(i)}) \pi(x_t|x_{t-1}^{(i)}) \delta_{x_{0:t-1}^{(i)}}$$

where $\hat{x}_t^{(i)}$ is a central value, such as the mean or the mode of $\pi(x_t|x_{t-1} = x_{t-1}^{(i)})$. For example $\hat{x}_t^{(i)} = \mathrm{E}[x_t|x_{t-1} = x_{t-1}^{(i)}]$, which may, for example, be $f(x_{t-1}^{(i)})$ if $\mathrm{E}[w_t] = 0$ in

transition equation (2.1), or $\theta_2(1-\exp(-\theta_1))+x_{t-1}^{(i)}\exp(-\theta_1)$ for the Vasicek Interest Rate model using transition equation (2.7). Then, there are two step to produce $k=1,...,N$ particle paths from this importance density:

1– Draw the latent variable I_k with

$$P(I_k=i)\propto w_{t-1}^{(i)}\pi(y_t|\hat{x}_t^{(i)}) \qquad i=1,2,...,N.$$

2– Draw $x_t^{(k)}\sim\pi(x_t|x_{t-1}^{(i)})$ for $I_k=i$.

and set $x_{0:t}^{(k)}=(x_{0:t-1}^{(I_k)},x_t^{(k)})$.

Thereafter, the auxiliary particle filter proceeds as follows:-

– At time $t=1,...,T$.

– Compute the importance weights via (2.13) for example as:-

$$\begin{aligned}\widetilde{w}^{(k)} &= \frac{w_{t-1}^{(I_k)}\cdot\pi(y_t|x_t^{(k)}).\pi(x_t^{(k)}|x_{t-1}^{(k)})}{w_{t-1}^{(I_k)}\cdot\pi(y_t|\hat{x}_t^{(I_k)})\cdot\pi(x_t^{(k)}|x_{t-1}^{(k)})}\\ &= \frac{\pi(y_t|x_t^{(k)})}{\pi(y_t|\hat{x}_t^{(I_k)})}, \qquad k=1,...,N.\end{aligned}$$

– Normalize the weights:-

$$w_t^{(i)}=\frac{\widetilde{w}_t^{(i)}}{\sum_{j=1}^N\widetilde{w}_t^{(j)}}, \quad i=1,...,N.$$

– Compute the N_{eff}:-

$$N_{eff}=\left[\sum_{i=1}^N(w_t^{(i)})^2\right]^{-1}.$$

– If $N_{eff}\leq N_0$, re-sample :-

Draw a sample of size N from the discrete distribution $P(x_{0:t} = x_{0:t}^{(i)}) = w_t^{(i)}$, for $i = 1, ..., N$, and relabel this sample $x_{0:t}^{(1)}, ..., x_{0:t}^{(N)}$.

Reset the weights: $w_t^{(i)} = N^{-1}$ for $i = 1, ..., N$.

– Set $\hat{\pi}_t = \sum_{i=1}^{N} w_t^{(i)} \delta_{x_{0:t}^{(i)}}$.

Therefore, the main advantage of the auxiliary particle filter is to improve the efficient of drawing in the sense that the more highly weighted particles at time $t-1$ that are consistent with new data y_t progress to be particles at time t. Effectively, particles that have a lot of information about the density of interest will survive and progress.

2.5 Auxiliary Particle Filter with Unknown Parameters

In practical situations the target density π_t may involve unknown parameters ψ. We will now see how to include the estimation of ψ in our particle filter algorithm. We often drop the subscript t for ease of notation. Liu and West (2001) suggested estimating unknown parameters by using the auxiliary particle filter. We will assume ψ is a vector of unknown parameters. The general idea of this method is that the target distribution at time t is continuous not only in x_t, but also in ψ. We will draw values of ψ from a continuous importance density based on a discrete approximation at time $t-1$. The discrete approximation at time $t-1$ is

$$\begin{aligned} \hat{\pi}_{t-1}(x_{0:t-1}, \psi) &= \sum_{i=1}^{N} w_{t-1}^{(i)} \delta_{(x_{0:t-1}^{(i)}, \psi^{(i)})} \\ &\approx \pi(x_{0:t-1}, \psi | y_{0:t-1}) \end{aligned}$$

from which the discrete marginal distribution follows as

$$\begin{aligned} \hat{\pi}_{t-1}(\psi) &= \sum_{i=1}^{N} w_{t-1}^{(i)} \delta_{\psi^{(i)}} \\ &\approx \pi(\psi | y_{0:t-1}). \end{aligned}$$

Basically, Liu and West (2001) suggest using a Normal distribution centred at $\psi^{(i)}$ with covariance matrix Λ instead of the point masses $\delta_{\psi^{(i)}}$. The target distribution for ψ therefore becomes:-

$$\hat{\pi}_{t-1}(\psi) = \sum_{i=1}^{N} w_{t-1}^{(i)} N(\psi; \psi^{(i)}, \Lambda); \tag{2.15}$$

where $N(\psi; \psi^{(i)}, \Lambda)$ is the probability density function of a multivariate normal random variable with mean $\psi^{(i)}$ and variance matrix Λ. Now, we will find the mean vector $\bar{\psi}$ and variance matrix of ψ under this $\hat{\pi}_{t-1}$:-

$$\begin{aligned}
\bar{\psi} &= \mathrm{E}(\psi) \\
&= E(E(\psi|I)) \\
&= \mathrm{E}(\psi^{(I)}) \\
&= \sum_{i=1}^{N} w_{t-1}^{(i)} \psi^{(i)}, \\
\mathrm{var}(\psi) &= \mathrm{E}(\mathrm{var}(\psi|I)) + \mathrm{var}(\mathrm{E}(\psi|I)) \\
&= \mathrm{E}(\Lambda) + \mathrm{var}(\psi^{(I)}) \\
&= \Lambda + \Sigma
\end{aligned}$$

where I is a latent classification variable for the components of the mixture distribution with $P(I = i) = w_{t-1}^{(i)}$ and $\Sigma = \mathrm{var}(\psi^{(I)})$ is the variance matrix of ψ under the original discrete approximation.

We will now modify the definition of $\hat{\pi}_{t-1}$ given in (2.15) to

$$\hat{\pi}_{t-1}(\psi) = \sum_{i=1}^{N} w_{t-1}^{(i)} N(\psi; m^{(i)}, h^2\Sigma) \tag{2.16}$$

where $m^{(i)} = a\psi^{(i)} + (1-a)\overline{\psi}$, in which $a \in (0,1)$ and $a^2 + h^2 = 1$.

We will re-calculate the expected value and covariance matrix of ψ under (2.16) as

follows:-

$$\begin{aligned}
\mathrm{E}(\psi) &= \mathrm{E}(\mathrm{E}(\psi|I)) \\
&= \mathrm{E}(a\psi^{(I)} + (1-a)\overline{\psi}) \\
&= a\overline{\psi} + (1-a)\overline{\psi} \\
&= \overline{\psi}; \qquad (2.17) \\
\mathrm{var}(\psi) &= \mathrm{E}(\mathrm{var}(\psi|I)) + \mathrm{var}(\mathrm{E}(\psi|I)) \\
&= \mathrm{E}(h^2\Sigma) + \mathrm{var}(a\psi^{(I)} + (1-a)\overline{\psi}) \\
&= h^2\Sigma + a^2\mathrm{var}(\psi^{(I)}) \\
&= h^2\Sigma + a^2\Sigma \qquad (2.18) \\
&= \Sigma.
\end{aligned}$$

We note that $\mathrm{var}(\psi) = \Sigma$ is better than the previous variance $\Lambda + \Sigma$ because, for example, for any non-zero vector a, $\mathrm{var}(a^T\psi) = a^T\Sigma a \ll a^T\Lambda a + a^T\Sigma a = a^T(\Lambda+\Sigma)a$ which would be the variance of $a^T\psi$ under the distribution (2.15). The resulting joint distribution for $x_{0:t-1}$ is discrete and for ψ is continuous,

$$\hat{\pi}_{t-1}(x_{0:t-1}, \psi) = \sum_{t=1}^{N} w_{t-1}^{(i)} N(\psi; m^{(i)}, h^2\Sigma)\delta_{x_{0:t-1}^{(i)}}.$$

Now, the target distribution at time t becomes:-

$$\begin{aligned}
\pi(x_{0:t}, \psi|y_{1:t}) &\propto \pi(x_{0:t}, \psi, y_t|y_{1:t-1}) \\
&= \pi(y_t|x_{0:t}, \psi, y_{1:t-1}) \cdot \pi(x_{0:t}|x_{0:t-1}, \psi, y_{1:t-1}) \cdot \pi(x_{0:t-1}, \psi|y_{1:t-1}) \\
&= \pi(y_t|x_t, \psi) \cdot \pi(x_t|x_{t-1}, \psi) \cdot \pi(x_{0:t-1}, \psi|y_{1:t-1}) \\
&\approx \pi(y_t|x_t, \psi) \cdot \pi(x_t|x_{t-1}, \psi) \cdot \hat{\pi}_{t-1}(x_{0:t-1}, \psi) \\
&= \sum_{i=1}^{N} w_{t-1}^{(i)} \pi(y_t|x_t, \psi)\pi(x_t|x_{t-1}^{(i)}, \psi)N(\psi, m^{(i)}, h^2\Sigma)\delta_{x_{0:t-1}^{(i)}}. \qquad (2.19)
\end{aligned}$$

We can now apply the auxiliary particle filter, in a slightly modified form, to sample from (2.19). This is Lui and West's algorithm (Liu and West, 2001):-

– Draw $(x_0^{(1)}, \psi^{(1)}), ..., (x_0^{(N)}, \psi^{(N)})$ independently from $\pi(x_0)\pi(\psi)$ and set $w_0^{(i)} = N^{-1}, \quad i = 1, 2, ..., N$, so

$$\hat{\pi}_0 = \sum_{i=1}^{N} w_0^{(i)} \delta_{(x_0^{(i)}, \psi^{(i)})}.$$

– For $t = 1, 2, ..., T$

– Compute $\overline{\psi} = \mathrm{E}_{\hat{\pi}_{t-1}}(\psi)$ and $\Sigma = \mathrm{var}_{\hat{\pi}_{t-1}}(\psi)$.

For $i = 1, 2, ..., N$, set

$$\begin{aligned} m^{(i)} &= a\psi^{(i)} + (1-a)\overline{\psi} \\ \hat{x}_t^{(i)} &= \mathrm{E}(x_t | x_{t-1} = x_{t-1}^{(i)}, \psi = m^{(i)}). \end{aligned}$$

– For $k = 1, 2, ..., N$

– Draw I_k with

$$P(I_k = i) \propto w_{t-1}^{(i)} \pi(y_t | x_t = \hat{x}_t^{(i)}, \psi = m^{(i)}).$$

– Draw $\psi^{(k)}$ from $N(m^{(I_k)}, h^2\Sigma)$.

– Draw $x_t^{(k)}$ from $\pi(x_t | x_{t-1} = x_{t-1}^{(I_k)}, \psi = \psi^{(k)})$ and set $x_{0:t}^{(k)} = (x_{0:t-1}^{(I_k)}, x_t^{(k)})$.

– Set

$$\widetilde{w}_t^{(k)} = \frac{\pi(y_t | x_t = x^{(k)}, \psi = \psi^{(k)})}{\pi(y_t | x_t = \hat{x}_t^{(I_k)}, \psi = m^{(I_k)})}$$

– Normalize the weights:-

$$w_t^{(i)} = \frac{\widetilde{w}_t^{(i)}}{\sum_{j=1}^{N} \widetilde{w}_t^{(i)}}$$

– Compute $N_{eff} = \left[\sum_{i=1}^{N} (w_t^{(i)})^2\right]^{-1}$

– If $N_{eff} \leq N_0$ re-sample:-

– Draw a sample of size N from the discrete distribution $p((x_{0:t}, \psi) =$

$(x_{0:t}^{(i)}, \psi^{(i)})) = w_t^{(i)}, i = 1, ..., N$, and relabel this sample

$$(x_{0:t}^{(1)}, \psi^{(1)}), ..., (x_{0:t}^{(N)}, \psi^{(N)})$$

– Reset the weights:-

$$w_t^{(i)} = N^{-1}, \qquad i = 1, 2, ..., N.$$

– Set $\hat{\pi}_t = \sum_{i=1}^{N} w_t^{(i)} \delta_{(x_{0:t}^{(i)}, \psi^{(i)})}$

2.5.1 Dealing with Positive Parameters

We shall briefly indicate how to deal with the case when elements of ψ are constrained to be positive by explaining how we can estimate the parameters $\psi = (\theta_1, \theta_2, \theta_3, H)$ of the Vasicek Interest Rate Model expressed as a Hidden Markov model through equations (2.7) and (2.8). Let us suppose that each of the positive parameters in ψ is indexed by the subscript j, so that in the Vasicek Interest Rate model case the four components of ψ are $\psi_j, j = 1, ..., 4$. Each component ψ_j has associated with it a gamma distribution with shape parameters α_j and scale parameters β_j. The multivariate normal distribution used in (2.15) is then replaced with the product of these gamma densities. The associated gamma mean $\mu_j(\alpha_j, \beta_j)$ and variance $\sigma_j^2(\alpha_j, \beta_j)$ are the following:-

$$\begin{aligned} \mu_j^{(i)} &= \mu_j(\alpha_j^{(i)}, \beta_j^{(i)}) \\ &= a\psi_j^{(i)} + (1-a)\overline{\psi_j} \\ \sigma_j^{2(i)} &= \sigma_j^2(\alpha_j^{(i)}, \beta_j^{(i)}) \\ &= h^2 \Sigma_{jj}, \end{aligned}$$

where Σ_{jj} is the j^{th} diagonal element of Σ. These equations can be solved for $\alpha_j^{(i)}$ and $\beta_j^{(i)}$. The resulting mixture distribution is

$$\sum_{i=1}^{N} w_{t-1}^{(i)} \pi(\psi; \alpha^{(i)}, \beta^{(i)}) \tag{2.20}$$

where

$$\pi(\psi; \alpha^{(i)}, \beta^{(i)}) = \prod_{j=1}^{4} \pi(\psi_j; \alpha_j^{(i)}, \beta_j^{(i)}), \tag{2.21}$$

in which $\alpha^{(i)}$ and $\beta^{(i)}$ are vectors of parameters and $\pi(\cdot; \alpha_j^{(i)}, \beta_j^{(i)})$ is the gamma probability density function with parameters $\alpha_j^{(i)}$ and $\beta_j^{(i)}$. We can work out the mean and variance of the vector ψ under this mixture distribution by using arguments similar to those that led to (2.17) and (2.18). In particular $h^2\Sigma$ is replaced by h^2(diag) Σ, where diag(Σ) is a matrix formed from Σ by setting all except the diagonal elements to zero, since by (2.21) the parameters are independent give I. Hence, under (2.20), ψ has the same mean $\overline{\psi}$ and variance matrix $h^2\text{diag}(\Sigma) + a^2\Sigma \approx \Sigma$ when a is close to 1.

2.6 Simulation Studies

2.6.1 Stochastic Volatility model

Tayler (1982) suggested a Hidden Markov model for univariate stochastic volatility x_t as follows:-

$$x_t = \alpha x_{t-1} + \sigma w_t \tag{2.22}$$

$$y_t = \beta \exp(x_t/2) v_t \tag{2.23}$$

where

$w_t \sim^{iid} N(0,1)$,

$v_t \sim^{iid} N(0,1)$ and

x_t really represents Volatility.

This model is written in the same forms as (2.1) and (2.2). We have already seen that the Hidden Markov model is formulated as follows:-

- We have an underlying process $x_0, x_1, ...$, assumed to follow a first order Markov process, such that $x_k|x_{k-1} \sim \pi_{x_k|x_{k-1}}(x_k|x_{k-1})$, with initial distribution $\pi(x_0)$;
- we have observations $y_1, y_2, ...$, assumed to be conditionally independent provided that $x_0, x_1, ...$, are known, allowing us to write $y_k|x_k \sim \pi_{y|x}(y_k|x_k)$.

If we use the notation $N(x; \mu, V)$ for the probability density function of an univariate normal random variable with mean μ and variance V, the stochastic volatility model has :

$$\begin{aligned} \pi(x_0) &= N\left(x_0; 0, \frac{\sigma^2}{1-\alpha^2}\right), \\ \pi(x_t|x_{t-1}) &= N(x_t; \alpha x_{t-1}, \sigma^2), \\ \pi(y_t|x_t) &= N(y_t; 0, \beta^2 \exp(x_t)). \end{aligned}$$

This prior is chosen so that the unconditional variance of x_t remains constant.

We will apply to the Stochastic Volatility model the particle filter assuming that the parameters α, β and σ are known, and the auxiliary particle filter assuming that these parameters are unknown.

First, Figure 2.1 shows the results of the particle filter assuming known parameters. In particular the value x_t of the process which represent volatility, the observation y_t, the approximate posterior mean of x_t under $\pi(x_t|y_{1:t})$ and approximate 90% credible intervals are shown across time t. The true values of the parameters are $\alpha = 0.91, \beta = 0.5, \sigma = 1$. The 90% credible intervals contain the true volatilities indicating that this is estimated well.

Secondly, Figure 2.2 shows the estimation of the parameters (α, β, σ) across time using the auxiliary particle filter. In particular the approximate posterior mean and approximate 90% credible intervals of each parameter are presented together with the known parameter values. Again, good estimation is achieved, especially for α and σ.

Figure 2.3 is an equivalent plot to Figure 2.1 for the auxiliary particle filter. It is quite similar to Figure 2.1, although the credible intervals are rather different perhaps due to the fact that the parameters (α, β, σ) are estimated.

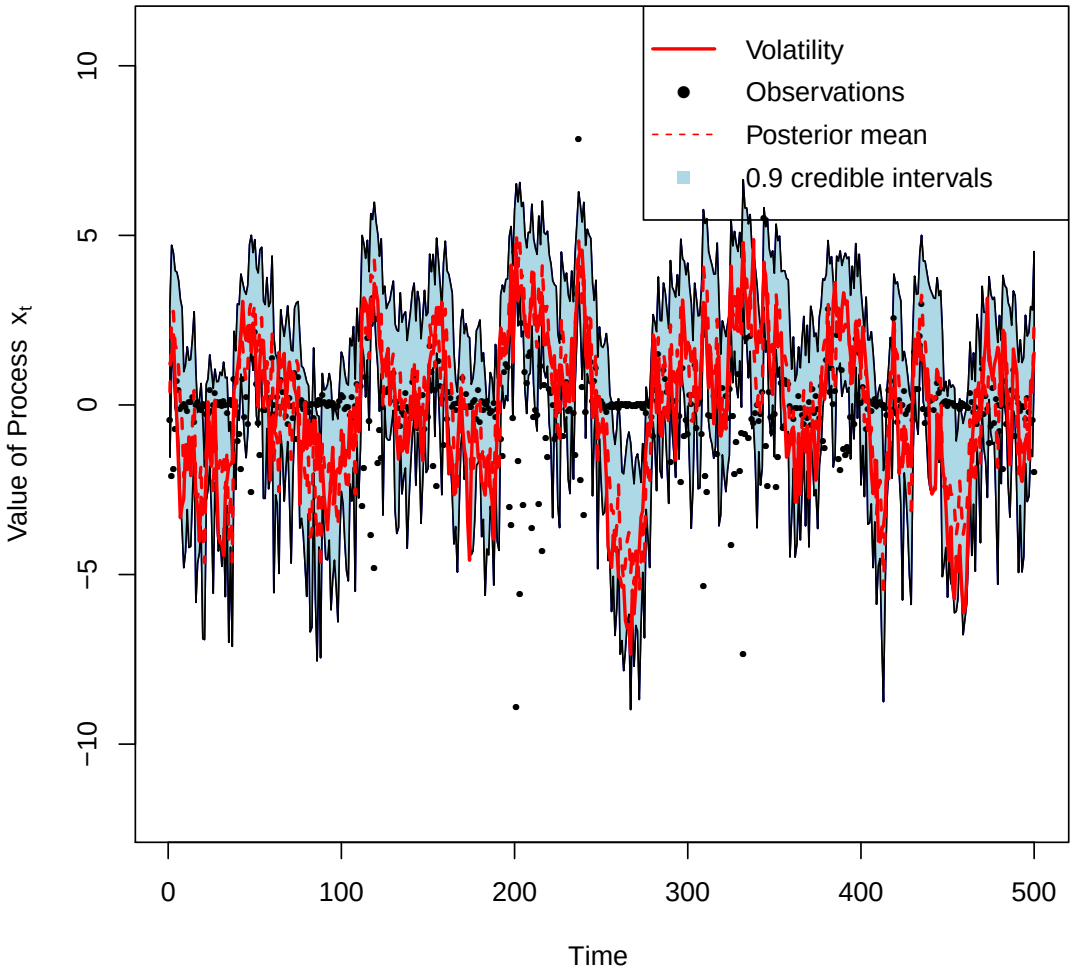


Figure 2.1: Results of the particle filter assuming known parameters $\alpha = 0.91, \beta = 1, \sigma = 0.5$. The true volatility $x_t, t = 1, ..., T, T = 500$, assumed known in our simulation study, is indicated by the red/unbroken line. The data y_t are represented by the dots. The approximate mean of the posterior distribution $\pi(x_t|y_{1:t})$ is represented using the red/broken line, while associated 90% credible intervals are shaded light blue/grey.

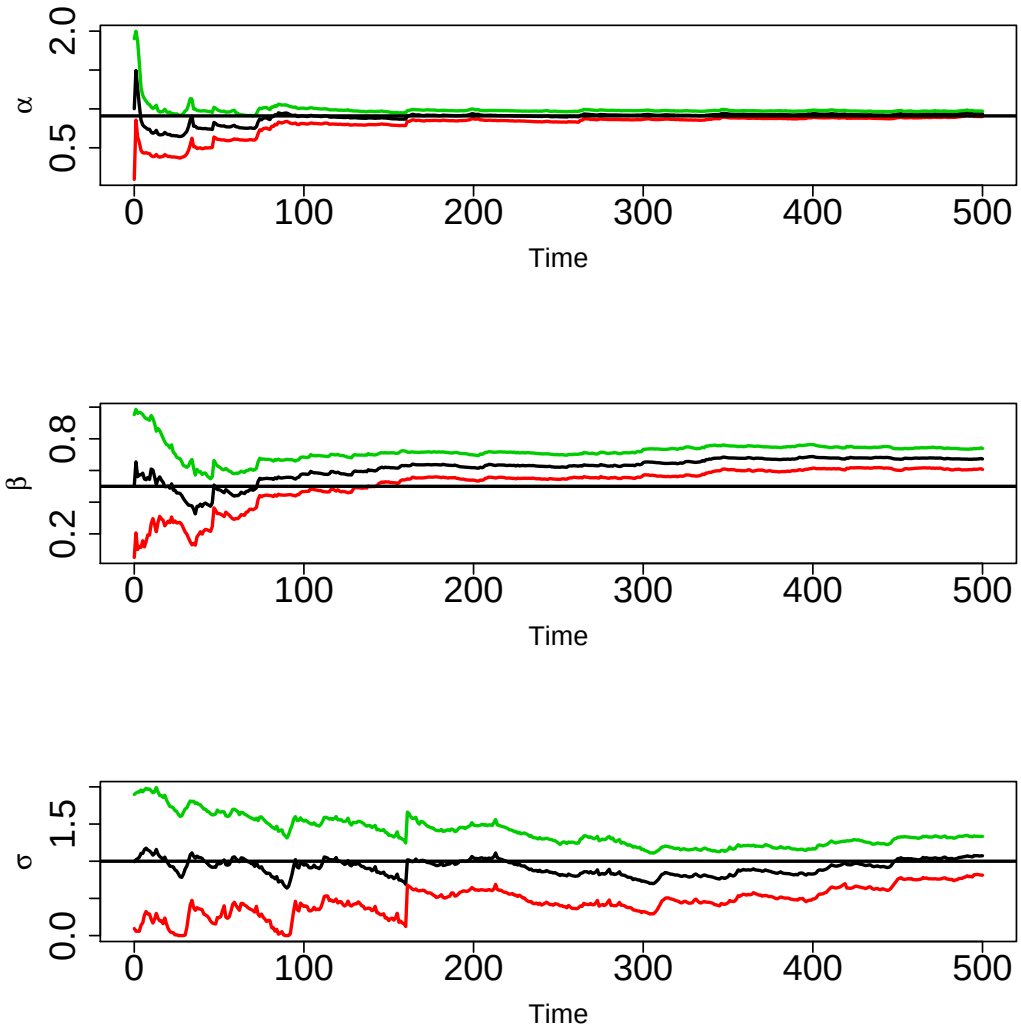


Figure 2.2: Results for the auxiliary particle filter for the parameters (α, β, σ). The true values of the parameters, which are known in this simulation study, are indicated by the thick horizontal line. The first graph is for parameter α, defined though equation (2.22). The posterior mean is shown by the black curve, while 90% credible intervals are shown by the outer lines. The second graph is for parameter β, defined through equation (2.23), while the third graph is for parameter σ, defined through equation (2.22). The credible intervals become narrower as more data become available.

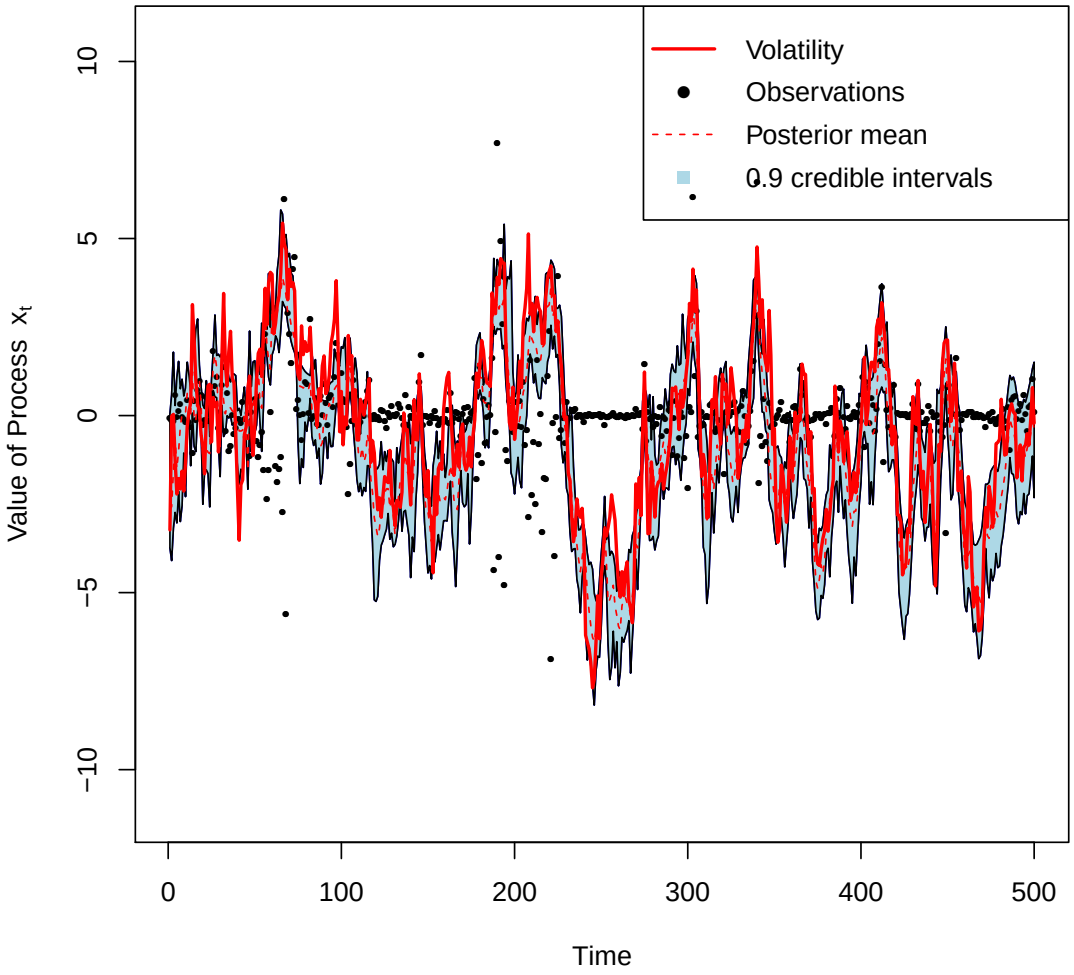


Figure 2.3: Results of the auxiliary particle filter assuming unknown parameters parameters α, β, σ. The estimate volatility $x_t, t = 1, ..., T, T = 500$, is indicated by the red/unbroken line. The data y_t are represented by the dots. The estimated mean of the posterior distribution $\pi(x_t|y_{1:t})$ is represented using the red/broken line, while associated 90% credible intervals are shaded light blue/grey.

2.6.2 Vasicek Interest Rate Model

Next, we will apply the Sequential Monte Carlo Bayesian Algorithm using Weight Updating (outlined in Section 2.2.2), the particle filter assuming known parameters (outlined in Section 2.3,) and the auxiliary particle filter assuming unknown parameters (outlined in Section 2.4) to the Vasicek Interest Rate Model discussed in Section 2.1 in order to illustrate further this methodology.

Results from the Sequential Monte Carlo Bayesian Algorithm are given in Figure 2.4. In particular, the estimation of each parameter $\beta_1 = \theta_1 \times \theta_2$, θ_1 and θ_3 based on data $y_t, t = 1, 2, \ldots, T, T = 500$, is shown over time by presenting approximate posterior means and 90% credible intervals. The maximum likelihood estimate based on all the data and the true value of each parameter is also shown. The credible intervals become narrower as more data become available.

The results from the particle filter are given Figure 2.5. In particular, the true value of the Interest Rate $x_t, t = 1, ..., T, T = 500$, the data y_t, the approximate posterior mean of x_t under $\pi(x_t|y_{1:t})$ and approximate 90% credible intervals are presented. Good estimation is achieved.

The results of the auxiliary particle filter assuming unknown parameters $(\beta_1, \theta_1, \theta_3, H)$ are presented in Figure 2.6. The first graph for parameter β_1(top left) shows the true value of the parameter, the maximum likelihood estimation, the approximate posterior mean and associated approximate 90% credible intervals. The second graph for θ_1 (top right), the third graph for θ_3 (bottom left) and the fourth graph H (bottom right) all show similar behaviour with the credible intervals containing the maximum likelihood estimates based on the data. Again, the credible intervals tend to become narrower as more data become available.

Figure 2.7 is an equivalent plot to Figure 2.5 for the auxiliary particle filter. An interest rate estimate of similar quality has been achieved.

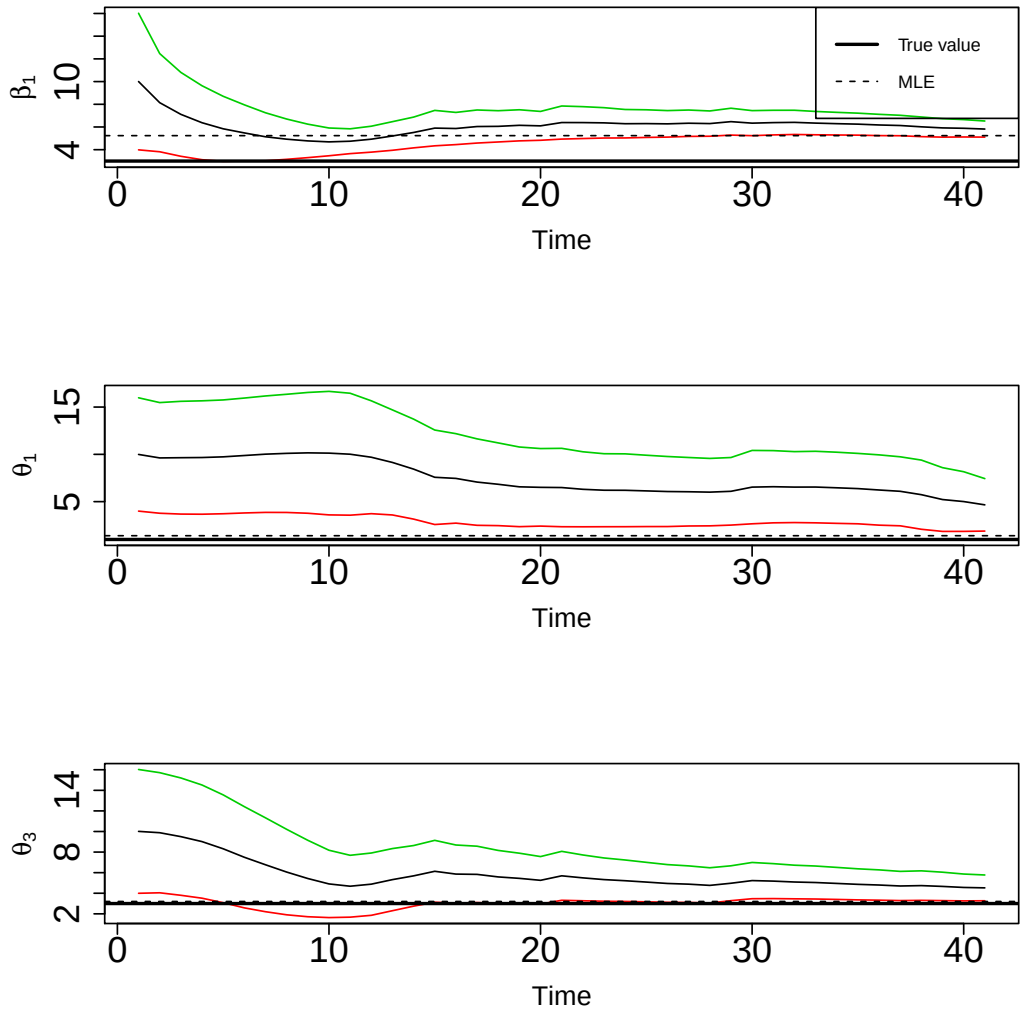


Figure 2.4: Result for the Sequential Monte Carlo Bayesian Algorithm Weight Updating for estimating the parameters β_1, θ_1 and θ_3, with $\beta_1 = 3, \theta_1 = 3, \theta_3 = 3$. The first graph is for parameter $\beta_1 = \theta_1 \times \theta_2$. The approximate posterior mean is shown by the black curve, while approximate 90% credible intervals are shown by the outer lines. The second graph is for parameter θ_1, while the third graph is for parameter θ_3. The true value of each parameter is also shown. Note that we plot time $= \delta t, \delta = \frac{1}{12}$, on the horizontal axis as we are thinking of the data as monthly with time being in years.

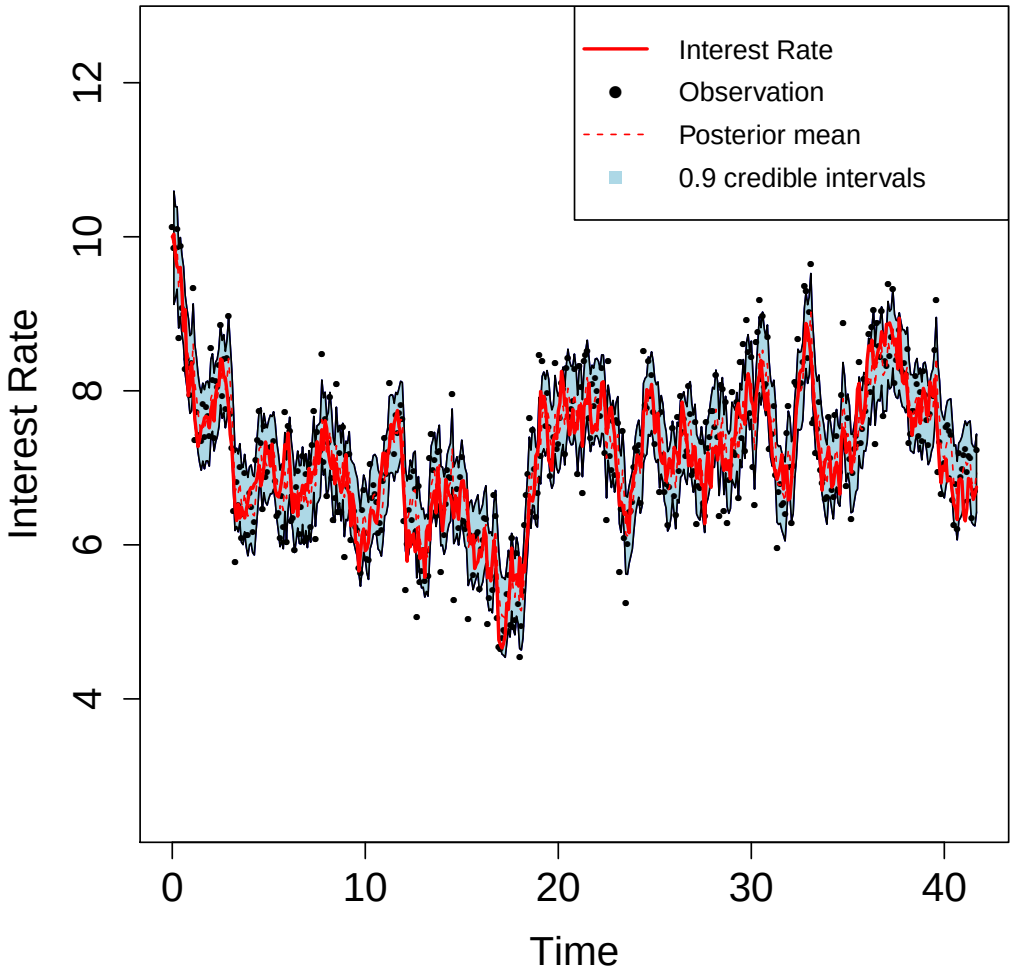


Figure 2.5: Result of the particle filter assuming known parameters $\beta_1 = 3, \theta_1 = 3, \theta_3 = 3$ where $\beta_1 = \theta_1 \times \theta_2$. The true value of the Interest Rate $x_t, t = 1, ..., T, T = 500$, are indicated by the unbroken line. The data y_t are represented by the dots. The approximate mean of the posterior distribution $\pi(x_t|y_{1:t})$ is represented by the broken line and is close to the true values x_t, while associated 90% credible intervals are shaded light blue/grey. Note that we plot time $= \delta t, \delta = \frac{1}{12}$, on the horizontal axis as we are thinking of the data as monthly with time being in years.

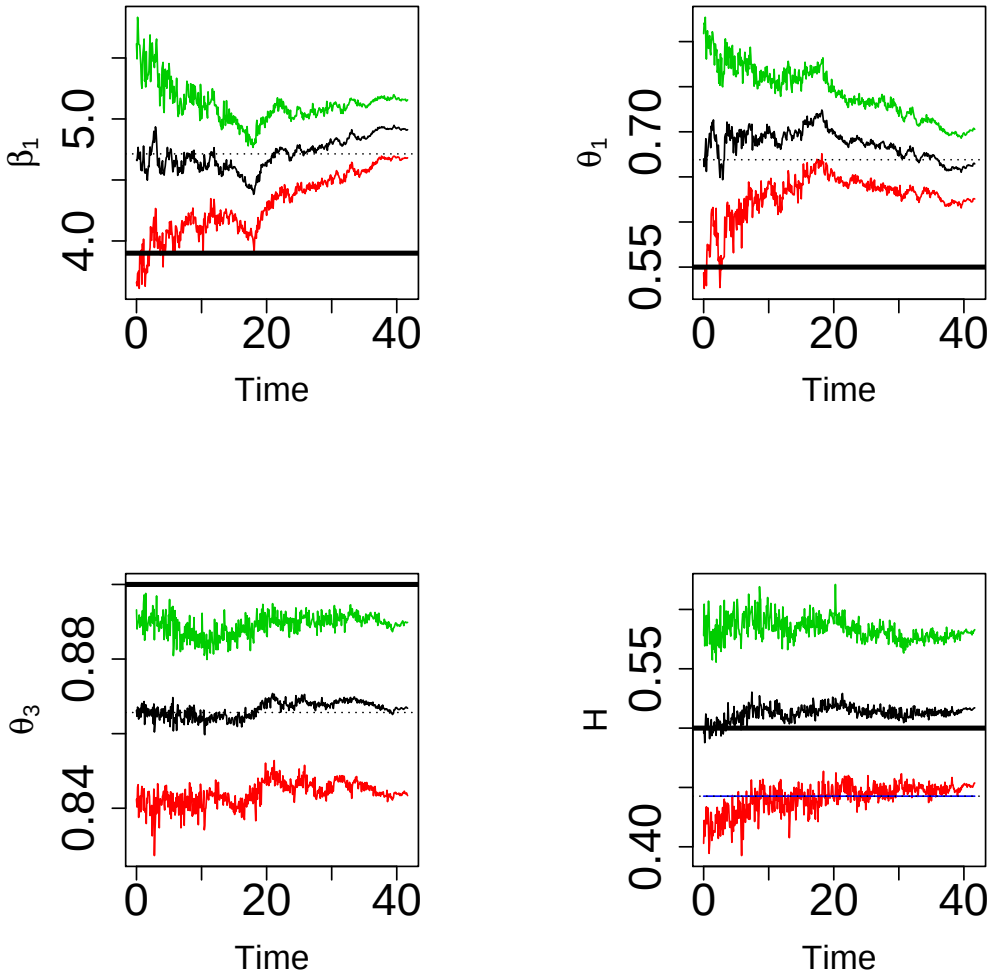


Figure 2.6: Result of the auxiliary particle filter for the parameters $\beta_1 = \theta_1 \times \theta_2, \theta_1, \theta_3$ and H, where the standard deviation H of the assumed noise is defined through equation (2.8). The true parameter values, which are known in this simulation study, are shown by the thick black horizontal lines, while the dashed horizontal lines show the maximum likelihood estimates based on all the data $y_t, t = 1, ..., T, T = 500$. The approximate posterior mean of each parameter is shown by the black line, while the associated approximate 90% confidence intervals are shown by the outer lines. The graphs are for parameters β_1 (top left), θ_1 (top right), θ_3 (bottom left) and H (bottom right). The credible intervals contain the maximum likelihood estimate. Note that we plot time $= \delta t, \delta = \frac{1}{12}$, on the horizontal axis as we are thinking of the data as monthly with time being in years.

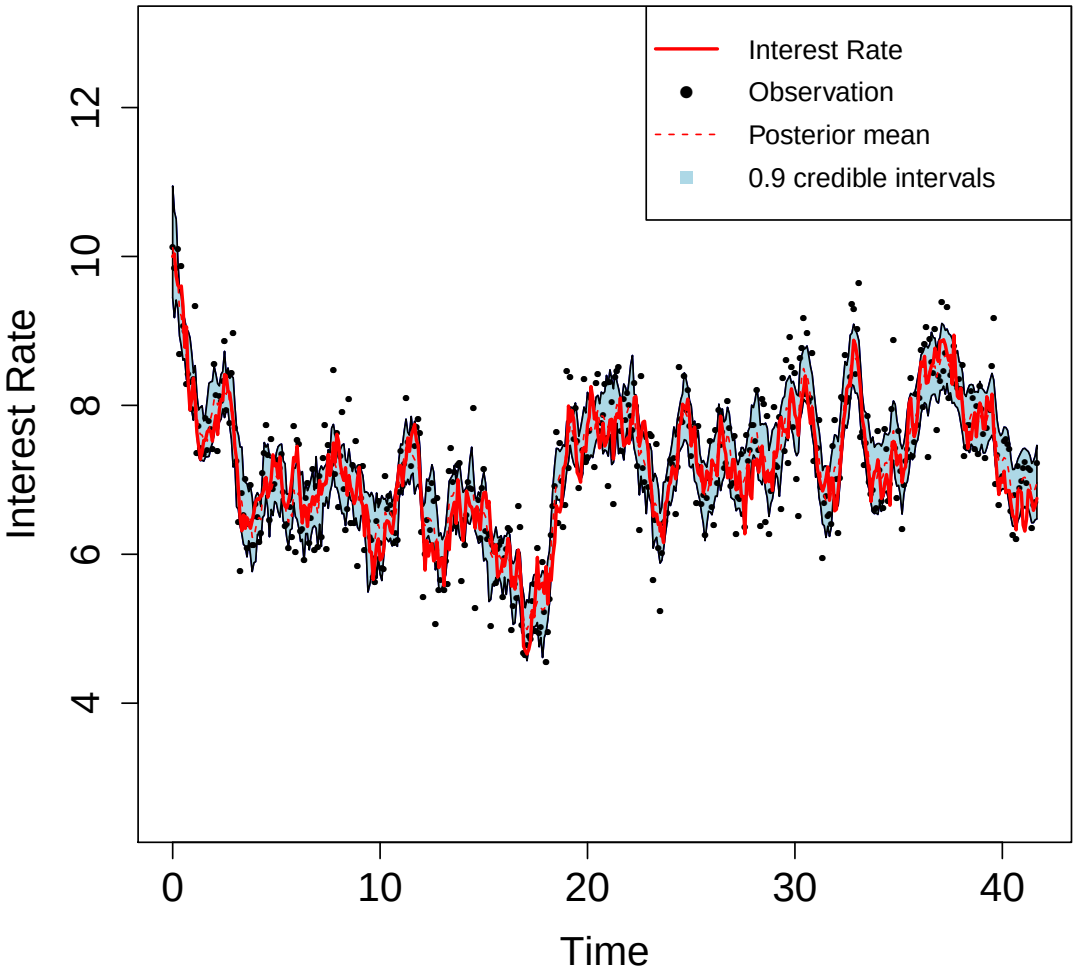


Figure 2.7: Result of the auxiliary particle filter assuming unknown parameters $\beta_1, \theta_1, \theta_3$ where $\beta_1 = \theta_1 \times \theta_2$. The true value of the Interest Rate $x_t, t = 1, ..., T, T = 500$, are indicated by the unbroken line. The data y_t are represented by the dots. The approximate mean of the posterior distribution $\pi(x_t|y_{1:t})$ is represented by using broken line and is close to the true values x_t, while associated 90% credible intervals are shaded light blue/grey. Note that we plot time $= \delta t, \delta = \frac{1}{12}$, on the horizontal axis as we are thinking of the data as monthly with time being in years.

2.6.3 The prior distribution on the parameter H of the noisy Vasicek Interest Rate Model

When we implemented the auxiliary particle filter (Petris et al., 2009) for the noisy Vasicek Interest Rate Model, we assumed that the prior distribution on the noise standard deviation H was a uniform distribution. Informal experiments led us to believe that the performance of the auxiliary particle filter was more sensitive to the choice of the prior on H that to the choice of the prior on the other parameters of the noisy Vasicek Interest Rate Model.

Therefore, we conducted a simulation study to assess the sensitivity of posterior inference to the prior assumptions made about H. We generated data sets from the noisy Vasicek Interest Rate Model of length 133 (including the initial value x_0), thought of as being eleven years of interest rate data. We considered various uniform priors on H, with small minimum value 0.005 and different maximum values. We also considered a range of true values of H. There are many ways to assess the effect of the choice of the prior on H on posterior inference. We decided to use the number out of fifty 95% credible intervals for the underlying interest rate at the final time point that contained the true interest rate. So we simulated fifty data sets from the noisy Vasicek Interest Rate Model, ran the auxiliary particle filter on each data set, and counted the number of 95% credible intervals at the final time point that contain the underlying interest rate. In Table 2.1 we represent our results when the true values of the other parameters are $\theta_1 = 3$, $\theta_2 = 2$ and $\theta_3 = 1$. We immediately see that if the true value of H lies well within the interval that defines the uniform prior relatively good performance is achieved. In Table 2.2 the results are for different true values of the other parameters in particular $\theta_1 = 3.87$, $\theta_2 = 0.75$ and $\theta_3 = 1$. Again, relatively good performance is achieved if the true value of H lies well within the prior interval. We can also see from Table 2.1 and Table 2.2 that posterior inference can be rather sensitive to the choice of the prior distribution on H in that a misspecified prior can lead to poor posterior inference. In particular, for small true values of H $(0.01, 0.05$ and $0.1)$, the auxiliary particle filter performs well, with almost all the credible intervals

containing the true interest rate value, across all uniform prior distribution on H that we considered. For large values of H $(0.5$ and $0.75)$, the performance of the auxiliary particle filter is less good, especially when the interval for the uniform prior on H is narrow and considerably below the true value of H.

In conclusion, the performance of the auxiliary particle filter seems consistent over prior interval width provided that the true value of H lies well within the interval for the uniform prior on H. A uniform prior distribution across a wide range of possible value of H, such as $[0.05, 2]$ or $[0.05, 10]$, seems to offer good posterior performance for a range of realistic true H values.

	True value of H				
Prior distribution on H	0.01	0.05	0.1	0.5	0.75
[0.005,0.25]	49	50	50	36	28
[0.005,0.5]	50	50	49	40	34
[0.005,0.75]	49	49	48	47	44
[0.005,1]	49	50	49	43	44
[0.005,1.5]	50	50	50	40	42
[0.005,2]	50	50	50	46	39
[0.005,10]	48	50	48	43	43

Table 2.1: The number out of fifty 95% credible intervals for the underlying interest rate at the final time point that contain the true interest rate for the noisy Vasicek Interest Rate Model where $\theta_1 = 3$, $\theta_2 = 2$ and $\theta_3 = 1$. The columns correspond to different true values of H, while the rows correspond to different uniform prior distributions on H, with the minimum value being set to 0.005 and with different maximum values.

	True value of H				
Prior distribution on H	0.01	0.05	0.1	0.5	0.75
[0.005,0.25]	50	50	50	30	29
[0.005,0.5]	50	49	50	45	40
[0.005,0.75]	50	50	49	43	40
[0.005,1]	50	50	49	44	45
[0.005,1.5]	50	49	49	44	40
[0.005,2]	50	50	50	43	44
[0.005,10]	50	50	48	41	42

Table 2.2: The number out of fifty 95% credible intervals for the underlying interest rate at the final time point that contain the true interest rate for the noisy Vasicek Interest Rate Model where $\theta_1 = 3.87$, $\theta_2 = 0.75$ and $\theta_3 = 1$. The columns correspond to different true values of H, while the rows correspond to different uniform prior distributions on H, with the minimum value being set to 0.005 and with different maximum values.

