ESTIMATING STOCHASTIC VOLATILITY USING PARTICLE FILTERS

by

HUAIZHI CHEN

Submitted in partial fulfillment of requirement

For the degree of MS in Applied Mathematics

Thesis Advisor: Dr. Daniela Calvetti

Department of Mathematics

CASE WESTERN RESERVE UNIVERSITY

August, 2009

CASE WESTERN RESERVE UNIVERSITY

SCHOOL OF GRADUATE STUDIES

We hereby approve the thesis/dissertation of

Huaizhi Chen

candidate for the Master of Science degree *.

(signed) Daniela Calvetti
(chair of the committee)

Erkki Somersalo

Peter Kotelenez

(date) 6/25/2009

*We also certify that written approval has been obtained for any proprietary material contained therein.

Estimating Stochastic Volatility Using Particle Filters

Abstract

by

HUAIZHI CHEN

Abstract: The value of financial derivatives such as options depends, among other things, on the volatility of the underlying asset. Estimating volatility from historic data on asset returns with respect to models of stochastic volatility is inherently difficult due to the fact that volatility states cannot be directly measured. In order to investigate a solution to this problem, we use a sequential method based on particle filters to infer historic volatility from simulated data for a specific discrete approximation of the Hull-White model on stochastic volatility.

Contents

List of Tables

List of Figures

Acknowledgement

I would like to thank my advisor, Prof. Daniela Calvetti, for her direction and her unyielding patience throughout the years. Without her thoughtful advice and insight, this thesis would have never gotten beyond the first page. Moreover, her counsel and enthusiasm toward teaching and research have made important imprints on me as a student and as an individual.

The number of hours that Professor Erkki Somersalo has taken to aid me in this project is so large that I am embarrassed at the thought of his other duties. My only consolation is in knowing his enthusiasm for teaching as well as for research. His thoughtful suggestions and guidance are what helped me pull through during the toughest parts of the project, and I must express my enduring gratitude.

It has been an unforgettable experience being a student of these two brilliant Professors. No matter where my professional development takes me, I hope that our paths will cross again in the future.

1 Introduction

Stochastic volatility models are frequently used to describe the dispersiveness of asset prices with respect to an underlying random process. However, it is difficult to assess past volatility from historic data with respect to these models because such process is not directly measurable. In order to investigate a solution to this problem, we use a sequential Monte Carlo method proposed by Liu and West [11] to infer historic volatility as simulated by a specific discrete approximation of the Hull-White model on asset prices with stochastic volatility.

This thesis will try to approach the problem from the ground up. In Chapter 2, we will first review some basic concepts and results of probability, then introduce the concept of volatility through option pricing models, specifically, the Black-Scholes option model and the Hull-White stochastic volatility model in Chapter 3. Following that, we will build a repertoire of particle filter methods for state estimation in Chapter 4. Finally, in Chapter 5, a specific Liu and West particle filter will be introduced to approach the problem of parameter estimation, and be adapted to investigate historic volatility as simulated by an approximation of the Hull-White model of stochastic volatility.

2 Introduction to Probability

This chapter reviews some of the basic ideas and notions in probability that will be used in the methodology for volatility estimation. Before starting with a review of the definitions and results of importance, we present a rather informal, motivational discussion. The common starting point of probability theory is the sample space set, Ω, whose members are put into a one to one correspondence with all the possible outcomes of some experiment. For example, Ω could be the result of a flip of a coin ($\Omega = \{H, T\}$), or the roll of an ordered pair of dices, ($\Omega = \{(1,1), ..., (1,6), (2,1), ..., (6,6)\}$). An "event" in the forementioned experiments is commonly defined as a subset of Ω that corresponds to the answer of a "yes" or "no" question. For example, prior to a roll of a single dice, we may ask the question "Will the top facing value be greater than 3?" If we call A the subset of Ω associated with the answer "yes", we can deduce that $A = \{4, 5, 6\}$ and that the complement for "no" is $A^c = \{1, 2, 3\}$. Furthermore, we may assign A a probability, $\mathbb{P}(A)$, that must satisfy certain properties. As a mathematical construct, we would require that $\mathbb{P}(\Omega) = 1$ and countable additivity (Condition 3 in Definition 1.1) for a certain events set F, which we shall soon formalize, on Ω. [1, 4]

While it may be tempting to define the set of events in terms of all possible subsets of Ω, this may cause problems with measurability, as would be the case for the Ω that corresponds to $\mathbb{R}^n$. Therefore, we shall require

that the event set be a σ-algebra defined over the subsets of Ω. If Ω is finite then the probability space simplifies and F is the set of all measurable subsets of Ω; see [1, 4].

Definiton 1.1: Probability Space

Given a nonempty set, Ω, let F be a σ-algebra of subsets of Ω. A probability measure $\mathbb{P}$ is a function with the properties that:

(1) $\forall A \in F, \quad \mathbb{P}(A) \in [0, 1]$,

(2) $\mathbb{P}(\Omega) = 1$,

(3) For any sequence of disjoint sets $A_1, A_2, ... \in F$ we have

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n). \tag{1}$$

The triple, $(\Omega, F, \mathbb{P})$, is called a probability space. [4]

There are two interpretations of the measurement $\mathbb{P}(A)$ assigned to the event $A \in F$. In the frequentist perspective, $\mathbb{P}(A)$ would correspond to the relative frequency of the occurence of A in a large number of independent trials of the same random experiment modeled by our probability space. A common illustrative example of the interpretation is the probability assigned to the outcome of head or tail from a single flip of a fair coin. Since this experiment has been performed many times and is scientifically re-

peatable, it is commonly justified by assigning the value 0.5 to each of the possible events. While this intuitive approach has certainly its appeal, since many situations that we deem random, in the sense that they are not fully predictable, do not have the benefit of empirical histories, we are commonly confronted with the task of assigning probabilities to events for which we inherently lack enough information to follow the frequentist perspective. For example, what is the chance of Ohio State beating Michigan this year? Or that the growth in the Gross Domestic Product of Norway would exceed two percent per annum for the next decade? Given our definition of a probability space, probability measurements can still be assigned consistently without the sample information needed for the frequentist interpretation. In these cases, we assume an inherently Bayesian perspective and assign the probability values according to our subjective beliefs about the outcome of the situation [2].

2.1 Random Variables

Random variables and the functions which describe them are at the core of our methodology. We introduce here their formal definitions and the most important related concepts which will be used throughout this thesis. Our notation follows closely [1].

Definition 1.2

Let $(\Omega, F, \mathbb{P})$ be a probability space. A random variable is a real-valued function X defined on Ω with the property that for every Borel subset B of $\mathbb{R}$, $B \in B(\mathbb{R})$, the subset of Ω given by

$$X^{-1}(B) = \{\omega \in \Omega | X(\omega) \in B\} \subset \Omega \tag{2}$$

is in the σ-algebra F. [4]

For conciseness in the following, we will refer to $X(\omega)$ simply as the random variable X.

Since the Borel set B can be thought of in terms of events for which we have probability measurements, this leads naturally to the concept of the random events $X^{-1}(B)$, which we will denote by $\mathbb{P}(X \in B)$. To simplify the notation, we will sometimes write $\mathbb{P}(B)$ with the understanding that we are refering to the underlying events in Ω.

Definition 1.3: Discrete Random Variables

A random variable X defined on $(\Omega, F, \mathbb{P})$ is simple if and only if it can take on only finitely many possible values in $\mathbb{R}$; X is discrete if and only if set the of the set of possible values of X is finite or countably infinite.

The probability mass function of a discrete random variable is defined as

$$\mathbb{P}(X \in B) = \sum_{X(\omega) \in B} \mathbb{P}(\omega).$$

The summation is sometimes written in the form $\sum_{x \in B} \mathbb{P}(x)$ [1].

Definition 1.4: Distribution Function

The distribution function of a real valued random variable X, $\mathbf{F}_X = \mathbf{F}$, is a mapping from [0,1] such that

$$\mathbf{F}(x) = \mathbb{P}(\omega : X(\omega) < x).$$

Definition 1.5: Continuous Random Variables

A random variable X is continuous if and only if its distribution function $\mathbf{F}(x)$ is continuous on $x \in \mathbb{R}$.

Definition 1.6: Absolutely Continuous Random Variables and Density Functions

A random variable X is absolutely continuous if and only if there is a nonnegative real-valued Borel measurable function π on $\mathbb{R}$ such that:

$$\mathbf{F}_X(x) = \int_{-\infty}^{x} \pi(t)dt.$$

Here, π is called the density function of X.

Since under these conditions it is possible to compute the difference $\mathbf{F}(b) - \mathbf{F}(a)$, there is a Lebesque-Stieltjes measure determined by $\mathbf{F}$ in $\mathbb{P}$ which satisfies

$$\mathbb{P}_X(B) = \int_B \pi(x)dx, \qquad \forall B \in B(\mathbb{R}).$$

In this thesis, all random variables will be assumed to be absolutely continuous with a given probability density unless otherwise stated.

2.2 Random Vectors

The previous definitions for univariate random variables can be extended to the multivariate case as follows; see [1] for details.

Definition 1.7: Random Vectors

An n-dimensional random vector X on a probability space $(\Omega, F, \mathbb{P})$ is a Borel measurable map from Ω to $\mathbb{R}^n$, such that for every Borel subset B of $\mathbb{R}^n$, the subset of Ω given by

$$X^{-1}(B) = \{\omega \in \Omega; X(\omega) \in B\} \subset \Omega \tag{3}$$

is in the σ-algebra F.

Similiarly, the univariate case $\mathbb{P}(X \in B)$ can be thought of as as the probability measure of the random vector X; see [4].

Definition 1.8: Discrete Random Vectors

X is discrete if and only if the set of the values of X is finite or countably infinite, which is equivalent to saying that all the components $X_1, ..., X_n$ are discrete. The probability mass of a discrete random vector is defined similarly as in the univariate case.