2.6.4 The prior distribution on the β of the Stochastic Volatility Model

As we discussed above our implementation of the auxiliary particle filter for the stochastic volatility model assumes a uniform prior distribution on β. We now investigate how the performance of the auxiliary particle filter depends on the specification of this prior distribution. We focus attention on the prior distribution on β because our experience is that the performance of the algorithm is more sensitive to the choice of prior distribution on this parameter than to the choice of the other prior distributions.

Therefore, we conducted a simulation study to assess the sensitivity of posterior inference to the prior assumptions made about β by generating data sets from the stochastic volatility model of length 500 (including the initial value x_0). We considered various uniform priors on β with small minimum value 0.005 and different maximum values. We also considered a range of the true values of β and true values of α. We decided to use the number out of fifty 95% credible intervals for the underlying volatility at the final time point that contained the true volatility value as a measure of the quality of posterior inference. So, we simulated fifty data sets from the noisy stochastic volatility model, ran the auxiliary particle filter on each data set, and counted the number of 95% credible intervals at the final time point that contained the underlying volatility process.

In Table 2.3 we present our results when the other parameters are $\alpha = 0.3$ and $\sigma = 0.01$. We immediately see that when the true of value β lies well within the interval that defines the uniform prior, good performance is achieved, while bad performance occurs when the true value of β was out of the prior interval. In this case the algorithm often fails because the likelihood term effectively assigns zero probability to the data. Performance is also not good when the true value of β is very small and when the uniform prior interval for β is very large.

In Table 2.4 results are presented for a different true values of α, in particular $\alpha = 0.6$, and the same true value of σ. Again, relatively good performance is achieved if the true

value of β lies well within the prior interval and bad performance occurs if the true value of β is outside the prior interval.

In Table 2.5 results are presented for another different true values of α in particular $\alpha = 0.8$, and the same true value of σ. The basic conclusion from Table 2.5 are the same as those obtained from Tables 2.3 and 2.4: good performance is achieved if the true value of β lies well within the prior interval, while bad performance occurs it the true value of β is outside the prior interval. We can therefore see from Tables 2.3, 2.4 and 2.5 that posterior inference can be rather sensitive to the choice of the distribution on β in that a misspecified prior can lead to bad posterior inference.

For small true values of β (for example 0.01, 0.05 and 0.1), the auxiliary particle filter generally performs well, unless the true value of β is outside of the prior interval, in which case failure is due to the fact that the likelihood term effectively assigns zero probability to the data. For large values of β (for example 0.6 and 0.7), the performance of the Auxiliary Particle Filter is again generally good, especially when the interval for the uniform prior is wider. Good performance can sometimes be achieved in this case even when the prior interval does not contain the true parameter value.

In conclusion, the auxiliary particle filter seems to perform in a consistently good way over the choice of uniform prior intervals for β prior width provided that the true value of β is well supported by the interval. A uniform prior distribution for β such as $[0.005, 2]$ or $[0.005, 10]$ seems to offer good posterior performance for a wide range of realistic true β values. Of course, the prior distribution for β should reflect specific prior knowledge about β, if such knowledge is available.

	True value of β					
Prior distribution on β	0.01	0.05	0.1	0.5	0.6	0.7
[0.005,0.01]	50	–	–	–	–	–
[0.005,0.05]	50	49	2	–	–	–
[0.005,0.1]	50	50	49	–	–	–
[0.005,0.5]	49	50	49	50	50	50
[0.005,0.75]	48	50	50	50	50	50
[0.005,1]	49	50	50	50	50	50
[0.005,2]	43	50	49	50	50	50
[0.005,10]	28	43	50	50	50	49

Table 2.3: The number out of fifty 95% credible intervals for the underlying volatility process at the final time that contain the true volatility value. The columns correspond to different true values of β, while the rows correspond to different uniform prior distributions on β, with the minimum value being set to 0.005 and with different maximum values. Data sets were simulated with $\alpha = 0.3$ and $\sigma = 0.01$. Dashes mean that the algorithm failed in some way.

	True value of β					
Prior distribution on β	0.01	0.05	0.1	0.5	0.6	0.7
[0.005,0.01]	49	0	–	–	–	–
[0.005,0.05]	50	49	1	–	–	–
[0.005,0.1]	50	50	50	0	–	0
[0.005,0.5]	50	49	50	50	50	48
[0.005,0.75]	50	50	49	49	50	50
[0.005,1]	50	49	50	50	50	50
[0.005,2]	44	49	49	50	50	50
[0.005,10]	30	43	48	50	50	50

Table 2.4: The number out of fifty 95% credible intervals for the underlying volatility process at the final time that contain the true volatility value. The columns correspond to different true values of β, while the rows correspond to different uniform prior distributions on β, with the minimum value being set to 0.005 and with different maximum values. Data sets were simulated with $\alpha = 0.6$ and $\sigma = 0.01$. Dashes mean that the algorithm failed in some way.

	True value of β					
Prior distribution on β	0.01	0.05	0.1	0.5	0.6	0.7
[0.005,0.01]	50	–	–	–	–	–
[0.005,0.05]	49	50	–	–	–	–
[0.005,0.1]	49	50	49	–	–	–
[0.005,0.5]	50	49	49	50	50	50
[0.005,0.75]	50	50	50	50	49	50
[0.005,1]	48	49	49	50	48	49
[0.005,2]	46	50	50	50	48	50
[0.005,10]	38	47	44	50	49	50

Table 2.5: The number out of fifty 95% credible intervals for the underlying volatility process at the final time that contain the true volatility value. The columns correspond to different true values of β, while the rows correspond to different uniform prior distributions on β, with the minimum value being set to 0.005 and with different maximum values. Data sets were simulated with $\alpha = 0.8$ and $\sigma = 0.01$. Dashes mean that the algorithm failed in some way.

2.7 Summary

In this chapter, we have discuss:

1. The definition and implementation of various sequential Monte Carlo algorithms such as the particle filter and the auxiliary particle filter;

2. Applications of these algorithms to various models, namely the Vasicek Interest Rate Model and the Stochastic Volatility model;

3. Studies concerning the choice of the prior distribution for some of the parameters of our models.

Following on from our results, we will be applying these particles filter based on algorithms to the Heston model, a popular model for financial data, in Chapter 3, and to an extension of the Heston model in Chapter 4.

Chapter 3

The Heston Model

In this chapter we discuss the Heston model and how to make inference about its unknown quantities. In Section 3.1 we introduce the Heston model, before discussing inference methodology for a special case in Section 3.2. In Section 3.3 we extend this methodology to the general Heston model, while in Section 3.4 we present all our results using both simulated and real data.

3.1 Introduction

Heston (1993) presented a new model to improve and extend the Black-Scholes model (Fisher and Scholes, 1973) by adding some features such as non-constant volatility, a non-log-normal distribution for the assets prices, a potential correlation between the stock price process and the volatility process. The volatility is assumed to be a stochastic process defined through a stochastic differential equation which gives it the important mean-reverting property (Bauer, 2012). One form of this model can be written as follows:-

$$\frac{dS_t}{S_t} = \mu dt + \sqrt{v_t} dw_t^s$$

$$dv_t = \beta(\kappa - v_t)dt + \sigma\sqrt{v_t} dw_t^v$$

$$\mathrm{E}(dw_t^s dw_t^v) = \rho dt,$$

where

S_t is an asset price at continuous time t,

v_t is a volatility at continuous time t,

μ is a rate of the asset returns,

β is a mean-reversion rate,

κ is a long run volatility,

σ is a volatility of the volatility process,

ρ is a correlation coefficient between the two underlying processes and

dw_t^s and dw_t^v are standard Brownian motion increments over a time interval dt.

In this model, the volatility remains greater than or equal to zero. Also, if the Feller condition $2\kappa\beta \geqslant \sigma^2$ holds, volatility will be strictly non-zero (Heston, 1993).
We will study the Heston model in the two cases $\rho = 0$ (uncorrelated underlying processes) and $\rho \neq 0$ (correlated underlying processes) as follows:-

3.2 The Heston Model when $\rho = 0$

In this case, the Heston model will be as follows:-

$$\frac{dS_t}{S_t} = \mu dt + \sqrt{v_t} dw_t^s \tag{3.1}$$

$$dv_t = \beta(\kappa - v_t)dt + \sigma\sqrt{v_t} dw_t^v, \tag{3.2}$$

in which w_t^s and w_t^v are independent Brownian motion processes.
Equations (3.1) and (3.2) can be thought of as a form of the general stochastic volatility model.
Our aim is to estimate the volatility process v_t assuming that the asset price process S_t is known.
It is more convenient to work with $y_t = \log(S_t/S_0)$. Using the Itô Lemma (Mikosch, 1998), we can transform (3.1) into an equation for y_t:

$$dy_t = \left[\frac{\partial y_t}{\partial S_t}\mu S_t + \frac{\partial y_t}{\partial t} + \frac{1}{2}v_t S_t^2 \frac{\partial^2 y_t}{\partial S_t^2}\right] dt + \frac{\partial y_t}{\partial S_t}\sqrt{v_t} S_t dw_t^s$$

where

$\frac{\partial y_t}{\partial S_t} = \frac{1}{S_t}$, $\frac{\partial^2 y_t}{\partial S_t^2} = -\frac{1}{S_t^2}$, and $\frac{\partial y_t}{\partial t} = 0$.

This leads to

$$\begin{aligned} dy_t &= \left[\frac{1}{S_t}\mu S_t + 0 + \frac{1}{2}v_t S_t^2\left(-\frac{1}{S_t^2}\right)\right]dt + \frac{1}{S_t}\sqrt{v_t}S_t dw_t^s \\ &= \left(\mu - \frac{1}{2}v_t\right)dt + \sqrt{v_t}dw_t^s. \end{aligned} \tag{3.3}$$

The advantage is that the right side of equation (3.3) depends only on the process v_t, and not on y_t.

Now, we will use the Euler scheme to transform the Heston model (Korn et al., 2010) from continuous to discrete time. The aim of applying the Euler scheme is that it is easier to deal with discrete rather than continuous time.

Firstly, the Euler scheme of y_t is

$$y_{t+\triangle} = y_t + \int_t^{t+\triangle}\left(\mu - \frac{1}{2}v_u\right)du + \int_t^{t+\triangle}\sqrt{v_u}dw_u^s,$$

by integrating the above equation. We now let $\triangle \to 0$. When we move from continuous to discrete time, positivity of the volatility process is unfortunately no longer guaranteed. Hence, we replace v_t with $|v_t|$ under the square root:-

$$y_{t+\triangle} = y_t + \left(\mu - \frac{1}{2}v_t\right)\triangle + \sqrt{|v_t|}\sqrt{\triangle}Z_t^s, \quad Z_t^s \sim N(0,1),$$

since $dw_t^s \sim N(0,\triangle)$ when $dw_t^s = w_{t+\triangle}^s - w_t^s$.

Secondly, the Euler scheme of v_t is

$$v_{t+\triangle} = v_t + \int_t^{t+\triangle}\beta(\kappa - v_u)du + \int_t^{t+\triangle}\sigma\sqrt{v_u}dw_u^v$$

so

$$v_{t+\triangle} = v_t + \beta(\kappa - v_t)\triangle + \sigma\sqrt{v_t}\sqrt{\triangle}Z_t^v, \quad Z_t^v \sim N(0,1), \tag{3.4}$$

since $dw_t^v \sim N(0, \triangle)$.

As before, because of the possibility of a negative variance, we rewrite equation (3.4) as follows:-

$$v_{t+\triangle} = v_t + \beta(\kappa - v_t)\triangle + \sigma\sqrt{|v_t|}\sqrt{\triangle}Z_t^v.$$

In conclusion, the Heston model when $\rho = 0$ in discrete time can be written as :-

$$y_{t+\triangle} = y_t + \left(\mu - \frac{1}{2}v_t\right)\triangle + \sqrt{|v_t|}\sqrt{\triangle}Z_t^s \tag{3.5}$$

$$v_{t+\triangle} = v_t + (\alpha - \beta v_t)\triangle + \sigma\sqrt{|v_t|}\sqrt{\triangle}Z_t^v, \tag{3.6}$$

where $y_t = \log(S_t/S_0)$ and $\alpha = \beta \times \kappa$.

3.2.1 Maximum Likelihood Estimation

To write the likelihood function of the Heston model when $\rho = 0$, we will use equations (3.5) and (3.6). We will assume that the initial values y_0 and v_0 are known. We will indicate the values of y and v at time $\triangle, 2\triangle, \ldots, t\triangle$ as $y_1, \ldots, y_t$ and $v_1, \ldots, v_t$, and we will write $y_{1:t}$ and $v_{1:t}$ for the collection of all these values. The volatilities $v_0, v_{1:t}$ are of course positive. Mathematically, the likelihood function can be expressed as

$$\begin{aligned} L((y_t, v_t), (y_{t-1}, v_{t-1}), \ldots, (y_1, v_1)|(y_0, v_0), \theta) &= L((y_t, v_t)|(y_{t-1}, v_{t-1}), \ldots, (y_1, v_1), (y_0, v_0), \theta) \times \\ & L((y_{t-1}, v_{t-1})|(y_{t-2}, v_{t-2}), \ldots, (y_1, v_1), (y_0, v_0), \theta) \\ & \times \cdots \times L((y_2, v_2)|(y_1, v_1), (y_0, v_0), \theta) \\ & \times L((y_1, v_1)|(y_0, v_0), \theta) \end{aligned}$$

where

$\theta = (\mu, \alpha, \beta, \sigma)$ is a vector of parameters.

In addition, since here the $y_{1:t}$ and $v_{1:t}$ processes satisfy the Markov property, we can write the likelihood function as

$$\begin{aligned} L((y_t,v_t),(y_{t-1},v_{t-1}),\ldots,(y_1,v_1)|(y_0,v_0),\theta) &= L((y_t,v_t)|(y_{t-1},v_{t-1}),\theta) \\ &\quad \times L((y_{t-1},v_{t-1})|(y_{t-2},v_{t-2}),\theta) \\ &\quad \times \cdots \times L((y_2,v_2)|(y_1,v_1),\theta) \\ &\quad \times L((y_1,v_1)|(y_0,v_0),\theta). \end{aligned}$$

From equations (3.5) and (3.6), it can be seen that each process does not depend on all of the parameters, so we can rewrite the above equation for the likelihood as

$$\begin{aligned} L((y_t,v_t),(y_{t-1},v_{t-1}),\ldots,(y_1,v_1)|(y_0,v_0),\theta) &= L(y_t|(y_{t-1},v_{t-1}),\mu) \times L(v_t|v_{t-1},\alpha,\beta,\sigma) \times \\ &\quad L(y_{t-1}|(y_{t-2},v_{t-2}),\mu) \times L(v_{t-1}|v_{t-2},\alpha,\beta,\sigma) \\ &\quad \times \cdots \times L(y_1|(y_0,v_0),\mu) \times L(v_1|v_0,\alpha,\beta,\sigma). \end{aligned}$$

Since the conditional distribution of each process is normal with mean and variance as follows:-

$$\begin{aligned} y_t|(y_{t-1},v_{t-1},\mu) &\sim N(y_{t-1}+(\mu-0.5v_{t-1})\triangle,\triangle v_{t-1}) \\ v_t|v_{t-1},\alpha,\beta,\sigma &\sim N(v_{t-1}+(\alpha-\beta v_{t-1})\triangle,\sigma^2\triangle v_{t-1}), \end{aligned}$$

the likelihood function of the Heston model when $\rho = 0$ takes the following form :-

$$\begin{aligned} L((y_t, v_t), (y_{t-1}, v_{t-1}), \ldots, (y_1, v_1) | (y_0, v_0), \theta) &= N(y_t; y_{t-1} + (\mu - 0.5 v_{t-1}) \triangle, \triangle v_{t-1}) \times \\ & N(v_t; v_{t-1} + (\alpha - \beta v_{t-1}) \triangle, \sigma^2 \triangle v_{t-1}) \times \cdots \times \\ & \times N(y_2; y_1 + (\mu - 0.5 v_1) \triangle, \triangle v_1) \\ & \times N(v_2; v_1 + (\alpha - \beta v_1) \triangle, \sigma^2 \triangle v_1) \\ & \times N(y_1; y_0 + (\mu - 0.5 v_0) \triangle, \triangle v_0) \\ & \times N(v_1; v_0 + (\alpha - \beta v_0) \triangle, \sigma^2 \triangle v_0), \end{aligned}$$

in which $N(x; m, s^2)$ is the normal probability density function with mean m and variance s^2 evaluated at x.

From the above equation, we can use the likelihood function to estimate the parameters μ, α, β and σ using maximum likelihood estimation when the values of volatility process and asset price process are assumed known. In general, our aim here is to estimate the current volatility v_t from the current and past asset prices and past volatilities.

We remark that we will use the full likelihood form $L((y_t, v_t), (y_{t-1}, v_{t-1}), \ldots, (y_1, v_1) | (y_0, v_0), \theta)$ when we discuss the fractional Heston Model in Section 4.3 and Section 4.4 of the next chapter. We also remark that we sometimes write $L((y_t, v_t), (y_{t-1}, v_{t-1}), \ldots, (y_1, v_1) | (y_0, v_0), \theta)$ as $L(y_{1:t}, v_{1:t} | y_0, v_0, \theta)$.

3.2.2 Bayesian Inference for the Heston Model when $\rho = 0$

The aim of Bayesian Inference (Brooks et al., 2011) is to use data to update our prior belief about parameters into a posterior belief. In the Heston model when $\rho = 0$, we will sample the posterior distribution of the parameters and the volatility process. Therefore,

we can write our posterior distribution as follows:-

$$\begin{aligned}\pi(\theta, v_{1:t}|y_{1:t}, v_0, y_0) &\propto \pi(\theta, y_{1:t}, v_{1:t}|v_0, y_0) \\ &= L(y_{1:t}, v_{1:t}|v_0, y_0, \theta)\pi(\theta),\end{aligned}$$

where we assume prior $\theta = (\mu, \alpha, \beta, \sigma)$ independence of y_0, v_0 so that $\pi(\theta|v_0, y_0) = \pi(\theta)$, $\pi(\theta, v_{1:t}|y_{1:t}, v_0, y_0)$ is the posterior distribution of the parameter vector θ and the $v_{1:t}$ processes,

$L(y_{1:t}, v_{1:t}|v_0, y_0, \theta)$ is the likelihood function of both the $y_{1:t}$ and $v_{1:t}$ processes, and

$\pi(\theta)$ is the prior distribution of θ.

Here our ultimate aim is to estimate the historical volatility v_t up to t and the parameters θ. Because of the complexity of the posterior distribution for the parameters and the volatility process, we understand it by sampling from it. We now describe three methods for sampling from this posterior distribution. We sometimes simplify notation by dropping unnecessary variables.

3.2.2.1 Markov chain Monte Carlo

We will apply a Markov chain Monte Carlo (MCMC) algorithm to sample from the posterior distribution of the parameters and the volatility process. In particular, we will use a popular method called the Metropolis-Hasting algorithm to sample from this posterior distribution.

The Metropolis-Hasting algorithm consists of several steps. In general, these steps involve simulating a candidate value from a proposal distribution, and then deciding whether or not to accept the candidate value. This algorithm can be described as follows:-

- We sample using the framework of the Gibbs sampler. More precisely, we update the parameters in the order $\mu, \alpha, \beta, \sigma$ and $v_{1:t}$. For example, when updating μ, we generate a candidate μ, called $\mu^{(\text{cand})}$, from a probability density function k,

- We calculate

$$R_1 = \frac{L(y_{1:t}, v_{1:t}|v_0, y_0, \theta^{(\text{cand})})\,\pi(\theta^{(\text{cand})})\,k(\theta)}{L(y_{1:t}, v_{1:t}|v_0, y_0, \theta)\,\pi(\theta)\,k(\theta^{(\text{cand})})},$$

where $\theta^{(\text{cand})} = (\mu^{(\text{cand})}, \alpha, \beta, \sigma)$, and we then accept $\mu^{(\text{cand})}$ as the new value of μ with probability $\alpha_1 = \min[R_1, 1]$.

- We update all the components of $v_{1:t}$ together by sampling $v_{1:t}^{(\text{cand})}$ from a probability density function g, calculating

$$R_2 = \frac{L\left(y_{1:t}, v_{1:t}^{(\text{cand})}|y_0, v_0, \theta\right)\,g(v_{1:t})}{L(y_{1:t}, v_{1:t}|y_0, v_0, \theta)\,g\left(v_{1:t}^{(\text{cand})}\right)},$$

and accepting $v_{1:t}^{(\text{cand})}$ as the new value of $v_{1:t}$ with probability $\alpha_2 = \min[R_2, 1]$.

Note that the probability density functions k and g can depend on previous values.

3.2.2.2 The Particle Filter assuming θ is known

Before starting our discussion of the particle filter, we will write the Heston model when $\rho = 0$ as a state space model

$$v_t = v_{t-1} + (\alpha - \beta v_{t-1})\triangle + \sigma\sqrt{|v_{t-1}|}\sqrt{\triangle}Z_t^v \quad \text{(unobserved equation)} \quad (3.7)$$

$$y_t = y_{t-1} + \left(\mu - \frac{1}{2}v_{t-1}\right)\triangle + \sqrt{|v_{t-1}|}\sqrt{\triangle}Z_t^s \quad \text{(observed equation)}, \quad (3.8)$$

where $t = 1, \ldots, T$.

The particle filter is an example of a sequential importance sampling algorithm and aims to produce a good estimate of the unobserved v_t from information about the observed process up to time t. It is used for filtering and smoothing. The main task of filtering is to update the target distribution every iteration when the distribution is computed approximately. The main task of smoothing is to sample from the joint distribution, i.e, $\pi(v_{1:t}|y_{1:t})$, and

approximate the associated marginals $\pi(v_k|y_{1:t})$ where $k = 1,\ldots,t$. The target distribution $\pi(v_{1:t}|y_{1:t})$ is approximated by means of particles $v_{1:t}^{(i)}$ with associated weights $w_t^{(i)}$.

When applying the particle filter, the parameters and initial values of each process are assumed known.

Loosely speaking, our approach follows that in Chapter 5 of Giovanni Petris, Sonia Petrone and Patrizia Campagnoli (2009) (Petris et al., 2009). The particle filter comprises two steps. The first step draws new observation $v_t^{(i)} i = 1,\ldots,N$, to obtain $v_{1:t}^{(i)}$, and the second step updates the weights $w_{t-1}^{(i)}$ to $w_t^{(i)}$.

Therefore, we can say that the particle filter updates a discrete approximation of $\pi_{t-1}(v_{1:t-1}|y_{1:t-1})$ when new data are available to produce a discrete approximation of $\pi_t(v_{1:t}|y_{1:t})$, where π_t represents the posterior distribution at time t.

Mathematically, our importance function is

$$g_t(v_{1:t}|y_{1:t}) = g_{t|t-1}(v_t|v_{1:t-1},y_{1:t})g_{t-1}(v_{1:t-1}|y_{1:t-1})$$

where $g_{t|t-1}$ is an importance transition function.

Now, the weight $w_t^{(i)}$ of the i^{th} particle, $i = 1,\ldots,N$, will be updated as follows:-

$$\begin{aligned}
\widehat{w}_t^{(i)} &\propto \frac{\pi_t(v_{1:t}^{(i)}|y_{1:t})}{g_t(v_{1:t}^{(i)}|y_{1:t})} \propto \frac{\pi_t(v_{1:t}^{(i)},y_t|y_{1:t-1})}{g_t(v_{1:t}^{(i)}|y_{1:t})} \\
&= \frac{\pi_t(v_t^{(i)},y_t|v_{1:t-1}^{(i)},y_{1:t-1})\pi_{t-1}(v_{1:t-1}^{(i)}|y_{1:t-1})}{g_{t|t-1}(v_t^{(i)}|v_{1:t-1}^{(i)},y_{1:t-1})g_{t-1}(v_{1:t-1}^{(i)}|y_{1:t-1})} \\
&\propto \frac{\pi_t(v_t^{(i)},y_t|v_{1:t-1}^{(i)},y_{1:t-1})}{g_{t|t-1}(v_t^{(i)}|v_{1:t-1}^{(i)},y_{1:t-1})} \times \widehat{w}_{t-1}^{(i)} \\
&= \frac{\pi_t(v_t^{(i)}|v_{t-1}^{(i)})\pi_t(y_t|y_{t-1},v_{t-1}^{(i)})}{g_{t|t-1}(v_t^{(i)}|v_{1:t-1}^{(i)},y_{1:t-1})} \times \widehat{w}_{t-1}^{(i)},
\end{aligned}$$

because to be changed, applied to v_t, y_t in equations (3.7) and (3.8) for example.

We finally normalized the weights

$$w_t^{(i)} = \frac{\widehat{w}_t^{(i)}}{\sum_{j=1}^N \widehat{w}_t^{(j)}}.$$

A problem that can occur with the normalized weights is that some values are very close to 1 or to 0; effectively such values are outliers. So, we will calculate the effective sample size to quantity and later overcome this problem

$$N_{eff} = \left[\sum_{i=1}^{N} (w_t^{(i)})^2 \right]^{-1}$$

– If $w_t^{(i)} = \frac{1}{N}$, then the weights are distributed evenly over the particles and N_{eff} takes its maximum value of N.

– If $w_t^{(i)} = 1$, for example, and $w_t^{(i)} = 0, i = 2, \ldots, N$, then $N_{eff} = 1$ as only one particle is important.

– If N_{eff} drops below a certain value N_0, then we re-sample the particles path $v_{1:t}^{(i)}, i = 1, \ldots, N$. The goal of the effective sample size is to keep the particle filter in control.

Let's give a short summary of the particle filter for the Heston model when $\rho = 0$:

- Initialize at time $t = 1$. Generate $v_1^{(i)} \sim g_1(v_1|v_0, y_0)$, where $g_1(v_1|v_0, y_0)$ is a probability density function chosen by the user, and set $w_0^{(i)} = N^{-1}, \quad i = 1, \ldots, N$.

- Generating, computing and normalizing at time $t = 2, \ldots, T$:

  - Generate $v_t^{(i)} \sim g_{t|t-1}(v_t|v_{1:t-1}^{(i)}, y_{1:t-1})$;

  - Update the weights

$$\widehat{w}_t^{(i)} = \frac{\pi(v_t^{(i)}|v_{t-1}^{(i)})\pi(y_t|y_{1:t-1}, v_{1:t-1}^{(i)})}{g_{t|t-1}(v_t^{(i)}|v_{1:t-1}^{(i)}, y_{1:t-1})} \times \widehat{w}_{t-1}^{(i-1)};$$

  - Normalize the weights

$$w_t^{(i)} = \frac{\widehat{w}_t^{(i)}}{\sum_{j=1}^{N} \widehat{w}_t^{(j)}}.$$

- Compute the effective sample size

$$N_{eff} = \left[\sum_{i=1}^{N} (w_t^{(i)})^2 \right]^{-1}.$$

- If $N_{eff} \leqslant N_0$, where N_0 is a certain preset fraction of the total number of particles, re-sample

    - Draw a sample of size N from a discrete distribution $P(v_{1:t} = v_{1:t}^{(i)}) = w_t^{(i)}, i = 1,\ldots,N$, and relabel this sample $v_{1:t}^{(1)},\ldots,v_{1:t}^{(N)}$.
    - Reset the weights:- $w_t^{(i)} = N^{-1}, i = 1,\ldots,N$.

3.2.2.3 Auxiliary Particle Filter with unknown parameters

The aim of the auxiliary particle filter is to overcome some of the problems associated with the particle filter by using an auxiliary variable to identify potentially good particles. In fact, these problems are concerned with the efficiency of the sample and the often poor quality of the particle filter approximation in the tails of the density of interest. The reason for the inefficiency of the sampler is that some of the weights may take very large values, known as outliers, leading to degeneracy.

An additional advantage of the auxiliary particle filter is that it allows us to sample from the posterior distribution of both the volatility process at time t and the parameters in (3.7) and (3.8) sequentially. We follow Chapter 5 of Giovanni Petris, Sonia Petrone and Patrizia Campagnoli (2009) (Petris et al., 2009).

Mathematically, at time $t-1$, a value of θ can be drawn from a continuous importance density based on a discrete approximate at time $t-1$:

$$\widehat{\pi}_{t-1}(v_{1:t-1},\theta) \approx \pi(v_{1:t-1},\theta|y_{1:t-1})$$

so that the distribution of θ is

$$\widehat{\pi}_{t-1}(\theta) \approx \pi(\theta|y_{1:t-1}).$$

We will assume that particles $\theta^{(i)}$, $i = 1,\ldots,N$, have been drawn with associated weights $w_{t-1}^{(i)}$.

We will now assume that the target distribution of θ, at time $t-1$, is a mixture of multivariate normal distributions with mean $\theta^{(i)}$ and variance matrix Λ, as suggested by

Lui and West (2001)

$$\widehat{\pi}_{t-1}(\theta) = \sum_{i=1}^{N} w_{t-1}^{(i)} N(\theta; \theta^{(i)}, \Lambda). \tag{3.9}$$

Under the continuous distribution (3.9), the mean vector $\overline{\theta}$ is

$$\begin{aligned} \overline{\theta} &= \mathrm{E}(\theta^{(I)}) \\ &= \sum_{i=1}^{N} w_{t-1}^{(i)} \theta^{(i)}, \end{aligned}$$

where I is an auxiliary variable, with $Pr(I = i) = w_{t-1}^{(i)}$, $i = 1, \ldots, N$, and the covariance matrix of θ is

$$\begin{aligned} \mathrm{var}(\theta) &= \mathrm{E}(\mathrm{var}(\theta|I)) + \mathrm{var}(\mathrm{E}(\theta|I)) \\ &= \Lambda + \Sigma > \Sigma \end{aligned} \tag{3.10}$$

where $\Sigma = \mathrm{var}(\theta^{(I)})$ is the covariance matrix of θ under the original discrete approximation $\hat{\pi}_{t-1}$, and the inequality holds since Λ is a positive definite matrix.

In order to reduce $\mathrm{var}(\theta)$ to its original Σ, we re-defined equation (3.9) at time $t-1$ as follows:-

$$\hat{\pi}_{t-1}(\theta) = \sum_{i=1}^{N} w_{t-1}^{(i)} N(\theta; m^{(i)}, h^2\Sigma)$$

where $m^{(i)} = a\theta^{(i)} + (1-a)\overline{\theta}$, in which $a \in (0,1)$, and $a^2 + h^2 = 1$.

We compute the new expected value and covariance matrix as follows

$$\begin{aligned} \mathrm{E}(\theta) &= \mathrm{E}(\mathrm{E}(\theta|I)) \\ &= a\overline{\theta} + (1-a)\overline{\theta} \\ &= \overline{\theta}; \\ \mathrm{var}(\theta) &= \mathrm{E}(\mathrm{var}(\theta|I)) + \mathrm{var}(\mathrm{E}(\theta|I)) \\ &= h^2\Sigma + a^2\mathrm{var}(\theta^{(I)}) \\ &= h^2\Sigma + a^2\Sigma \\ &= \Sigma. \end{aligned} \tag{3.11}$$

By comparing (3.10) and (3.11), it is obvious that (3.11) is better than (3.10) because the variance in (3.11) is smaller.

Therefore, the joint distribution for the discrete $v_{1:t-1}$ and the continuous θ takes the form

$$\hat{\pi}_{t-1}(v_{1:t-1}, \theta) = \sum_{t=1}^{N} w_{t-1}^{(i)} N(\theta; m^{(i)}, h^2\Sigma)\delta_{v_{1:t-1}^{(i)}}$$

where $\delta_{v_{1:t-1}^{(i)}}$ is the unit mass at $v_{1:t-1}^{(i)}$.

Now, the target distribution at time t will become

$$\begin{aligned} \pi(v_{1:t}, \theta|y_{1:t}) &\propto \pi(v_{1:t}, \theta, y_t|y_{1:t-1}) \\ &= \pi(v_t, y_t|v_{1:t-1}, y_{1:t-1}, \theta)\pi(v_{1:t-1}, \theta|y_{1:t-1}) \\ &\approx \pi(y_t|v_{t-1}, y_{t-1}, \theta)\pi(v_t|v_{t-1}, \theta)\hat{\pi}_{t-1}(v_{1:t-1}, \theta) \\ &= \sum_{i=1}^{N} w_{t-1}^{(i)} \pi(y_t|v_{t-1}^{(i)}, y_{t-1}, \theta)\pi(v_t|v_{t-1}^{(i)}, \theta)N(\theta; m^{(i)}, h^2\Sigma)\delta_{v_{1:t-1}^{(i)}}. \end{aligned}$$

The auxiliary particle filter algorithm for sequentially sampling from the posterior distribution of the volatility process and the parameters is therefore:-

- Initialize at time $t = 1$.

Generate $(v_1^{(1)},\ldots,v_1^{(N)}) \sim \pi^v(v_1)$ and $(\theta^{(1)},\ldots,\theta^{(N)}) \sim \pi^\theta(\theta)$, where $\pi^v(v_1)$ and $\pi^\theta(\theta)$ are prior probability density functions chosen by the user. Set $w_1^{(i)} = N^{-1}, i = 1,\ldots,N$.

- Computing and generating for $t = 2,\ldots,T$ and $i = 1,\ldots,N$:

  - Compute the mean of $\overline{\theta} = \mathrm{E}_{\hat{\pi}_{t-1}}(\theta)$ and the variance $\Sigma = \mathrm{var}_{\hat{\pi}_{t-1}}(\theta)$. The particle $\theta_t^{(i)}$ will depend on the new mean of θ :

$$m^{(i)} \quad = \quad a\theta^{(i)} + (1-a)\overline{\theta}.$$

  - Generate an auxiliary variable I_k for $k = 1,\ldots,N$ where

$$p(I_k = i) \propto w_{t-1}^{(i)} \pi(y_t | v_{t-1} = v_{t-1}^{(i)}, y_{t-1}, \theta = m^{(i)}).$$

  - Generate

$$\begin{aligned} \theta^{(k)} &\sim N(m^{(I_k)}, h^2\Sigma) \\ v_t^{(k)} &\sim \pi(v_t | v_{t-1} = v_{t-1}^{(I_k)}, \theta = \theta^{(k)}) \end{aligned}$$

    and set $v_{1:t}^{(k)} = (v_{1:t-1}^{(I_k)}, v_t^{(k)})$.

- Compute the weights

$$\widehat{w}_t^{(k)} = \frac{\pi(y_t | v_{t-1}^{(k)}, y_{t-1}, \theta = \theta^{(k)})}{\pi(y_t | v_{t-1}^{(I_k)}, y_{t-1}, \theta = m^{(I_k)})}$$

  and normalize them

$$w_t^{(i)} = \frac{\widehat{w}_t^{(i)}}{\sum_{j=1}^N \widehat{w}_t^{(j)}}.$$

- Compute the effective sample size N_{eff}.

  If $N_{eff} \leqslant N_0$, where N_0 is a certain preset fraction of the total number of particles, re-sample

- Draw a sample of size N from a discrete distribution $P((v_{1:t},\theta)=(v_{1:t}^{(i)},\theta^{(i)}))=w_t^{(i)}, i=1,\ldots,N$, and relabel this sample $(v_{1:t}^{(1)},\theta^{(1)}),\ldots,(v_{1:t}^{(N)},\theta^{(N)})$.

- Reset the weights:-

$$w_t^{(i)}=N^{-1},\quad i=1,\ldots,N.$$

3.3 The Heston Model with Correlation Coefficient ρ Between the Stochastic Processes

The Heston model including the correlation coefficient ρ can be written as follows:-

$$\frac{dS_t}{S_t}=\mu dt+\sqrt{v_t}\left(\rho dW_t^v+\sqrt{1-\rho^2}dW_t^s\right) \tag{3.12}$$

$$dv_t=(\alpha-\beta v_t)dt+\sigma\sqrt{v_t}dW_t^v, \tag{3.13}$$

where $\alpha=\beta\times\kappa$, in which κ is the long run volatility, and dW_t^s and dW_t^v are independent Brownian motion increments. Note that $\mathrm{E}\left[\rho dW_t^v+\sqrt{1-\rho^2}dW_t^s\right]=0$,
that $\mathrm{var}\left[\rho dW_t^v+\sqrt{1-\rho^2}dW_t^s\right]=dt$, and that $\mathrm{E}\left[\left(\rho dW_t^v+\sqrt{1-\rho^2}dW_t^s\right)dW_t^v\right]=\rho dt$, so that correlation $\left[\rho dW_t^v+\sqrt{1-\rho^2}dW_t^s,dW_t^v\right]=\rho$.
Our aim is to estimate the historical volatility v_t up to time t, and the parameters $\theta=(\mu,\alpha,\beta,\sigma,\rho)$.
The data that we will use will be the logarithms of the asset price process, i.e, $y_t=\log(S_t/S_0), t=1,\ldots,T$, where S_0 is the initial asset price value.
As we have seen in Section 3.2, an easier way of dealing with the Heston model is to use discrete time.
Therefore, the Euler scheme of the above Heston model with non-zero ρ is

$$y_{t+\triangle}=y_t+\left(\mu-\frac{1}{2}v_t\right)\triangle+\sqrt{|v_t|}\sqrt{\triangle}\left(\rho Z_t^v+\sqrt{1-\rho^2}Z_t^s\right)$$

$$v_{t+\triangle}=v_t+(\alpha-\beta v_t)\triangle+\sigma\sqrt{|v_t|}\sqrt{\triangle}Z_t^v,$$

in which $Z_t^s,Z_t^v\sim N(0,1)$ independently.

3.3.1 Maximum Likelihood Estimation

In this more general case, the likelihood function will include the correlation coefficient ρ. So, to obtain the likelihood function of the Heston model with non-zero ρ, we begin

$$\begin{aligned} L((y_t,v_t),(y_{t-1},v_{t-1}),\ldots,(y_1,v_1)|(y_0,v_0),\theta) &= L((y_t,v_t)|(y_{t-1},v_{t-1}),\ldots,(y_1,v_1),(y_0,v_0),\theta)\times \\ & L((y_{t-1},v_{t-1})|(y_{t-2},v_{t-2}),\ldots,(y_1,v_1),(y_0,v_0),\theta) \\ & \times\cdots\times L((y_2,v_2)|(y_1,v_1),(y_0,v_0),\theta) \\ & \times L((y_1,v_1)|(y_0,v_0),\theta) \end{aligned}$$

where $\theta=(\mu,\alpha,\beta,\sigma,\rho)$ as above.

Thus, letting

$$\mu_t=\begin{pmatrix} E[y_t] \\ E[v_t] \end{pmatrix} \text{ and } \Sigma_t=\begin{pmatrix} \sigma_{y_t}^2 & \rho\sigma_{y_t}\sigma_{v_t} \\ \rho\sigma_{y_t}\sigma_{v_t} & \sigma_{v_t}^2 \end{pmatrix},$$

in which $\sigma_{y_t}^2$ and $\sigma_{v_t}^2$ are the conditional variances of y_t and v_t as we shall shortly specify, we have

$$\begin{aligned} L((y_t,v_t),(y_{t-1},v_{t-1}),\ldots,(y_1,v_1)|(y_0,v_0),\theta) &= N\left(\begin{pmatrix} y_t \\ v_t \end{pmatrix};\mu_t,\Sigma_t\right)\times \\ & N\left(\begin{pmatrix} y_{t-1} \\ v_{t-1} \end{pmatrix};\mu_{t-1},\Sigma_{t-1}\right)\times\cdots\times \\ & N\left(\begin{pmatrix} y_1 \\ v_1 \end{pmatrix};\mu_1,\Sigma_1\right) \end{aligned} \tag{3.14}$$

in which $N(x;M,S)$ is the bivariate normal probability density function with 2×1 mean M and 2×2 variance covariance matrix S evaluated at the bivariate x.

In general, for $t = 1, \ldots, T$, we have

$$\begin{aligned} E(y_t) &= y_{t-1} + \triangle\left(\mu - \frac{1}{2}v_{t-1}\right) \\ E(v_t) &= v_{t-1} + \triangle(\alpha - \beta v_{t-1}) \\ \sigma_{y_t} &= \sqrt{|v_{t-1}|}\sqrt{\triangle} \\ \sigma_{v_t} &= \sigma\sqrt{|v_{t-1}|}\sqrt{\triangle} \end{aligned}$$

We can therefore estimate the parameters and volatility process using maximum likelihood estimation.

3.3.2 Bayesian Inference for the Heston Model with Correlation Coefficient ρ

Bayes theorem is a very important tool in statistical inference. Briefly, Bayes theorem updates previous information about the parameters and volatility process, called prior information, to obtain the current information about the parameters and volatility process, called posterior information, as we explained in Section 3.2.2, and 2; see (Lunn et al., 2013).

The difference here is to extend our previous work to include the correlation coefficient ρ in our model.

So, we can represent the posterior distribution as follows:-

$$\pi(\theta, v_{1:t}|y_{1:t}, v_0, y_0) \propto L(y_{1:t}, v_{1:t}|v_0, y_0, \theta)\pi(\theta)$$

where

$\pi(\theta, v_{1:t}|y_{1:t}, v_0, y_0)$ is the posterior distribution of the parameters θ and the $v_{1:t}$ process,

$L(y_{1:t}, v_{1:t}|y_0, v_0, \theta)$ is the likelihood function of the parameters defined through equation (3.14),

$\pi(\theta)$ is the prior distribution of the parameters; this is similar to what we saw in Section 3.2.2.

We will use three methods to sample from the posterior distribution of the parameters and the volatility process as follows:-

3.3.2.1 Markov chain Monte Carlo

As we explained in Section 3.2.2.1, in general, the Markov chain Monte Carlo algorithm is as follows:-

- Sample $\theta^{(\text{cand})} \sim k(\theta^{(\text{cand})})$.

- Sample $v_{1:t}^{(\text{cand})} \sim g(v_{1:t}^{(\text{cand})})$.

- Compute the following quantities

$$\begin{aligned}
\alpha_1 &= \min\left[1, \frac{L\left(y_{1:t}, v_{1:t}|y_0, v_0, \theta^{(\text{cand})}\right)\pi\left(\theta^{(\text{cand})}\right)k(\theta)}{L(y_{1:t}, v_{1:t}|y_0, v_0, \theta)\,\pi(\theta)\,k(\theta^{(\text{cand})})}\right], \\
\alpha_2 &= \min\left[1, \frac{L\left(y_{1:t}, v_{1:t}^{(\text{cand})}|y_0, v_0, \theta\right)g(v_{1:t})}{L(y_{1:t}, v_{1:t}|y_0, v_0, \theta)\,g\left(v_{1:t}^{(\text{cand})}\right)}\right]
\end{aligned}$$

- Set $\theta = \theta^{(\text{cand})}$ or $v_{1:t} = v^{(\text{cand})}$ with probability α_1 and α_2 respectively; otherwise the values of θ or $v_{1:t}$ do not change.

3.3.2.2 The Particle Filter

We will set up the Heston model with non-zero ρ as a state space model taking the following form

$$v_t = v_{t-1} + (\alpha - \beta v_{t-1}) \triangle + \sigma\sqrt{|v_{t-1}|}\sqrt{\triangle}Z_t^v \qquad \text{(unobserved equation)} \tag{3.15}$$

$$y_t = y_{t-1} + \left(\mu - \frac{1}{2}v_{t-1}\right)\triangle + \sqrt{|v_{t-1}|}\sqrt{\triangle}\left(\rho Z_t^v + \sqrt{1-\rho^2}Z_t^s\right) \text{(observed equation)} \tag{3.16}$$

where $t = 1, \ldots, T$

In general, the particle filter is an extension of the Monte Carlo method into the two elements of filtering and smoothing. The aim of filtering is to find a good estimation of the

volatility process and the parameters in the Heston model with non-zero ρ sequentially. The aim of smoothing is to sample from the joint distribution and then approximate the marginals of the volatility process and the parameters at time t. Again, we follow Chapter 5 of Giovanni Petris, Sonia Petrone and Patrizia Campagnoli (2009) (Petris et al., 2009).

Following Section 3.2.2.2, the difference in this case is that the weight $w_t^{(i)}$ of the i^{th} particle will be updated as follows:-

$$\begin{aligned}
\widehat{w}_t^{(i)} &\propto \frac{\pi(v_{1:t}^{(i)}|y_{1:t})}{g_t(v_{1:t}^{(i)}|y_{1:t})} \propto \frac{\pi(v_{1:t}^{(i)}, y_t|y_{1:t-1})}{g_t(v_{1:t}^{(i)}|y_{1:t})} \\
&= \frac{\pi(v_t^{(i)}, y_t|v_{1:t-1}^{(i)}, y_{1:t-1})\pi(v_{1:t-1}^{(i)}|y_{1:t-1})}{g_{t|t-1}(v_t^{(i)}|v_{1:t-1}^{(i)}, y_{1:t-1})g_{t-1}(v_{1:t-1}^{(i)}|y_{1:t-1})} \\
&\propto \frac{\pi(v_t^{(i)}, y_t|v_{1:t-1}^{(i)}, y_{1:t-1})}{g_{t|t-1}(v_t^{(i)}|v_{1:t-1}^{(i)}, y_{1:t-1})} \times \widehat{w}_{t-1}^{(i)} \\
&= \frac{L(v_{1:t}^{(i)}, y_{1:t})}{L(v_{1:t-1}^{(i)}, y_{1:t-1})g_{t|t-1}(v_t^{(i)}|v_{1:t-1}^{(i)}, y_{1:t-1})} \times \widehat{w}_{t-1}^{(i)}
\end{aligned}$$

where

$g_{t|t-1}$ is an importance transition function,

$L(v_{1:t}^{(i)}, y_{1:t})$ is the likelihood function at time t derived in Section 3.3.1

$L(v_{1:t-1}^{(i)}, y_{1:t-1})$ is the equivalent likelihood function at time $t-1$.

Finally normalize the weights

$$w_t^{(i)} = \frac{\widehat{w}_t^{(i)}}{\sum_{j=1}^{N} \widehat{w}_t^{(j)}}.$$

Therefore, the algorithm of the particle filter for the Heston model with non-zero ρ is

- At time $t = 1$,
  generate $v_1^{(i)} \sim g_1(v_1|v_0, y_0, \theta)$, where $g_1(v_1|v_0, y_0, \theta)$ is a probability density function chosen by the user, and set $w_1^{(i)} = N^{-1}, \qquad i = 1, \ldots, N.$

- At time $t = 2, \ldots, T$,

  - Generate $v_t^{(i)} \sim g_{t|t-1}(v_t|v_{1:t-1}^{(i)}, y_{1:t-1}, \theta)$;

- Update the weights

$$\widehat{w}_t^{(i)} = \frac{L(v_{1:t}^{(i)}, y_{1:t}|v_0, y_0, \theta)}{L(v_{1:t-1}^{(i)}, y_{1:t-1}|v_0, y_0, \theta) g_{t|t-1}(v_t^{(i)}|v_{1:t-1}^{(i)}, y_{1:t-1}, v_0, y_0, \theta)} \times \widehat{w}_{t-1}^{(i-1)};$$

- Normalize the weights

$$w_t^{(i)} = \frac{\widehat{w}_t^{(i)}}{\sum_{j=1}^{N} \widehat{w}_t^{(j)}}.$$

- Compute the effective sample size N_{eff}

$$N_{eff} = \left[\sum_{i=1}^{N} (w_t^{(i)})^2 \right]^{-1}.$$

- if $N_{eff} \leqslant N_0$, where N_0 is a certain preset fraction of the total number of particles, re-sample

  - Draw a sample of size N from a discrete distribution $P(v_{1:t} = v_{1:t}^{(i)}) = w_t^{(i)}, i = 1, \ldots, N$, and relabel this sample $v_{1:t}^{(1)}, \ldots, v_{1:t}^{(N)}$.

  - Reset the weights:- $w_t^{(i)} = N^{-1}, i = 1, \ldots, N$.

3.3.2.3 Auxiliary Particle Filter with unknown parameters

If we follow the treatment of Section 3.2.2.3, it turns out that the difference is that the weights depend on the likelihood function as shown in equation (3.14). Then, the algorithm for the auxiliary particle particle of the Heston model with non-zero ρ is:-

- Initialize at time $t = 1$

  generate $(v_1^{(1)}, \ldots, v_1^{(N)}) \sim \pi^v(v_1)$ and $(\theta^{(1)}, \ldots, \theta^{(N)}) \sim \pi^\theta(\theta)$, where $\pi^v(v_1)$ and $\pi^\theta(\theta)$ are prior probability density functions chosen by the user. Set $w_1^{(i)} = N^{-1}, i = 1, \ldots, N$.

- Computing and generating at time $t = 2, \ldots, T$ and $i = 1, \ldots, N$:

  - Compute the mean of the $\theta^{(i)}$

$$m^{(i)} = a\theta^{(i)} + (1-a)\overline{\theta}$$

where $\overline{\theta} = \mathrm{E}_{\hat{\pi}_{t-1}}(\theta)$.

- Generate an auxiliary variable I_k for $k = 1, \ldots, N$, where

$$p(I_k = i) \propto w_{t-1}^{(i)} \pi(y_t | v_{t-1} = v_{t-1}^{(i)}, y_{t-1}, \theta = m^{(i)}).$$

- Generate

$$\begin{aligned} \theta^{(k)} &\sim N(m^{(I_k)}, h^2 \Sigma) \\ v_t^{(k)} &\sim \pi(v_t | v_{t-1} = v_{t-1}^{(I_k)}, \theta = \theta^{(k)}) \end{aligned}$$

and set $v_{1:t}^{(k)} = (v_{1:t-1}^{(I_k)}, v_t^{(k)})$.

- Compute the weights

$$\widehat{w}_t^{(i)} = \frac{L(v_{1:t}^{(k)}, y_{1:t} | y_0, v_0, \theta = \theta^{(k)})}{L(v_{1:t-1}^{(k)}, y_{1:t-1} | y_0, v_0, \theta = \theta^{(k)}) \, \pi(y_t | v_{t-1}^{(I_k)}, y_{t-1}, \theta = m^{(I_k)}) \pi(v_t^{(i)} | v_{t-1}^{(I_k)}, \theta = m^{(I_k)})}$$

and normalize them

$$w_t^{(i)} = \frac{\widehat{w}_t^{(i)}}{\sum_{j=1}^{N} \widehat{w}_t^{(j)}}.$$

- Compute the effective sample size N_{eff}.

If $N_{eff} \leqslant N_0$, re-sample

- Draw a sample of size N from a discrete distribution $P((v_{1:t}, \theta) = (v_{1:t}^{(i)}, \theta^{(i)})) = w_t^{(i)}, i = 1, \ldots, N$, and relabel this sample $(v_{1:t}^{(1)}, \theta^{(1)}), \ldots, (v_{1:t}^{(N)}, \theta^{(N)})$.

- Reset the weights:-

$$w_t^{(i)} = N^{-1}, \quad i = 1, \ldots, N.$$

3.4 Results

We apply our MCMC and particle filter based methodology to simulated data in Section 3.4.1 and to real data in Section 3.4.2.

3.4.1 Applications to Simulated Data

3.4.1.1 The Heston Model when $\rho = 0$

We will begin by estimating the parameters $\theta = (\mu, \alpha, \beta, \sigma)$ of the Heston model with $\rho = 0$ as defined through (3.5) and (3.6) using maximum likelihood estimation as explained in Section 3.2.1.

We begin by considering a simulated dataset of size $T = 525$ with

$$\theta = (\mu, \alpha, \beta, \sigma) = (0.1, 0.2497, 11.35, 0.618).$$

The initial values of the asset price process S_0 and volatility process v_0 are 1 and 0.04, respectively. We used `R`'s function `optim` to implement maximum likelihood estimation. Figure 3.1 shows the estimation of the parameters θ using maximum likelihood estimation with approximate 95% confidence intervals. The parameters in θ are clearly well estimated. Figure 3.2 shows the estimation of the volatility process at a particular time using maximum likelihood estimation with an approximate 95% confidence interval. Again, good estimation of the volatility process is achieved.

Next, we applied our MCMC methodology for estimating the parameters θ and the volatility process as explained in Section 3.2.2.1. We assume that the prior distributions for μ, α and β are log-normal and that the prior distribution of σ is inverse gamma. We also assume that the volatility process is known.

Figure 3.3 represents the results form our MCMC algorithm for sampling from the posterior distribution of $\theta = (\mu, \alpha, \beta, \sigma)$ assuming that the volatility process is known. The first column shows time series plots of sampled values $\theta^{(t)}$, where the superscript represents MCMC iteration number. The second column shows the histogram of these parameter values. In general, we can say that the results from the MCMC algorithm provide a good

estimation for the parameters because the black and red lines are closer to each other especially in the cases of μ, α and β, meaning that the true values of these parameters are close to posterior means which therefore provide good estimates of the parameters.

Now we assume that the volatility process is unknown and needs to be estimated. Figure 3.4 represents the results from our MCMC algorithm applied to the posterior distribution of θ and the volatility process. The first column shows the time series plots of μ, β and the middle value of the volatility process. The second column shows the associated histograms. The third column shows the time series plots of α, σ and value of the volatility at the three quarter time point. The fourth column shows the associated histograms. When we assume that the volatility process is unknown, our results are less good, and values of the volatility process itself can be poorly estimated, although the other parameters can be well estimated.

Next, we applied the particle filter to the Heston model when $\rho = 0$, as explained in Section 3.2.2.2. The true values of parameters $\theta = (\mu, \alpha, \beta, \sigma)$, and the initial values of each process and the number of observations for the Heston model remain as given above. Figure 3.5 shows the results for the volatility process assuming known θ parameters. In particular, the true volatility process, the approximate posterior mean of the volatility process and approximate 90% credible intervals are shown across time t. In general, the posterior mean almost follows the true volatility process. This indicates that we can get a good estimation of the volatility process using the particle filter when the parameters are known.

Figure 3.6 shows the estimation of the parameters $\theta = (\mu, \alpha, \beta, \sigma)$ across time using the auxiliary particle filter, assuming that the volatility process is unknown. In particular, the posterior mean and approximate 90% credible intervals for each parameter are presented together. In general, good results are achieved using the auxiliary particle filter, although it turns out to be more difficult to make inference about σ than about the other parameters. Figure 3.7 shows the results of the auxiliary particle filter for estimating the volatility process v_t, assuming that the parameters are unknown. In particular, we see the volatility process itself, the approximate posterior mean of the volatility process and associated

approximate 90% credible intervals. Again, good estimation of the volatility process is achieved using the auxiliary particle filter when the parameters are unknown.

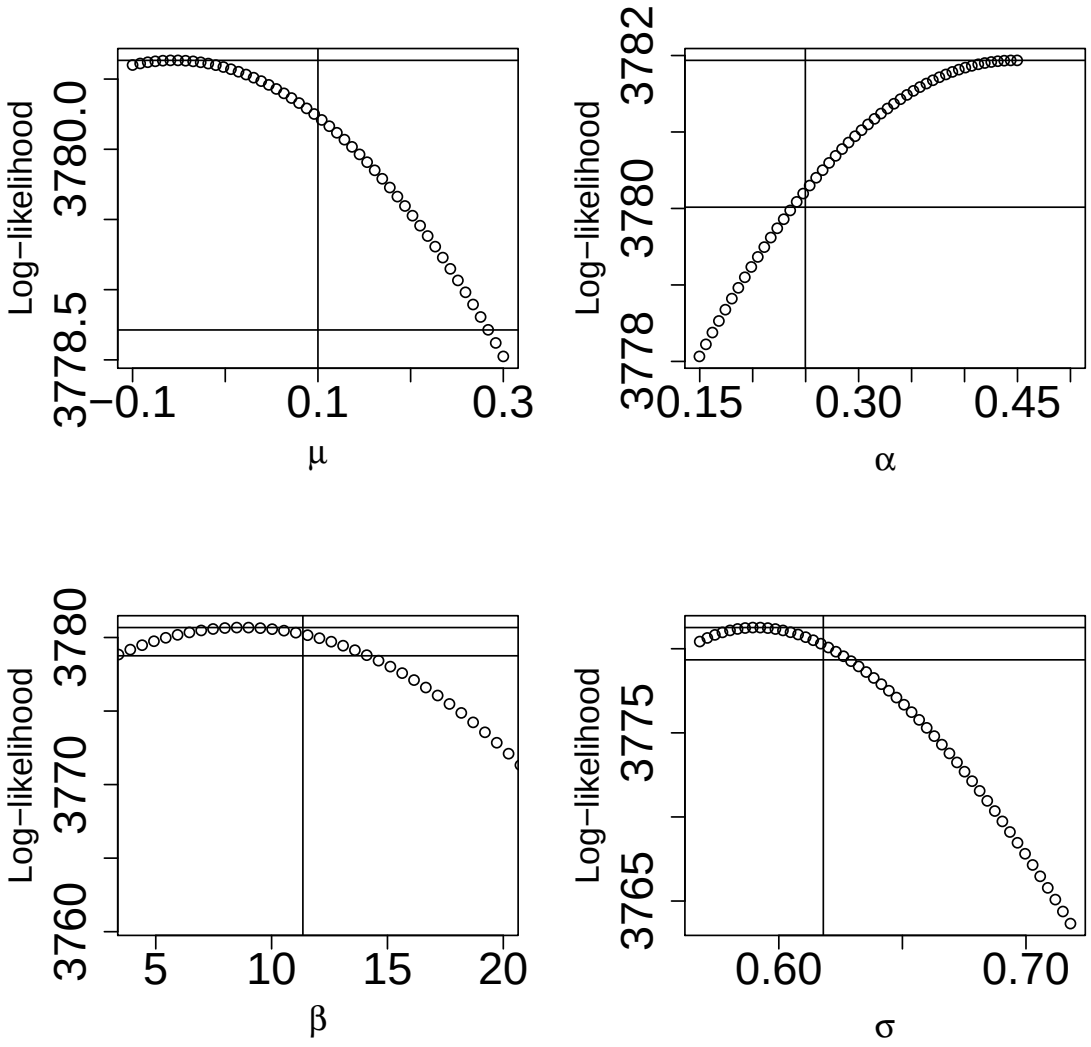


Figure 3.1: Profile log-likelihood function for the parameters θ based on likelihood function derived in Section 3.2.1. The lower horizontal line provides an approximate 95% confidence interval. In all cases the true parameter value shown by the vertical line is consistent with this approximate 95% confidence interval.

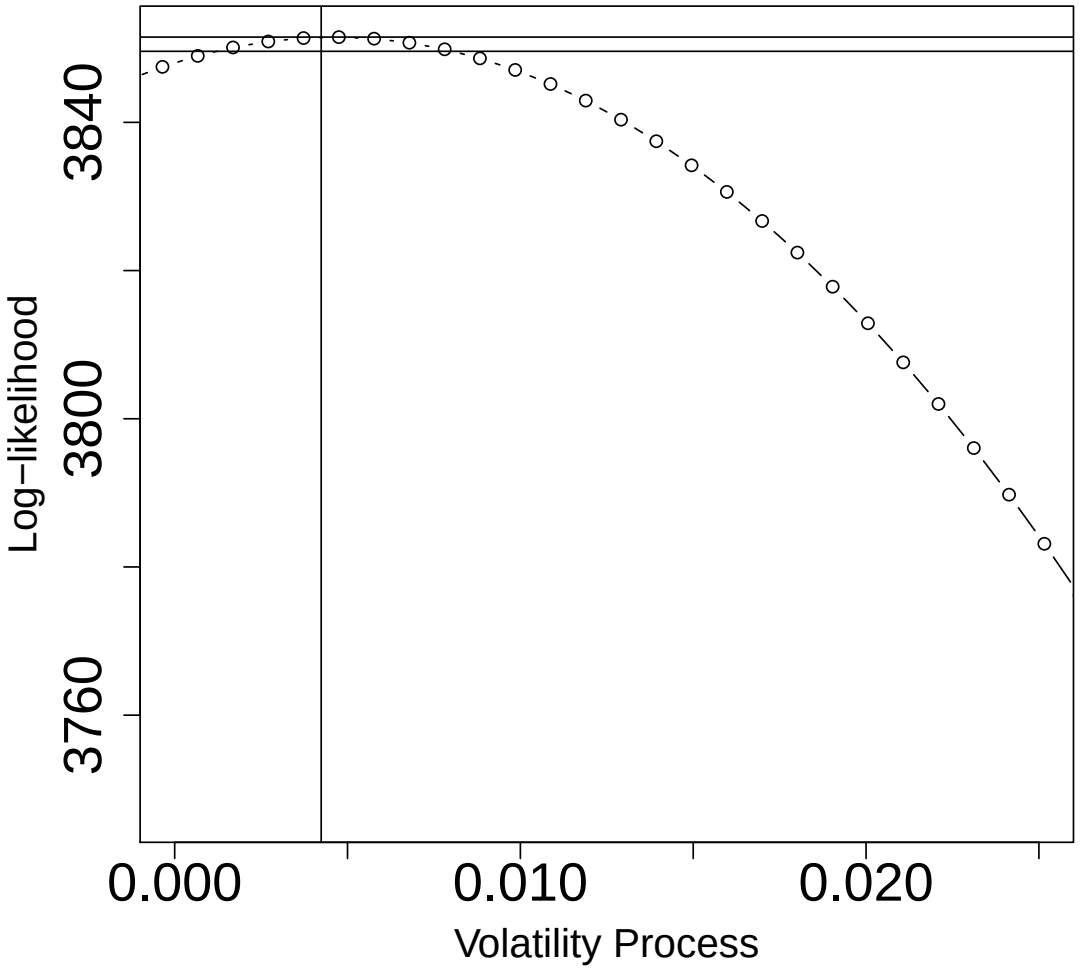


Figure 3.2: Profile log-likelihood function for the volatility process v_t at a particular time t based on the likelihood function derived in Section 3.2.1. The lower horizontal line indicates an approximate 95% confidence interval. The true volatility value shown by the vertical line is consistent with this approximate 95% confidence interval.

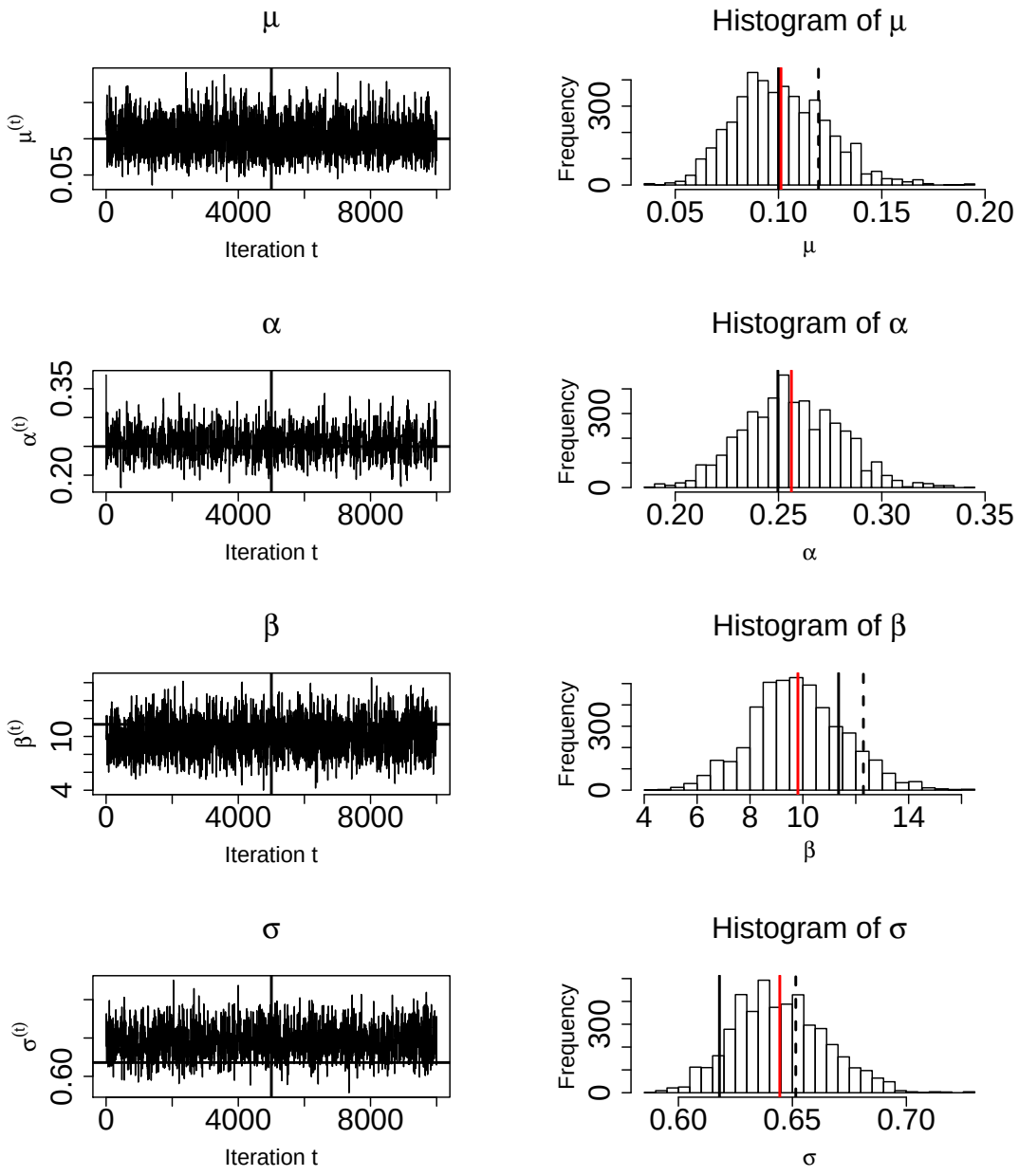


Figure 3.3: Graphs showing simulations from the posterior distribution of the parameters of the Heston model when $\rho = 0$ using the MCMC algorithm when the number of MCMC iterations $N = 10000$. The first column shows time series plots of sampled value of $\theta = (\mu, \alpha, \beta, \sigma)$. The second column shows the histogram of the sampled parameter values after burn–in. The vertical lines in the time series plots separates the burn-in phase (left) from samples that are used for the future inference (right). The horizontal lines in the times series plots and the black, red and dashed lines in the histograms show the true parameter, posterior mean and maximum likelihood values, respectively, where they can be shown on the scale.

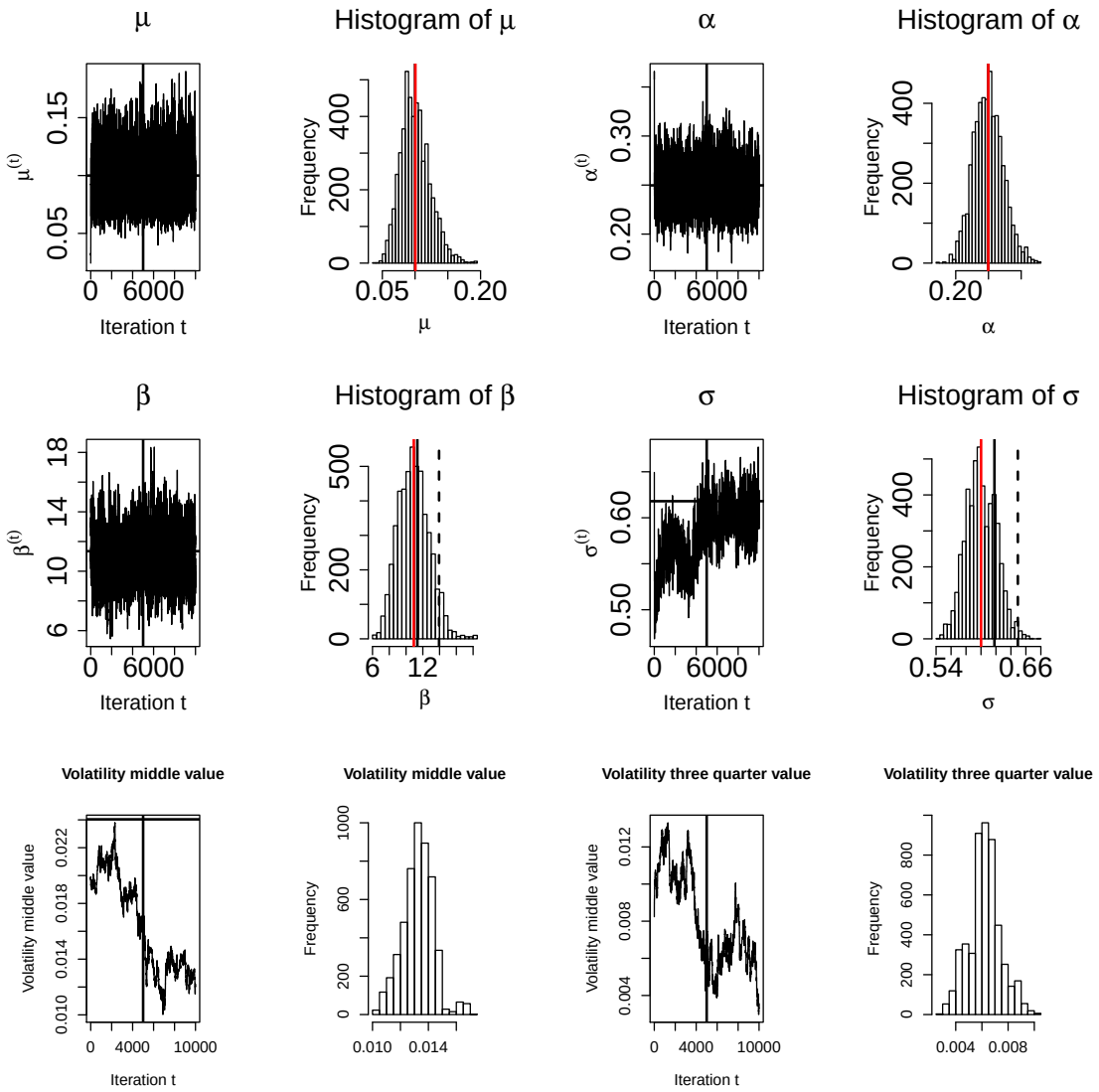


Figure 3.4: Graphs showing simulations from the posterior distribution of the parameters and underlying volatility process of the Heston model when $\rho = 0$ using the MCMC algorithm when the number of MCMC iteration $N = 10000$. The first and third columns show time series plots of sampled values of μ, β, the volatility middle value, α, σ, and volatility three quarter value, respectively. The second and fourth columns show the histogram of the sampled parameters and volatility middle and three quarter values, respectively, after burn–in. The vertical line in the time series plots separates the burn-in phase (left) from samples that are used for the future inference (right). The horizontal lines in the times series plots, and the black, red and dashed lines in the histograms show the true parameter, posterior mean and maximum likelihood values, respectively, where appropriate. In the histograms for μ and α the black and grey lines lie on top of each other.

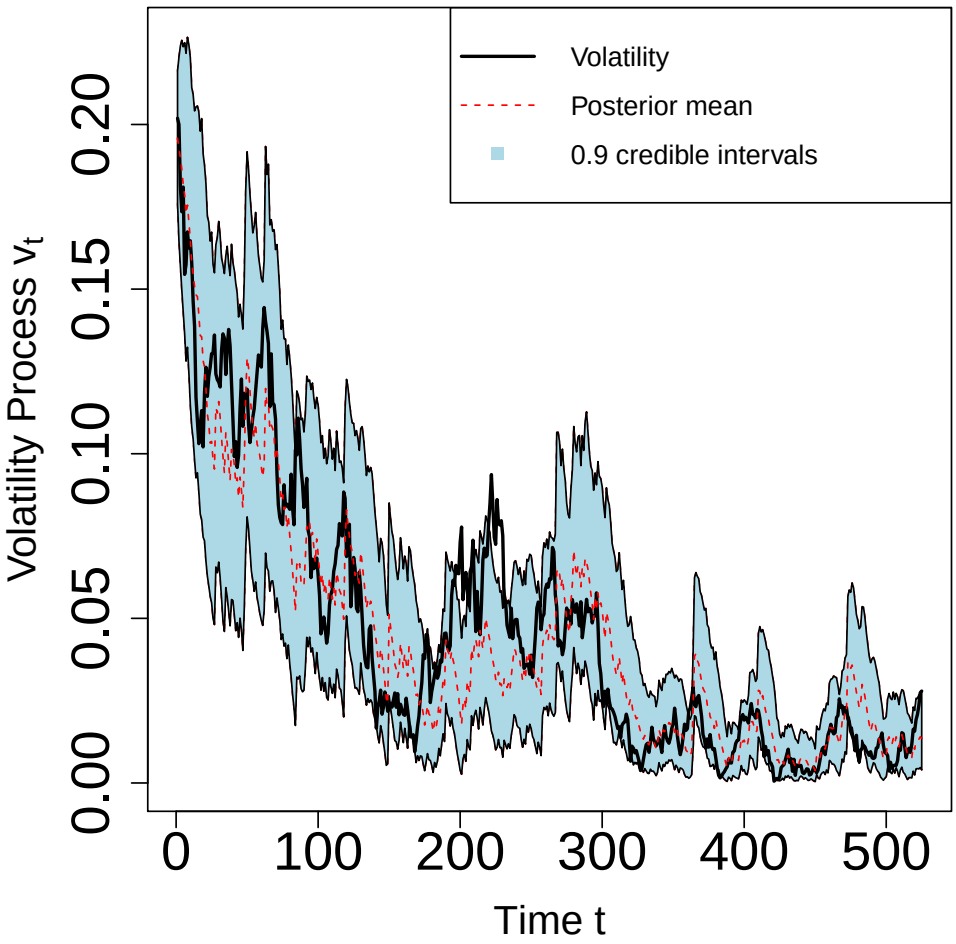


Figure 3.5: Estimation using the particle filter of the volatility process v_t of the Heston model when $\rho = 0$ assuming that the other parameters are known. The true volatility is represented by the black line. The posterior mean of the volatility process is indicated by the dashed red line and the associated 90% credible intervals are in light blue.

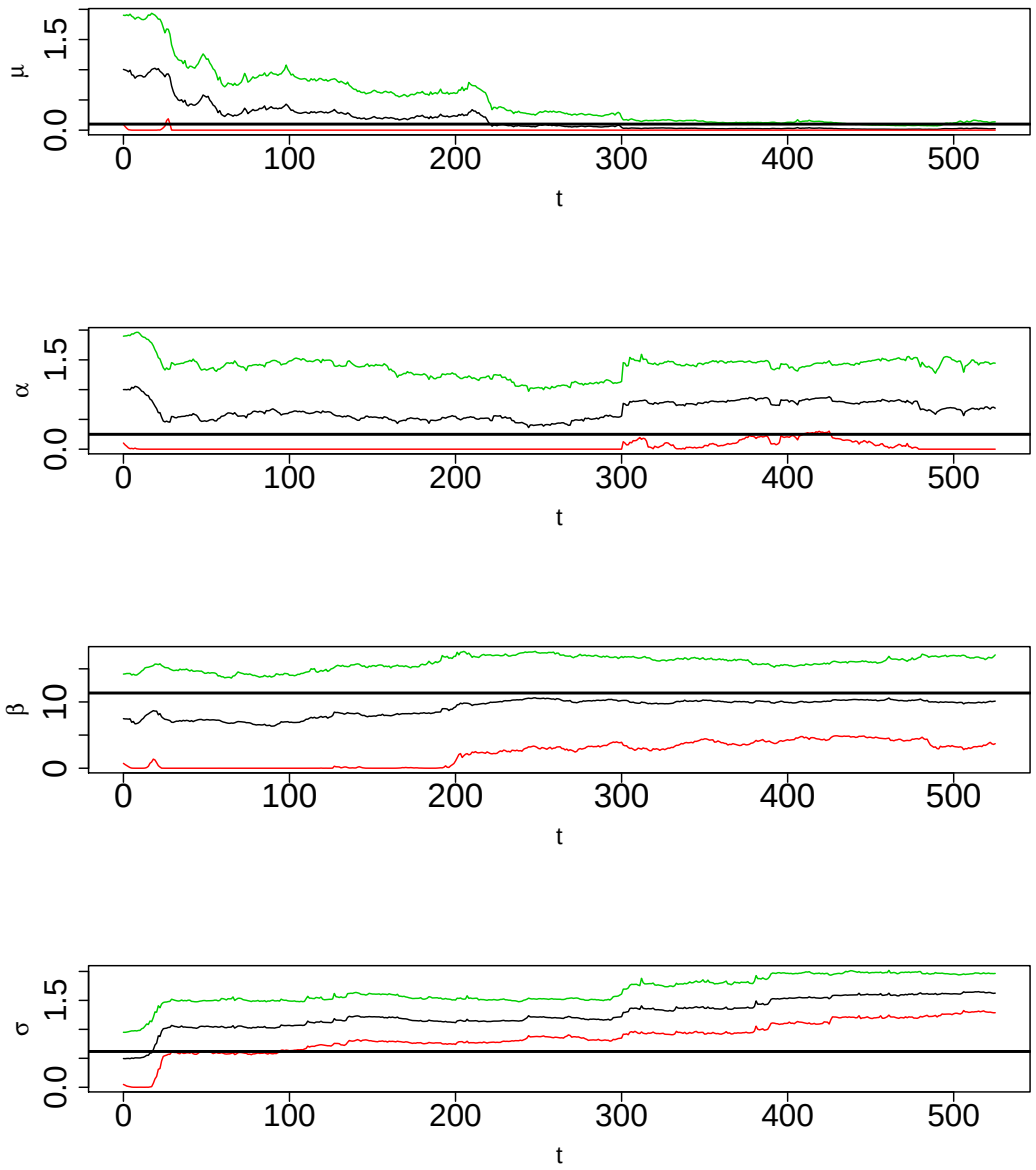


Figure 3.6: Estimation of the parameters of the Heston model when $\rho = 0$ using the auxiliary particle filter. In each graph, the posterior mean is shown using the black trace, while the associated approximate 90% credible intervals are shown by the outer traces. The first graph is for the parameter μ, the second graph is for the parameter α, the third graph is for the parameter β and the final graph is for the parameter σ. The true values are shown by the horizontal lines.

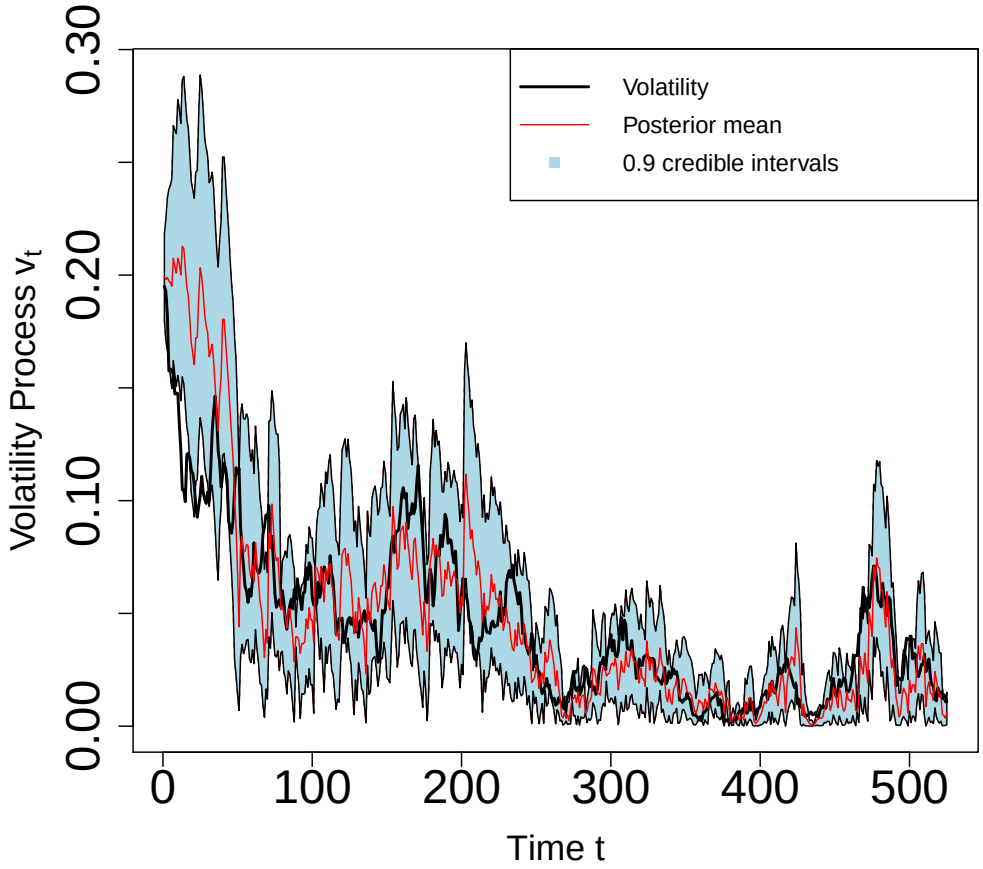


Figure 3.7: Estimation using the auxiliary particle filter of the volatility process v_t in the Heston model when $\rho = 0$ when the other parameters are unknown. The posterior mean of the volatility process is represented by the red line, and the associated 90% credible interval are in light blue.

3.4.1.2 The Heston Model with ρ Defined in Equations (3.12) and (3.13) of Section 3.3

We will illustrate the results of the Heston Model including the correlation coefficient ρ between the asset price and volatility processes. The aim here is to estimate the parameters $\theta = (\mu, \alpha, \beta, \sigma, \rho)$ and the volatility process v_t using the MCMC method and particle filter based methodology.

We consider simulated data sets of size $T = 525$ with

$\theta = (\mu, \alpha, \beta, \sigma, \rho) = (0.1, 0.2497, 11.35, 0.618, 0.2)$, or $(0.1, 0.2497, 11.35, 0.618, -0.2)$, so that we consider the cases when the correlation coefficient ρ takes either positive or negative values. The initial values of the asset price process and volatility process remain $S_0 = 1$ and $v_0 = 0.04$, respectively.

Figure 3.8 shows the estimation of the parameters θ in the Heston model with non-zero ρ using maximum likelihood with approximate 95% confidence intervals. The estimation of the parameters in θ is clearly good.

Figure 3.9 shows the estimation of the volatility process v_t at a particular time t in the Heston model with non-zero ρ using maximum likelihood estimation with an approximate 95% confidence interval, as discussed in Section 3.3.1. The estimation of the volatility process v_t is clearly good.

Next, we applied our MCMC methodology for estimating the parameters and the volatility process as explained in Section 3.3.2.1. We assume that the prior distribution for μ, α, β and σ are log-normal and the prior distribution of ρ is a uniform distribution. In our case, if $\rho = -0.2$, then the prior distribution is taken to be between $(-0.3, 0.2)$ and if $\rho = 0.2$, then the prior distribution is taken to be between $(-0.1, 0.3)$.

Figure 3.10 and Figure 3.11 show the results for the parameters θ in the Heston model with negative and positive ρ values, respectively. We are sampling from the posterior distribution of the parameters θ assuming that the volatility process is known. In each figure, we show the time series plots and the histograms for the parameters. We can say that the estimation of the parameters in both cases is acceptable.

Figure 3.12 and Figure 3.13 show the results from our MCMC algorithm for sampling from the posterior distribution of θ and the volatility process when the parameter ρ takes

the given negative and the given positive values, respectively. In each figure, we show the time series plots and the histograms for the parameters and the middle and three quarter values of the volatility process. In this case, the results are unsurprisingly less good than the previous case when the volatility is known.

Figure 3.14 and Figure 3.15 show the results of the particle filter for the Heston model with the true values of $\rho = -0.2$ and $\rho = 0.2$, respectively, assuming known parameters as discussed in Section 3.3.2.2. Here, we consider priors for ρ which are uniform $[-0.3, -0.1]$ and $[0.1, 0.3]$. In particular, the true volatility process, the approximate posterior mean of the volatility process and approximate 90% credible intervals are shown across time t. Here, the volatility process is generally well estimated, although estimation when $\rho = 0.2$ is better than when $\rho = -0.2$ because the volatility process in this case is sometimes outside the 90% credible intervals.

Figure 3.16 shows the estimation of the parameters θ of the Heston model with the given negative value of ρ across the time using the auxiliary particle filter as discussed in Section 3.3.2.3. In particular, the posterior mean and approximate 90% credible intervals of each parameters are presented together. The true value of ρ is -0.2. Here, estimation is less good than for the $\rho = 0$ case for all parameters except σ.

Figure 3.17 shows the results of the auxiliary particle filter for estimating the volatility process assuming that the θ parameters are unknown and have to be estimated. Here, the true value of $\rho = -0.2$. In particular, we see the volatility process itself, the approximate posterior mean of the volatility process and associated approximate 90% credible intervals. The volatility process seems well estimated.

Figure 3.18 shows the estimation of the parameters θ of the Heston model with the given positive value of ρ across the time using the auxiliary particle filter as discussed in Section 3.3.2.3. In particular, the posterior mean and approximate 90% credible intervals of each parameter are presented together. The true value of ρ is 0.2. Here, better estimation is achieved than previously when $\rho = -0.2$.

Figure 3.19 shows the results of the auxiliary particle filter for estimating the volatility process assuming that the θ parameters are unknown and have to be estimated. Here, the

true value of $\rho = 0.2$. In particular, we see the volatility process itself, the approximate posterior mean of the volatility process and associated approximate 90% credible intervals. Good estimation is achieved.

We performed other simulation studies with different true values of ρ, for example $\rho = -0.6$ and $\rho = 0.6$ similar results to the experiments when $\rho = -0.2$ and $\rho = 0.2$ were obtained, indicating that it is difficult to estimate the parameter ρ.

Overall, in this small simulation study the results for the Heston model with the given positive value of ρ are somewhat better than the results for the Heston model with the given negative ρ.

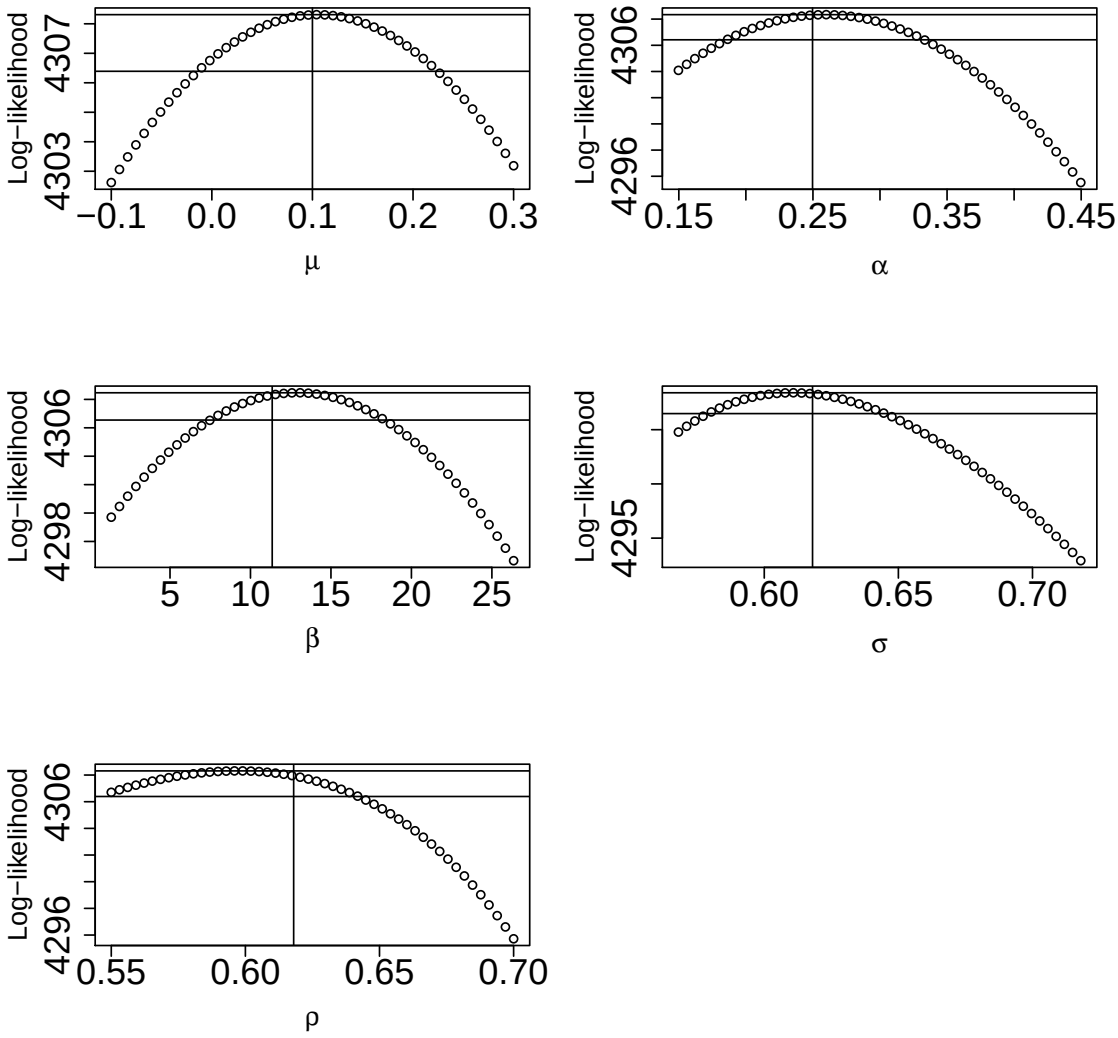


Figure 3.8: Profile log-likelihood function for the parameters θ of the Heston model with non-zero ρ, based on likelihood function derived Section 3.3.1. The lower horizontal line provides an approximate 95% confidence interval. In all cases the true parameter value shown by the vertical line is consistent with this approximate 95% confidence interval.

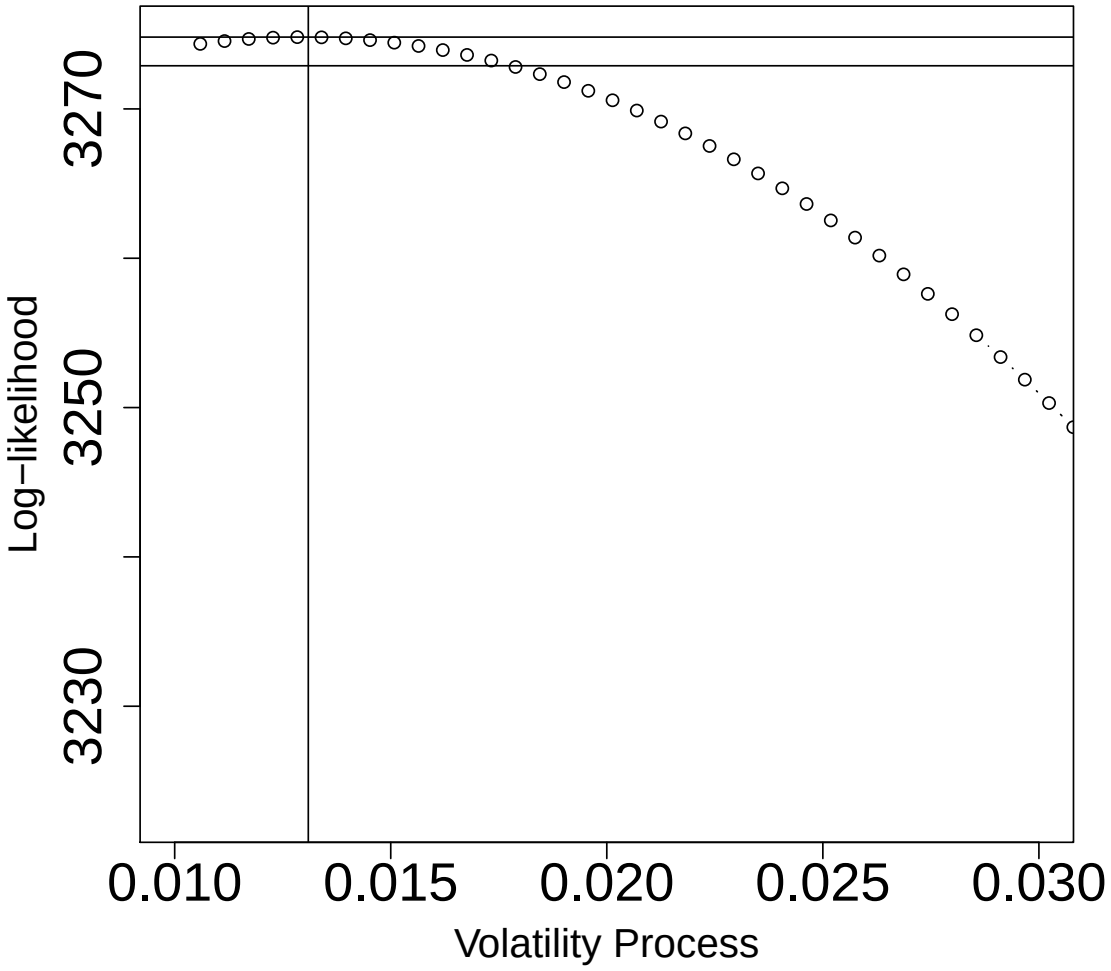


Figure 3.9: Profile log-likelihood function for the volatility process v_t at a particular time t of the Heston model with non-zero ρ based on the likelihood function derived in Section 3.3.1. The lower horizontal line indicates an approximate 95% confidence interval. The true value of v_t shown by the vertical line is consistent with this approximate 95% confidence interval.

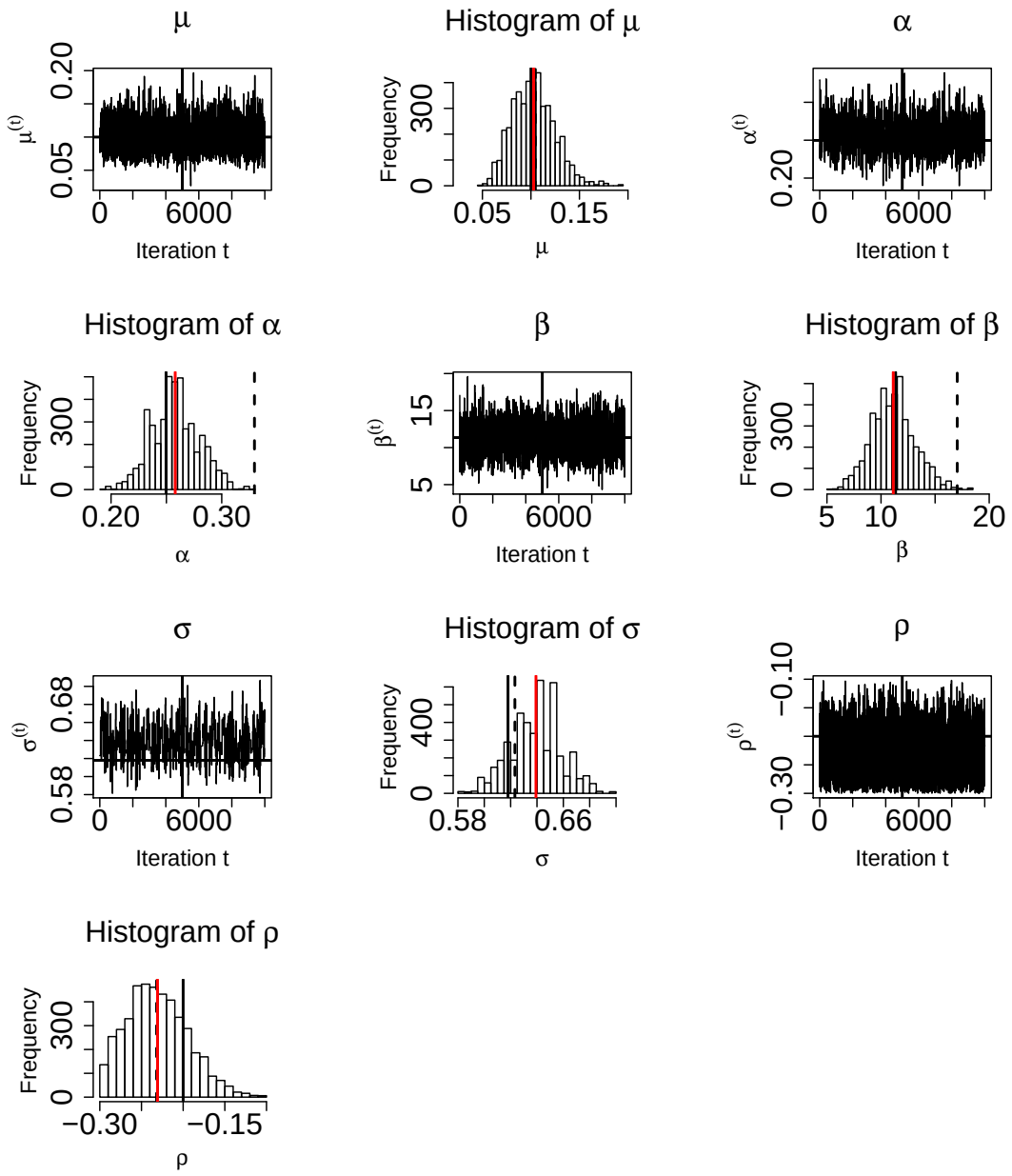


Figure 3.10: The graphs show simulations from the posterior distribution of the parameters θ of the Heston model with non-zero ρ using the MCMC algorithm when the number of iteration $N = 10000$. The first, third, fifth, seventh and ninth plots show time series plots of sampled values of $\theta = (\mu, \alpha, \beta, \sigma, \rho)$, where ρ takes a negative value. The second, forth, sixth, eighth and tenth plots show histograms of the sampled values of the parameters. The vertical line in the time series plots separates the burn-in phase (left) from samples that are used for the future inference (right). The horizontal lines in the times series plots and the black, red and dashed lines in the histogram show the true parameter, posterior mean and maximum likelihood values, respectively, where they can be shown on the scale.

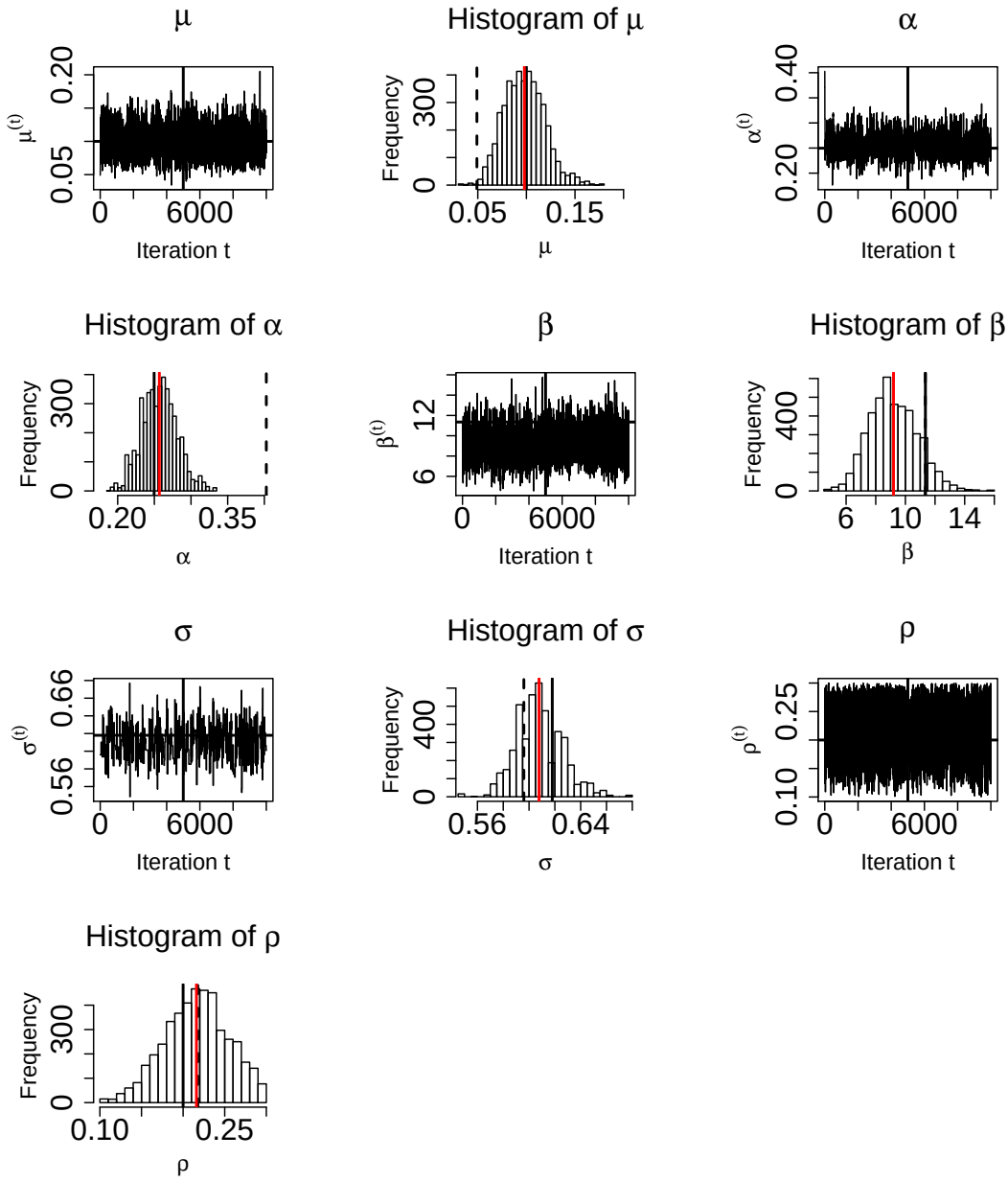


Figure 3.11: The graphs show simulations from the posterior distribution of the parameters of the Heston model with non-zero ρ using the MCMC algorithm when the number of iteration $N = 10000$. The first, third, fifth, seventh and ninth plots show time series plots of sampled values of $\theta = (\mu, \alpha, \beta, \sigma, \rho)$, where ρ takes a positive value. The second, forth, sixth, eighth and tenth plots show histograms of the sampled values of the parameters. The vertical line in the time series plot separates the burn-in phase (left) from samples that are used for the future inference (right). The horizontal lines in the times series plots and the black, red and dashed lines in the histogram show the true parameter, posterior mean and maximum likelihood values, respectively, where they can be shown on the scale.

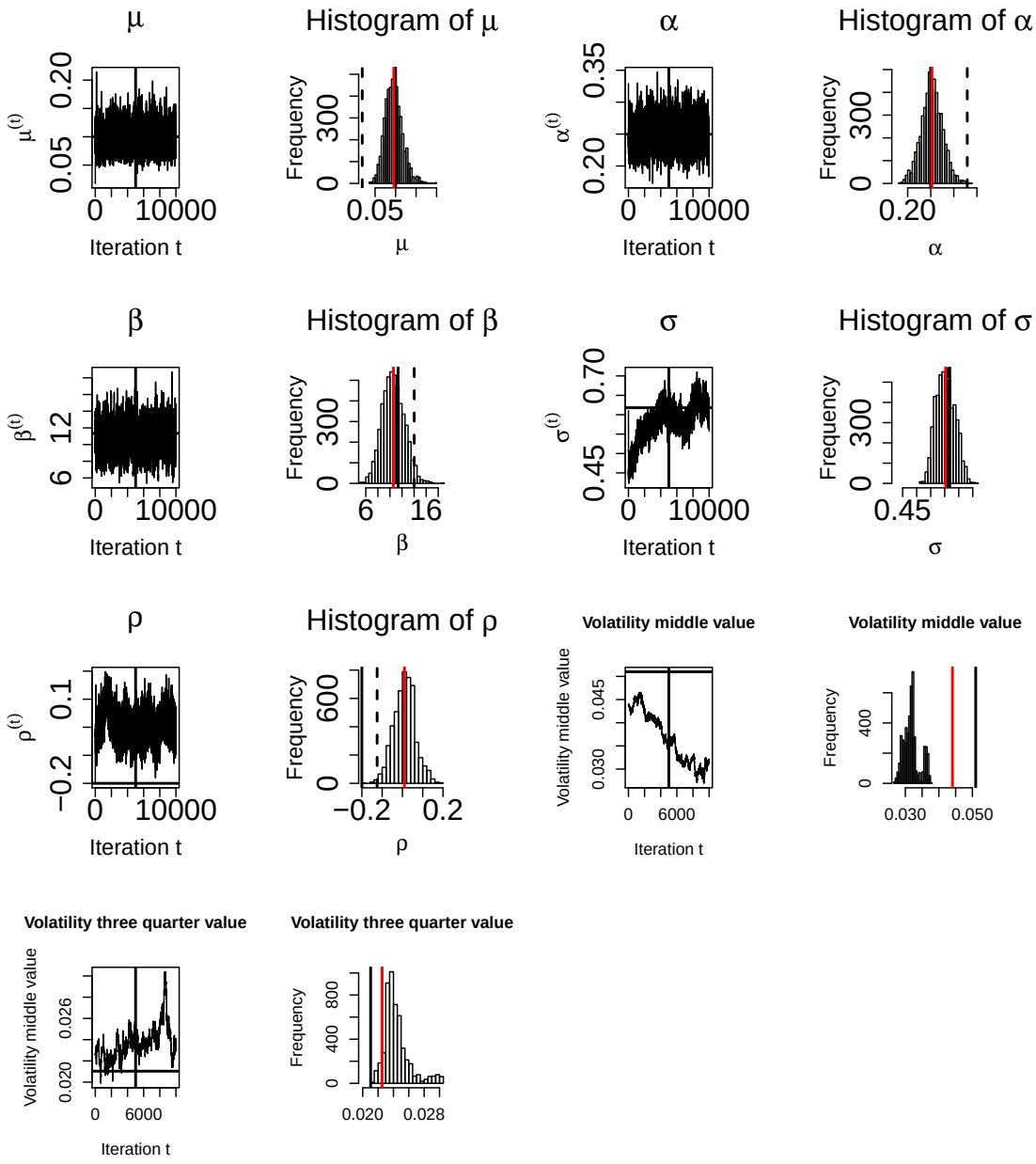


Figure 3.12: The graphs show simulations from the posterior distribution of the parameters and underlying volatility process of the Heston model with a negative ρ using the MCMC algorithm when the number of iteration $N = 10000$. The first and third columns show the time series plots of sampled values of the parameters $\theta = (\mu, \alpha, \beta, \sigma, \rho)$ and the middle and three quarter values of the volatility process, respectively. The second and fourth columns show the histogram of sampled values of the parameters and the volatility process values. The vertical line in the time series plot separates the burn-in phase (left) from samples that are used for the future inference (right). The horizontal lines in the times series plots and the black, red and dashed lines in the histogram show the true parameter, posterior mean and maximum likelihood values, respectively, where appropriate.

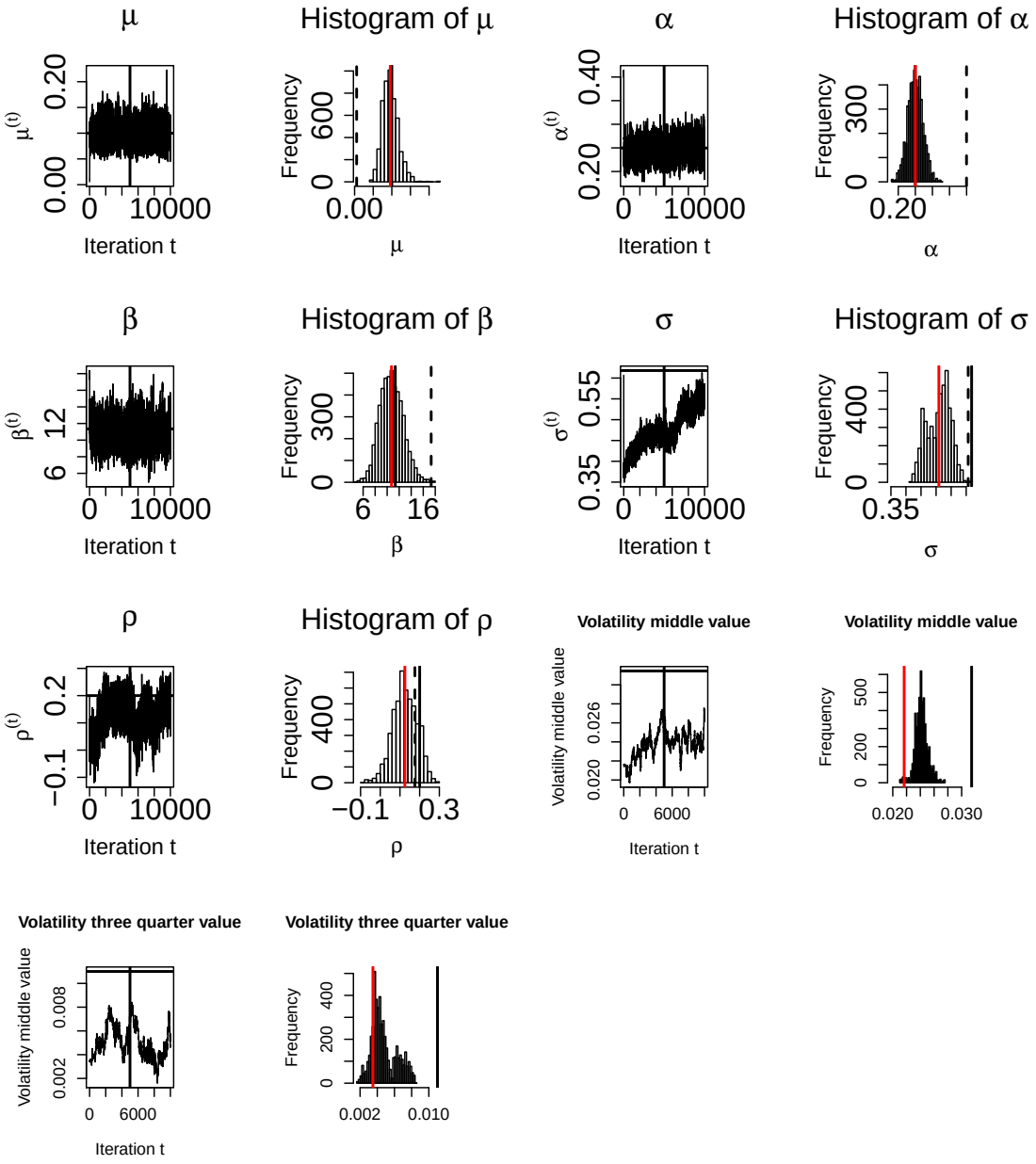


Figure 3.13: The graphs show simulations from the posterior distribution of the parameters and underlying volatility process of the Heston model with the positive value of ρ using the MCMC algorithm when the number of iteration $N = 10000$. The first and third columns show the time series plots of sampled values of the parameters $\theta = (\mu, \alpha, \beta, \sigma, \rho)$ and the middle and three quarter values of the volatility process, respectively. The second and fourth columns show the histogram of sampled values of the parameters and the volatility process values. The vertical line in the time series plot separates the burn-in phase (left) from samples that are used for the future inference (right). The horizontal lines in the times series plots and the black, red and dashed lines in the histogram show the true parameter, posterior mean and maximum likelihood values, respectively, where appropriate.

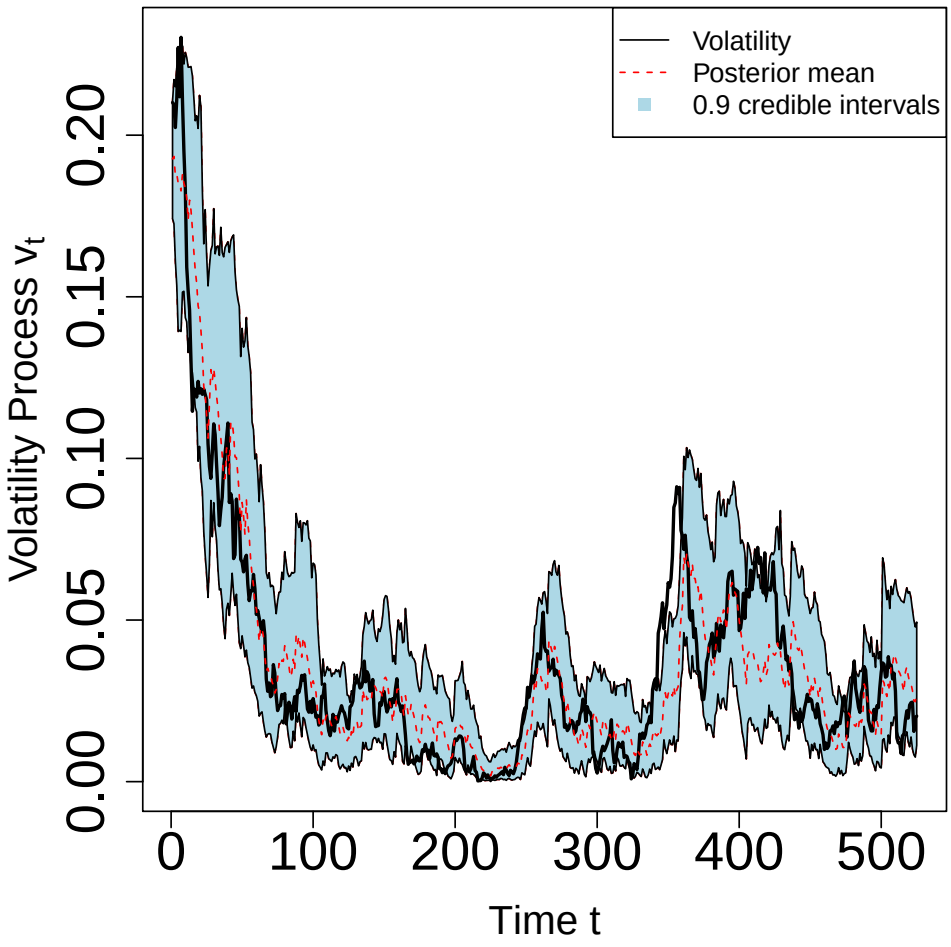


Figure 3.14: Estimation using the particle filter of the volatility process v_t of the Heston model with $\rho = -0.2$ assuming known parameters. The true volatility is represented by the black line. The posterior mean of the volatility process is represented by the dashed red line and the associated 90% credible intervals are in light blue.

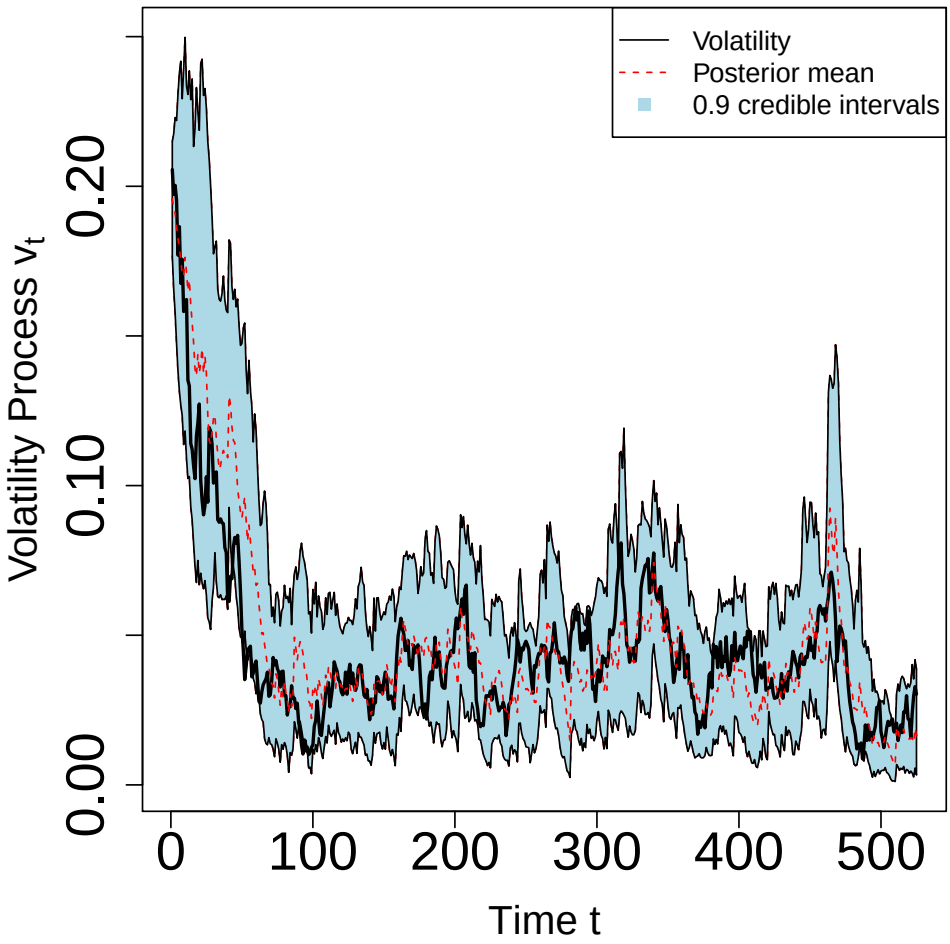


Figure 3.15: Estimation using the particle filter of the volatility process v_t of the Heston model with $\rho = 0.2$ assuming known parameters. The true volatility is represented by the black line. The posterior mean of the volatility process is represented by the dashed red line and the associated 90% credible intervals are in light blue.

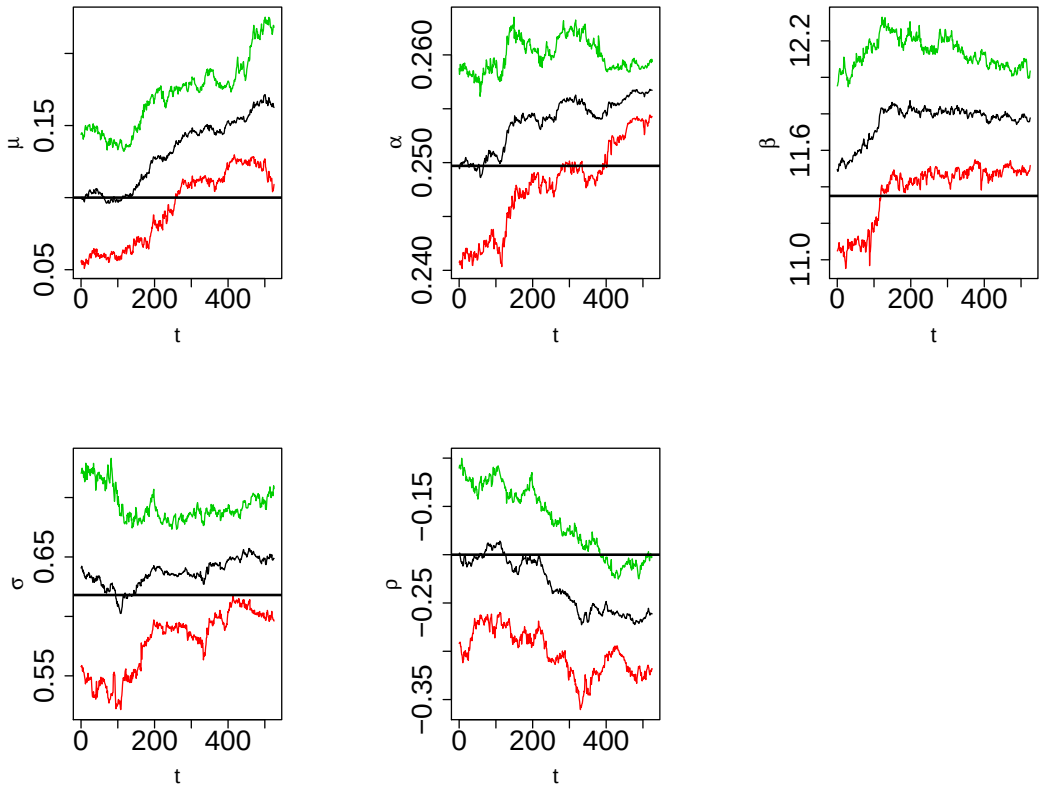


Figure 3.16: Estimation of the parameters of the Heston model with non-zero ρ using the auxiliary particle filter. In each graph, the posterior mean is shown using the black trace, while the associated approximate 90% credible intervals are shown by the outer traces. The first graph is for the parameter μ, the second graph is for the parameter α, the third graph is for the parameter β, the fourth graph is for the parameter σ, and final graph is for the parameter ρ. A uniform prior with support on $[-0.3, -0.1]$ was taken for ρ. The true values are shown by the horizontal line, e.g., $\rho = -0.2$.

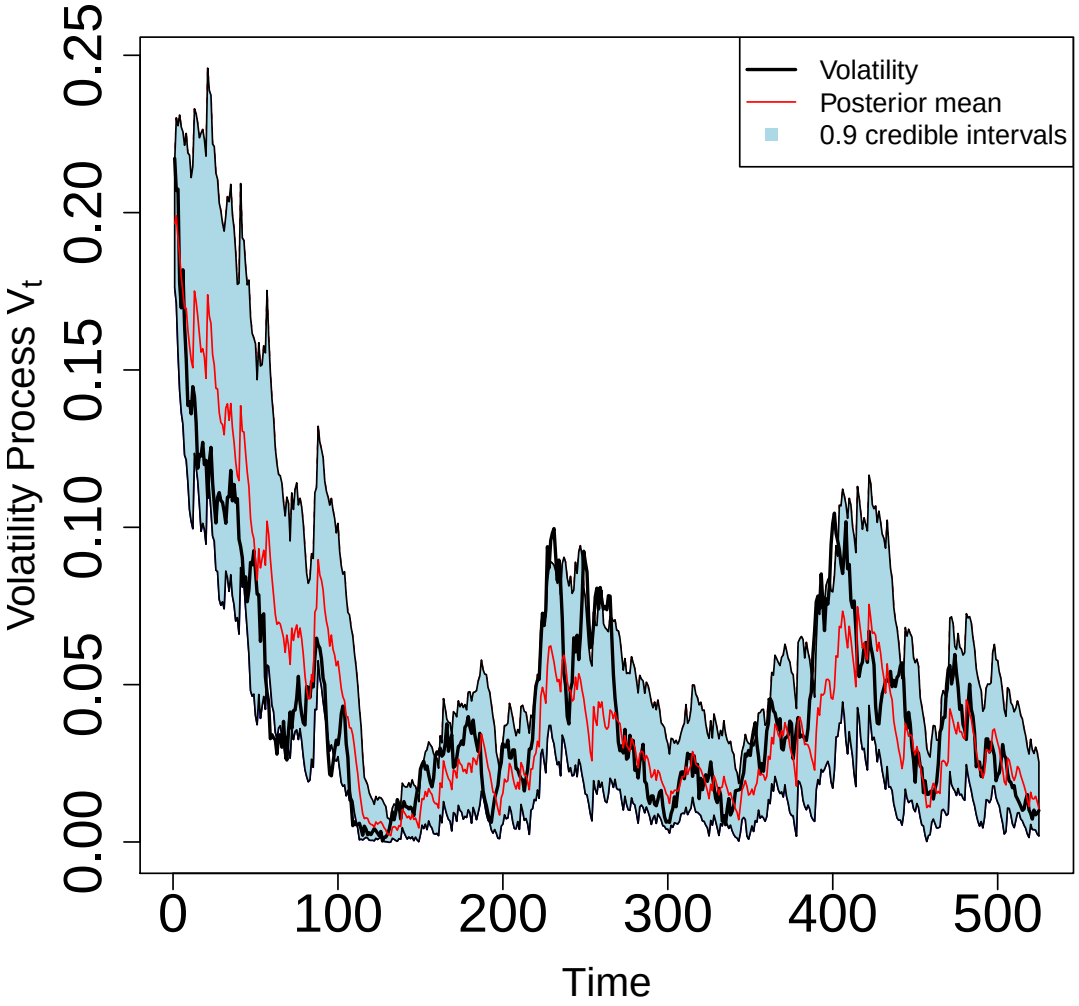


Figure 3.17: Estimation of the volatility process v_t in the Heston model with non-zero ρ using the auxiliary particle filter when the θ parameters are unknown, with a negative value -0.2 for ρ. The posterior mean of the volatility process is represented by the red line. The true volatility is represented by the black line, and the associated 90% credible intervals are in light blue.

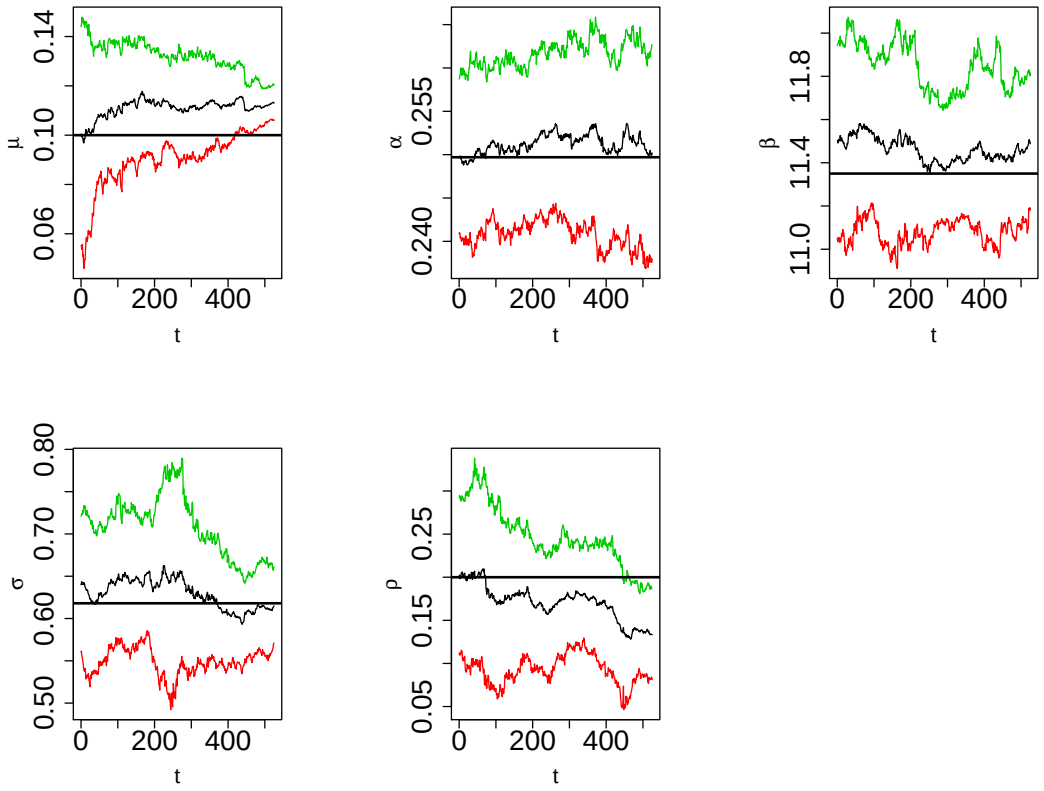


Figure 3.18: Estimation of the parameters of the Heston model with non-zero ρ using the auxiliary particle filter. In each graph, the posterior mean is shown using the black trace, while the associated approximate 90% credible intervals are shown by the outer traces. The first graph is for the parameter μ, the second graph is for the parameter α, the third graph is for the parameter β, the fourth graph is for the parameter σ, and final graph is for the parameter ρ. A uniform prior with support on $[0.1, 0.3]$ was taken for ρ. The true values are shown by the horizontal line, e.g., $\rho = 0.2$.

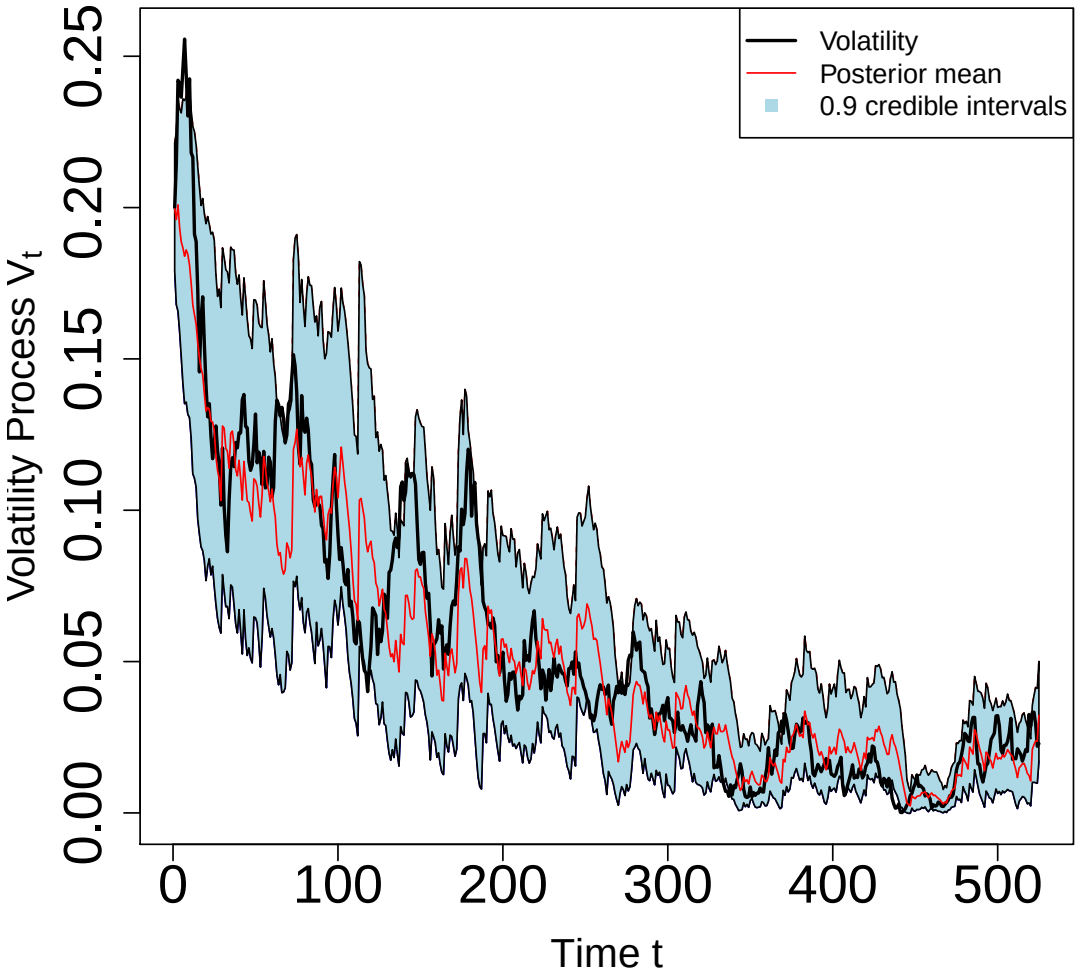


Figure 3.19: Estimation of the volatility process v_t in the Heston model with non-zero ρ using the auxiliary particle filter when the θ parameters are unknown with a positive value 0.2 for ρ. The posterior mean of the volatility process is represented by the red line. The true volatility is represented by the black line, and the associated 90% credible intervals are in light blue.

3.4.2 Standard & Poor's Index

Standard & Poor's Index is a popular index in markets, particularly in liquid stock markets around the world. Standard & Poor's European 350 (SP350) is the one of seven headline indices in the world. It is based on the European zone and it is drawn from 17 major markets. We will apply the Heston model to the data shown in Figure 3.20. The number of observation is $T = 308$.

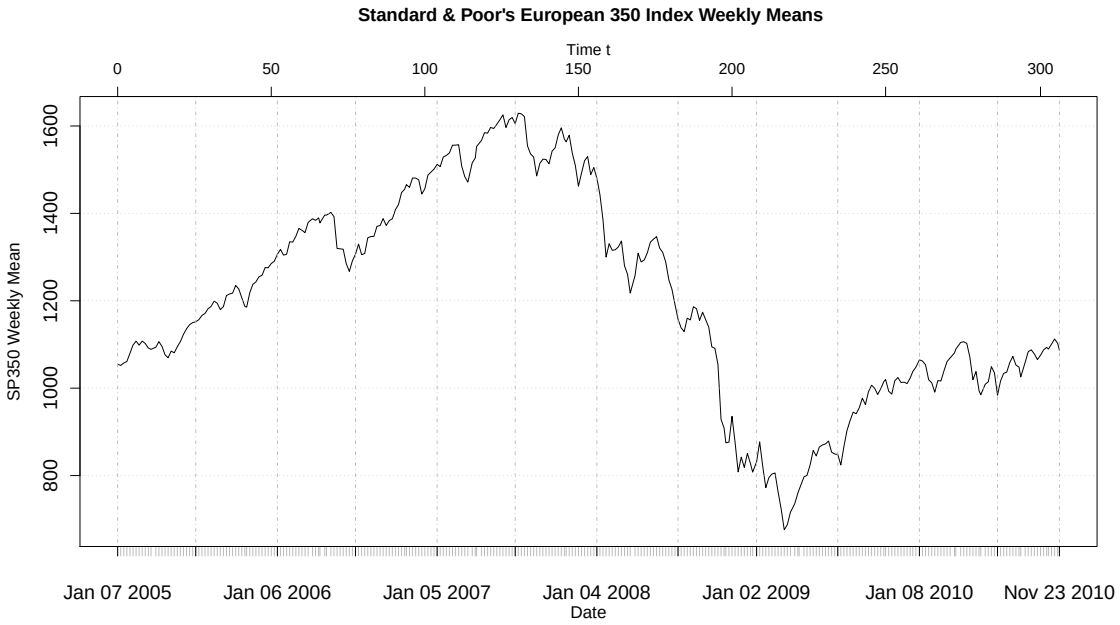


Figure 3.20: Graph of weekly means of the Standard & Poor's Index from January 2005 to December 2010.

3.4.2.1 The Heston model when $\rho = 0$

In this section, we will apply the Heston Model with $\rho = 0$, as defined though (3.5) and (3.6) in Section 3.2, to the SP350 data shown in Figure 3.20.

Figure 3.21 shows the results of applying the particle filter to the Heston model with $\rho = 0$, assuming that the other θ parameters are known. In particular, the posterior mean representing the estimation of the volatility process, and approximate 90% credible intervals, are shown across time. The values of the parameters are set to $\mu = 0.1$, $\alpha = 0.2497$, $\beta = 11.35$ and $\sigma = 0.618$, obtained from maximum likelihood estimation. We can say that the v_t process seems to reflect some of the features of the data.

Figure 3.22 shows the results of estimating the θ parameters of the Heston model with $\rho = 0$ across time t, using the auxiliary particle filter. In particular, the posterior mean and approximate 90% credible intervals of each parameter are presented together.

Figure 3.23 shows the results of the auxiliary particle filter for the estimated volatility process when the parameters are unknown. In particular, the estimated posterior mean of the volatility process and approximate 90% credible intervals are shown across time t. Again, the estimated v_t process seems to reflect some of the features of the data.

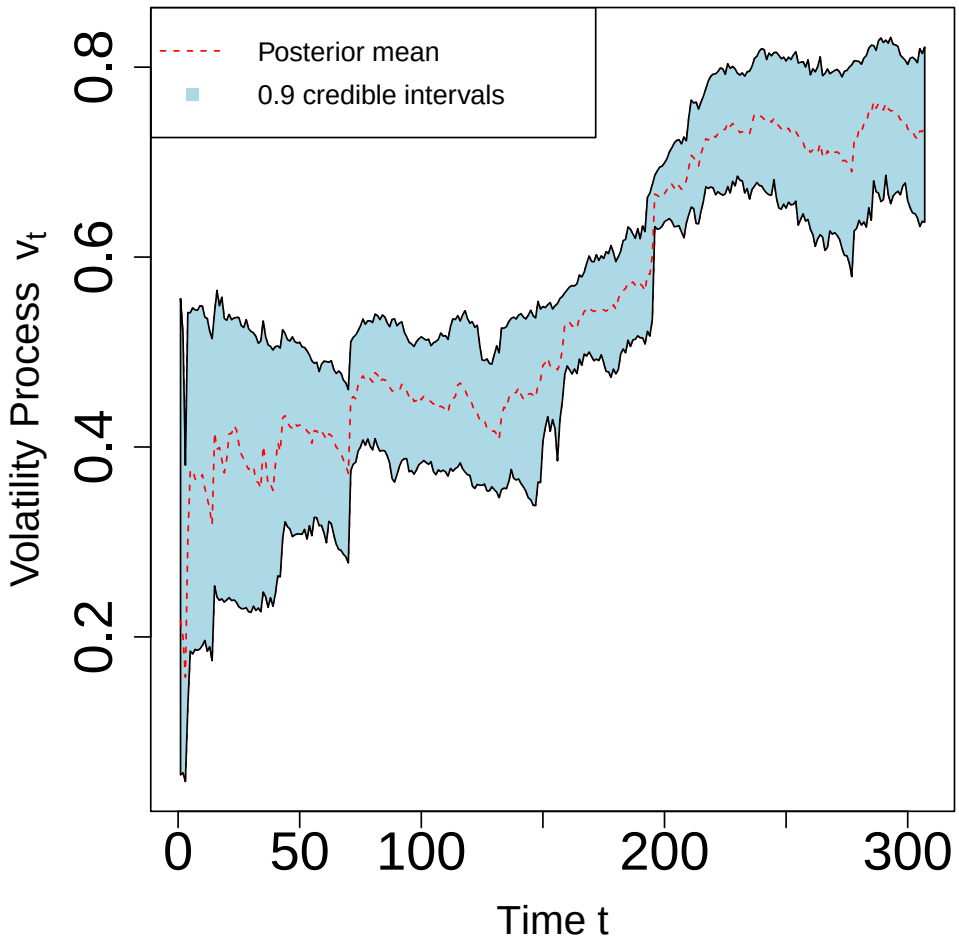


Figure 3.21: Estimation of the volatility v_t of the SP350 process using the particle filter for the Heston model with $\rho = 0$. The approximate posterior mean of the volatility process is represented by the dashed red line and the associated 90% credible intervals are in light blue.

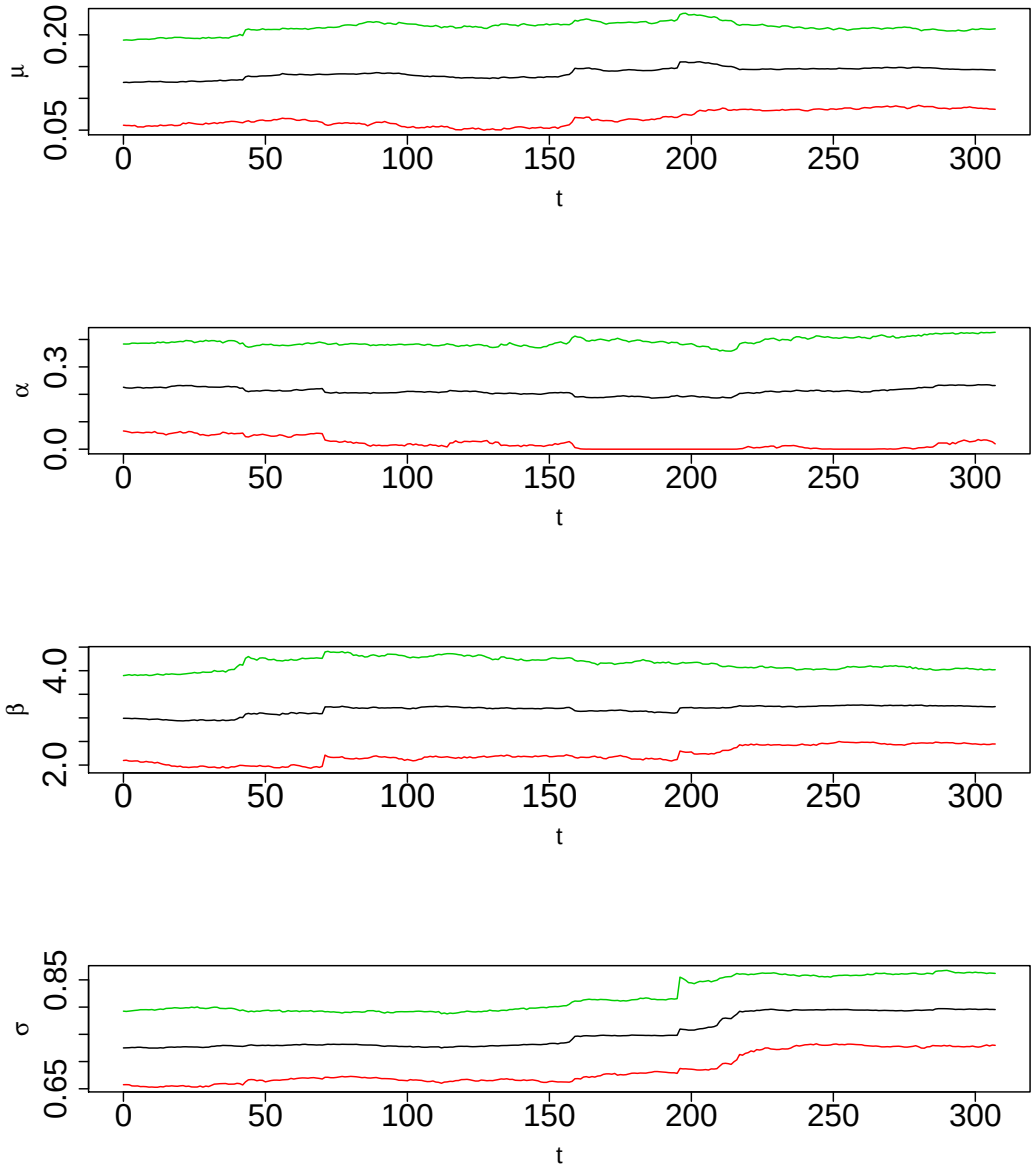


Figure 3.22: Estimation of the parameters of the Heston model with $\rho = 0$ across time t, using the auxiliary particle filter applied to the SP350 data. In each graph, the posterior mean is shown using the black trace, while the associated approximate 90% credible intervals are shown by the outer traces. The first graph is for the parameter μ, the second graph is for the parameter α, the third graph is for the parameter β and the final graph is for the parameter σ.

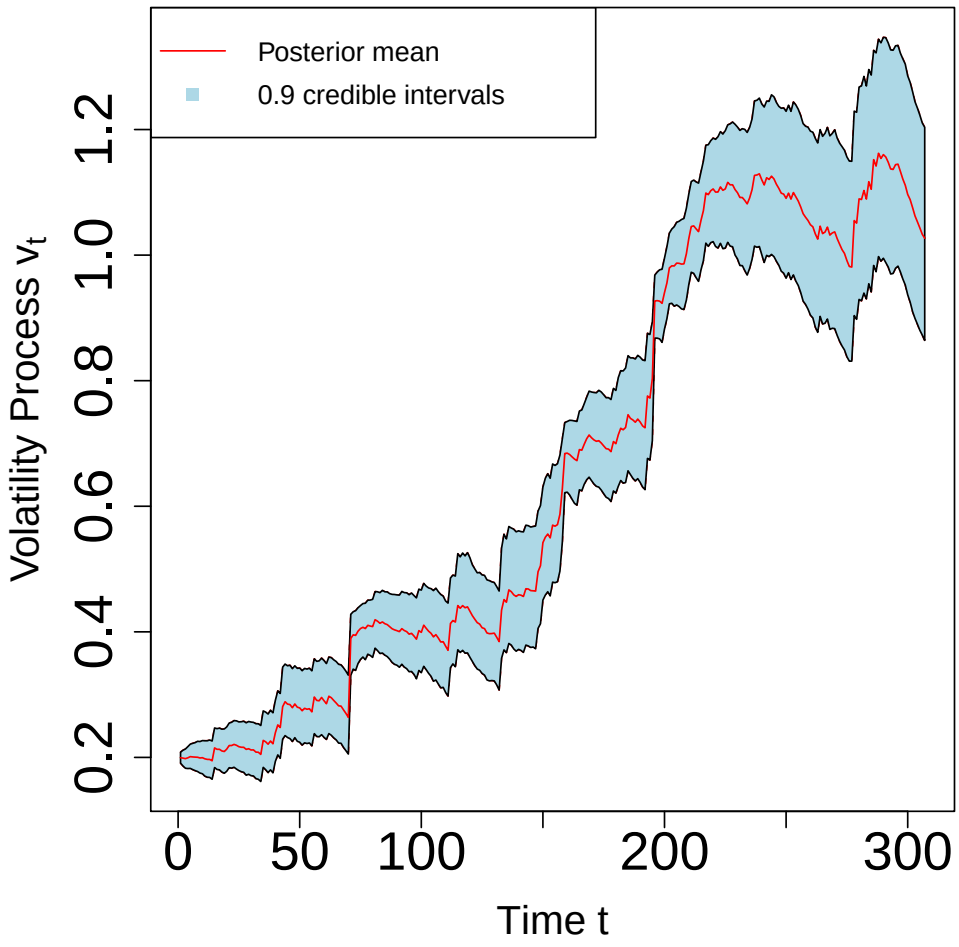


Figure 3.23: Estimation of the volatility v_t of SP350 process in the Heston Model with $\rho = 0$ when the other θ parameters are unknown. The estimated posterior mean of the volatility process is represented by the red line and the associated 90% credible intervals are in light blue.

3.4.2.2 The Heston model with non-zero ρ

We will now include the non-zero correlation coefficient ρ in the Heston model, as defined in equations (3.12) and (3.13) in Section 3.3. Again, we will use the SP350 data shown in Figure 3.20.

Figure 3.24 shows the results of applying the particle filter to the Heston model with non-zero ρ, assuming the other θ parameters are known. In particular, the posterior mean, representing the estimation of the volatility process, and approximate 90% credible intervals are shown across the time. The values of the parameters are set to $\mu = 0.1$, $\alpha = 0.2497$, $\beta = 11.35$, $\sigma = 0.618$ and $\rho = -0.2$. Again, the estimated v_t process seems to reflect some of the features of the data.

Figure 3.25 shows the results of applying the particle filter to the Heston model with non-zero ρ, assuming the other θ parameters are known. The only difference is that in this figure the value of ρ is set to 0.2.

Once more we can say that the v_t process is a sensible estimate as it coincides with some of the features that we see in the data.

Figure 3.26 shows the results of estimating the θ parameters of the Heston model with a uniform prior distribution with negative support for ρ, across for time t, using the auxiliary particle filter. In these plots the posterior mean and approximate 90% credible intervals of each parameter are shown.

Figure 3.27 shows the results of the auxiliary particle filter for the estimated volatility process of the Heston model with a uniform prior distribution with negative support for ρ. In particularly, the estimated posterior mean of the volatility process and approximate 90% credible intervals are shown across time. Figure 3.28 shows the results of estimating the θ parameters of the Heston model with a uniform prior distribution with positive support for ρ, across time t, using the auxiliary particle filter. In these plots the posterior mean and approximate 90% credible intervals of each parameter are shown. We can see from Figure 3.26 and 3.28 that the parameter ρ is hard to estimate in that the posterior distributions depend strongly on the choice of the prior.

Figure 3.29 shows the results of the auxiliary particle filter for the estimated volatility

process of the Heston model with a prior distribution with positive support for ρ. In particular, the estimated posterior mean of the volatility process and approximate 90% credible intervals are shown across time.

Again, the auxiliary particle filter yields a sensible estimate for the volatility process as it coincides with some of the features that we see in the data.

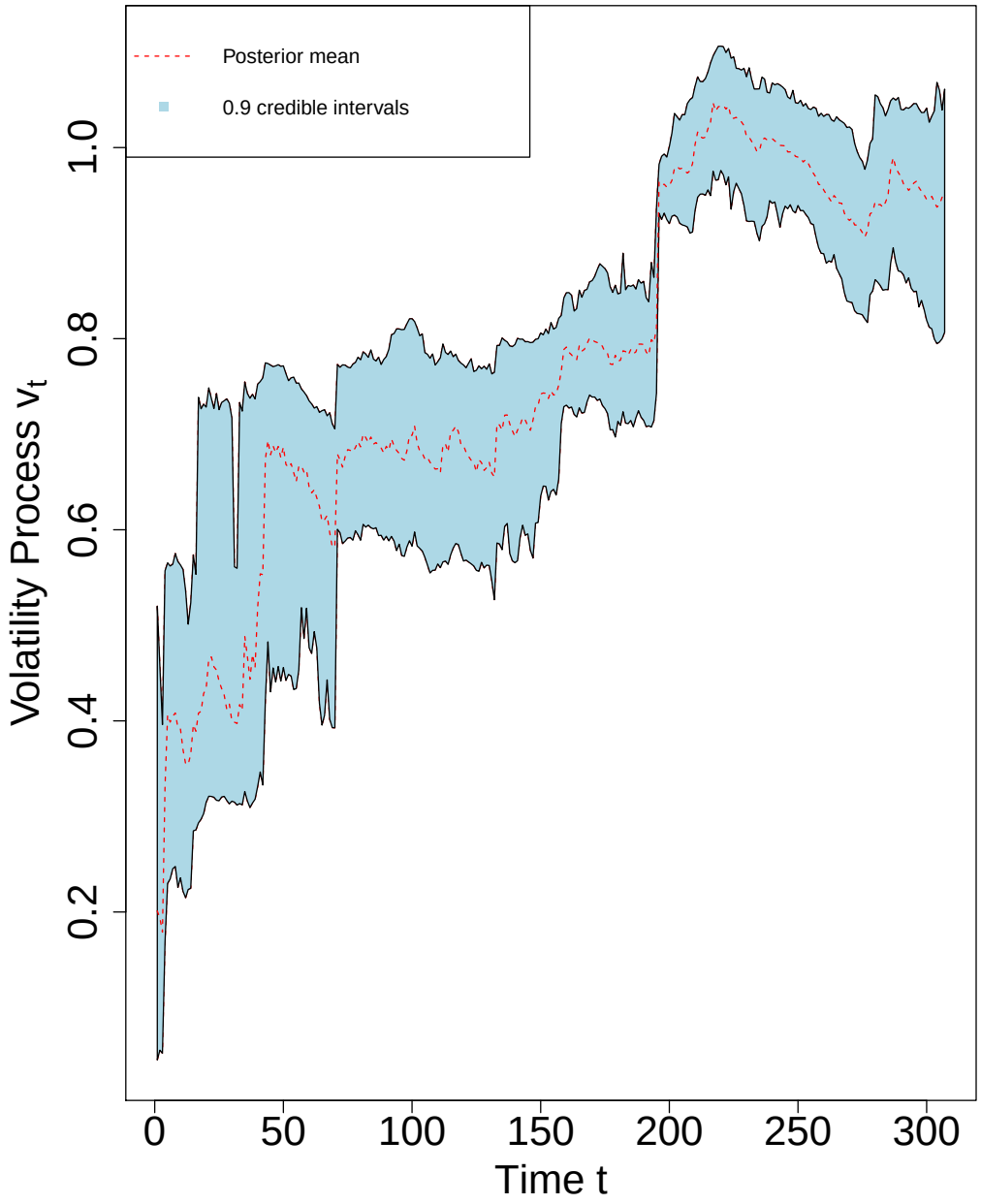


Figure 3.24: Estimation of the volatility v_t of the SP350 process using the particle filter for the Heston model with an assumed negative value of ρ, i.e., $\rho = -0.2$. The approximate posterior mean of the volatility process is represented by the dashed red line and the associated 90% credible intervals are in light blue.

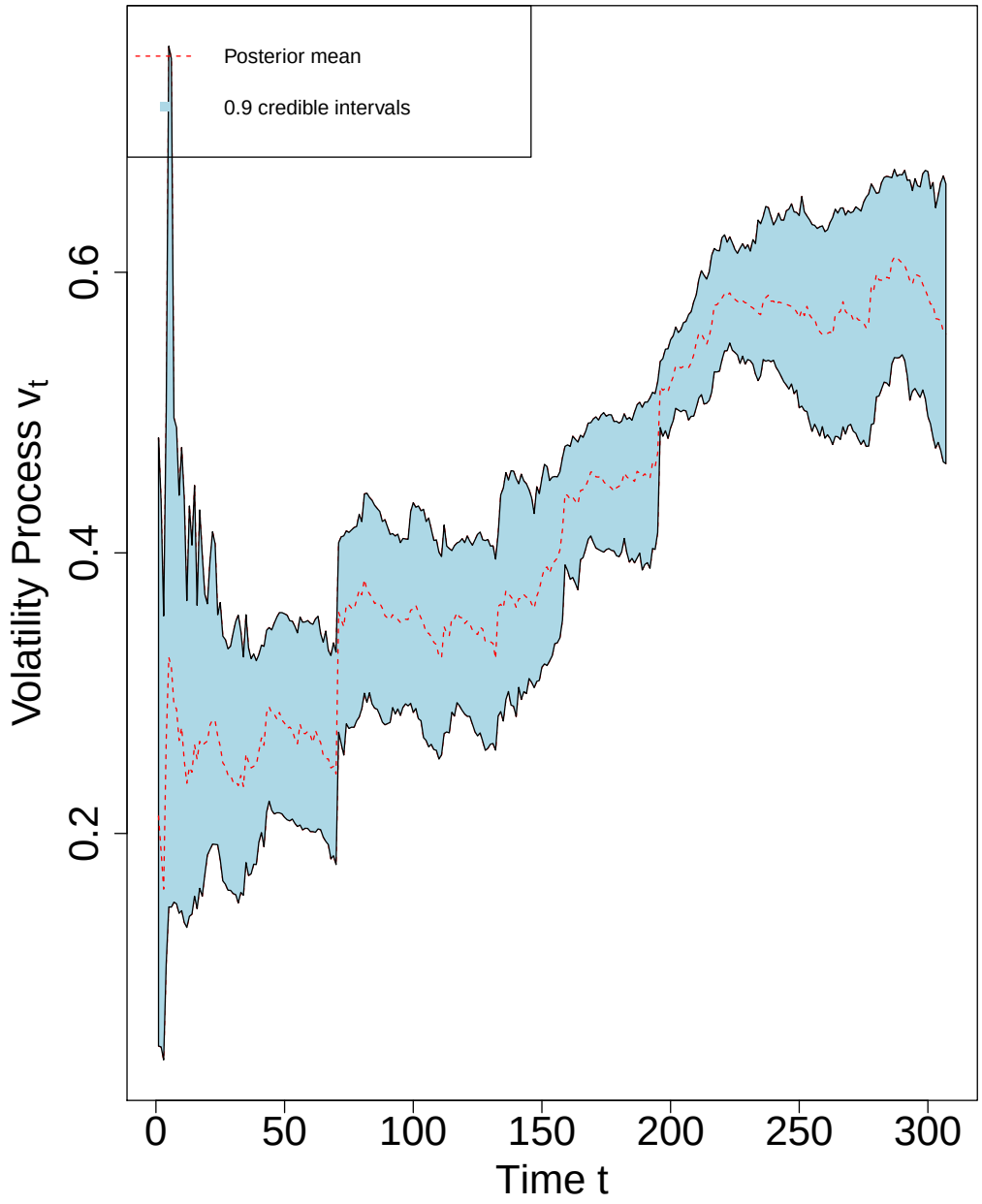


Figure 3.25: Estimation of the volatility v_t of the SP350 process using the particle filter for the Heston model with an assumed positive value of ρ, i.e., $\rho = 0.2$. The approximate posterior mean of the volatility process is represented the dashed red line and the associated 90% credible intervals are in light blue.

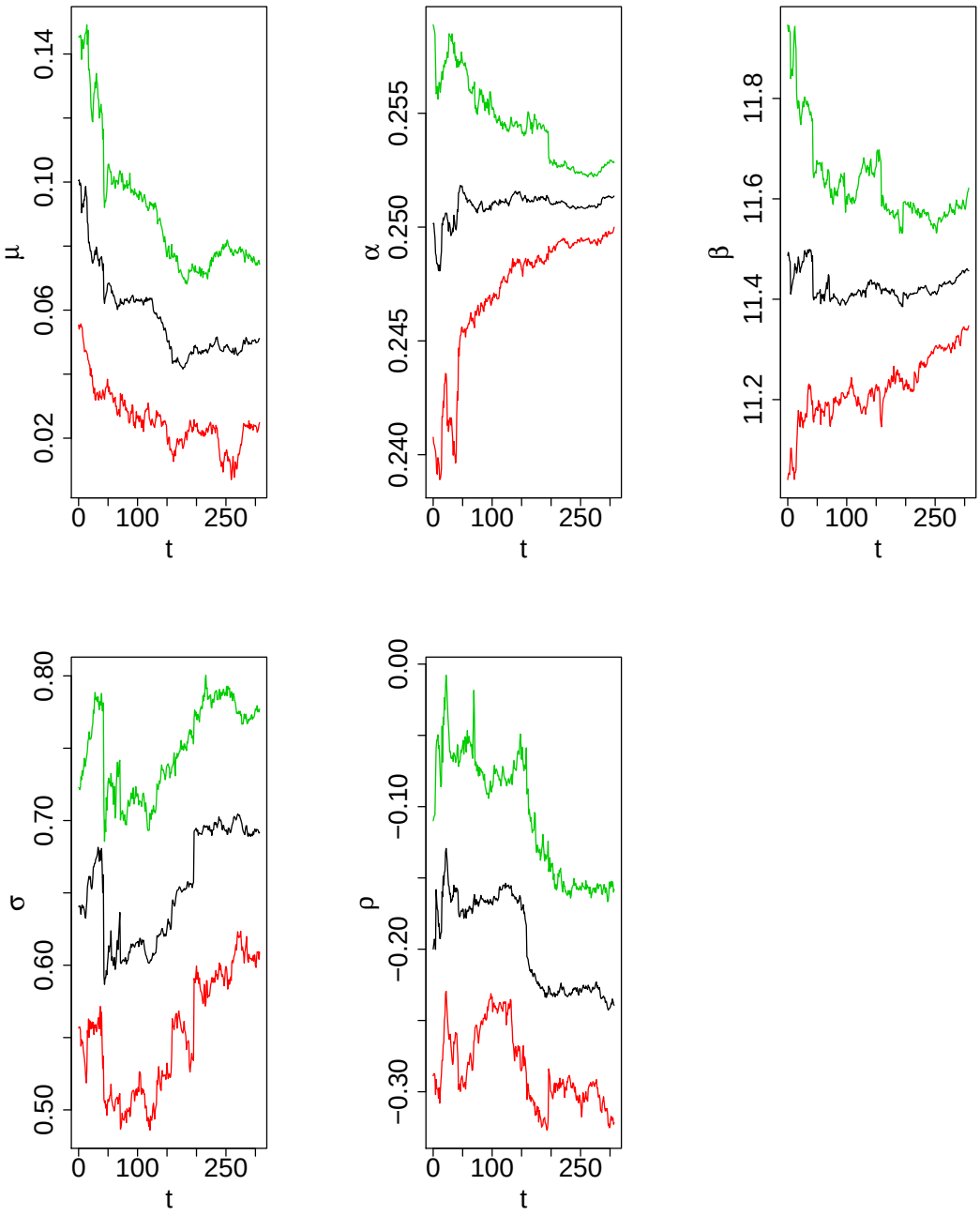


Figure 3.26: Estimation of the parameters of the Heston model with a uniform prior distrbution with negative support for ρ, across time t, using the auxiliary particle filter applied to the SP350 data. In each graph, the posterior mean is shown using the black trace, while the associated approximate 90% credible intervals are shown by the outer traces. The first graph is for the parameter μ, the second graph is for the parameter α, the third graph is for the parameter β, the fourth graph is for the parameter σ and the final graph is for the parameter ρ.

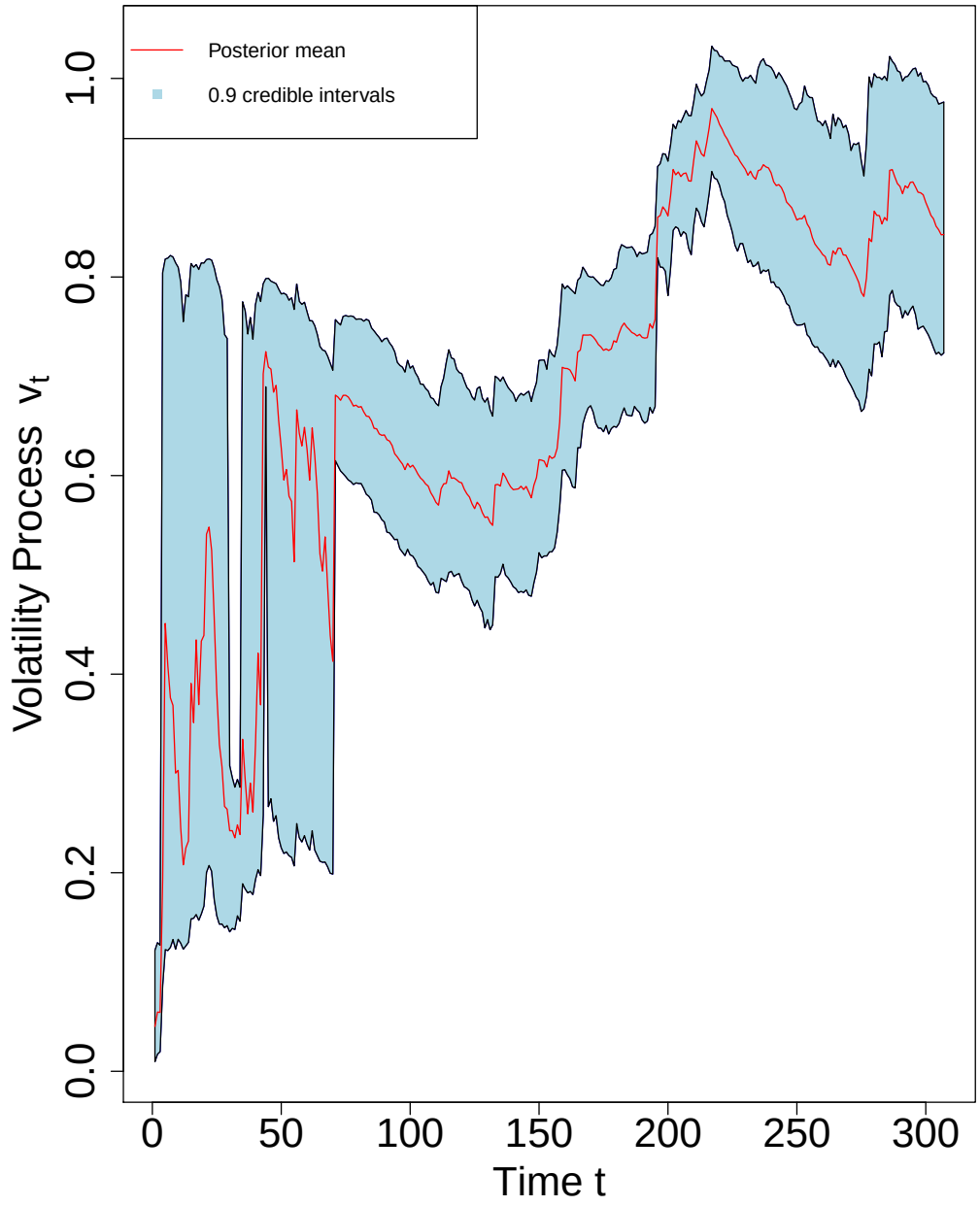


Figure 3.27: Estimation of the volatility v_t of the SP350 process in the Heston model with a uniform prior distribution with negative support for ρ, using the auxiliary particle filter when the θ parameters are unknown. The estimated posterior mean of the volatility process is represented by the red line and the associated 90% credible intervals are in light blue.

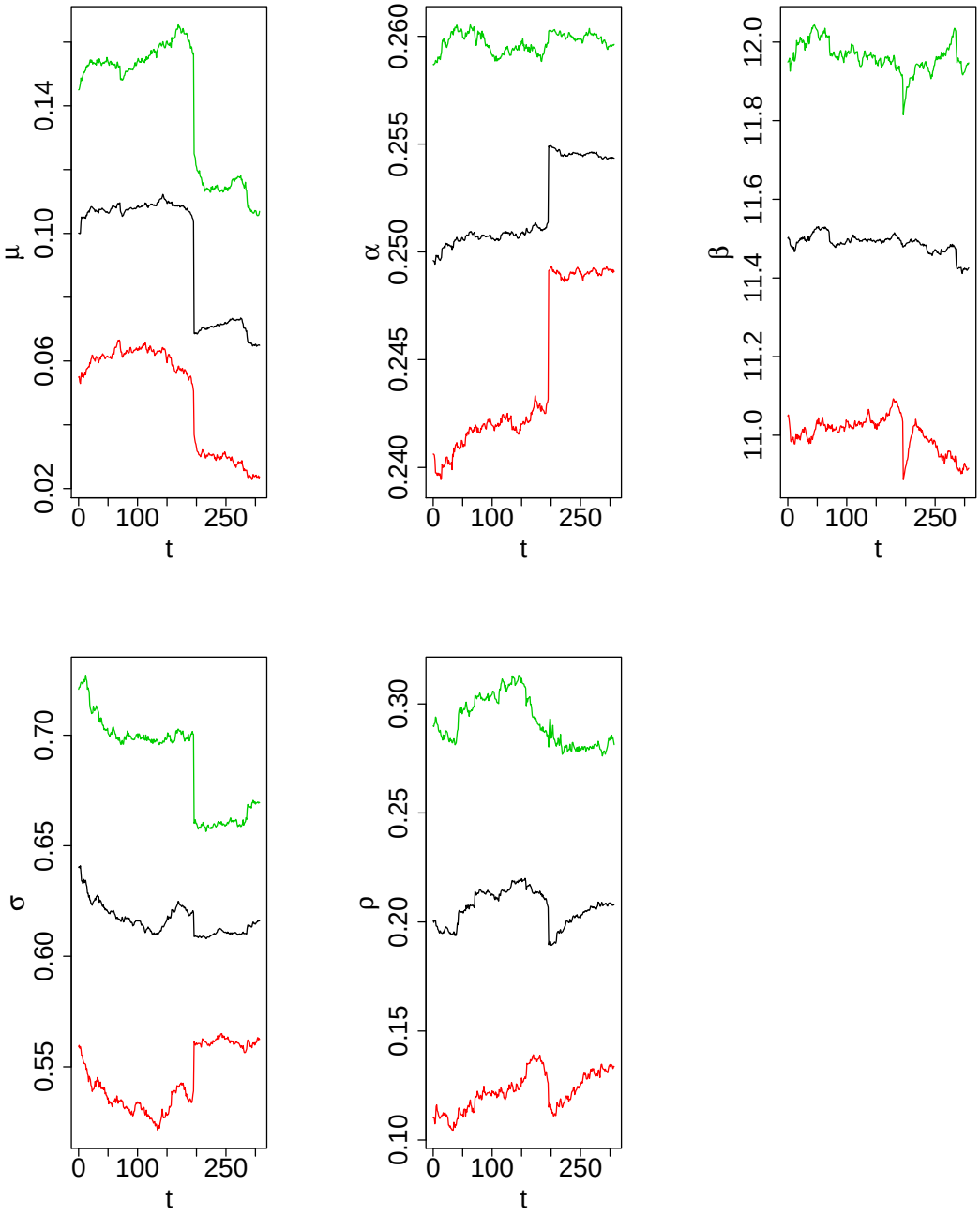


Figure 3.28: Estimation of the parameters of the Heston model with a uniform prior distribution with positive support for ρ, across time t, using the auxiliary particle filter applied to the SP350 data. In each graph, the posterior mean is shown using the black trace, while the associated approximate 90% credible intervals are shown by the outer traces. The first graph is for the parameter μ, the second graph is for the parameter α, the third graph is for the parameter β, the fourth graph is for the parameter σ and the final graph is for the parameter ρ.

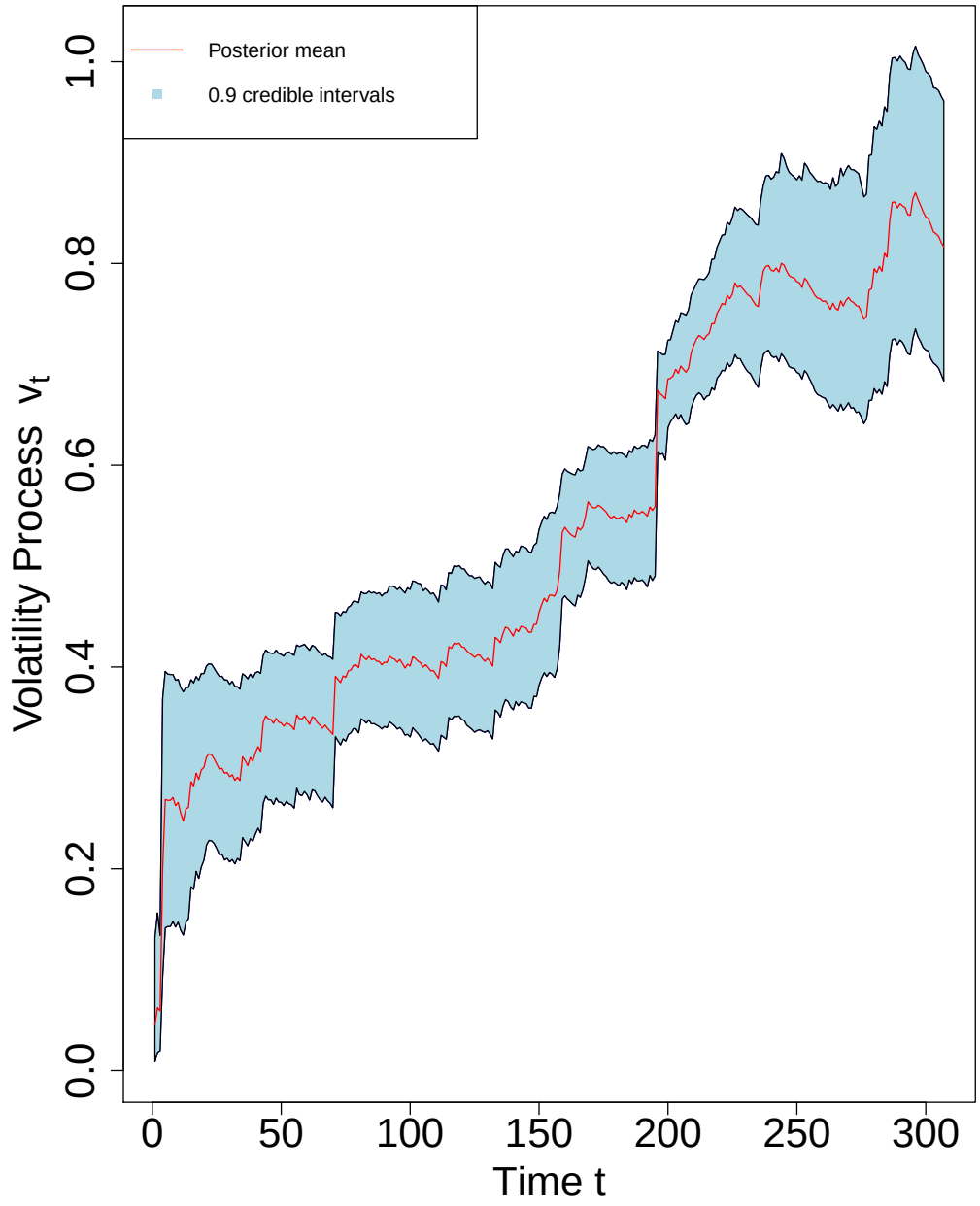


Figure 3.29: Estimation of the volatility v_t of the SP350 process in the Heston model with a uniform prior distribution with positive support for ρ, using the auxiliary particle filter when the θ parameters are unknown. The estimated posterior mean of the volatility process is represented by the red line and the associated 90% credible intervals are in light blue.

3.5 Summary

In this chapter, we have discussed:

1. The Heston model when $\rho = 0$, where ρ is the correlation between the driving stochastic processes. In particular,
   we have presented the model in discrete time, and maximum likelihood estimation and Bayesian inference using the MCMC and particle filter based algorithms for it.

2. The Heston model for general ρ.
   In particular, we have presented the model in discrete time, and maximum likelihood estimation and Bayesian inference using particle filter based algorithms for it.

3. Applications to simulated and real data when $\rho = 0$ and for non-zero ρ.

Following on from our results, we will extend the Heston model to the fractional Heston model by replacing the driving Brownian motions with fractional Brownian motions in Chapter 4. We will again use particle filter based algorithms to preform inference about the unknown quantities of these models.

Chapter 4

Fractional Heston Model

In this chapter we present and discuss our fractional Heston model. The Heston (1993) model has been widely used in finance and was presented and discussed in detail in Chapter 3. It is an extension of the Black-Scholes model (Fisher and Scholes, 1973) in

the following areas (Bauer, 2012):

1− In the Heston model, unlike the Black-Scholes model, it is assumed that the volatility process is non-constant, and that it is driven by a Cox-Ingersoll-Ross (CIR) process. The volatility process therefore has the important mean-reverting property.

2− It has non-log normal distribution of asset prices. This means that asset price is no longer assumed to follow a log normal distribution as in the Black-Scholes model as Fisher and Scholes (1973) suggested.

3− There is a correlation between the asset price and the volatility process. The correlation coefficient ρ between the driving stochastic processes is between -1 and 1, i.e., $\rho \in [-1,1]$. This relationship between the asset price and the volatility processes is also called a leverage effect (Bauwens et al., 2012).

The Heston model is defined using Brownian motion. We now extend the Heston model by using fractional Brownian motion instead of Brownian motion. Our extended model allows a richer correlation structure to be used. We show that our extension offers advantages when modelling and forecasting real data.

In Section 4.1 and Section 4.2 we introduce fractional Brownian motion and its properties. We discuss the fractional Heston model with zero correlation ρ between the driving stochastic processes in Section 4.3, while in Section 4.4 we extend this model to allow non-zero ρ. In Section 4.5 we discuss Bayesian inference for our models. In Section 4.6 we outline a forecasting method for the fractional Heston model, while in Sections 4.7 and 4.8 we present all our results.

4.1 Introduction to fractional Brownian motion

Fractional Brownian motion (fBm) is a generalization of Brownian motion B_t. It is a continuous Gaussian process B_t^H on $[0,T]$ as follows (Dieker, 2004; Coeurjolly, 2000):-

$$B_t^H = \frac{1}{\Gamma\left(H+\frac{1}{2}\right)}\left(\int_{-\infty}^{0}\left[(t-s)^{H-\frac{1}{2}}-(-s)^{H-\frac{1}{2}}\right]dB_s+\int_0^t(t-s)^{H-\frac{1}{2}}dB_s\right) \quad t,s\in\Re$$

where

Γ is the gamma function such that $\Gamma(\alpha)=\int_0^\infty x^{\alpha-1}e^{-x}dx$,

H is called the Hurst index and is a real number taking values in $(0,1)$. The Hurst index is sometimes referred to as an index of time series dependence. The main idea of the Hurst index is to model long memory in the underlying process (Giratis et al., 2012).

The fBm has expectation zero for all times $[0,T]$ and has covariance function given in Dieker (2004) as

$$E\left[B_t^H B_s^H\right]=\frac{1}{2}\left(|t|^{2H}+|s|^{2H}-|t-s|^{2H}\right), \quad t\neq s, \quad t,s\in\Re. \tag{4.1}$$

So that

$$B_t^H \sim N\left(0,t^{2H}\right)$$

An important concept in fBm is that of an increment. Let $s<t$. Then the **increment** of fBm over the interval $(s,t]$ is

$$B_t^H - B_s^H \sim B_{t-s}^H; \tag{4.2}$$

it can be shown that the increment $B_t^H - B_s^H$ has the same distribution as B_{t-s}^H, namely $N\left(0,(t-s)^{2H}\right)$.

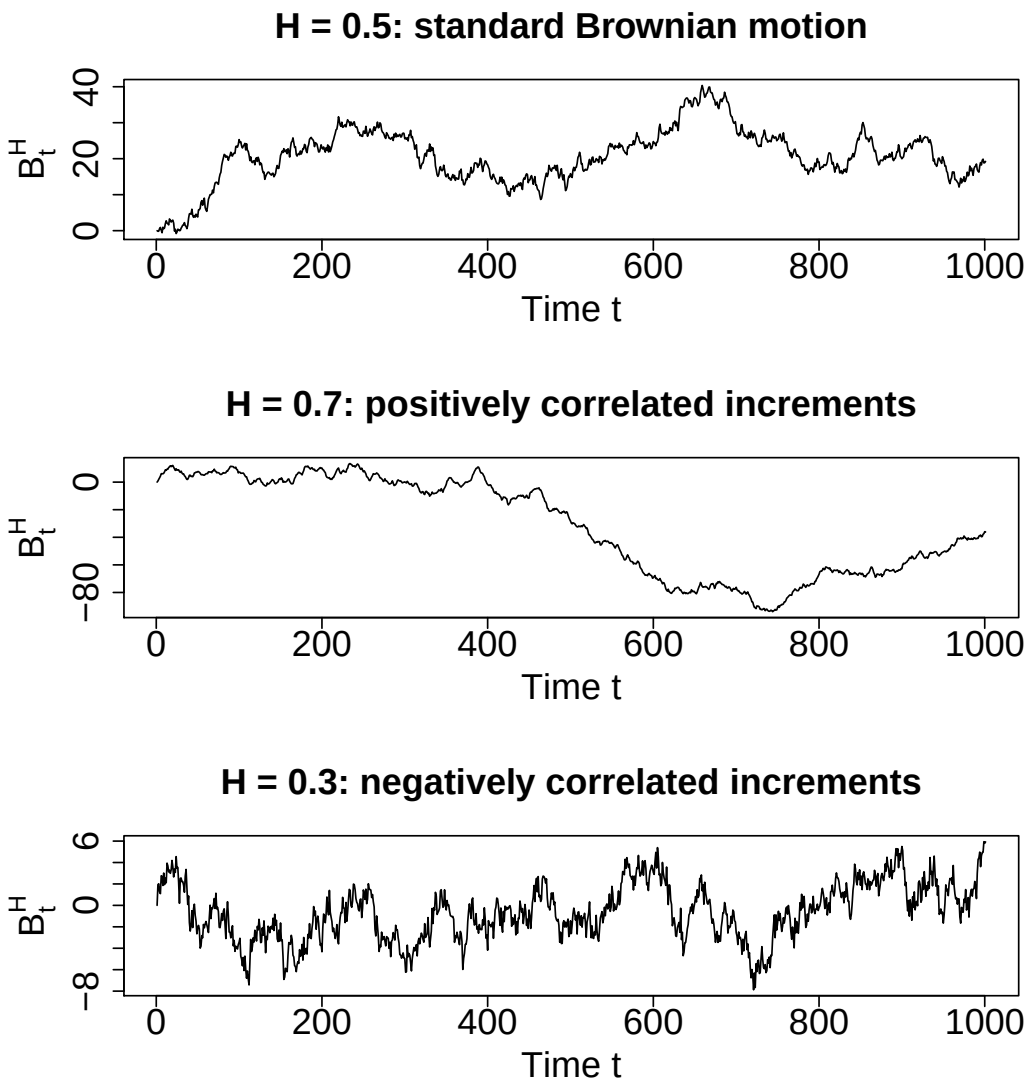


Figure 4.1: Fractional Brownian motion paths with different values of the Hurst index H

In general, we can describe three cases for the dependence of fBm in terms of H in Figure as follows:-

If $H = \frac{1}{2}$, the process is standard Brownian motion B and the covariance function $\gamma(k)$ between two increments is zero: $\gamma(k) = 0$, where k is the lag index.

If $H > \frac{1}{2}$, the process has positively correlated increments and is called a long memory processes. It mean that the covariance function $\gamma(k)$ between two increments is greater than zero: i.e., $\gamma(k) > 0$.

If $H < \frac{1}{2}$, the process has negatively correlated increments and is called a short memory

processes. It mean that the covariance function $\gamma(k)$ between two increments is less than zero: $\gamma(k) < 0$.

It can be seen that positive increment correlation leads to a process that seems visually less rough, while negative increment correlation leads a process that seems visually rougher.

The covariance function in the last two cases can be shown to have the property that:-

$$\gamma(k) = C^2 H(2H-1)|k|^{2H-2}, \quad \text{as } |k| \longrightarrow +\infty$$

where C is a constant.

Sample paths of fBm for these three cases are shown in Figure 4.1.

4.2 Properties of fractional Brownian motion

In this section, we present some important properties of fBm:

1– It is a self-similar stochastic process. A self-similar process has the same behaviour at different scales over time, i.e.

$$B_{at}^H \sim |a|^H B_t^H,$$

which means that the two sides in the above equation have the same distribution. This property tells us that when the process is considered on a different time scale, it is a scaled version of itself on the original time scale.

2– It has stationary increments as defined in (4.2) above. Therefore, fBm is a self-similar Gaussian process with stationary increments.

3– fBm has long-range dependence when $H > \frac{1}{2}$. Here long-range dependence can be defined in terms of the behaviour of the autocorrelation function across time t:-

$$\sum_{t=1}^{\infty} E\left[B_1^H\left(B_{t+1}^H - B_t^H\right)\right] = \infty.$$

The summand is the covariance between increments $B_{t+1}^H - B_t^H$ and $B_1^H - B_0^H$ over time intervals of size 1 at lag t. So the sum of this covariance is infinite.

There are two processes which are related to fBm. The advantage of these processes are in the areas of analysis and simulation. The processes are as follows:-

1− A fractional Gaussian noise (fGn) process is a stationary Gaussian process that represents the increments of fractional Brownian motion

$$X_t^H = B_t^H - B_{t-1}^H.$$

Immediately, X_t^H is a random variable that follows a $N(0,1)$ distribution.
Rose (1996) has shown that fGn is obtained by taking the difference of fBm, and the spectrum of fGn is given
as

$$f(\lambda) = \frac{\sigma^2}{\pi}\sin(\pi H)\Gamma(H+1)(1-\cos\lambda)\sum_{j=-\infty}^{\infty}|\lambda+2\pi j|^{-2H-1} \quad 0<H<1.$$

2− A Fractional Autoregressive Integrated Moving Average (FARIMA) process is a generalization of a ARMA(p,q) with a fractional difference parameter d, where $d = H - 0.5$. We can define a FARIMA$(0,d,0)$ process as:-

$$(1-\beta)^d X_t = a_t \quad -0.5 < d < 0.5$$

where

a_t is a white noise process at time t;
$(1-\beta)^d$ is a fractional difference operator, which is defined as $(1-\beta)^d = \sum_{k=0}^{\infty}\binom{k}{d}(-\beta)^k$, in which β is the backward shift operator $\beta Y_t = Y_{t-1}$.
The spectrum of the FARIMA process is simpler than the spectrum of fGn and takes

the form (Rose, 1996):

$$f(\lambda) \;=\; 2\sin\left(\frac{\lambda}{2}\right)^{-2d} = |1-\exp(i\lambda)|^{-2d},$$

in which $\exp(i\lambda) = \cos\lambda + i\sin\lambda$, since $2\sin^2\left(\frac{\lambda}{2}\right) = 1-\cos\lambda$.

4.3 The Fractional Heston Model when the Driving Stochastic Processes are Uncorrelated

The fractional Heston model when the driving stochastic processes are uncorrelated ($\rho = 0$) can be written as follows:-

$$\frac{dS_t}{S_t} = \mu dt + \sqrt{v_t} dB_t^{H^s} \tag{4.3}$$

$$dv_t = \beta(\kappa - v_t)dt + \sigma\sqrt{v_t} dB_t^{H^v} \tag{4.4}$$

where

S_t is the fractional asset price at continuous time t,

v_t is the fractional volatility at continuous time t,

μ is the rate of the asset returns,

β is the mean-reversion rate,

κ is a long run volatility,

σ is the volatility of the volatility process,

$dB_t^{H^s}$ and $dB_t^{H^v}$ are fractional Brownian motion increments and therefore fractional Gaussian noise.

In this case, $B_t^{H^s}$ and $B_t^{H^v}$ are independent fractional Brownian motion processes.

Our ultimate aim is to estimate the fractional volatility process v_t assuming that the fractional Black-Scholes process S_t defined through (4.3) is known. It is more convenient to work with $y_t = \log(S_t/S_0)$, as we mentioned in Chapter 3. Using Wick calculus (Biagini

et al., 2010) we obtain after some manipulation a solution to (4.3) as

$$S_t = S_0 \exp\left(\mu t - \frac{1}{2} v_t t^{2H^s} + \sqrt{v_t} B_t^{H^s}\right),$$

using the result at the end of Section 3.4 of Biagini et al. (2010). The only condition is that $H^s \in \left(\frac{1}{2}, 1\right)$ as suggested in Biagini et al. (2010). Thus,

$$\frac{S_t}{S_0} = \exp\left(\mu t - \frac{1}{2} v_t t^{2H^s} + \sqrt{v_t} B_t^{H^s}\right).$$

As we mentioned, $y_t = \log(S_t/S_0)$, and so we can rewrite the above equation as:

$$y_t = \mu t - \frac{1}{2} v_t t^{2H^s} + \sqrt{v_t} B_t^{H^s}.$$

This therefore motivates our fractional Heston model when $\rho = 0$ in discrete time which we define as:

$$y_{t+\triangle} = y_t + \mu \triangle - \tfrac{1}{2} v_t (\triangle)^{2H^s} + \sqrt{|v_t|}\sqrt{\triangle} Z_t^{H^s} \tag{4.5}$$

$$v_{t+\triangle} = v_t + (\alpha - \beta v_t) \triangle + \sigma \sqrt{|v_t|}\sqrt{\triangle} Z_t^{H^v}, \tag{4.6}$$

where $y_t = \log(S_t/S_0)$ and $\alpha = \beta \times \kappa$.

The absolute value of v_t is used to avoid a negative fractional volatility process arising from the discretization, as we have already discussed in Section 3.2.

4.3.1 Maximum likelihood estimation

We shall use maximum likelihood estimation defined thorough equations (4.5) and (4.6) to estimate the parameters $\theta = (\mu, \alpha, \beta, \sigma, H^s, H^v)$ of the fractional asset price process and the fractional volatility process. We will indicate the values of y and v at times $\triangle, 2\triangle, \ldots, t\triangle$ as $y_1, \ldots, y_t$ and $v_1, \ldots, v_t$, and we will write $y_{1:t}$ and $v_{1:t}$ for the collection of all these values. The volatilities $v_0, v_{1:t}$ are of course positive. Mathematically, the

likelihood function can be expressed as

$$\begin{aligned} L((y_t,v_t),(y_{t-1},v_{t-1}),\ldots,(y_1,v_1)|(y_0,v_0),\theta) &= L((y_t,v_t)|(y_{t-1},v_{t-1}),\ldots,(y_1,v_1),(y_0,v_0),\theta)\times \\ & L((y_{t-1},v_{t-1})|(y_{t-2},v_{t-2}),\ldots,(y_1,v_1),(y_0,v_0),\theta)\times\cdots \\ & \times L((y_2,v_2)|(y_1,v_1),(y_0,v_0),\theta)\times L((y_1,v_1)|(y_0,v_0),\theta) \end{aligned}$$

where

$\theta = (\mu,\alpha,\beta,\sigma,H^s,H^v)$ is a vector of parameters.

We can rewrite equations (4.5) and (4.6) as follows:-

$$Z_t^{H^s} = \frac{y_{t+\triangle} - y_t - \mu\triangle + \frac{1}{2}v_t\triangle^{2H^s}}{\sqrt{v_t}\sqrt{\triangle}} \tag{4.7}$$

$$Z_t^{H^v} = \frac{v_{t+\triangle} - v_t - (\alpha-\beta v_t)\triangle}{\sigma\sqrt{v_t}\sqrt{\triangle}} \tag{4.8}$$

$$\tag{4.9}$$

Thus, letting

$$\mu_k = \begin{pmatrix} E\left[Z_{1:k}^{H^s}\right] \\ E\left[Z_{1:k}^{H^v}\right] \end{pmatrix} \text{ and } \Sigma_k = \begin{pmatrix} \Sigma_{H^s} & 0 \\ 0 & \Sigma_{H^v} \end{pmatrix}, \tag{4.10}$$

in which Σ_k is the conditional variance matrices of $Z_{1:k}^H$ obtained using (4.1),

we can rewrite the likelihood function as follows:-

$$L((Z_t^{H^s},Z_t^{H^v}),(Z_{t-1}^{H^s},Z_{t-1}^{H^v}),\ldots,(Z_1^{H^s},Z_1^{H^v})|(y_0,v_0),\theta) = \prod_{k=1}^{t} MVN\left(\begin{pmatrix} Z_{1:k}^{H^s} \\ Z_{1:k}^{H^v} \end{pmatrix};\mu_k,\Sigma_k\right)$$

in which $MVN(x;M,S)$ is the multivariate normal probability density function with mean M and variance matrix S evaluated at x, and $Z_{1:t}^{H^s}$ and $Z_{1:t}^{H^v}$ are calculated through (4.7) and (4.8)

In general, for $k = 1, \ldots, t$, we have

$$\begin{aligned} E\left[Z_{1:k}^{H^s}\right] &= 0 \\ E\left[Z_{1:k}^{H^v}\right] &= 0. \end{aligned}$$

We compute Σ_k as described before, using equation (4.1) and find a Toeplitz covariance matrix for each process, where a Toeplitz matrix is such that each descending and ascending diagonal contains constant elements.

We can therefore estimate the parameters and fractional volatility process using maximum likelihood estimation.

4.4 The Fractional Heston Model with correlation ρ between the Driving Fractional Brownian Motion Processes

The fractional Heston Model with correlation ρ between the driving fractional Brownian motion processes can be written as follows:-

$$\begin{aligned} \frac{dS_t}{S_t} &= \mu dt + \sqrt{v_t}\left(\rho dB_t^{H^v} + \sqrt{1-\rho^2} dB_t^{H^s}\right) \\ dv_t &= \beta(\kappa - v_t)dt + \sigma\sqrt{v_t} dB_t^{H^v}, \end{aligned}$$

where the correlation parameter $\rho \in (-1, 1)$. As explained in Section 3.3, the correlation between $\rho B_t^{H^v} + \sqrt{1-\rho^2} B_t^{H^s}$ and $B_t^{H^v}$ is ρ.

As also mentioned in Section 4.3, we can solve the fractional Black-Scholes model as:

$$\frac{S_t}{S_0} = \exp\left(\mu t - \frac{1}{2} v_t t^{2H^s} + \sqrt{v_t}\left(\rho B_t^{H^v} + \sqrt{1-\rho^2} B_t^{H^s}\right)\right).$$

As we mentioned, $y_t = \log(S_t/S_0)$, and so we can rewrite the above equation as:

$$y_t = \mu t - \frac{1}{2} v_t t^{2H^s} + \sqrt{v_t}\left(\rho B_t^{H^v} + \sqrt{1-\rho^2} B_t^{H^s}\right).$$

This therefore motivates our fractional Heston model with general ρ in discrete time which we define as:

$$y_{t+\triangle} = y_t + \mu \triangle - \frac{1}{2} v_t (\triangle)^{2H^s} + \sqrt{|v_t|}\sqrt{\triangle}\left(\rho Z_t^{H^v} + \sqrt{1-\rho^2} Z_t^{H^s}\right) \quad (4.11)$$

$$v_{t+\triangle} = v_t + (\alpha - \beta v_t)\triangle + \sigma\sqrt{|v_t|}\sqrt{\triangle} Z_t^{H^v}, \quad (4.12)$$

where $y_t = \log(S_t/S_0)$ and $\alpha = \beta \times \kappa$.

The absolute value of v_t is used to avoid a negative fractional volatility process arising from discretization, as we have already discussed in Section 3.3.

4.4.1 Maximum likelihood estimation

We shall use maximum likelihood estimation defined through equations (4.11) and (4.12) to estimate the parameters $\theta = (\mu, \alpha, \beta, \sigma, \rho, H^s, H^v)$ of the processes and the fractional volatility process. Mathematically, the likelihood function can be expressed as

$$\begin{aligned} L((y_t, v_t), (y_{t-1}, v_{t-1}), \ldots, (y_1, v_1) | (y_0, v_0), \theta) &= L((y_t, v_t) | (y_{t-1}, v_{t-1}), \ldots, (y_1, v_1), (y_0, v_0), \theta) \times \\ & L((y_{t-1}, v_{t-1}) | (y_{t-2}, v_{t-2}), \ldots, (y_1, v_1), (y_0, v_0), \theta) \\ & \times \cdots \times L((y_2, v_2) | (y_1, v_1), (y_0, v_0), \theta) \\ & \times L((y_1, v_1) | (y_0, v_0), \theta) \end{aligned}$$

where

$\theta = (\mu, \alpha, \beta, \sigma, \rho, H^s, H^v)$ is a vector of parameters.

We will work with $Z_{1:k}^{H^s}$ and $Z_{1:k}^{H^v}$ to write the likelihood function.

Thus, we let

$$\mu_k = \begin{pmatrix} E\left[Z_{1:k}^{H^s}\right] \\ E\left[Z_{1:k}^{H^v}\right] \end{pmatrix} \text{ and } \Sigma_k = \begin{pmatrix} \Sigma_{H^s} & \rho\Sigma_{(H^s+H^v)/2} \\ \rho\Sigma_{(H^s+H^v)/2} & \Sigma_{H^v} \end{pmatrix},$$

in which Σ_H are the conditional variance matrices of $Z^H_{1:k}$, following the work of Jiang and Zhou (2011).

We can rewrite the likelihood function as follows:-

$$L((Z_t^{H^s}, Z_t^{H^v}), (Z_{t-1}^{H^s}, Z_{t-1}^{H^v}), \ldots, (Z_1^{H^s}, Z_1^{H^v}) | (y_0, v_0), \theta) = \prod_{k=1}^{t} MVN\left(\begin{pmatrix} Z^{H^s}_{1:k} \\ Z^{H^v}_{1:k} \end{pmatrix}; \mu_k, \Sigma_k\right), \tag{4.13}$$

In general, for $k = 1, \ldots, t$, we have

$$\begin{aligned} E\left[Z^{H^s}_{1:k}\right] &= 0 \\ E\left[Z^{H^v}_{1:k}\right] &= 0, \end{aligned}$$

and variance matrices can be computed as described before in Section 4.3.1.

We can therefore estimate the parameters and fractional volatility process using maximum likelihood estimation.

4.5 Bayesian Inference for the Fractional Heston Model with General Correlation ρ Between the Driving Stochastic Processes

We adopt the Bayesian framework, as discussed in Section 3.3.2, and base our inference on the posterior distribution of the parameters θ and the fractional volatility process in the fractional Heston model. Here we work with general correlation ρ between the driving stochastic processes. As mentioned in Section 4.4, the value of ρ is between $(-1, 1)$. We will describe two methods as follows:

4.5.1 The particle filter

We will begin by writing the fractional Heston model with general ρ as a state space model as follows:-

$$v_t = v_{t-1} + (\alpha - \beta v_{t-1}) \triangle + \sigma \sqrt{|v_{t-1}|} \sqrt{\triangle} Z_t^{H^v} \qquad \text{(unobserved equation)} \tag{4.14}$$

$$y_t = y_{t-1} + \mu \triangle - \frac{1}{2} v_{t-1} \triangle^{2H^s} + \sqrt{|v_{t-1}|} \sqrt{\triangle} \left(\rho Z_t^{H^v} + \sqrt{1-\rho^2} Z_t^{H^s} \right) \text{(observed equation)} \tag{4.15}$$

$t = 1, \ldots, T.$

In the special case when $\rho = 0$, we can rewrite equations (4.14) and (4.15) as follows:-

$$v_t = v_{t-1} + (\alpha - \beta v_{t-1}) \triangle + \sigma \sqrt{|v_{t-1}|} \sqrt{\triangle} Z_t^{H^v} \quad \text{(unobserved equation)} \tag{4.16}$$

$$y_t = y_{t-1} + \mu \triangle - \frac{1}{2} v_{t-1} \triangle^{2H^s} + \sqrt{|v_{t-1}|} \sqrt{\triangle} Z^{H^s} \text{(observed equation)} \tag{4.17}$$

$t = 1, \ldots, T.$

The aim of the particle filter is to estimate sequentially the fractional volatility process when the other parameters are known.

The basic idea of the particle filter is to generate the particles from an importance function and to weight them appropriately. Effective sample size is used to keep particles in control. This is done by re-sampling according to associated weights, meaning that particles will small associated weights are unlikely to be re-samples.

We now present the general particle filter algorithm:-

- At time $t = 1$,
  generate $v_1^{(i)} \sim g_1(v_1|v_0, y_0, \theta)$, where $g_1(v_1|v_0, y_0, \theta)$ is a probability density function chosen by the user, and set $w_1^{(i)} = N^{-1}$.

- At time $t = 2, \ldots, T, \qquad i = 1, \ldots, N.$

  - Generate the i^{th} particle of the process as $v_t^{(i)} \sim g_{t|t-1}(v_t|v_{1:t-1}^{(i)}, y_{1:t-1}, \theta)$, where $g_{t|t-1}(v_t|v_{1:t-1}^{(i)}, y_{1:t-1}, \theta)$ is an importance transition function.

- Compute the weights

$$\begin{aligned}
\widehat{w}_t^{(i)} &\propto \frac{\pi(v_{1:t}^{(i)}|y_{1:t})}{g_t(v_{1:t}^{(i)}|y_{1:t})} \propto \frac{\pi(v_{1:t}^{(i)}, y_t|y_{1:t-1})}{g_t(v_{1:t}^{(i)}|y_{1:t})} \\
&= \frac{\pi(v_t^{(i)}, y_t|v_{1:t-1}^{(i)}, y_{1:t-1})\pi(v_{1:t-1}^{(i)}|y_{1:t-1})}{g_{t|t-1}(v_t^{(i)}|v_{1:t-1}^{(i)}, y_{1:t})g_{t-1}(v_{1:t-1}^{(i)}|y_{1:t-1})} \\
&\propto \frac{\pi(v_t^{(i)}, y_t|v_{1:t-1}^{(i)}, y_{1:t-1})}{g_{t|t-1}(v_t^{(i)}|v_{1:t-1}^{(i)}, y_{1:t})} \times \widehat{w}_{t-1}^{(i)} \\
&= \frac{L(v_{1:t}^{(i)}, y_{1:t})}{L(v_{1:t-1}^{(i)}, y_{1:t-1})g_{t|t-1}(v_t^{(i)}|v_{1:t-1}^{(i)}, y_{1:t-1})} \times \widehat{w}_{t-1}^{(i)}
\end{aligned}$$

where

$L(v_{1:t}^{(i)}, y_{1:t})$ is the likelihood function at time t, and

$L(v_{1:t-1}^{(i)}, y_{1:t-1})$ is the equivalent likelihood function at time $t-1$. We suppress unnecessary parameters for notational simplicity.

- Normalize the weights

$$w_t^{(i)} = \frac{\widehat{w}_t^{(i)}}{\sum_{j=1}^{N} \widehat{w}_t^{(j)}}.$$

- Compute the effective sample size N_{eff}

$$N_{eff} = \left[\sum_{i=1}^{N} (w_t^{(i)})^2\right]^{-1}. \tag{4.18}$$

- if $N_{eff} \leqslant N_0$, where N_0 is a certain preset fraction of the total number of particles, re-sample

  - Draw a sample of size N from a discrete distribution $P\left(v_{1:t} = v_{1:t}^{(i)}\right) = w_t^{(i)}, i = 1, \ldots, N$, and relabel this sample $v_{1:t}^{(1)}, \ldots, v_{1:t}^{(N)}$.
  - Reset the weights:- $w_t^{(i)} = N^{-1}, i = 1, \ldots, N$.

4.5.2 The Auxiliary Particle with unknown parameters

The aim of the auxiliary particle is to overcome some problems that occur with the particle filter. One of these problems is inefficiency of the sampler which occurs when some of

the weights may take extreme values, known as outliers. The auxiliary particle filter uses an auxiliary variable to identify "good" particles.

Our aim is to sample from the posterior distribution of both the fractional volatility process at time t and the parameters in the fractional Heston model in two cases (when $\rho = 0$ and $\rho \neq 0$ generally) sequentially.

The methodology of the auxiliary particle filter with unknown parameters θ is:-

- At time $t-1$, a value θ can be drawn from a continuous importance density based on a discrete approximate so the distribution is then

$$\widehat{\pi}_{t-1}(v_{1:t-1}, \theta) \approx \pi(v_{1:t-1}, \theta | y_{1:t-1})$$

and the marginal distribution of θ is

$$\widehat{\pi}_{t-1}(\theta) \approx \pi(\theta | y_{1:t-1}).$$

- In addition, at time $t-1$, assuming that the target distribution of θ is a mixture of multivariate normal distribution with mean $\theta^{(i)}$ and variance matrix Λ, as suggested by Lui and West (2001), we have

$$\widehat{\pi}_{t-1}(\theta) = \sum_{i=1}^{N} w_{t-1}^{(i)} N(\theta; \theta^{(i)}, \Lambda). \tag{4.19}$$

Under the continuous distribution (4.19), the mean vector $\overline{\theta}$ is

$$\begin{aligned} \overline{\theta} &= \mathrm{E}(\theta^{(I)}) \\ &= \sum_{i=1}^{N} w_{t-1}^{(i)} \theta^{(i)}, \end{aligned}$$

where I is an auxiliary variable with $\Pr(I = i) = w_{t-1}^{(i)}$ and the covariance matrix of

θ is

$$\begin{aligned} \text{var}(\theta) &= \text{E}(\text{var}(\theta|I)) + \text{var}(\text{E}(\theta|I)) \\ &= \Lambda + \Sigma > \Sigma \end{aligned} \tag{4.20}$$

where $\Sigma = \text{var}(\theta^{(I)})$ is the covariance matrix of θ under the original discrete approximation $\hat{\pi}_{t-1}$ and the inequality holds since Λ is a positive definite matrix.

In order to reduce $\text{var}(\theta)$ to Σ, we re-defined equation (4.19) at time $t-1$ as follows:-

$$\hat{\pi}_{t-1}(\theta) = \sum_{i=1}^{N} w_{t-1}^{(i)} N(\theta; m^{(i)}, h^2\Sigma)$$

where $m^{(i)} = a\theta^{(i)} + (1-a)\overline{\theta}$, in which $a \in (0,1)$, and $a^2 + h^2 = 1$.

We compute the new expected value and covariance matrix as follows

$$\begin{aligned} \text{E}(\theta) &= \text{E}(\text{E}(\theta|I)) \\ &= a\overline{\theta} + (1-a)\overline{\theta} \\ &= \overline{\theta}; \\ \text{var}(\theta) &= \text{E}(\text{var}(\theta|I)) + \text{var}(\text{E}(\theta|I)) \\ &= h^2\Sigma + a^2\text{var}(\theta^{(I)}) \\ &= h^2\Sigma + a^2\Sigma \\ &= \Sigma. \end{aligned} \tag{4.21}$$

By comparing (4.21) and (4.20), it is obvious that (4.21) is better than (4.20) because the variance in (4.21) is smaller.

Therefore, the joint distribution for the discrete $v_{1:t-1}$ and the continuous θ takes the form

$$\hat{\pi}_{t-1}(v_{1:t-1}, \theta) = \sum_{t=1}^{N} w_{t-1}^{(i)} N(\theta; m^{(i)}, h^2\Sigma) \delta_{v_{1:t-1}^{(i)}}$$

where $\delta_{v_{1:t-1}^{(i)}}$ is the unit mass at $v_{1:t-1}^{(i)}$.

- At time t, the target distribution will now become

$$\begin{aligned}
\pi(v_{1:t},\theta|y_{1:t}) &\propto \pi(v_{1:t},\theta,y_t|y_{1:t-1}) \\
&= \pi(v_t,y_t|v_{1:t-1},y_{1:t-1},\theta)\pi(v_{1:t-1},\theta|y_{1:t-1}) \\
&\approx \pi(y_t|v_{t-1},y_{t-1},\theta)\pi(v_t|v_{t-1},\theta)\hat{\pi}_{t-1}(v_{1:t-1},\theta) \\
&= \sum_{i=1}^{N} w_{t-1}^{(i)}\pi(y_t|v_{t-1}^{(i)},y_{t-1},\theta)\pi(v_t|v_{t-1}^{(i)},\theta)N(\theta;m^{(i)},h^2\Sigma)\delta_{v_{1:t-1}^{(i)}}.
\end{aligned}$$

The auxiliary particle filter algorithm for sequentially sampling from the posterior distribution of the fractional volatility process and the parameters is therefore:-

- Initialize at time $t = 1$.

  Generate $(v_1^{(1)},\ldots,v_1^{(N)}) \sim \pi^v(v_1)$ and $(\theta^{(1)},\ldots,\theta^{(N)}) \sim \pi^\theta(\theta)$, where $\pi^v(v_1)$ and $\pi^\theta(\theta)$ are prior probability density functions chosen by the user. Set $w_1^{(i)} = N^{-1}, i = 1,\ldots,N$.

- Computing and generating for $t = 2,\ldots,T$ and $i = 1,\ldots,N$:

  - Compute the mean of $\overline{\theta} = \mathrm{E}_{\hat{\pi}_{t-1}}(\theta)$ and the variance $\Sigma = \mathrm{var}_{\hat{\pi}_{t-1}}(\theta)$. The particle $\theta^{(i)}$ will depend on the new mean of θ :

$$m^{(i)} = a\theta^{(i)} + (1-a)\overline{\theta}.$$

  - Generate an auxiliary variable I_k for $k = 1,\ldots,N$ where

$$p(I_k = i) \propto w_{t-1}^{(i)}\pi(y_t|v_{t-1} = v_{t-1}^{(i)},y_{t-1},\theta = m^{(i)}).$$

  - Generate

$$\begin{aligned}
\theta^{(k)} &\sim N(m^{(I_k)},h^2\Sigma) \\
v_t^{(k)} &\sim \pi(v_t|v_{t-1} = v_{t-1}^{(I_k)},\theta = \theta^{(k)})
\end{aligned}$$

and set $v_{1:t}^{(k)} = (v_{1:t-1}^{(I_k)}, v_t^{(k)})$.

- Compute the weights

$$\widehat{w}_t^{(i)} = \frac{L(v_{1:t}^{(k)}, y_{1:t}|y_0, v_0, \theta = \theta^{(k)})}{L(v_{1:t-1}^{(k)}, y_{1:t-1}|y_0, v_0, \theta = \theta^{(k)})\,\pi(y_t|v_{t-1}^{(I_k)}, y_{t-1}, \theta = m^{(I_k)})\,\pi(v_t^{(i)}|v_{t-1}^{(I_k)}, \theta = m^{(I_k)})}$$

and normalize them

$$w_t^{(i)} = \frac{\widehat{w}_t^{(i)}}{\sum_{j=1}^{N} \widehat{w}_t^{(j)}}.$$

- Compute the effective sample size N_{eff} using (4.18).

If $N_{eff} \leqslant N_0$, where N_0 is a certain preset fraction of the total number of particles, re-sample

- Draw a sample of size N from a discrete distribution $P((v_{1:t}, \theta) = (v_{1:t}^{(i)}, \theta^{(i)})) = w_t^{(i)}, i = 1, \ldots, N$, and relabel this sample $(v_{1:t}^{(1)}, \theta^{(1)}), \ldots, (v_{1:t}^{(N)}, \theta^{(N)})$.

- Reset the weights:-

$$w_t^{(i)} = N^{-1}, \quad i = 1, \ldots, N.$$

4.6 Forecasting

Forecasting is a very useful tool to estimate values within and beyond an underlying process (Lunn et al., 2013). In our work, we will forecast the stock price s_t of the Heston and fractional Heston models.

Mathematically, if we let $s_{t^{\text{new}}}$ be the required forecast of the asset price at time t^{new}, then we base our estimate of $s_{t^{\text{new}}}$ on the predictive density

$$\pi(s_{t^{\text{new}}}|s_{1:t}) = \int_{\emptyset} \pi(s_{t^{\text{new}}}|\emptyset)\pi(\emptyset|s_{1:t})d\emptyset,$$

in which $s_{1:t}$ are the data available up to time t, the integration is over the vector $\emptyset = (\theta, v_t)$ of unknown parameter, where $\theta = (\mu, \alpha, \beta, \sigma, \rho, H^s, H^v)$ and v_t is the volatility at time t, and $\pi(\emptyset|s_t)$ is the posterior density.

For example, we will consider the expectation of a future value $s_{t^{\text{new}}}$ as follows:-

$$\begin{aligned} E_{\pi(s_{t^{\text{new}}}|s_{1:t})}\left[s_{t^{\text{new}}}\right] &= \int_{s_{t^{\text{new}}}} s_{t^{\text{new}}} \pi(s_{t^{\text{new}}}|s_{1:t}) ds_{t^{\text{new}}}, \\ &= \int_{s_{t^{\text{new}}}} s_{t^{\text{new}}} \int_{\emptyset} \pi(s_{t^{\text{new}}}|\emptyset)\pi(\emptyset|s_t) d\emptyset\, ds_{t^{\text{new}}}, \\ &= \int_{\emptyset}\int_{s_{t^{\text{new}}}} s_{t^{\text{new}}} \pi(s_{t^{\text{new}}}|\emptyset) ds_{t^{\text{new}}}\, \pi(\emptyset|s_{1:t})\, d\emptyset, \\ &= E_{\pi(\emptyset|s_{1:t})}\left[E_{\pi(s_{t^{\text{new}}}|\emptyset)}\left[s_{t^{\text{new}}}\right]\right]. \end{aligned}$$

We can use our posterior sample to approximate this quantity as follows

$$E_{\pi(s_{t^{\text{new}}}|s_{1:t})}\left[s_{t^{\text{new}}}\right] = \frac{1}{N}\sum_{i=1}^{N} E_{\pi(s_{t^{\text{new}}}|\emptyset^{(i)})}\left[s_{t^{\text{new}}}\right] w_t^{(i)},$$

where $\emptyset^{(i)} = (\theta^{(i)}, v_t^{(i)})$ is obtained from the i^{th} particle with associated weights $w_t^{(i)}$. Furthermore, we can use simulated value of $s_{t^{\text{new}}}$ to approximate $E_{\pi(s_{t^{\text{new}}}|\emptyset)}\left[s_{t^{\text{new}}}\right]$. Together we obtain this approximation:

$$E_{\pi(s_{t^{\text{new}}}|s_{1:t})}\left[s_{t^{\text{new}}}\right] = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{M}\sum_{j=1}^{M} s_{t^{\text{new}}}^{(i,j)}\right) w_t^{(i)}, \tag{4.22}$$

where $s_{t^{\text{new}}}^{(i,j)}$ is found through equations (4.11) and (4.12) using $\emptyset^{(i)}$.

To quantify the accuracy of our forecast, we will compute the Absolute and Square Root differences between our forecast and an omitted part of the real data. These differences can be written as follow:

$$\text{Absolute difference} = \sum_j |\, s_j - \text{forecast value of } s_j \,| \tag{4.23}$$

$$\text{Square Root difference} = \sqrt{\sum_j \left(s_j - \text{forecast value of } s_j\right)^2}, \tag{4.24}$$

in which s_j are the values of the process that we are trying to forecast and that here are assumed known.

4.7 Results

We apply the particle filter based methodology to simulated data in Section 4.7.1 and to real data in Section 4.8.

4.7.1 Application to Simulated Case

4.7.1.1 The fractional Heston Model when $\rho = 0$

We will begin by estimating the parameters $\theta = (\mu, \alpha, \beta, \sigma, H^s, H^v)$ of the fractional Heston model with $\rho = 0$ as defined through (4.16) and (4.17) using maximum likelihood estimation as explained in Section 4.3.1.

We begin by considering a simulated dataset of size $T = 300$ with

$$\theta = (\mu, \alpha, \beta, \sigma, H^s, H^v) = (0.2, 2.497, 11.35, 0.618, 0.6, 0.4).$$

The initial values of the fractional Black-Scholes process S_0 and fractional volatility process v_0 are 1 and 0.04, respectively. We used `R`'s function `optim` to implement maximum likelihood estimation.

We applied the particle filter to the fractional Heston model when $\rho = 0$, as explained in Section 4.5.1. The true values of the parameters $\theta = (\mu, \alpha, \beta, \sigma, H^s, H^v)$, and the initial values of each process and the number of observation for the fractional Heston model are given above.

Figure 4.2 shows the results for the fractional volatility process assuming that the θ parameters are known. In particular, the true fractional volatility process, the approximate posterior mean of the fractional volatility process and approximate 90% credible intervals are shown across time t. In general, the posterior mean follows the true fractional volatility process quite well. The true fractional volatility process is almost always the inside the 90% credible intervals indicating that our estimation methodology works well in this case.

Figure 4.3 shows the estimation of the parameters $\theta = (\mu, \alpha, \beta, \sigma, H^s, H^v)$ across time using the auxiliary particle filter, assuming that the fractional volatility process is unknown. In particular, the posterior mean and approximate 90% credible intervals for each parameter are presented together. In general, good results are achieved using the auxiliary particle filter, although it turns out here to be more difficult to make inference about μ and σ than about the other parameters. In particular, the estimate of H^s and H^v are seen to be good here.

Figure 4.4 shows the results of the auxiliary particle filter for estimating the fractional volatility process v_t, assuming that the parameters are unknown. In particular, we see the fractional volatility process itself, the approximate posterior mean of the fractional volatility process and associated approximate 90% credible intervals. Good estimation of the fractional volatility process is achieved here using the auxiliary particle filter when the parameters are unknown because the true fractional volatility process is almost always inside 90% credible intervals.

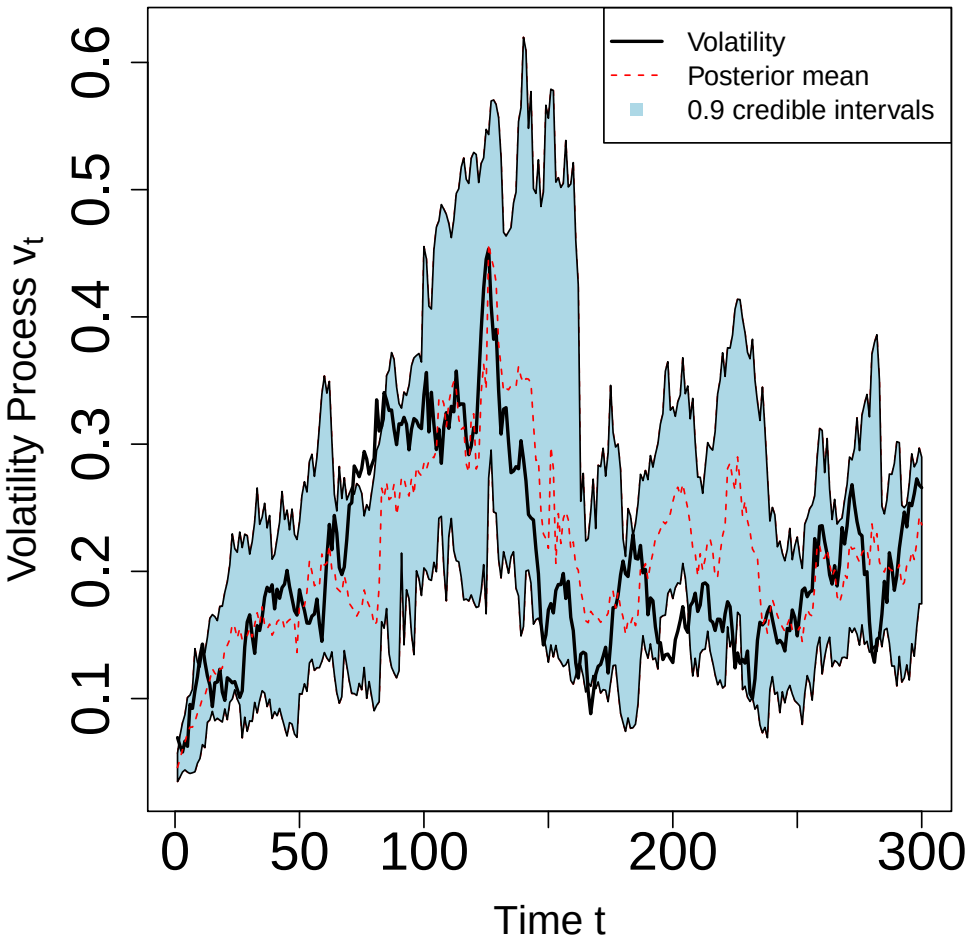


Figure 4.2: Estimation using the particle filter of the fractional volatility process v_t of the fractional Heston Model when $\rho = 0$ assuming that the other parameters are known. The true fractional volatility is represented by the black line. The posterior mean of the fractional volatility process is indicated by the dashed red line and the associated 90% credible intervals are in light blue.

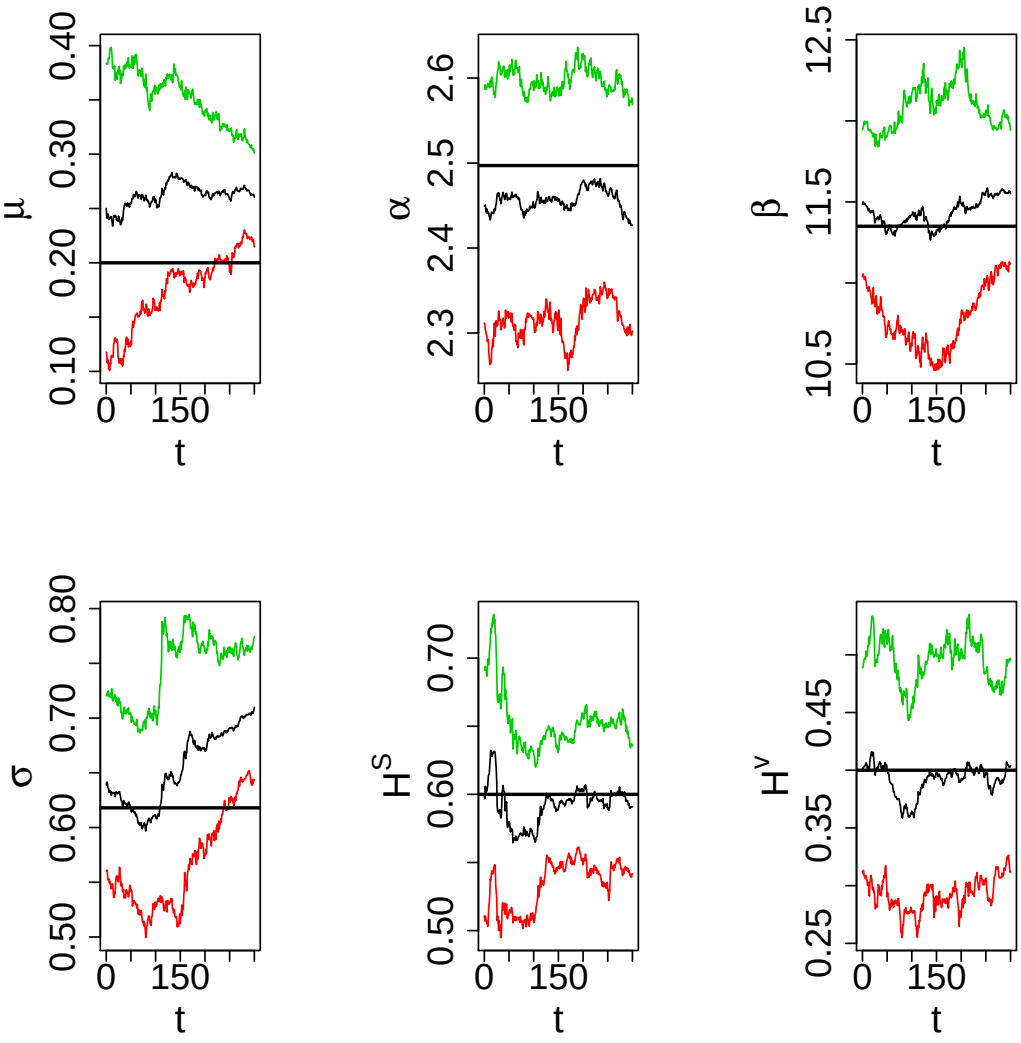


Figure 4.3: Estimation of the parameters of the fractional Heston model when $\rho = 0$ using the auxiliary particle filter. In each graph, the posterior mean is shown using the black trace, while the associated approximate 90% credible intervals are shown by the outer traces. The first graph is for the parameter μ, the second graph is for the parameter α, the third graph is for the parameter β, the fourth graph is for the parameter σ, the fifth graph is for the parameter H^s and the final graph is for the parameter H^v. The true values are shown by the horizontal lines.

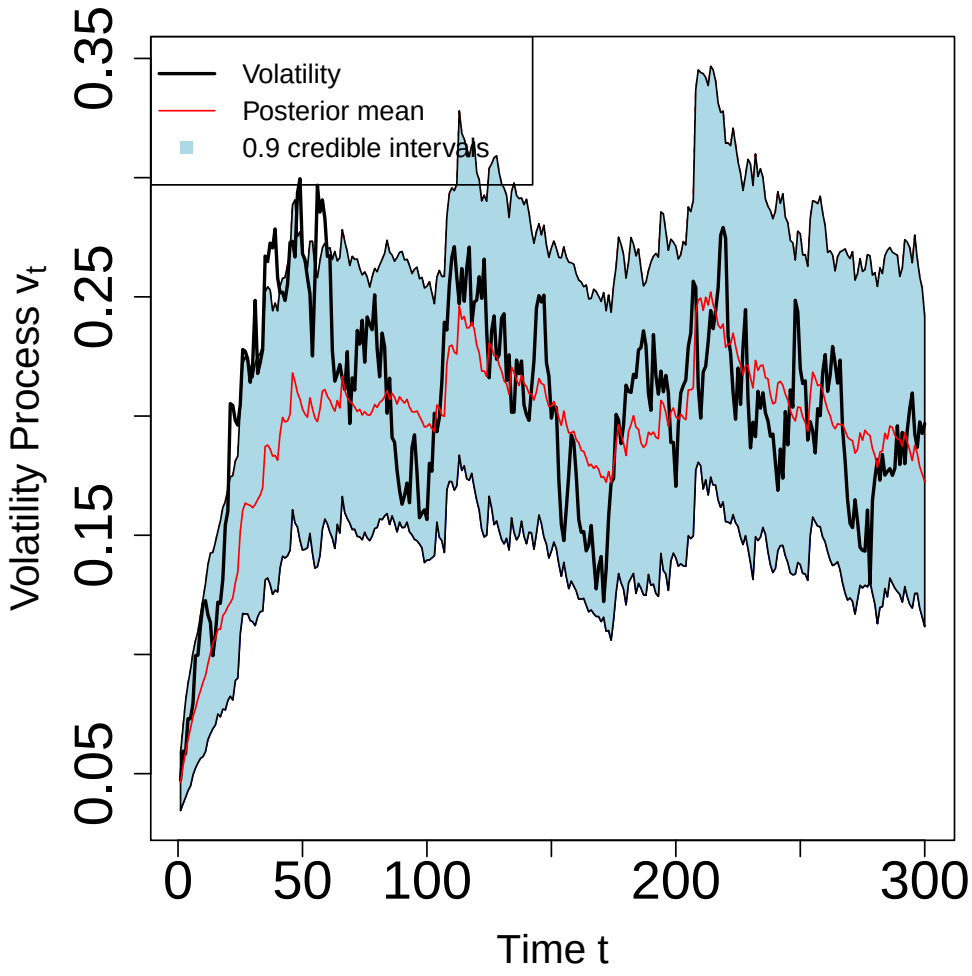


Figure 4.4: Estimation using the auxiliary particle filter of the fractional volatility process v_t in the fractional Heston model when $\rho = 0$ when the other parameters are unknown. The true fractional volatility process is represented by the black line. The posterior mean of the fractional volatility process is represented by the red line, and the associated 90% credible interval are in light blue.

4.7.1.2 The fractional Heston Model with ρ defined in equations (4.11) and (4.12) of Section 4.4

We will illustrate the results of the fractional Heston model including the correlation coefficient ρ between the stochastic processes that drive the fractional Black-Scholes and fractional volatility processes. The aim here is to estimate the parameters $\theta = (\mu, \alpha, \beta, \sigma, \rho, H^s, H^v)$ and the fractional volatility process v_t using particle filter based methodology.

We work with simulated data sets of size $T = 300$ with

$\theta = (\mu, \alpha, \beta, \sigma, \rho, H^s, H^v) = (0.2, 2.497, 11.35, 0.618, -0.2, 0.6, 0.4)$,

or $(0.2, 2.497, 11.35, 0.618, 0.2, 0.6, 0.4)$, so that we consider the cases when the correlation coefficient ρ takes either negative or positive values. The initial values of the fractional Black-Scholes process and fractional volatility process remain $S_0 = 1$ and $v_0 = 0.04$, respectively.

Figures 4.5 and 4.6 show the results of the particle filter for the fractional Heston model with the true values of $\rho = -0.2$ and $\rho = 0.2$, respectively, assuming known parameters as discussed in Section 4.5.1. In particular, the true fractional volatility process, the approximate posterior mean of the fractional volatility process and approximate 90% credible intervals are shown across time t. Here, the estimation of the fractional volatility process when $\rho = -0.2$ is better than the estimation of the fractional volatility process when $\rho = 0.2$ because the fractional volatility process in this case follows the posterior mean quite well.

Figure 4.7 shows the estimation of the parameters θ of the fractional Heston model with the given negative value -0.2 of ρ across time using the auxiliary particle filter as discussed in Section 4.5.2. In particular, the posterior mean and approximate 90% credible intervals of each parameters are presented together.

Figure 4.8 shows the results of the auxiliary particle filter for estimating the fractional volatility process assuming that the θ parameters are unknown with $\rho = -0.2$. In particular, we see the fractional volatility process itself, the approximate posterior mean of the fractional volatility process and associated approximate 90% credible intervals.

Figure 4.9, which is similar to Figure 4.7 shows the estimation of the parameters θ of the

fractional Heston model with the given positive value of ρ across time using the auxiliary particle filter as discussed in Section 4.5.2.

Figure 4.10, which is similar to Figure 4.8 shows the results of the auxiliary particle filter for estimating the fractional volatility process assuming that the θ parameters are unknown with $\rho = 0.2$.

From these results we can see that particle filter based algorithms can produce some good estimates of the unknown quantities of the fractional Heston model. The results for the fractional Heston model with the given positive value of ρ are somewhat better than the results for the fractional Heston model with the given negative ρ.

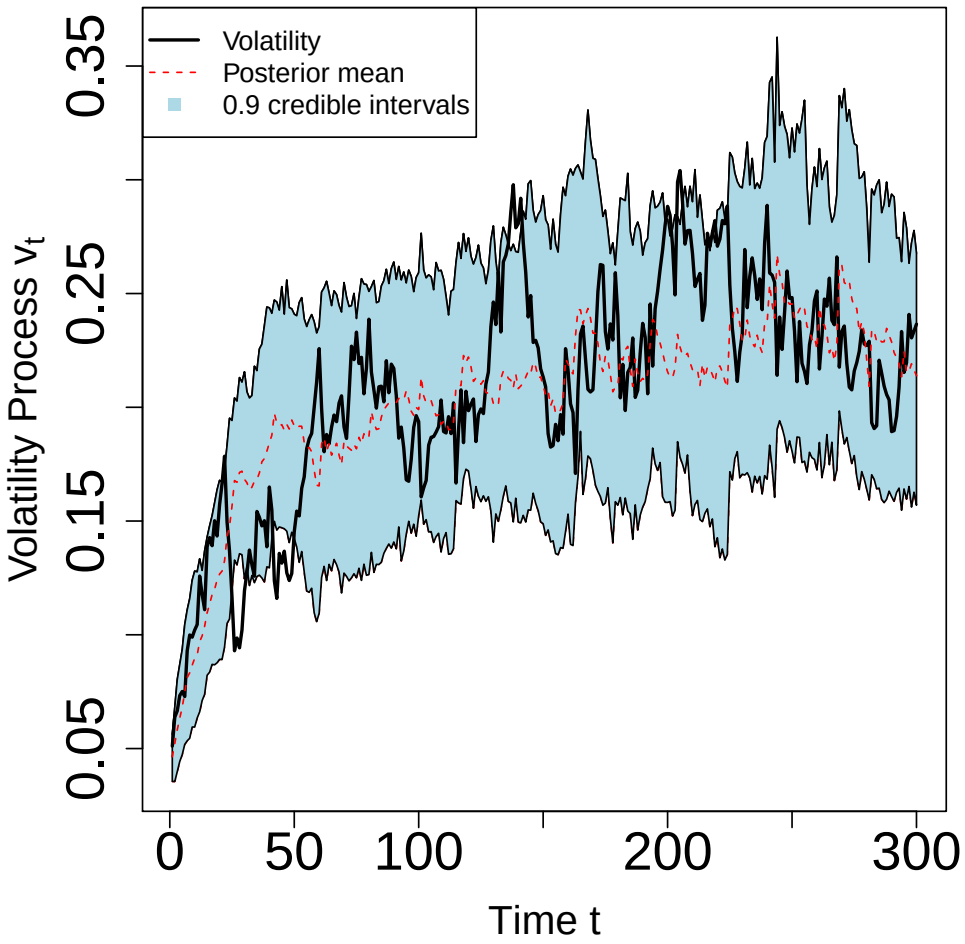


Figure 4.5: Estimation using the particle filter of the fractional volatility process v_t of the fractional Heston model with $\rho = -0.2$ assuming that the other parameters are known. The true fractional volatility is represented by the black line. The posterior mean of the fractional volatility process is indicated by the dashed red line and the associated 90% credible intervals are in light blue.

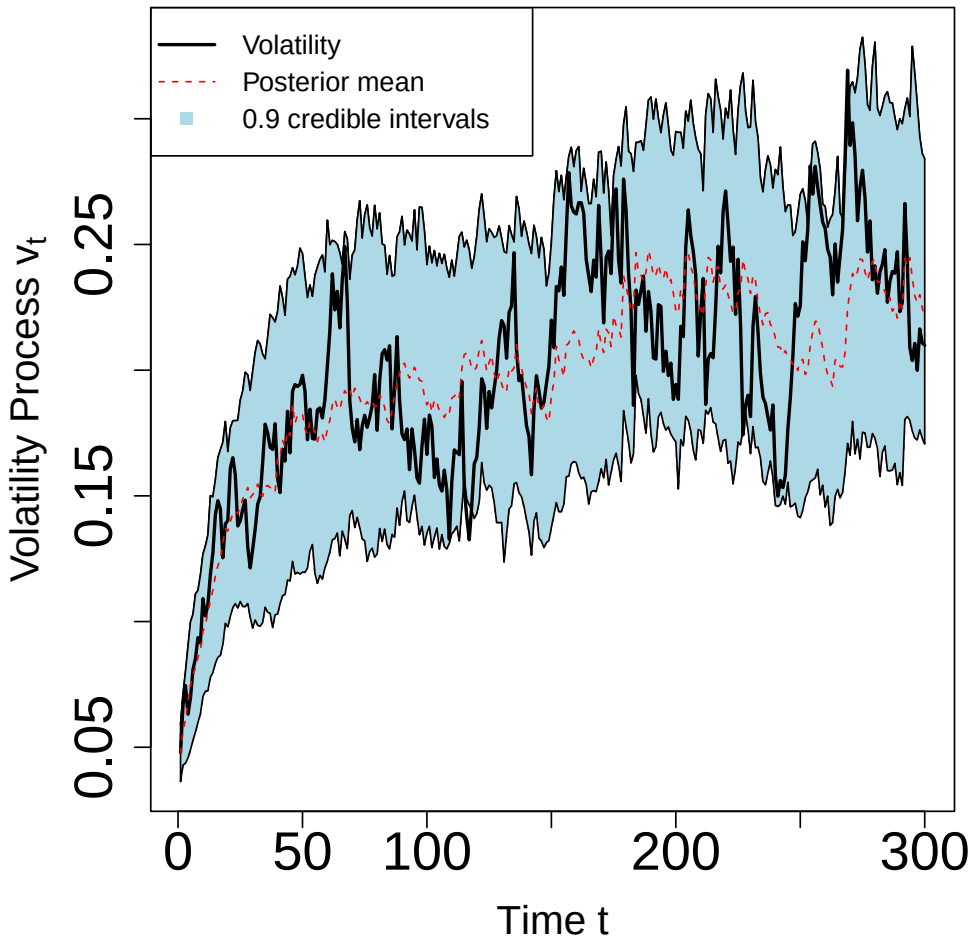


Figure 4.6: Estimation using the particle filter of the fractional volatility process v_t of the fractional Heston model with $\rho = 0.2$ assuming that the other parameters are known. The true fractional volatility is represented by the black line. The posterior mean of the fractional volatility process is indicated by the dashed red line and the associated 90% credible intervals are in light blue.

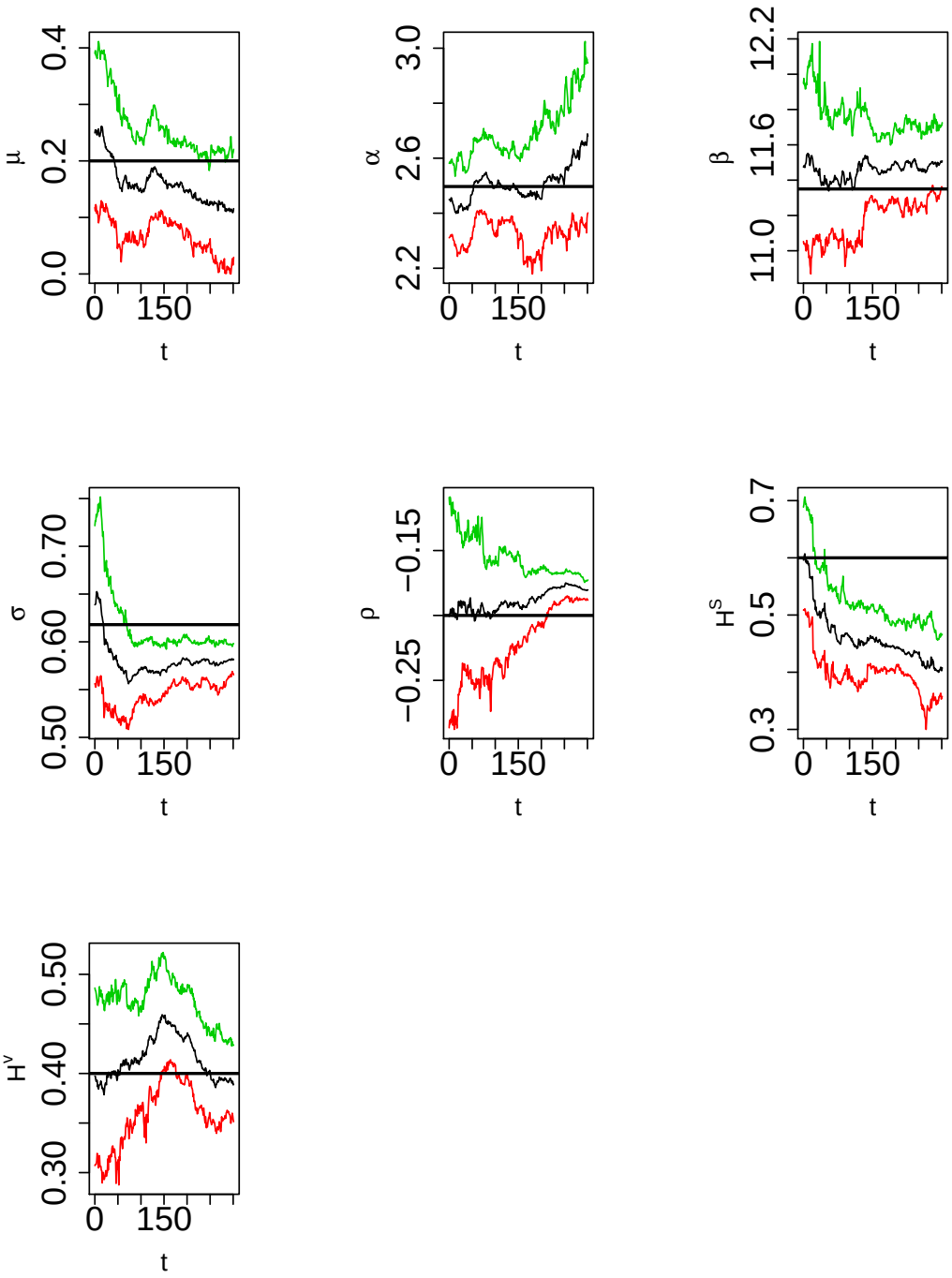


Figure 4.7: Estimation of the parameters of the fractional Heston model with non-zero ρ using the auxiliary particle filter. In each graph, the posterior mean is shown using the black trace, while the associated approximate 90% credible intervals are shown by the outer traces. The first graph is for the parameter μ, the second graph is for the parameter α, the third graph is for the parameter β, the fourth graph is for the parameter σ, the fifth graph is for the parameter ρ, the sixth graph is for the parameter H^s and the final graph is for the parameter H^v. The true values are shown by the horizontal line, e.g., $\rho = -0.2$.

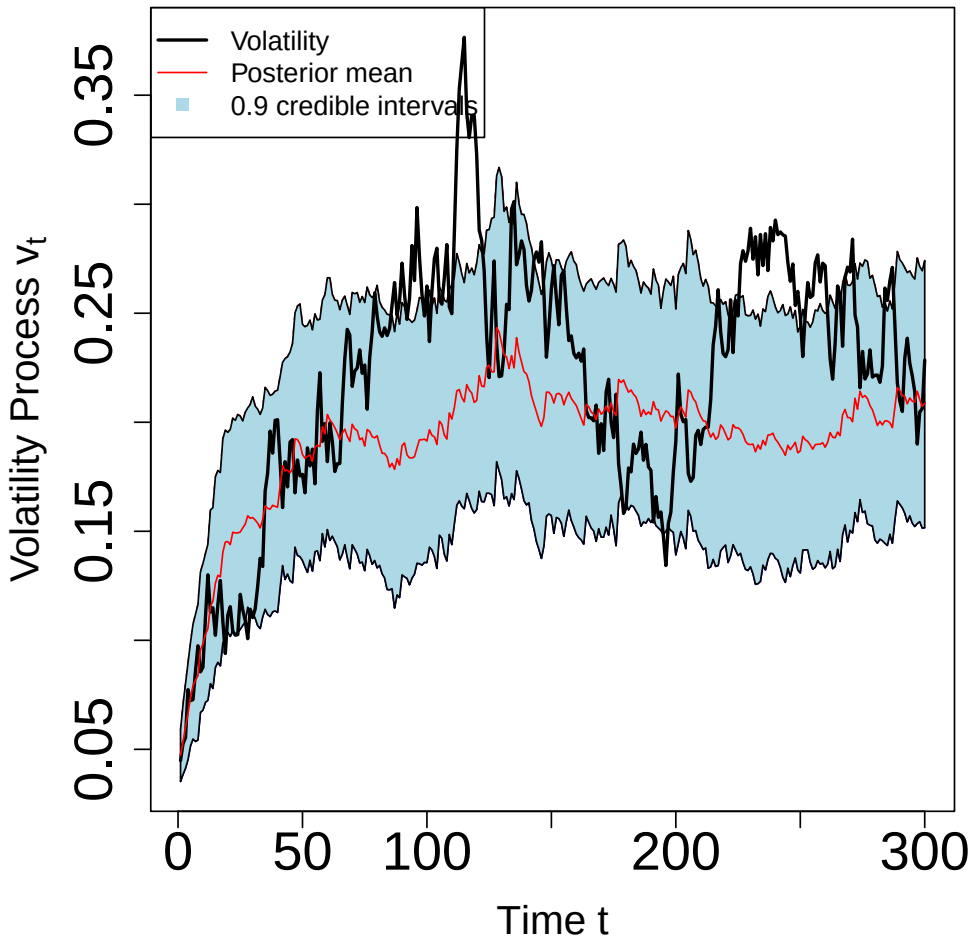


Figure 4.8: Estimation of the volatility process v_t in the fractional Heston model with non-zero ρ using the auxiliary particle filter when the θ parameters are unknown, with a negative value of ρ; $\rho = -0.2$. The posterior mean of the fractional volatility process is represented by the red line. The true fractional volatility is represented by the black line, and the associated 90% credible intervals are in light blue.

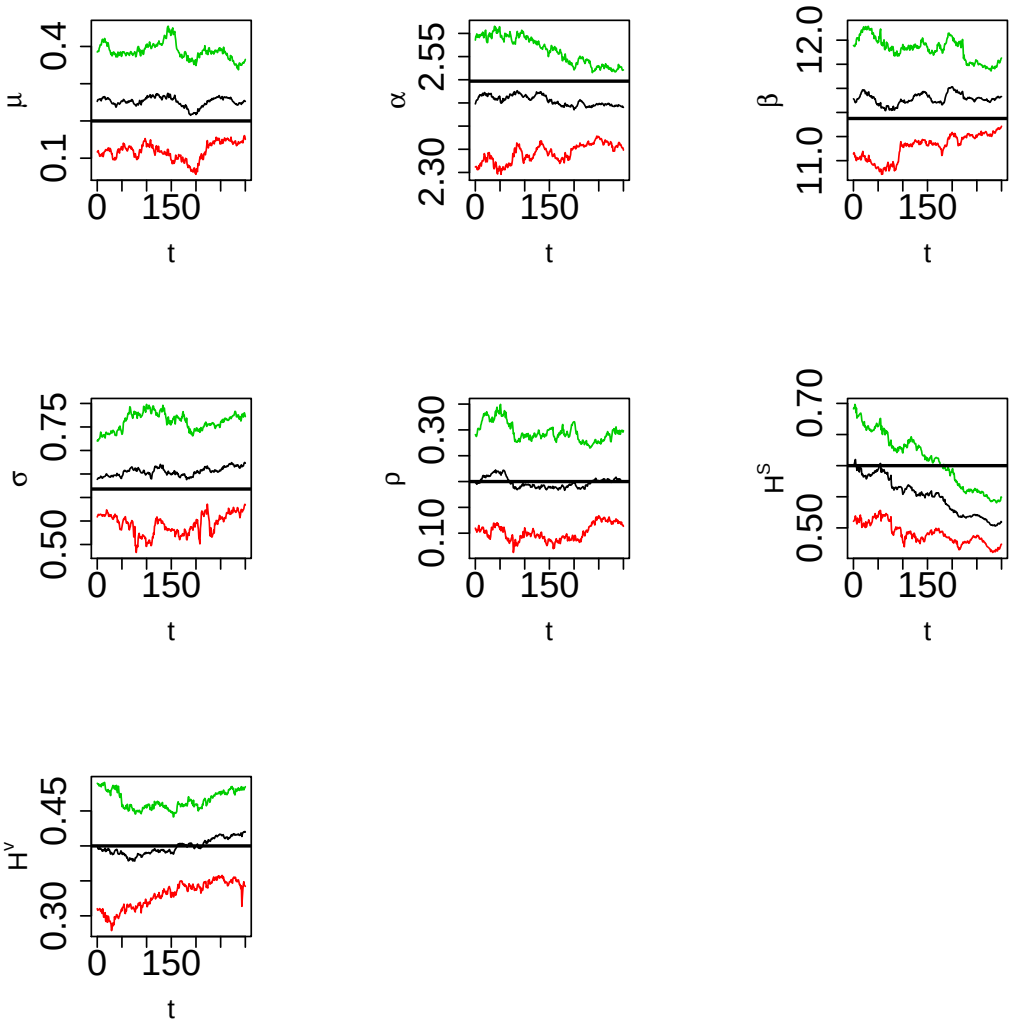


Figure 4.9: Estimation of the parameters of the Heston model with non-zero ρ using the auxiliary particle filter. In each graph, the posterior mean is shown using the black trace, while the associated approximate 90% credible intervals are shown by the outer traces. The first graph is for the parameter μ, the second graph is for the parameter α, the third graph is for the parameter β, the fourth graph is for the parameter σ, the fifth graph is for the parameter ρ, the sixth graph is for the parameter H^s and the final graph is for the parameter H^v. The true values are shown by the horizontal line, e.g., $\rho = 0.2$.

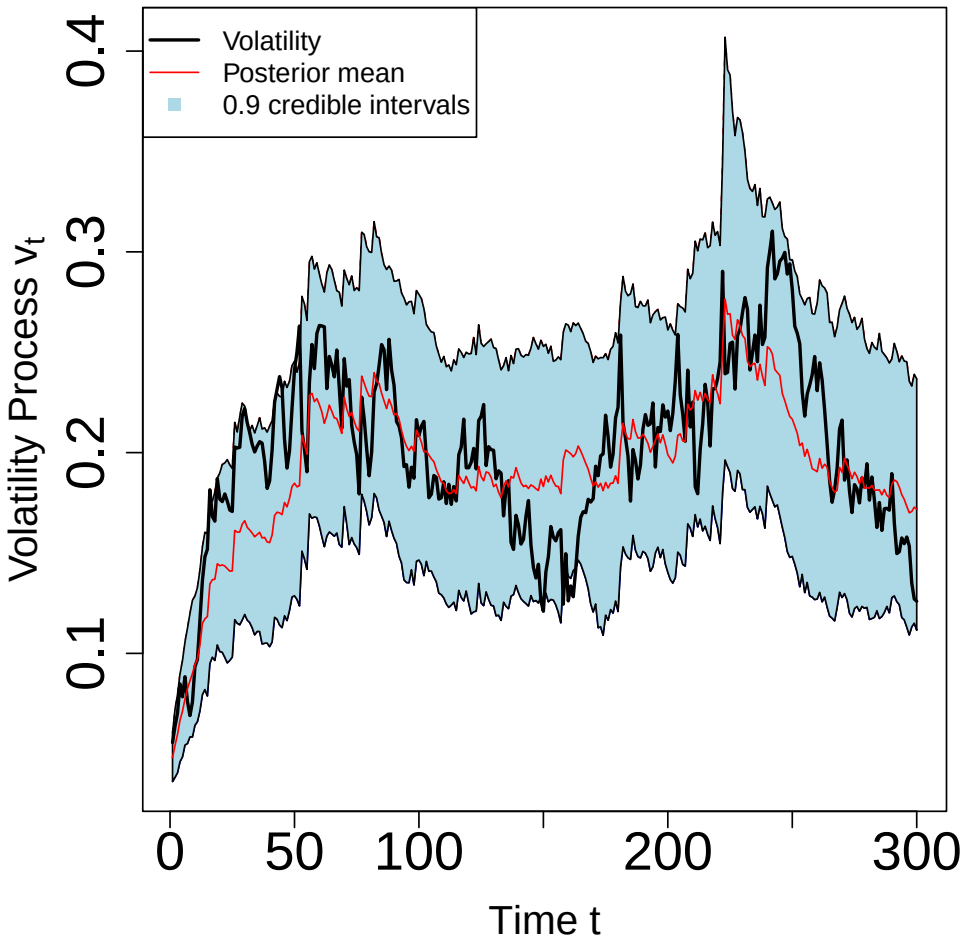


Figure 4.10: Estimation of the volatility process v_t in the fractional Heston model with non-zero ρ using the auxiliary particle filter when the θ parameters are unknown, with a positive value of ρ; $\rho = 0.2$. The posterior mean of the fractional volatility process is represented by the red line. The true fractional volatility is represented by the black line, and the associated 90% credible intervals are in light blue.

4.8 Application to Data from the Standard and Poor's Index

Similar to the work described in Section 3.4.2, we now apply the fractional Heston model when $\rho = 0$ and with non-zero ρ to the SP350 data shown in Figure 4.11. The number of observation is $T = 308$.

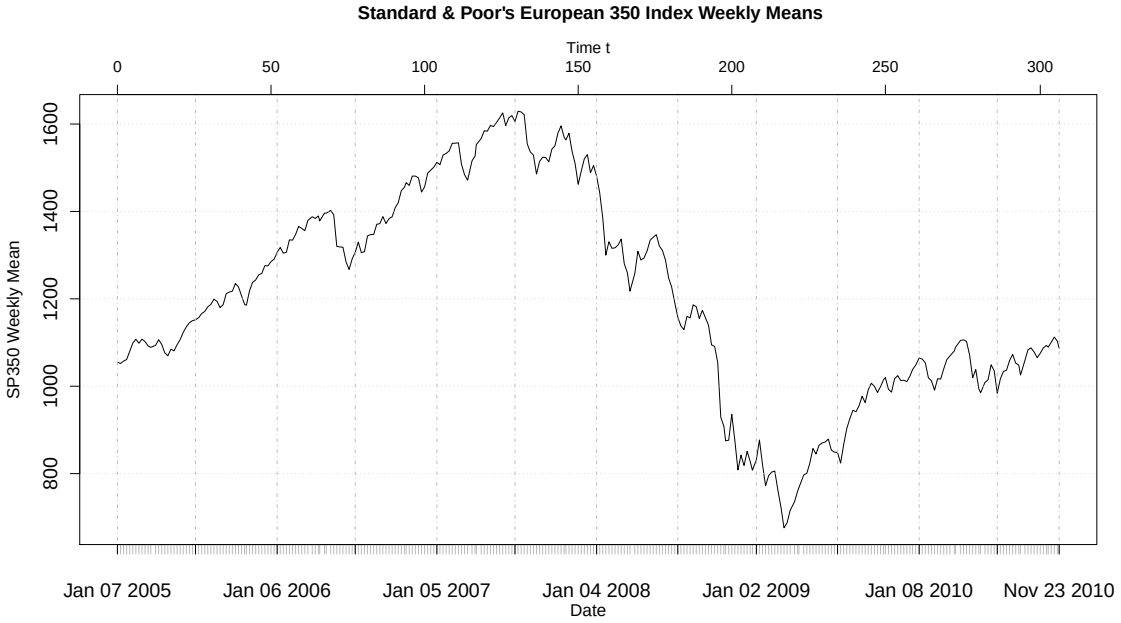


Figure 4.11: Graph of the Standard & Poor's weekly means Index from January 2005 to December 2010.

The fractional Heston Model when $\rho = 0$

In this section, we will apply the fractional Heston Model with $\rho = 0$, as defined though (4.16) and (4.17) in Section 4.5.1, to the SP350 data shown in Figure 4.11.

The values of the parameters are set to $\mu = 0.2$, $\alpha = 2.497$, $\beta = 11.35$, $\sigma = 0.618$, $H^s = 0.6$ and $H^v = 0.4$, obtained from maximum likelihood estimation as it seems to coincide with some of the features that you see in the data. Figure 4.12 shows the results of applying the particle filter to the fractional Heston model with $\rho = 0$, assuming that the other θ parameters are known. In particular, the posterior mean representing the estimation of the fractional volatility process together with approximate 90% credible intervals are shown across time. In general, we can say that the estimated fractional v_t

process is sensible when the other parameters are estimated using maximum likelihood estimation.

Figure 4.13 shows the results of estimating the θ parameters of the fractional Heston model with $\rho = 0$ across time t, using the auxiliary particle filter. In particular, the posterior mean and approximate 90% credible intervals of each parameter are presented together.

Figure 4.14 shows the results of the auxiliary particle filter for the estimated fractional volatility process when the parameters are unknown. In particular, the estimated posterior mean of the fractional volatility process and approximate 90% credible intervals are shown across time t. We note from Figure 4.13 that the posterior distribution for H^s does not contain 0.5 suggesting that the fractional Heston model provides some advantage over the standard Heston model discussed in Chapter 3.

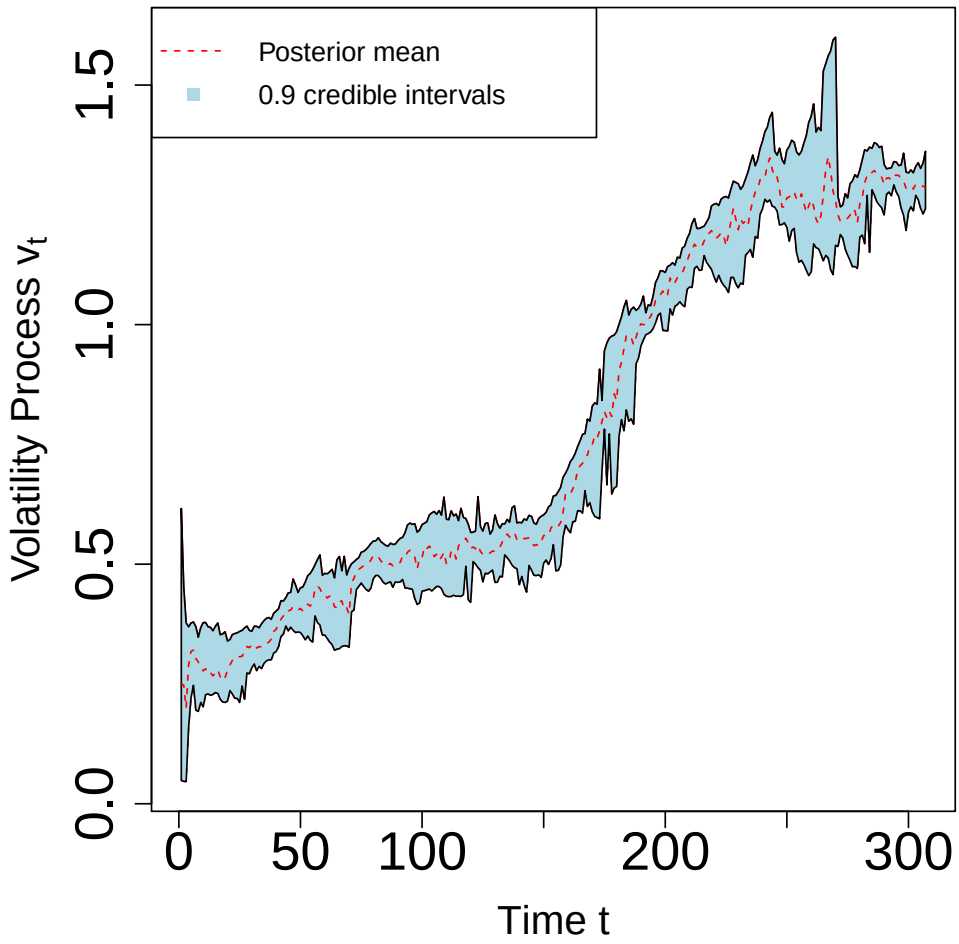


Figure 4.12: Estimation of the fractional volatility process v_t from the SP350 data using the particle filter for the fractional Heston model with $\rho = 0$ with the other θ parameters set to their maximum likelihood estimates. The approximate posterior mean of the fractional volatility process is represented by the dashed red line and the associated 90% credible intervals are in light blue.

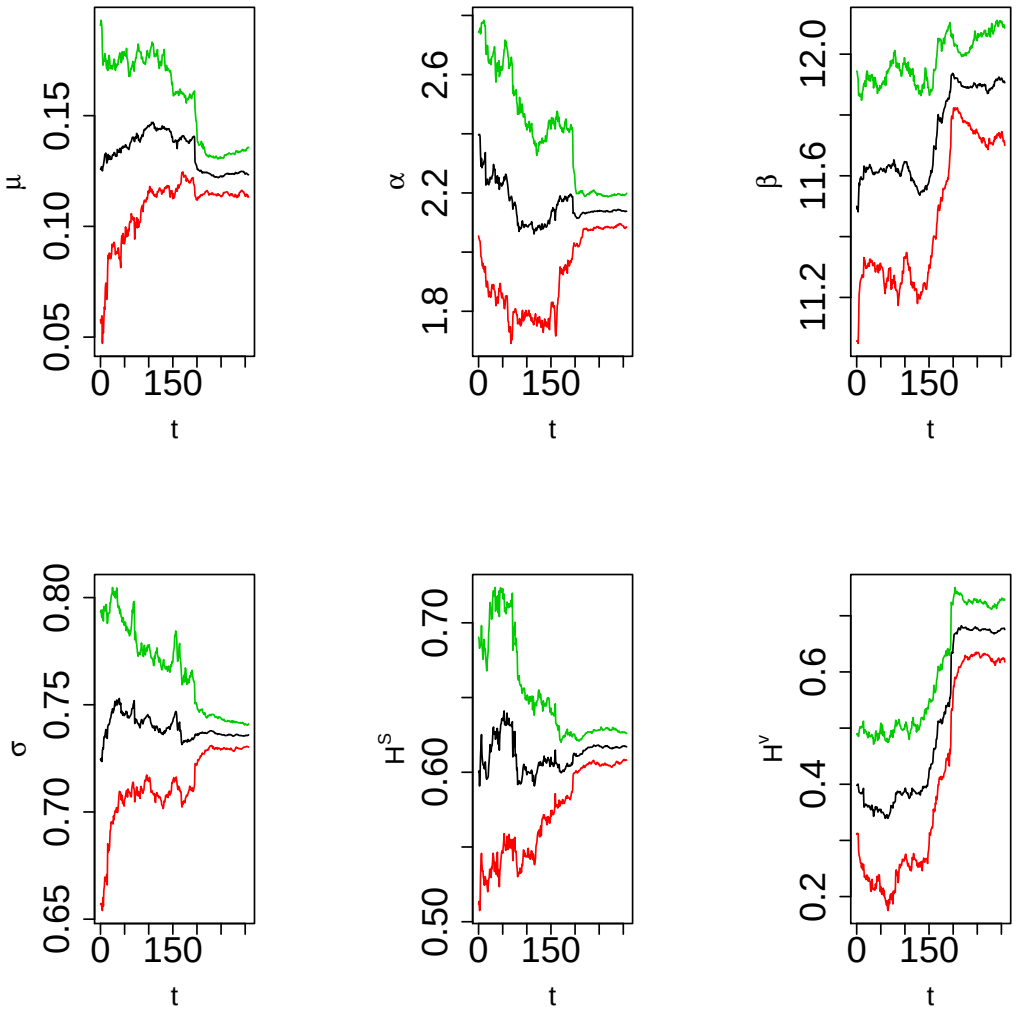


Figure 4.13: Estimation of the parameters of the fractional Heston model with $\rho = 0$ across time t, using the auxiliary particle filter applied to the SP350 data. In each graph, the posterior mean is shown using the black trace, while the associated approximate 90% credible intervals are shown by the outer traces. The first graph is for the parameter μ, the second graph is for the parameter α, the third graph is for the parameter β and the fourth graph is for the parameter σ, the fifth graph is for the parameter H^s and the final graph is for the parameter H^v.

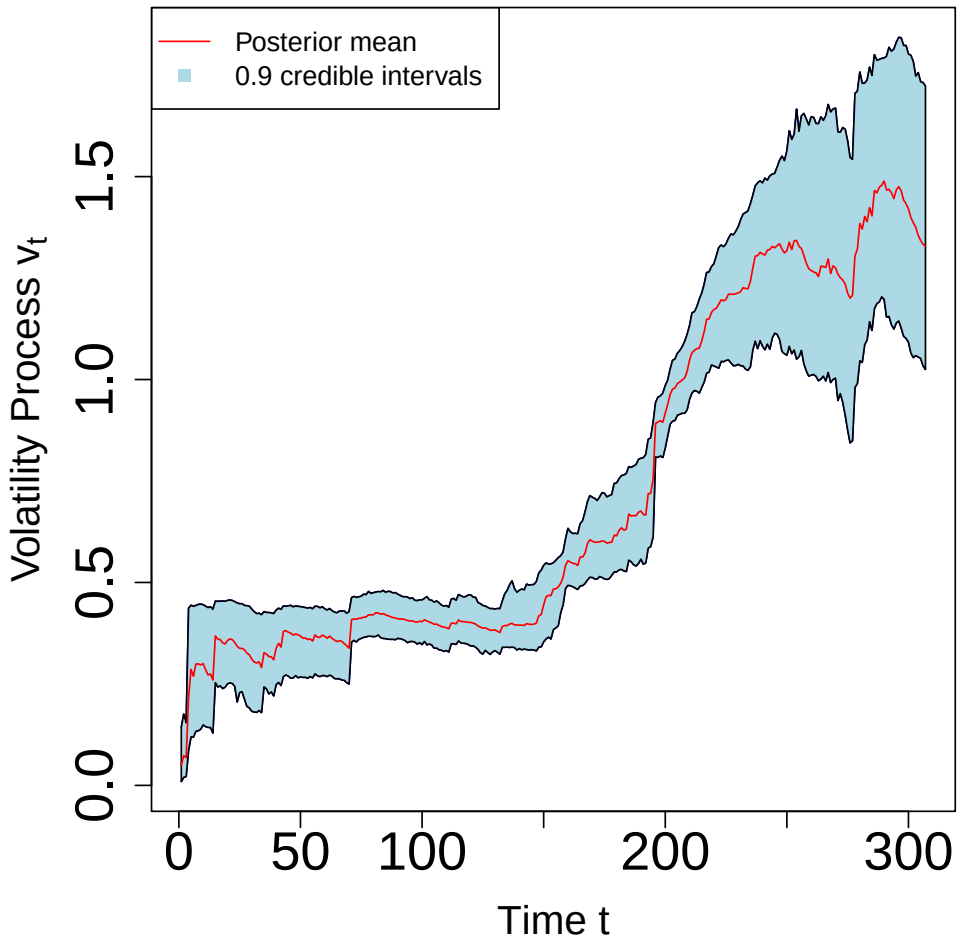


Figure 4.14: Estimation of the fractional volatility process v_t of SP350 data in the fractional Heston model with $\rho = 0$ when the other θ parameters are unknown. The estimated posterior mean of the volatility process is represented by the red line and the associated 90% credible intervals are in light blue.

The fractional Heston model with ρ

We will now include the non-zero correlation coefficient ρ in the fractional Heston model, as defined in equations (4.14) and (4.15) in Section 4.5.1. Again, we will use the SP350 data shown in Figure 4.11. The number of observation is $T = 308$.

Figure 4.15 shows the results of applying the particle filter to the fractional Heston model with non-zero ρ, assuming the other θ parameters are known. In particular, the posterior mean, representing the estimation of the fractional volatility process, and approximate 90% credible intervals are shown across time. The values of the parameters are set to $\mu = 0.2$, $\alpha = 2.497$, $\beta = 11.35$, $\sigma = 0.618$, $\rho = -0.2$, $H^s = 0.6$ and $H^v = 0.4$.

Figure 4.16 shows the results of applying the particle filter to the fractional Heston model with non-zero ρ, assuming the other θ parameters are known. The only difference is that in this figure the value of ρ is set to 0.2.

We can say that the fractional v_t process is a sensible estimate as it coincides with some of the features that we see in the data.

Figure 4.17 shows the results of estimating the θ parameters of the fractional Heston model with a uniform prior distribution with negative support for ρ, across time t, using the auxiliary particle filter. In these plots the posterior mean and approximate 90% credible intervals of each parameter are shown.

Figure 4.18 shows the results of the auxiliary particle filter for the estimated fractional volatility process of the fractional Heston model with a uniform prior distribution with negative support for ρ. In particular, the estimated posterior mean of the fractional volatility process and approximate 90% credible intervals are shown across time.

Figure 4.19 shows the results of estimating the θ parameters of the fractional Heston model with a uniform prior distribution with positive support for ρ, across time t, using the auxiliary particle filter. In these plots the posterior mean and approximate 90% credible intervals of each parameter is shown.

Figure 4.20 shows the results of the auxiliary particle filter for the estimated fractional volatility process of the fractional Heston model with a uniform prior distribution with positive support for ρ.

Again, the auxiliary particle filter for the estimated fractional v_t process provides a sensible estimate as it coincides with some of the features that we see in the data.

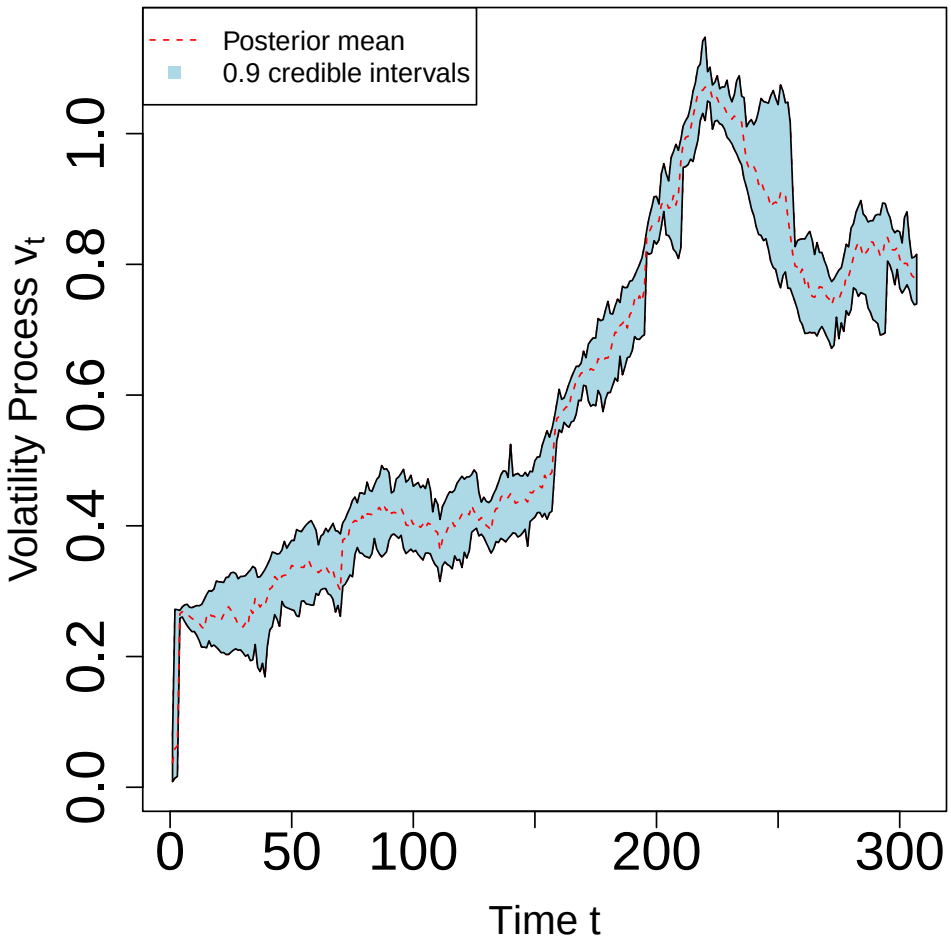


Figure 4.15: Estimation of the fractional volatility process v_t from the SP350 data using the particle filter for the fractional Heston model with a negative value of ρ, i.e., $\rho = -0.2$, and with the other θ parameters set at their maximum likelihood estimates. The approximate posterior mean of the fractional volatility process is represented by the dashed red line and the associated 90% credible intervals are in light blue.

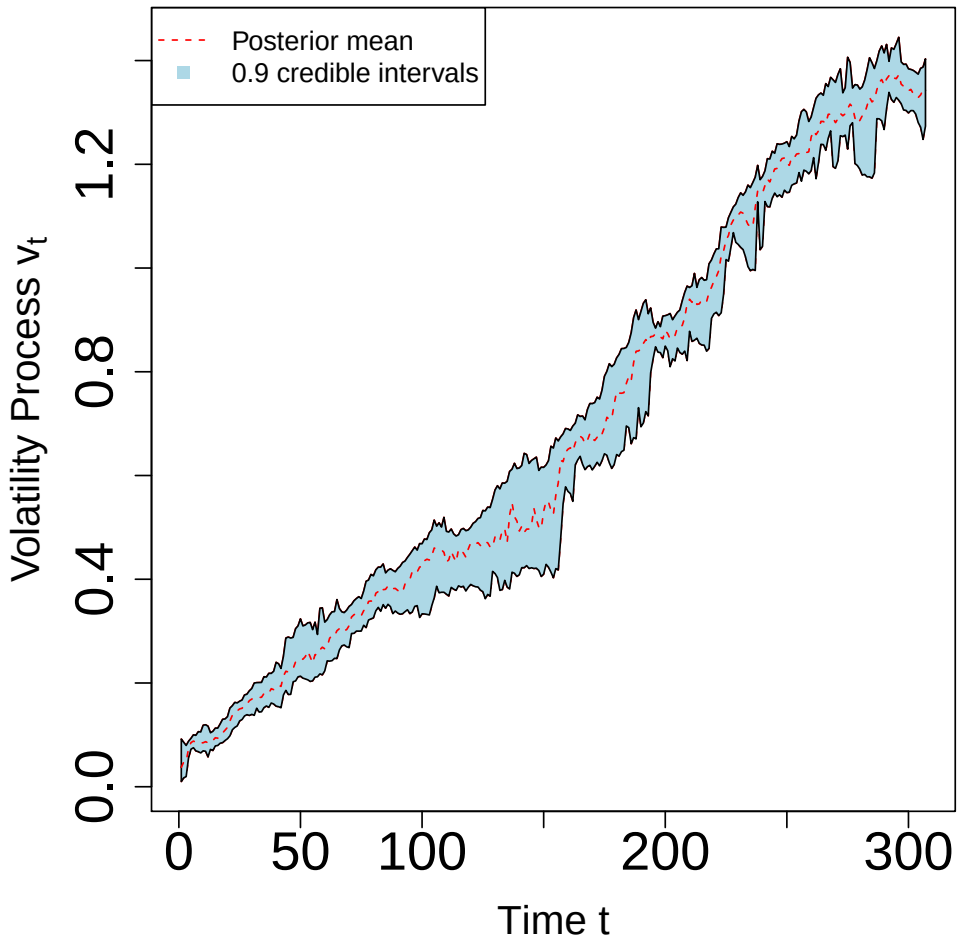


Figure 4.16: Estimation of the fractional volatility process v_t from the SP350 data using the particle filter for the fractional Heston model with a positive value of ρ, i.e., $\rho = 0.2$, and with the other θ parameters set at their maximum likelihood estimates. The approximate posterior mean of the fractional volatility process is represented by the dashed red line and the associated 90% credible intervals are light blue.

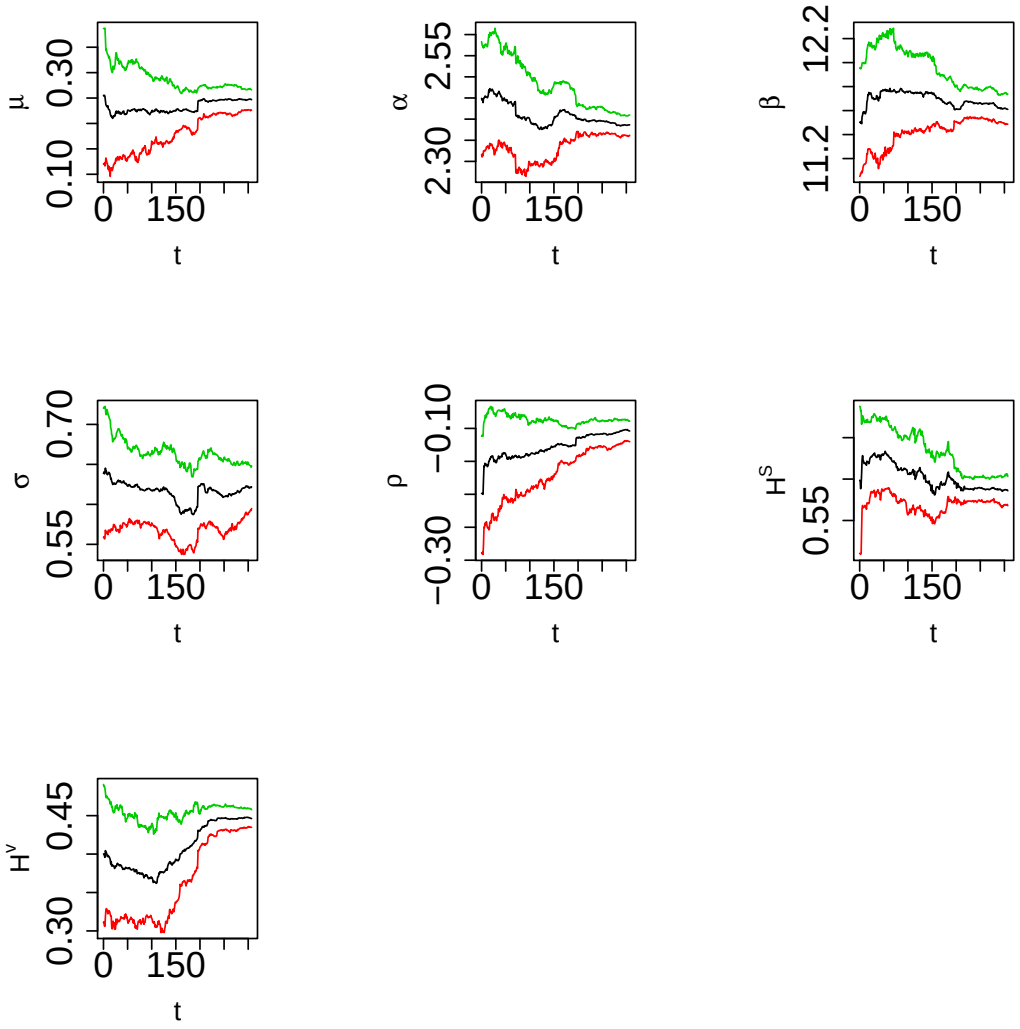


Figure 4.17: Estimation of the parameters of the fractional Heston model with a uniform prior distribution for ρ with negative support, across time t, using the auxiliary particle filter applied to the SP350 data. In each graph, the posterior mean is shown using the black trace, while the associated approximate 90% credible intervals are shown by the outer traces. The first graph is for the parameter μ, the second graph is for the parameter α, the third graph is for the parameter β, the fourth graph is for the parameter σ, the fifth graph is for the parameter ρ, the sixth graph is for the parameter H^s and the final graph is for the parameter H^v.

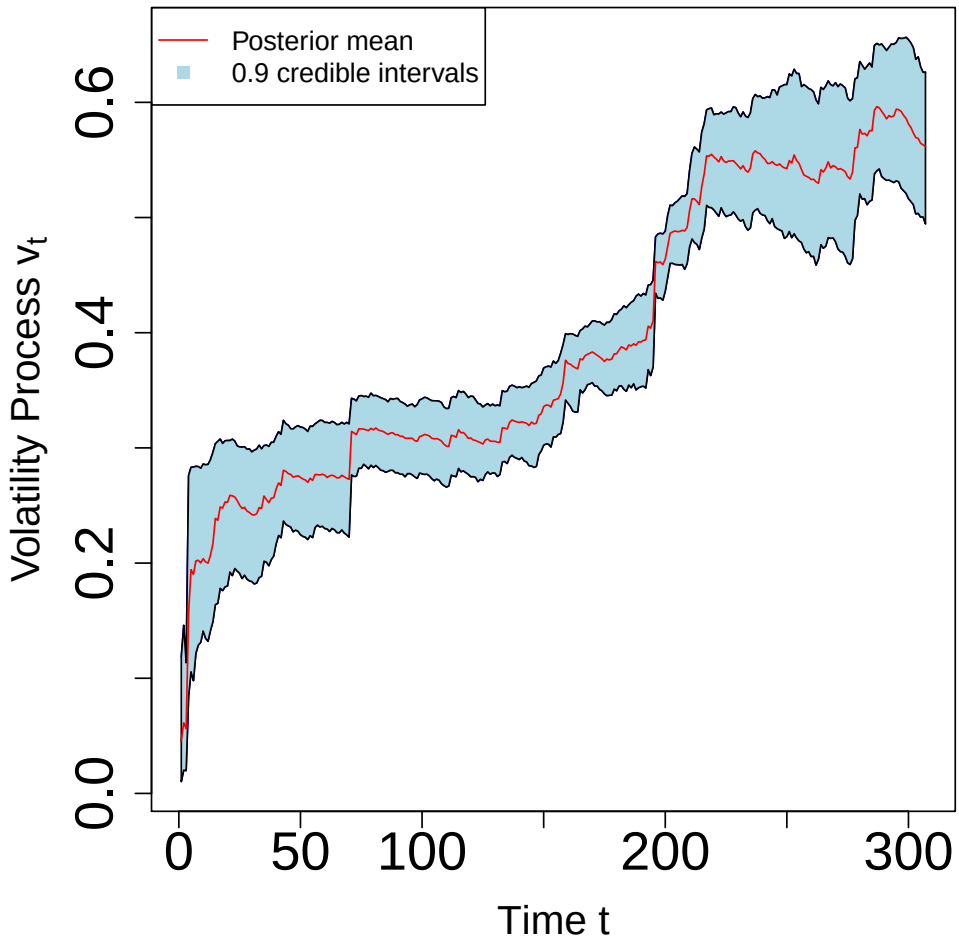


Figure 4.18: Estimation of the fractional volatility process v_t from the SP350 data in the fractional Heston model with a uniform prior distribution for ρ with negative support, using the auxiliary particle filter when the θ parameters are unknown. The estimated posterior mean of the volatility process is represented by the red line and the associated 90% credible intervals are in light blue.

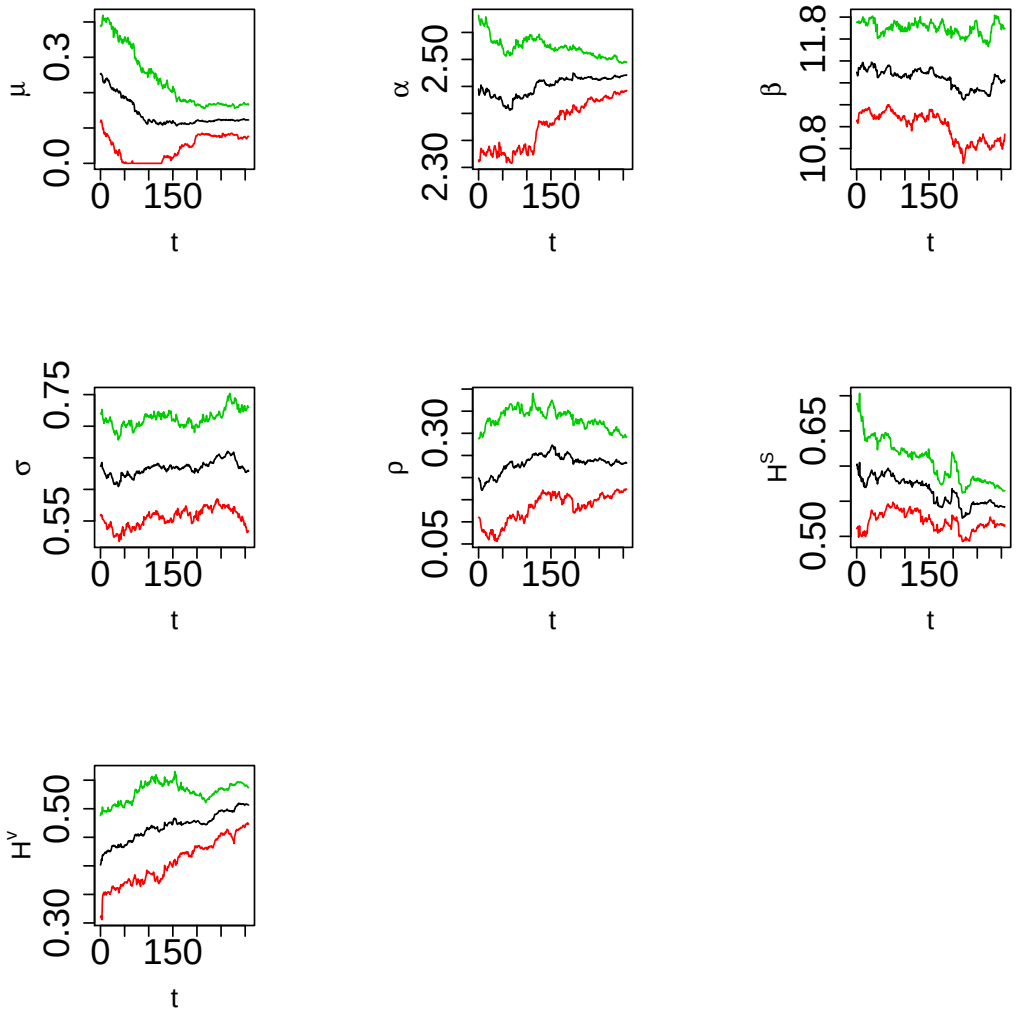


Figure 4.19: Estimation of the parameters of the fractional Heston model with a uniform prior distribution for ρ with positive support, across time t, using the auxiliary particle filter applied to the SP350 data. In each graph, the posterior mean is shown using the black trace, while the associated approximate 90% credible intervals are shown by the outer traces. The first graph is for the parameter μ, the second graph is for the parameter α, the third graph is for the parameter β, the fourth graph is for the parameter σ, the fifth graph is for the parameter ρ, the sixth graph is for the parameter H^s and the final graph is for the parameter H^v.

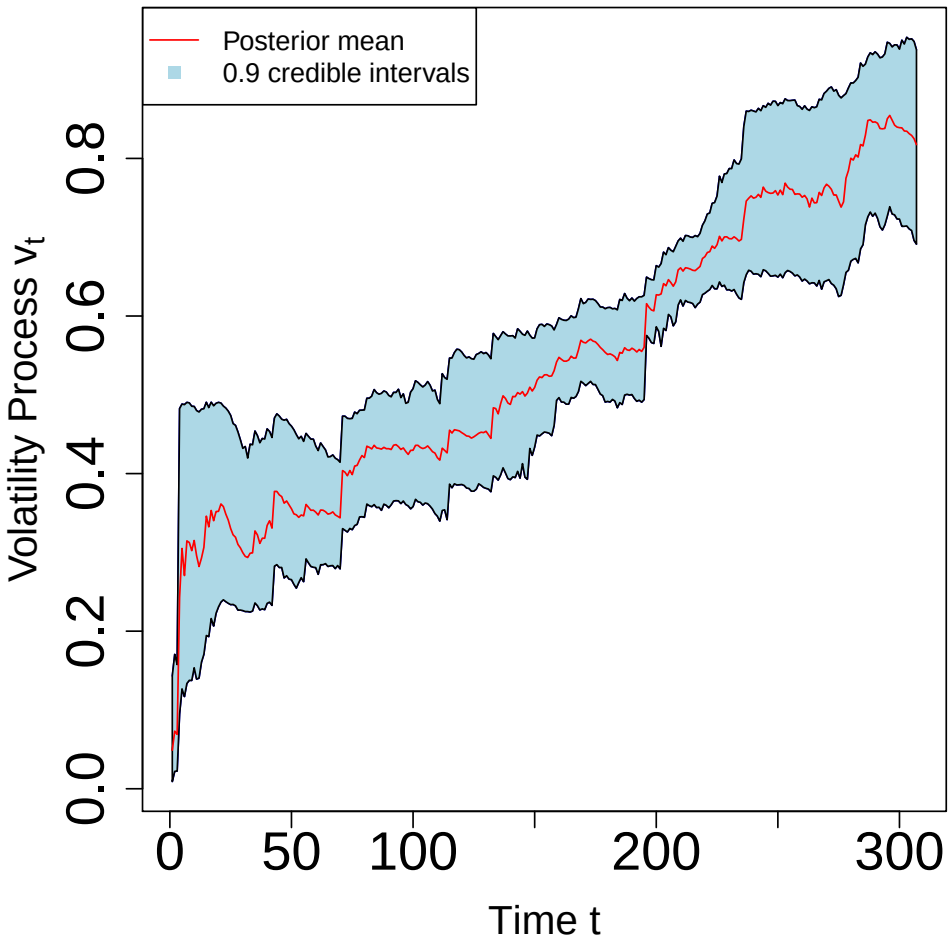


Figure 4.20: Estimation of the fractional volatility process v_t from the SP350 data in the fractional Heston model with a uniform prior distribution for ρ with positive support, using the auxiliary particle filter when the other θ parameters are unknown of SP350. The estimated posterior mean of the fractional volatility process is represented by the red line and the associated 90% credible intervals are in light blue.

4.8.1 Forecasting

In this section, we consider forecasts based on the Heston and the fractional Heston models with non-zero ρ. Our forecasts are generated according to the the method outlined in Section 4.6. We will forecast within and beyond the observed stock price $s_{1:T}$ of the SP350 data. The number of forecasting points is 58 in our cases.

4.8.1.1 Forecasting with negative prior support for ρ

As mentioned, we will consider both the Heston and the fractional Heston models. The posterior parameter distribution will be obtained using the auxiliary particle filter, with a uniform prior distribution with negative support for ρ. Forecasts will be generated in the way described in Section 4.6.

Figure 4.21 shows forecasting based on the Heston model beyond the final point of the SP350 data. The number of forecasts beyond the final point is 58. Figure 4.22 shows the weighted mean and the associated 90% forecast intervals beyond the final point.

Figure 4.23 shows forecasting based on the fractional Heston model beyond the final point of the SP350 data, while Figure 4.24 shows the weighted mean and associated 90% forecast intervals.

Figure 4.25 shows forecasting based on the Heston model of the last 58 points of the SP350 data, while Figure 4.26 shows the weighted mean and associated 90% forecast intervals.

Figure 4.27 shows forecasting based on the fractional Heston model of the last 58 points of the SP350 data, while Figure 4.28 shows the weighted mean and associated 90% forecast intervals.

Figure 4.29 provides a comparison between the forecasting of stock prices using the weighted mean and approximate 90% forecast intervals based on the Heston model and the actual data, namely the last 58 points of the SP350 series. Figure 4.30 provides a similar comparison based on the fractional Heston model.

We will use equations (4.23) and (4.24) to quantify the accuracy of the SP350 forecasts based on the Heston and the fractional Heston models when a uniform prior with negative

support for ρ is used. Table 4.1 shows the values of the Absolute and Square Root differences between the weighted mean of the forecasts and the observed data SP350 data for the Heston and the fractional Heston models with a uniform prior with negative support for ρ.

We conclude from Table 4.1 that the differences are smaller for the fractional Heston model than for the Heston model. This indicates that the fractional Heston model may offer an advantage when forecasting a real financial time series. As we would expect, the forecast intervals based on the fractional Heston model are wider than those based on the Heston model. This is because the fractional Heston model involves more parameters and is therefore more flexible than the Heston model.

	Absolute difference	Square Root difference
Heston model	2442.2	2240.9
Fractional Heston model	1655.3	1010.8

Table 4.1: The Absolute and Square Root differences between the weighted mean of the forecast paths within the SP350 data and the actual data for the Heston and fractional Heston models with a negative prior support for ρ.

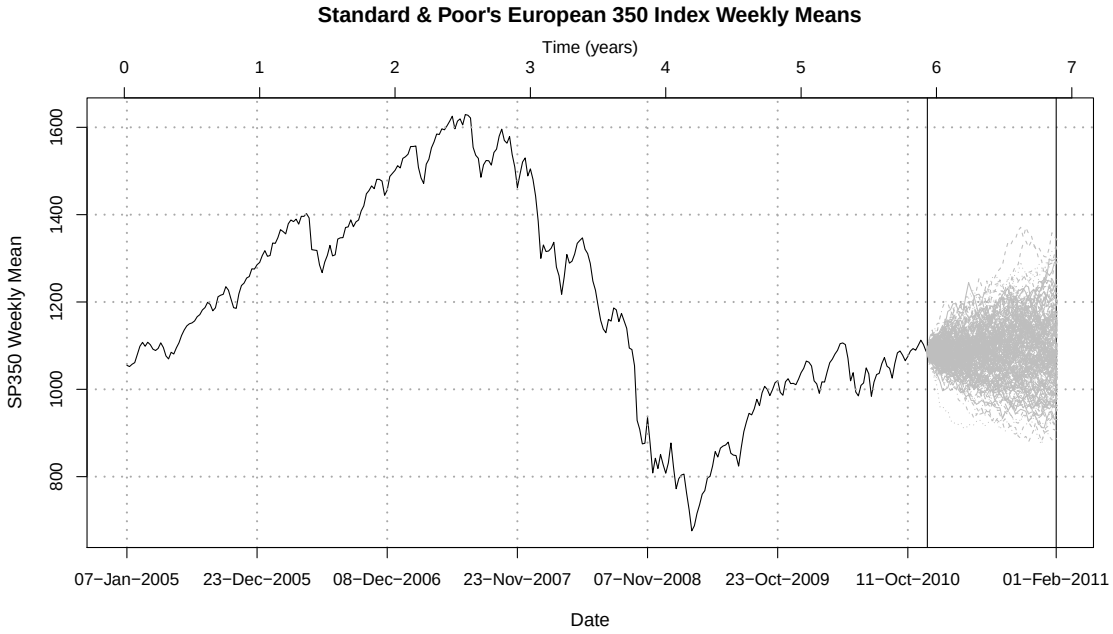


Figure 4.21: The Standard & Poor's European 350 Index together with forecast sample paths beyond the end of the data based on the Heston model. The number of points that have been forecast is 58. A uniofrm prior with negative support is used for ρ.

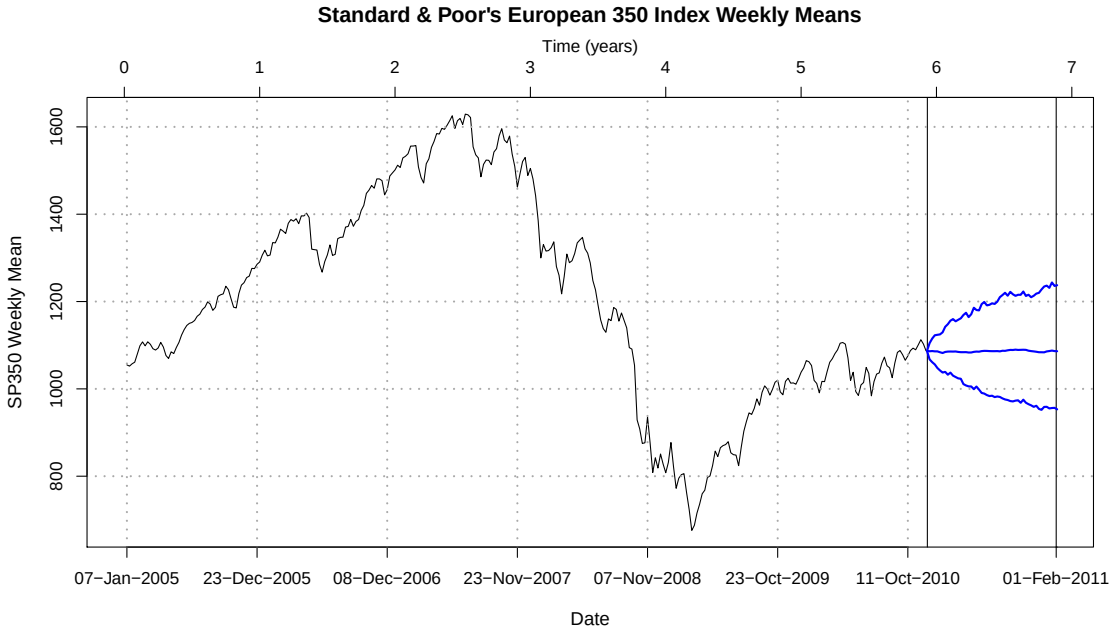


Figure 4.22: The Standard & Poor's European 350 Index together with the weighted mean of the forecast paths beyond the end of the data (blue line) and upper and lower 90% forecast interval based on the forecast sample paths shown in Figure 4.21. A uniform prior with negative support is used for ρ.

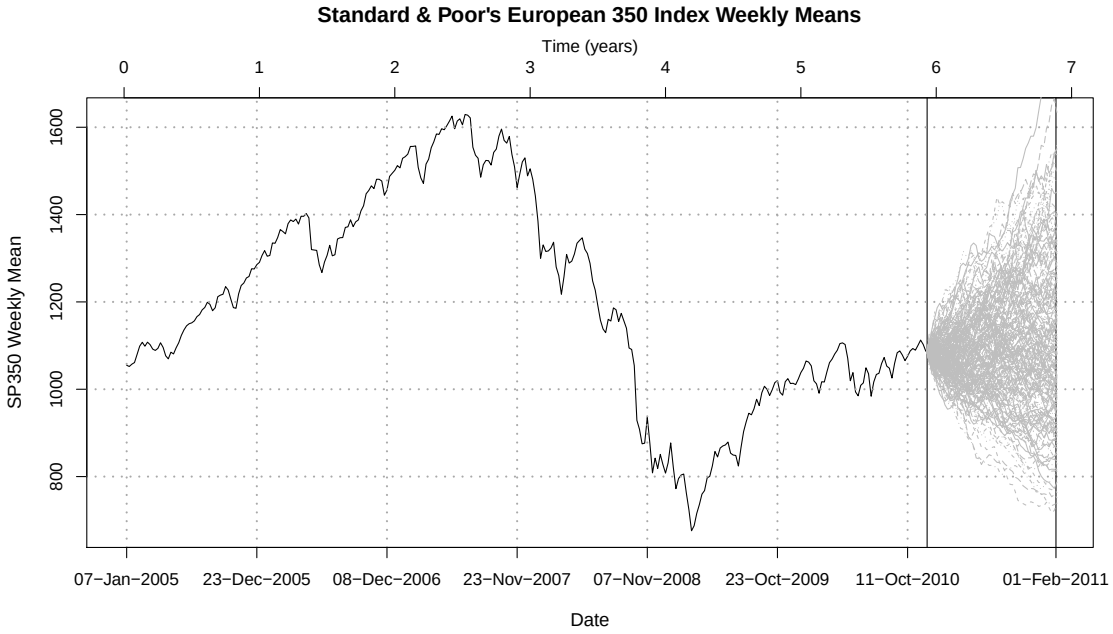


Figure 4.23: The Standard & Poor's European 350 Index together with forecast sample paths beyond the end of the data based on the fractional Heston model. The number of points that have been forecast is 58. A uniform prior with negative support is used for ρ.

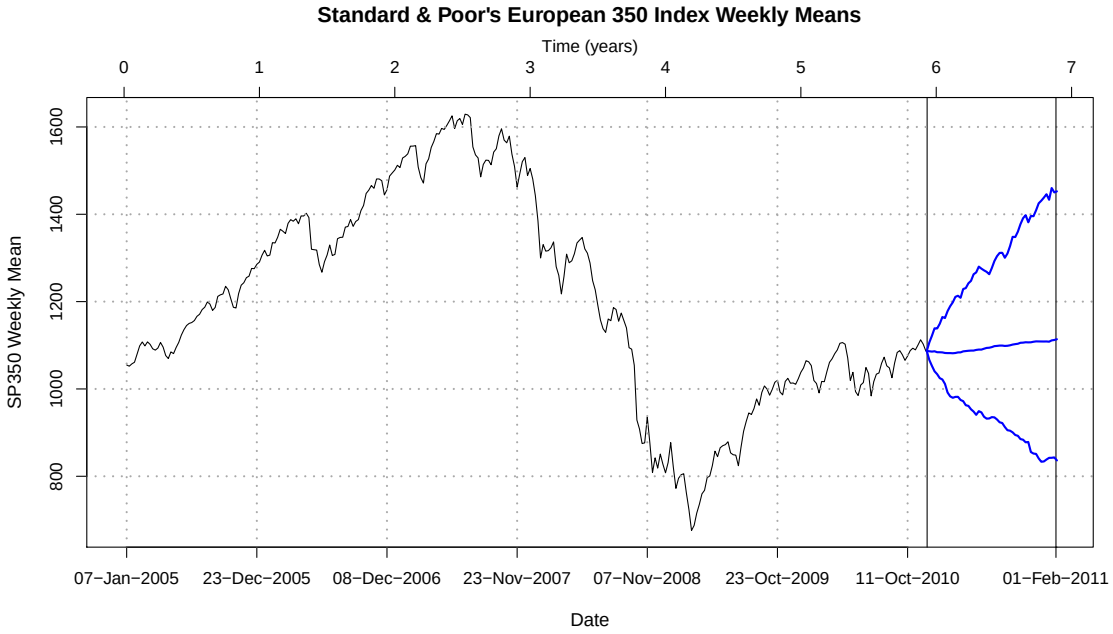


Figure 4.24: The Standard & Poor's European 350 Index together with the weighted mean of the forecast paths beyond the end of the data (blue line) and upper and lower 90% forecast interval based on the forecast sample paths shown in Figure 4.23. A uniform prior with negative support is used for ρ.

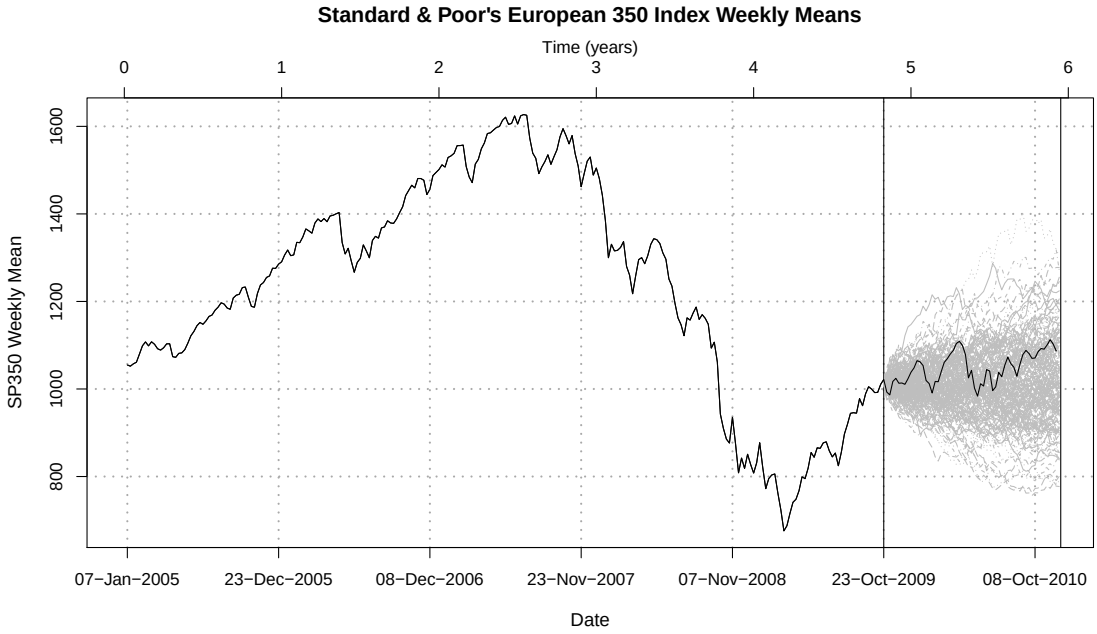


Figure 4.25: The Standard & Poor's European 350 Index together with forecast sample paths within the data based on the Heston model. A uniofrm prior with negative support is used for ρ.

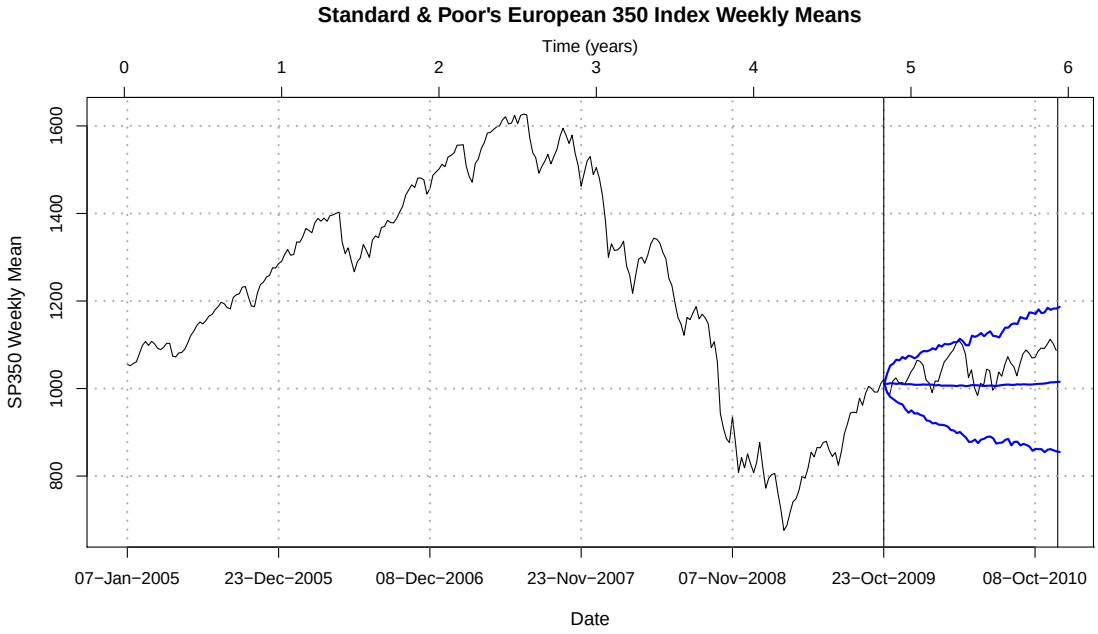


Figure 4.26: The Standard & Poor's European 350 Index together with the weighted mean of the forecast paths within the data (blue line) and upper and lower 90% forecast intervals based on the forecast sample paths shown in Figure 4.25. A uniform prior with negative support is used for ρ.

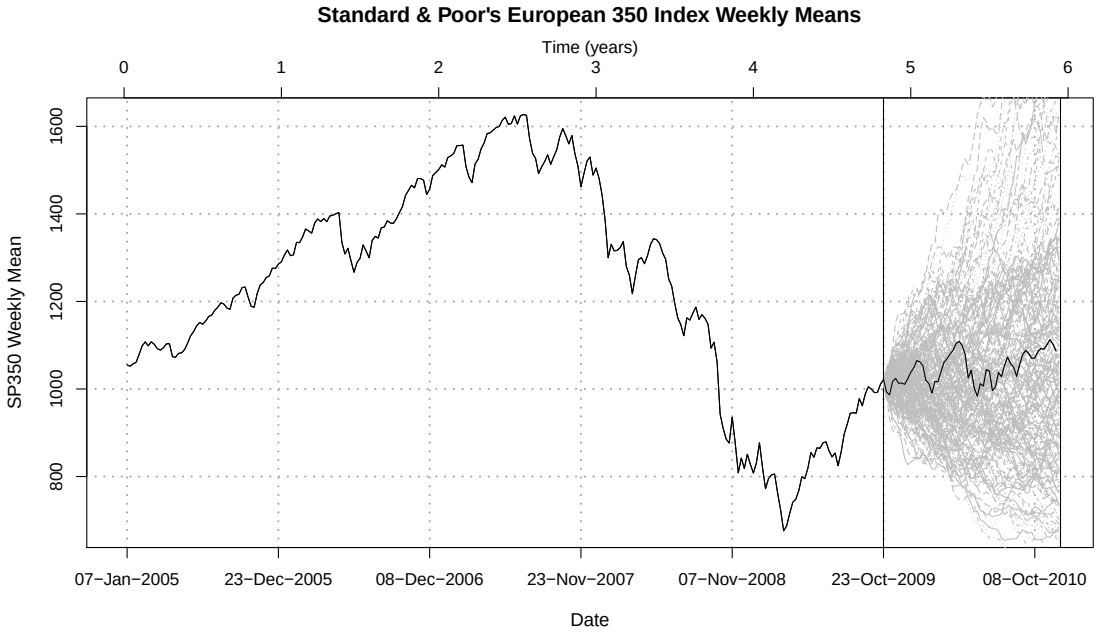


Figure 4.27: The Standard & Poor's European 350 Index together with forecast sample paths within the data based on the fractional Heston model. A uniform prior with negative support is used for ρ.

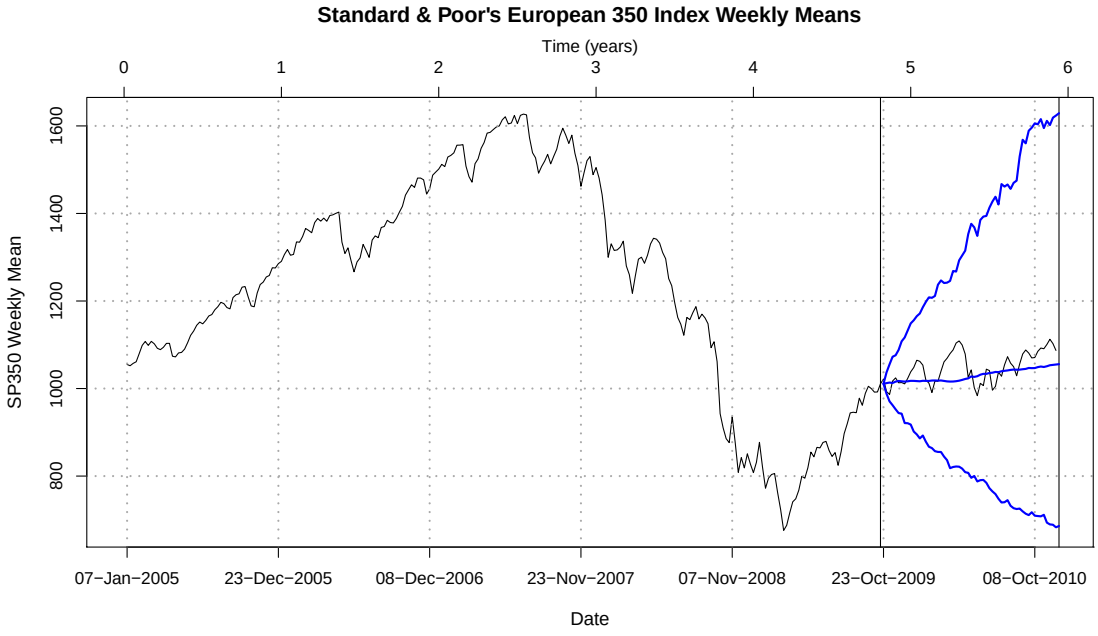


Figure 4.28: The Standard & Poor's European 350 Index together with the weighted mean of the forecast paths within the data (blue line) and upper and lower 90% forecast intervals based on the forecast sample paths shown in Figure 4.27. A uniform prior with negative support is used for ρ.

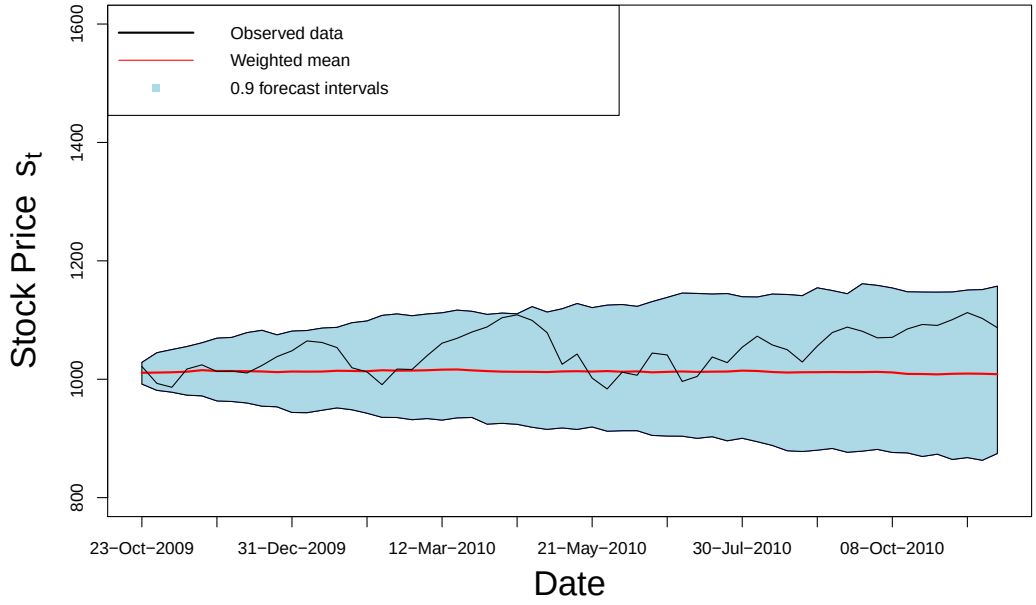


Figure 4.29: The forecast region of Figure 4.26 together with the observed data. The weighted mean forecast path together with 90% forecast intervals are based on the Heston model. A uniofrm prior with negative support is used for ρ.

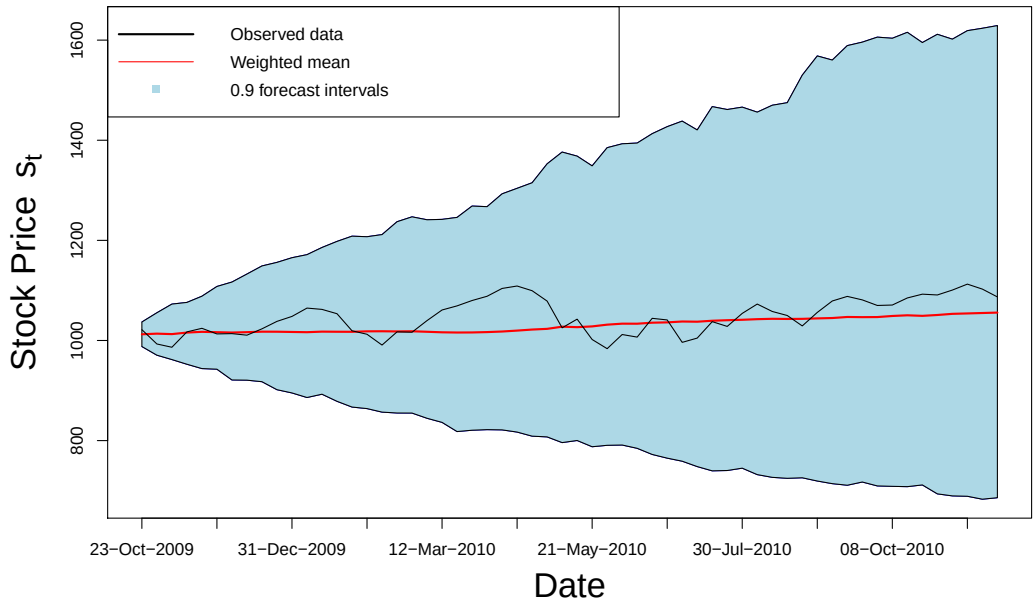


Figure 4.30: The forecast region of Figure 4.28 together with the observed data. The weighted mean forecast path together with 90% forecast intervals are based on the fractional Heston model. A uniform prior with negative support is used for ρ.

4.8.1.2 Forecasting with a positive prior support ρ

Next, we will use the auxiliary particle filter to forecast the stock price within and beyond the SP350 data. We consider both the Heston and the fractional Heston models with a positive support for ρ. The posterior parameter distribution is obtained using the auxiliary particle filter, with a uniform prior with positive support for ρ. Forecasts will be generated in the way described in Section 4.6.

Figure 4.31 shows forecasting based on the Heston model beyond the final point of the SP350 data. The number of forecasts beyond the final point is 58. Figure 4.32 shows the weighted mean and the associated 90% forecast intervals beyond the final point.

Figure 4.33 shows forecasting based on the fractional Heston model beyond the final point of the SP350 data, while Figure 4.34 shows the weighted mean and the associated 90% forecast intervals.

Figure 4.35 shows forecasting based on the Heston model of the last 58 points of the SP350 data, while Figure 4.36 shows the weighted mean and 90% forecast intervals.

Figure 4.37 shows forecasting based on the fractional Heston model of the last 58 points of the SP350 data, while Figure 4.38 shows the weighted mean and 90% forecast intervals.

Figure 4.39 provides a comparison between the forecasting of stock price using the weighted mean and approximate 90% forecast intervals based on the Heston model and the actual data, namely the last 58 points of the SP350 series. Figure 4.40 provides a similar comparison based on the fractional Heston model.

We will use equations (4.23) and (4.24) to quantify the accuracy of the SP350 forecasts based on the Heston and the fractional Heston models when a uniform prior with a positive support for ρ.

Table 4.2 shows the values of Absolute and Square Root differences between the weighted mean of the forecasting and the observed SP350 data for the Heston and the fractional Heston models with a uniform prior with positive support for ρ is used.

We conclude from Table 4.2 that the differences are smaller for the fractional Heston model than for the Heston model. This indicates that the fractional Heston model may offer an advantage when forecasting a real financial time series.

	Absolute difference	Square Root difference
Heston model	2476.6	2258.6
Fractional Heston model	1466.7	782.1

Table 4.2: The Absolute and Square Root differences between the weighted mean of the forecast paths within the SP350 data and the actual data itself for the Heston and fractional Heston model with a positive prior support for ρ.

We can compare Table 4.2 and Table 4.1, we see that the Absolute difference and Square Root difference for the Heston model are smaller when adopt a uniform prior for ρ with positive support than they are when the prior for ρ has negative support. On the other hand, for the fractional Heston model the differences are larger when the prior for ρ has positive support than when it has negative support.

We conclude form there and other experiments that the fractional Heston model is more flexible than the Heston model and that it may provide us with better forecasts of real financial time series data. The prior density adopted for the parameter ρ can have some effect on forecast quality.

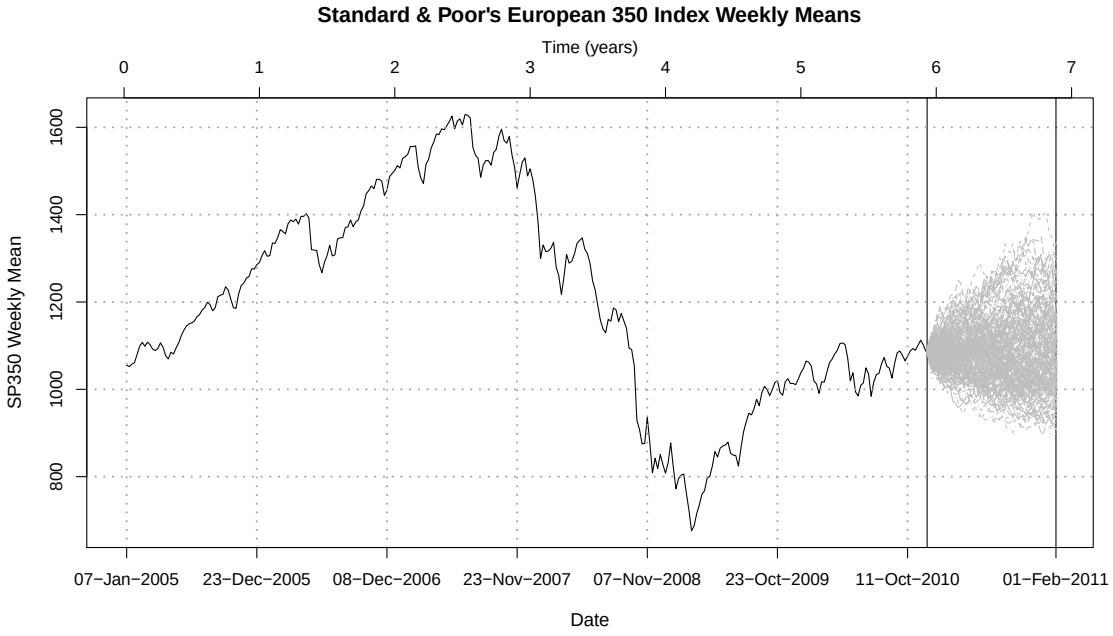


Figure 4.31: The Standard & Poor's European 350 Index together with forecast sample paths beyond the end of the data based on the Heston model. The number of points that have been forecast is 58. A uniform prior with positive support is used for ρ.

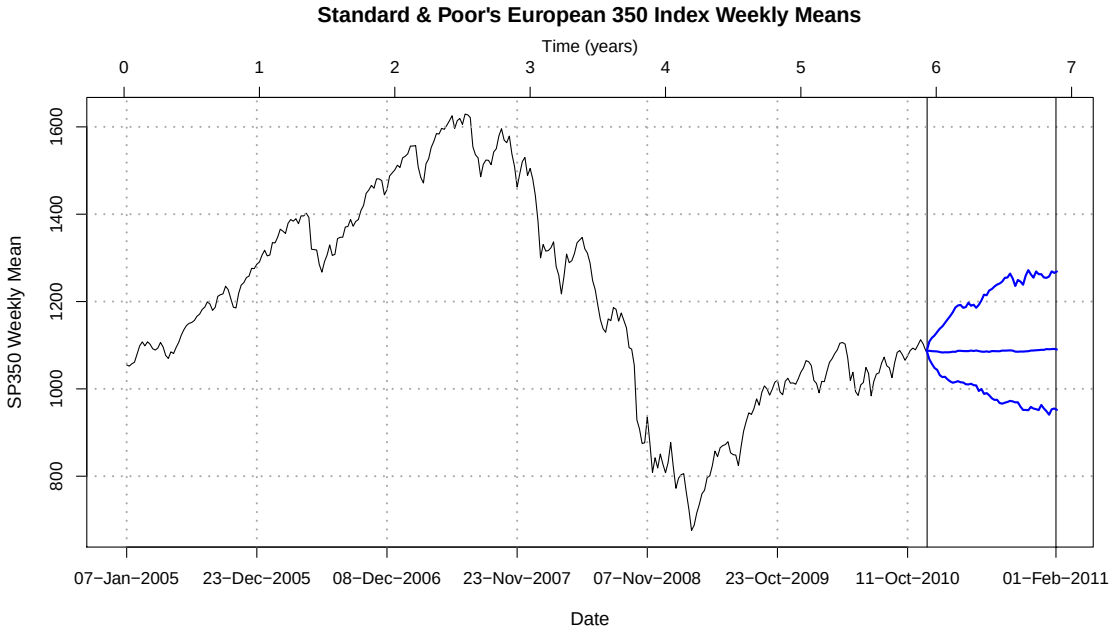


Figure 4.32: The Standard & Poor's European 350 Index together with the weighted mean of the forecast paths beyond the end of the data (blue line) and upper and lower 90% forecast interval based on the forecast sample paths shown in Figure 4.31. A uniform prior with positive support is used for ρ.

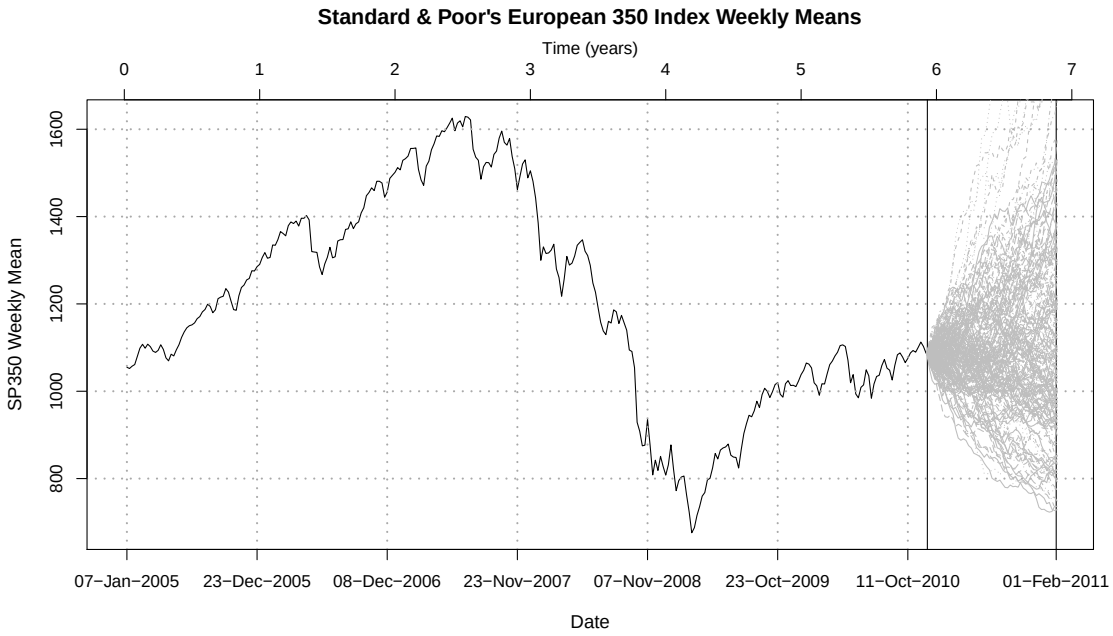


Figure 4.33: The Standard & Poor's European 350 Index together with forecast sample paths beyond the end of the data based on the fractional Heston model. The number of points that have been forecast is 58. A uniform prior with positive support is used for ρ.

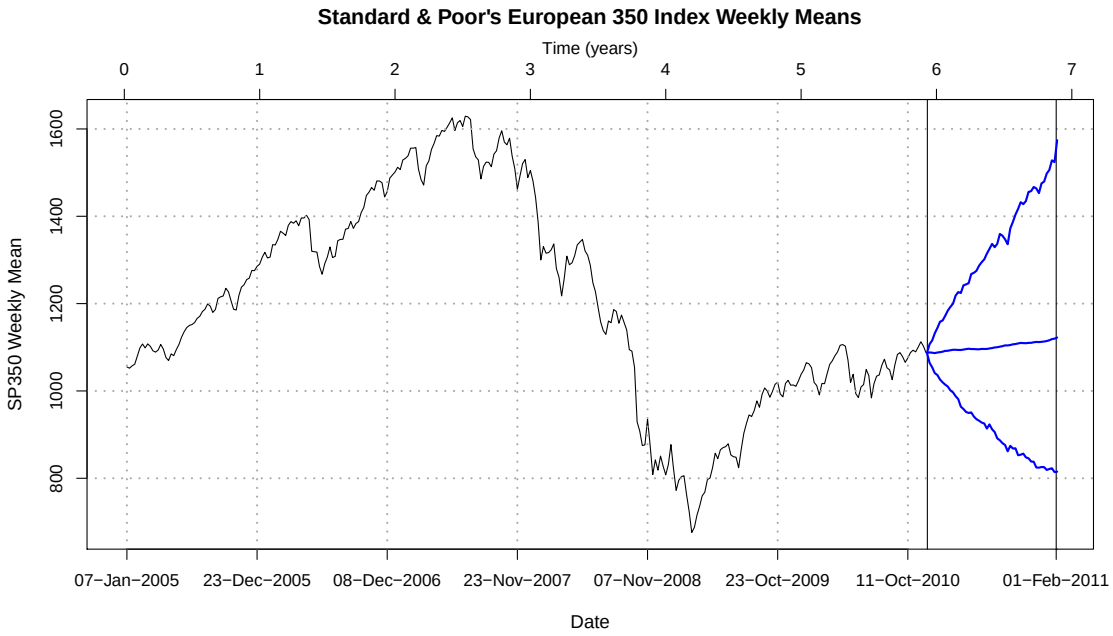


Figure 4.34: The Standard & Poor's European 350 Index together with the weighted mean of the forecast paths beyond the end of the data (blue line) and upper and lower 90% forecast interval based on the forecast sample paths shown in Figure 4.33. A uniform prior with positive support is used for ρ.

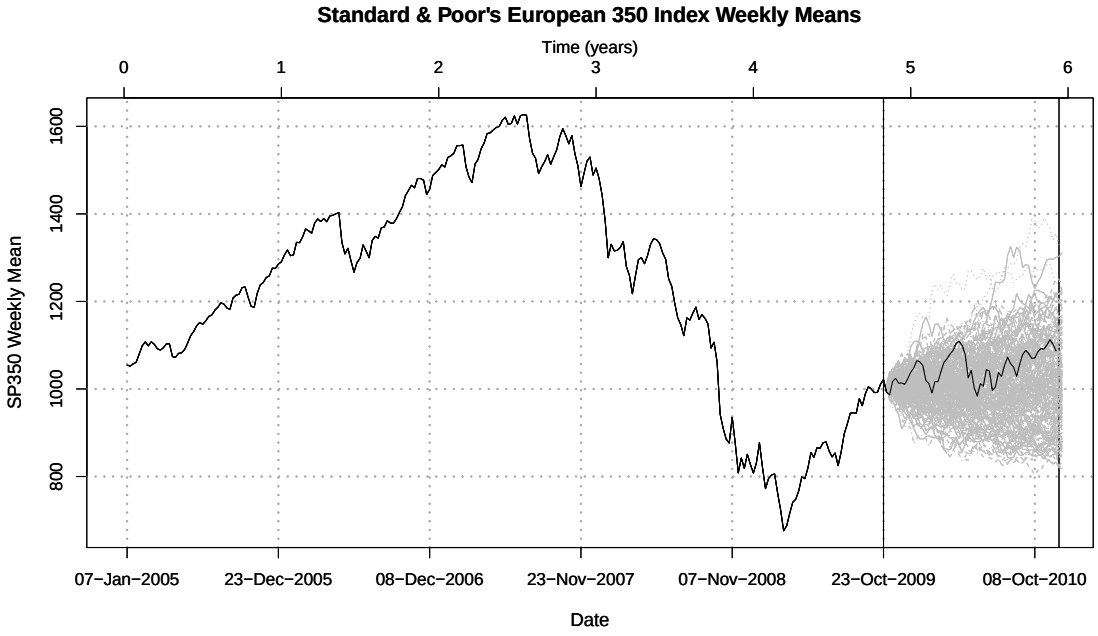


Figure 4.35: The Standard & Poor's European 350 Index together with forecast sample paths within the data based on the Heston model. A uniform prior with positive support is used for ρ.

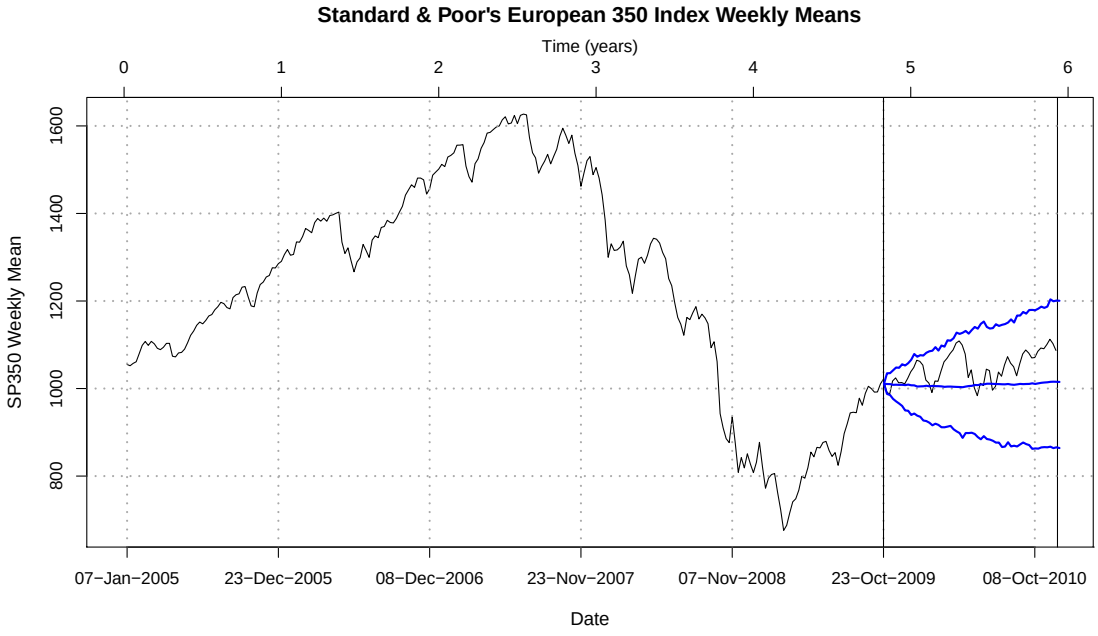


Figure 4.36: The Standard & Poor's European 350 Index together with the weighted mean of the forecast paths within the data (blue line) and upper and lower 90% forecast intervals based on the forecast sample paths shown in Figure 4.35. A uniform prior with positive support is used for ρ.

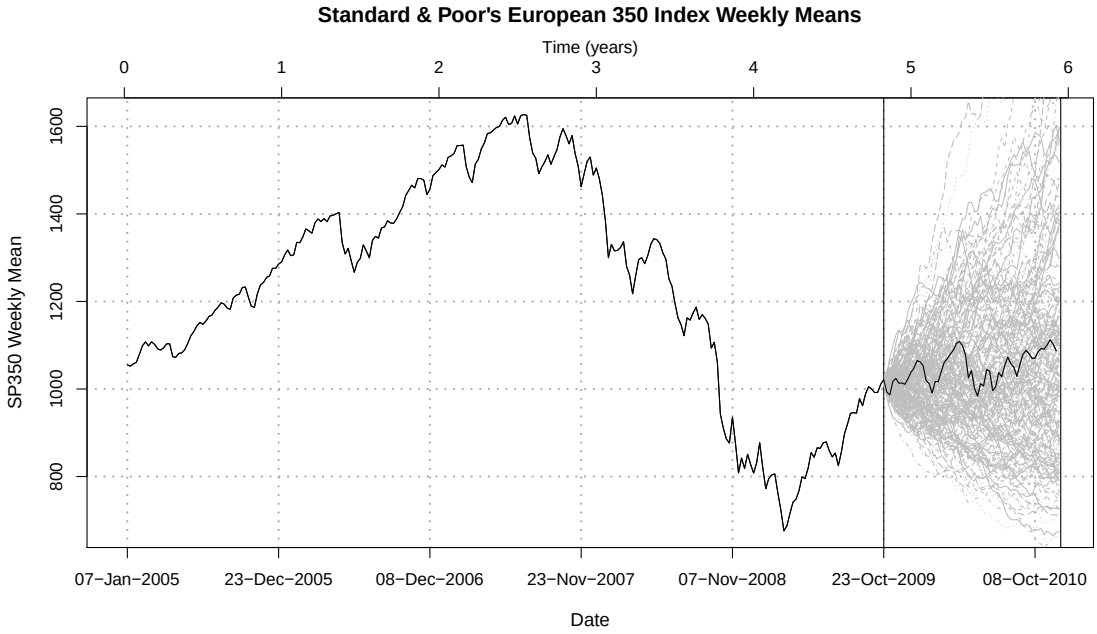


Figure 4.37: The Standard & Poor's European 350 Index together with forecast sample paths within the data based on the fractional Heston model. A uniform prior with positive support is used for ρ.

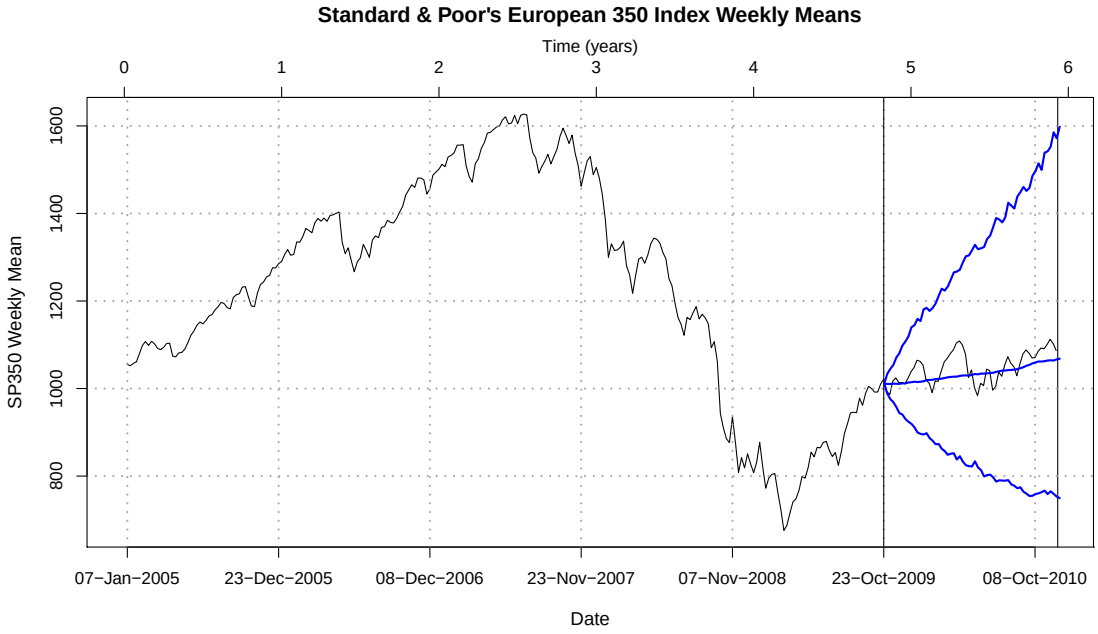


Figure 4.38: The Standard & Poor's European 350 Index together with the weighted mean of the forecast paths within the data (blue line) and upper and lower 90% forecast intervals based on the forecast sample paths shown in Figure 4.37. A uniform prior with positive support is used for ρ.

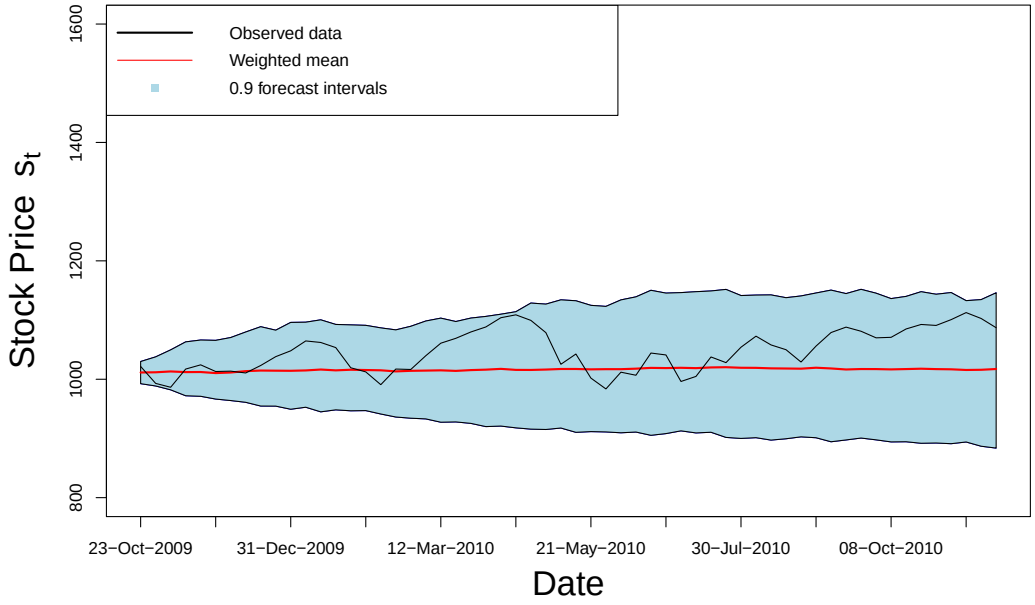


Figure 4.39: The forecast region of Figure 4.36 together with the observed data. The weighted mean forecast path together with 90% forecast intervals are based on the Heston model. A uniform prior with positive support is used for ρ.

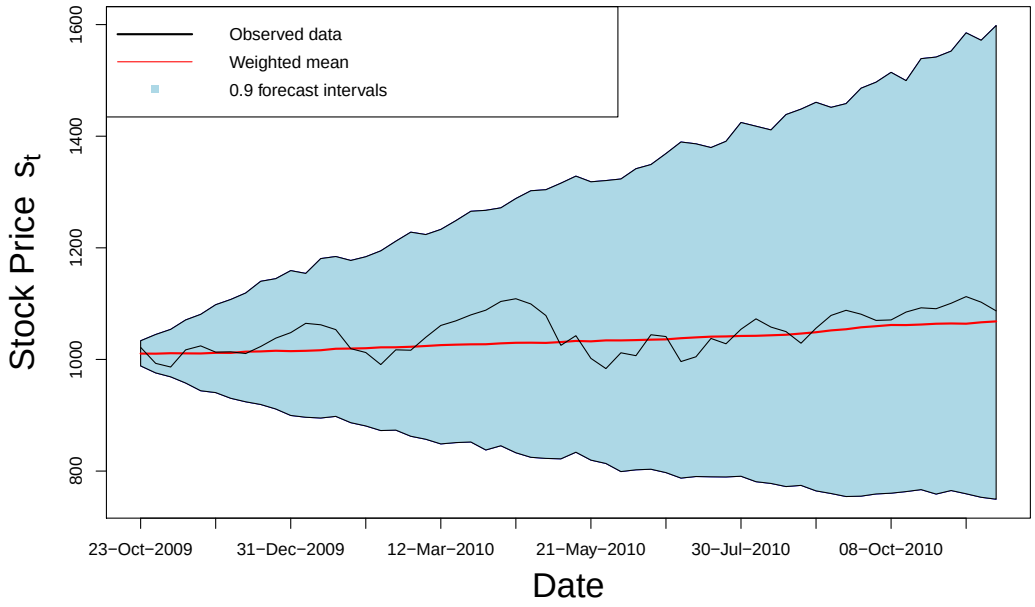


Figure 4.40: The forecast region of Figure 4.38 together with the observed data. The weighted mean forecast path together with 90% forecast intervals are based on the fractional Heston model. A uniform prior with positive support is used for ρ.

4.9 Summary

In this chapter, we have discussed:

1. The fractional Heston model with $\rho = 0$ and with non-zero ρ, where ρ is the correlation between the driving stochastic processes, and we have presented a discrete time version of the model.

2. Maximum likelihood estimation and Bayesian inference using particle filter based algorithms.

3. Forecasting

4. Applications to simulated and real data.

In Chapter 5 we briefly present our conclusion and discuss how we can extend our model into a fractional Lévy Heston model that allows jumps in the underlying stochastic processes.

Chapter 5

Conclusion

5.1 Summary

In this thesis we have discussed inference techniques for the unknown quantities of standard and fractional stochastic differential equation models within the Bayesian framework.

Our main contribution has been to extend the Heston model to the fractional Heston model by replacing Brownian motion by fractional Brownian motion in the definition of the model so allowing for a more flexible correlation structure across time. Particle filter based methodologies have been developed to perform inference about the unknown quantities of the fractional Heston model, sequentially across time as new data arrive. In addition to making inference about the model parameters, our methodology sequentially estimates the fractional volatility process.

Our fractional Heston model offers an advantage over the standard Heston model when there is limited posterior support for the additional parameters H^s and H^v used in the definition of the fractional Brownian motion processes being 0.5, a value that corresponds to standard Brownian motion. This can indeed be the case for financial time series data.

In addition, we have shown that the fractional Heston model can offer an advantage when forecasting a real financial series.

Here is a summary of what has been covered in this thesis:

In Chapter 2 we have made comparisons between some sequential importance sampling methods. In particular we have implemented and experimented with particle filter based

methodologies. We have applied these methods to the Vasicek Interest Rate model and to Taylor's formula for stochastic volatility as a Hidden Markov model. We have computed sequentially posterior inferences for the underlying volatility process and the model parameters. Then, we have assessed the sensitivity of those posterior inferences to our prior assumptions. For example, for the Vasicek Interest Rate model we have explored the effect of choosing different prior distribution for the parameters θ_1, θ_2 and θ_3 or β_1, θ_1 and θ_3 that define the model and that control particular features of interest rate behaviour.

Next we have extended the methodology to make sequential inference for processes defined using more than one stochastic differential equation. One such process is the Heston model where stock price and volatility are each assumed to follow a stochastic differential equation.

In Chapter 3 we have performed inference for the parameters and the volatility process of the Heston model when the correlation ρ between the driving stochastic processes is zero and when $\rho \neq 0$ using Markov chain Markov Carlo (MCMC) techniques and particle filter based methodologies.

We discussed by means of a simulation study that unsurprisingly the MCMC method provides better estimation of the model parameters $\theta = (\mu, \alpha, \beta, \sigma)$ or $\theta = (\mu, \alpha, \beta, \sigma, \rho)$ when the volatility is known than when it is unknown. Moreover, when the particle filter is applied to the Heston model with $\rho = 0$, good estimation of the volatility process results. When ρ is non-zero, the volatility process can often be better estimated when ρ is positive than when it is negative.

Next, when the auxiliary particle filter is applied to the Heston model when $\rho = 0$, the model parameters are well estimated, although it can be more difficult to make inference about the parameter σ that controls the variability of the volatility process than about the other parameters. When the auxiliary particle filter is applied to the Heston model with non-zero ρ, better results were obtained when the prior density for ρ had positive support than when it had negative support. Sensible estimates of the underlying volatility process and the model parameters can be achieved.

In Chapter 4 we have extended the Heston model to the fractional Heston model by

replacing Brownian motion in the driving stochastic equations by fractional Brownian motion. The resulting model allows for a more flexible correlation structure across time. We have performed inference for the parameters of the fractional Heston model and the fractional volatility process when $\rho = 0$ and when $\rho \neq 0$ using our particle filter based methodologies.

In particular, we have applied the particle filter to the fractional Heston model to estimate the fractional volatility assuming the other parameters are known. Better estimation can be achieved when ρ is negative than when ρ is zero or positive.

Next, we have applied the auxiliary particle filter to the fractional Heston model when $\rho = 0$. Good estimation of the model parameters and of the fractional volatility process can be achieved.

Moreover, we have found that when $\rho \neq 0$ better inference for the fractional Heston model can be achieved when the prior density for ρ has positive support than when it has negative support. Sometimes better inference can be achieved for the fractional Heston model than for the original model.

A core finding of this thesis is that for real data the posterior density of the Hurst indices H^s and H^v that control correlation across time do not support 0.5, a value which corresponds to standard Brownian motion and hence to the original Heston model. Moreover, we have implemented forecasting methodology for the fractional Heston model, and have obtained better forecasts than with the Heston model for real stock price data.

5.2 Future work

We have recently found that it can be very useful to extend the fractional Heston model to the fractional Lévy Heston model.

A fractional Lévy process is a generalization of fractional Brownian motion. In particular, a fractional Lévy process is a linear term which consists of a fractional Brownian part and a jump part.

Mathematically, in our case a fractional Lévy process can be written as follows:

$$W_t^{fL} = W_t^f + L_t,$$

where

W_t^{fL} is a fractional Lévy process,

W_t^f is a fractional Brownian process,

L_t is a Lévy process as defined in Korn et al. (2010).

A fractional Lévy process has the same properties as fractional Brownian motion as explained in Sections 4.2 and 4.3. The only difference is that a jump is allowed between two increments sometimes called a pure jump. These jumps can be modelled by the symmetric $\alpha-$stable Lévy process, where

$\alpha \in (0,2]$ (Tsoi et al., 2011). Once again the idea is to produce a more flexible model that can take better account of some of the features of real financial time series data and indeed other data such as environmental or weather data.

In our future work we will generalize the Heston model so that it is driven by a fractional Lévy processes instead of by fractional Brownian motion. We will attempt to estimate the parameters and the fractional Lévy volatility process using particle filter based methodologies. Forecasts will be produced and compared using real data.

List of references

Aït-Sahalia, Y. and R. Kimmel (2007). Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics 83*, 413–452.

Amblard, P.-O. and J.-F. Coeurjolly (2011). *Identification of the Multivariate Fractional Brownian Motion*. `http://hal.archives-ouvertes.fr/docs/00/49/76/39/PDF/HAL-amblard-coeurjolly-lavancier-philippe.pdf`.

Atiya, A. F. and S. Wall (2009, April). An analytic approximation of the likelihood function for the Heston model volatility estimation problem. *Journal of Quantitative Finance 9*(3), 289–296.

Bauer, P. (2012, February). Fast Calibration in the Heston Model. Master's thesis, Vienna University of Technology, `http://www.fam.tuwien.ac.at/~sgerhold/pub__files/theses/bauer.pdf`.

Bauwens, L., C. Hafner, and S. Laurent (2012). *Handbook Of Volatility Models and their Applications* (First ed.). Canada: Wiley Handbook in Financial Engineering and Econometrics. ISBN–978–0–470–87251–2.

Benhamou, E., E. Gobet, and M. Miri (2010). Time dependent Heston model. *SIAM Journal of Financial Mathematics 1*, 289–325.

Bhar, R. (2010). *Stochastic Filtering with Applications in Finance* (First ed.). Sydney: World Scientific Publishing. ISBN–13 978–981–4304–85–6.

Biagini, F., Y. Hu, B. Øksendal, and T. Zhang (2010). *Stochastic Calculus for Fractional Brownian Motion and Applications* (First ed.). London: Springer. ISBN 978–1–84996–994–9.

Brooks, S., A. Gelman, G. Jones, and X. Meng (2011). *Handbook of Markov Chain Monte Carlo*. USA: Chapman & Hall/CRC Handbooks of Modern Statistical Methods. ISBN 978–1–4200–7941–8.

Buchen, P. (2012). *An Introduction to Exotic Option Pricing*. Chapman & Hall/CRC Financial Mathematics series. 978–1–4200–9100–7.

Cappé, O., E. Mouline, and T. Rydén (2005). *Inference in Hidden Markov Models* (First ed.). France: Springer. ISBN 978–0–387–40264–0.

Chang, Y.-C. and S. Chang (2002, March). A fast estimation algorithm on the Hurst parameter of discrete–time fractional Brownian motion. *IEEE Transactions on Signal Processing 50*(3), 554–559.

Chronopoulou, A. and F. G. Viens (2011). *Stochastic Volatility and Option Pricing with Long–Memory in Discrete and Continuous Time*. http://kolmogorov.math.stevens.edu/conference2011/images/stories/ViensChronLMSV_article.pdf.

Coeurjolly, J.-F. (2000). Simulation and Identification of the Fractional Brownian Motion: A Bibliographical and Comparative Study. http://www-ljk.imag.fr/membres/Jean-Francois.Coeurjolly/documents/Coe00.pdf.

Coeurjolly, J.-F., P.-O. Amblard, and S. Achard (2010, 23–27 August). On multivariate fractional Brownian motion and multivariate fractional Gaussian noise. In *European Signal Processing Conference*, Number 18th, http://www.eurasip.org/Proceedings/Eusipco/Eusipco2010/Contents/papers/1569291919.pdf. EUSIPCO: Aalborg, Denmark.

Date, P. and K. Ponomareva (2011). Linear and non-linear filtering in mathematical finance: A review. *Journal of Management Mathematics 22*, 195–211.

Dieker, T. (2004). *Simulation of Fractional Brownian Motion*. http://www2.isye.gatech.edu/~adieker3/fbm/thesis.pdf.

Doucet, A. and M. Johansen (2008, December). *A Tutorial on Particle Filter and Smoothing: Fifteen Years Later*. http://www.cs.ubc.ca/~arnaud/doucet_johansen_tutorialPF.pdf.

Drăgulescu, A. A. and V. M. Yakovenko (2002). Probability distrbiution of returns in the Heston model with stochastic volatility. *Quantitative Finance 2*, 443–453.

Durbin, J. and S. Koopman (2012). *Time Series Analysis by State Space Methods* (Second ed.). Number 38 in Oxford Statistical Science Series. London: Oxford Press. ISBN–987–0–19–964117–8.

Faff, R., K. Hamza, and B. Do (2006). *A New Approach to Modeling and Estimation for Pairs Trading*. Ph. D. thesis, Department of Accounting and Finance, Monash University.

Fisher, B. and M. Scholes (1973). The pricing of options and corporate liabilities. *Journal of Political Economy 81*(3), 637–654.

Giratis, L., H. L. Koul, and D. Surgailis (2012). *Large Sample Inference for Long Memory Processes*. Imperial College Press. ISBN 978–1–84816–278–5.

Gloter, A. and M. Hoffmann (2004). Stochastic volatility and fractional Brownian motion. *Stochastic Processes and their Applications 113*, 143–172.

Gloter, A. and M. Hoffmann (2007, Octeber). Estimation of the Hurst Parameter from Discrete Noisy Data. *The Annals of Statistics 35*(5), 1947–1974.

Heston, S. (1993). A closed-form solution for options with stochastic volatility with application to bond and currency options. *The Review of Financial Studies 6*(2), 327–343.

Hyndman, R. J., A. B. Koehler, and R. D. Snyder (2008). *Forecasting with Exponential Smoothing: The State Space Approach*. Springer Series in Statistics. ISBN 978–540–7191–8.

Iacus, S. M. (2007). *Simulation and Inference for Stochastic Differential Equations with R Examples* (First ed.). Italy: Springer. ISBN 978–0–387–75838–1.

Iacus, S. M. (2010). *Option Pricing and Estimation of Financial Models with R* (First ed.). Italy: Wiley. ISBN 978–0–470–74584–7.

Javaheri, A., D. Lauiter, and A. Galli (2002, December). *Filtering in Finance*. `http://www.cis.upenn.edu/~mkearns/finread/filtering_in_finance.pdf`.

Jiang, Z.-Q. and W.-X. Zhou (2011, March). *Multivariate detrending moving average cross–correlation analysis*. `http://arxiv.org/pdf/1103.2577.pdf`.

Jianhui, H. (2008). *Microstructure Filtering in Financial Market*. Ph. D. thesis, Department of Mathematical and Statistical Science, University of Alberta.

Kessler, M., A. Lindner, and Sørensen (2012). *Statistical Methods for Stochastic Differetial Equations*. Monographs on Statistical and Applied Probability 124. Boca Raton: Chapman & Hall/CRC Book. ISBN 978–1–4398–4940–8.

Klebaner, F. C. (2012). *Introduction to Stochastic Calculus with Applications* (Third ed.). Australia: Imperial College Press. ISBN–13 987–1–84816–831–2.

Korn, R., E. Korn, and G. Kroisandt (2010). *Monte Carlo Methods and Models in Finance and Insurance* (First ed.). Chapman & Hall/CRC Financial Mathematics Series. ISBN 978–1–4200–7618–9.

Lavancier, F., A. Philippe, and D. Surgailis (2009). *Covariance function of vector Self-Similar Process*. http://hal.archives-ouvertes.fr/docs/00/41/10/01/PDF/lavancier-etal-pfractionnaire-v2.pdf.

Liu, J. and M. West (2001). Combined parameter and state estimation in simulation-based filtering. In A. Doucet, N. De Freitas, and N. Gordon (Eds.), *Sequential Monte Carlo Methods in Practice*. Springer, New York.

Lunn, D., C. Jackson, N. Best, A. Thomas, and D. Spiegelhalter (2013). *The BUGS BOOK: A PRACTICAL Introduction to Bayesian Analysis*. Texts in Statistical Science Series. Chapman & Hall CRC Press. ISBN 978–1–58488–849–9.

Mikosch, T. (1998). *Elementary Stochastic Calculus with Finance in View*, Volume 6 of *Advanced Series on Statistical Science & Applied Probability*. World Scientific. ISBN 981–02–3543–7.

Miyahara, Y. (2012). *Option Pricing in Incomplete Markets: Modelling Based on Geometric Lévy Processes and Minimal Entropy Martingale Measures* (First ed.), Volume 3 of *Series in Quantitative Finance*. Japan: Imperial College Press. ISBN–13 978–1–84816–347–8.

Nielsen, O. E. and A. Shiryaev (2010). *Change of Time and Change of Measure* (First ed.), Volume 13 of *Advanced Series on Statistical Science and Applied Probability*. Denmark: World Scientific Publishing. ISBN–13 978–981–4338–47–2.

Petris, G., S. Petrone, and P. Campagnoli (2009). *Dynamic Linear Models with R* (First ed.). Italy: Springer. ISBN 978–0–387–77238–7.

Pitt, M. and N. Shepard (1999). Filtering via simulation: Auxiliary particle filter. *Journal of the American Statistical Association 94*, 590–599.

Pitt, M. and N. Shepard (2001). Auxiliary variable based particle filter. In A. Doucet, N. De Freitas, and N. Gordon (Eds.), *Sequential Monte Carlo Methods in Practice*. Springer, New York.

Prakasa Rao, B. (2010). *Statistical Inference for Fractional Diffusion Processes* (First ed.). India: Wiely series in Probability and Statistics. ISBN 978–0–470–66568–8.

R Core Team (2012). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Rose, O. (1996, February). Estimation of the Hurst Parameter of Long–Range Dependent Time Series. Technical Report 137, Institute of Computer Science, University of Wűrzburg, http://phobos.informatik.uni-wuerzburg.de/TR/tr137.pdf.

Tsoi, A., D. Nualart, and G. Yin (2011). *Stochastic Analysis, Stochastic Systems, and Applications to Finance*. USA: World Scientific Publishing. ISBN-13 978–981–4355–70–4.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics 5*, 177–188.

White, S. I. (2006, July). *Stochastic Volatility: Maximum Likelihood Estimation and Specification Testing*. Ph. D. thesis, Economic and Finance, Queensland University of Techology, Brisbane, Australia.