Definition 1.9: Joint Distribution Function

The joint distribution function $\mathbf{F}_X$ of $X = (X_1, ..., X_n)$ in $\mathbb{R}^n$ is the mapping from [0,1] such that:

$$\mathbf{F}(x) = \mathbb{P}(\omega : X_i(\omega) \leq x_i, \;\; i = 1, ..., n).$$

Definition 1.10: Continuous Random Variables

An n-dimensional random vector X is continuous if and only if its distribution function $\mathbf{F}(x)$ is continuous on $x \in \mathbb{R}^n$.

Definition 1.11: Absolutely Continuous Random Vectors

An n-dimensional random vector X is absolutely continuous if and only if there exists a nonnegative Borel measurable function π on $\mathbb{R}^n$, such that:

$$\mathbf{F}(x) = \int_{-\infty}^{x_1} ... \int_{-\infty}^{x_n} \pi(t) dt_1 ... dt_n.$$

Since the previous equation allows the calculation of the difference $\mathbf{F}(b) - \mathbf{F}(a)$, there is a Lebesque-Stieltjes measure determined by $\mathbf{F}$ in $\mathbb{P}$ which satisfies

$$\mathbb{P}_X(B) = \int_B \pi(x)dx, \qquad \forall B \in B(\mathbb{R}^n).$$

Additionally, for each X_i, $1 \leq i \leq n$, there is a density function π_i and its associated measure $\mathbb{P}(X_i)$, which is obtained from the joint measure by simply integrating out the other components of the random vector [1].

All random vectors that will be considered in this thesis are assumed to be absolutely continuous unless otherwise stated.

2.3 Independence, Conditioning, and Expectations

Definition 1.12: Independence

Random variables $X_1, ..., X_n$ are independent if and only if for any collection of set $B_1, ..., B_n \in B(\mathbb{R})$,

$$\mathbb{P}(X_1 \in B_1, ..., X_n \in B_n) = \mathbb{P}(X_1 \in B_1) \cdots \mathbb{P}(X_n \in B_n).$$

Definition 1.13: Conditional Probability

Given a pair of random vectors, X and Y, the conditional probability

of X given that Y takes on a specific value y is:

$$\mathbb{P}(X|Y = y) = \frac{\mathbb{P}(X \cap Y = y)}{\mathbb{P}(Y = y)},$$

provided that the denominator does not vanish. Unfortunately this formula is not applicable when we have a continuous joint distribution because the distribution function $\mathbb{P}(Y = y)$ is equal to zero. In this case, since we assume absolute continuity of our random variables, we use the probability density of X given $Y = y$:

$$\pi(x|y) = \frac{\pi(x, y)}{\pi(y)},$$

where $\pi(y)$ is value of the density function of the random variable Y at $Y = y$. Swapping the roles of x and y in the last formula and solving for the joint density yields the famous Bayes' identity:

$$\pi(x|y) = \frac{\pi(x)\pi(y|x)}{\pi(y)}.$$

Definition 1.14: Expectation, Variance, and Covariance

The expectation and variance of an univariate random variable X are defined as

$$E[X] = \int_{\Omega} X d\mathbb{P},$$

and

$$V[X] = E[X^2] - E[X]^2 = E[(X - E[X])^2],$$

respectively.

The covariance of two random variables X and Y with finite expectations is

$$C[X, Y] = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y].$$

If X is an absolutely continuous random variables with density function π,

$$E[X] = \int_{\Omega} X d\mathbb{P} = \int_{\mathbb{R}} x \cdot \pi(x) dx.$$

Furthermore, for joint probability density,

$$E[XY] = \int_{\Omega} XY d\mathbb{P} = \int_{\mathbb{R}\times\mathbb{R}} x \cdot y \cdot \pi(x, y) dx dy.$$

If X and Y are two independent random variables then

$$E[XY] = E[X]E[Y], \tag{4}$$

therefore their covariance is

$$C[X,Y] = E[XY] - E[X]E[Y] = E[X]E[Y] - E[X]E[Y] = 0.$$

It must be noted that uncorrelated varaibles are not necessarily independent.

2.4 Stochastic Process

A stochastic process on a probability space $(\Omega, F, \mathbb{P})$ is a family of random variables $(X_t)_{t\in T}$, where T is an arbitrary index set. Often, t denotes time and T is the set of natural numbers for discrete time random variables, and of the positive real $\mathbb{R}^+$ for continuous time random variables [1]. We refer to X_t, for a given $t \in T$, as the state of the stochastic process. An example of a stochastic process in discrete time would be the wealth of a gambler under successive plays at the roulette table. In that case, X_0 is the gambler's initial wealth and X_n the state of his fortune after the n-th spin of the wheel.

2.5 Markov Process

A Markov process is a stochastic process defined over a well-ordered set T, such that for a given $t \in T$ and any $h > 0$ and $B \in B(\mathbb{R})$:

$$\mathbb{P}(X_{t+h} \in B | X_s = x_s, \forall s \leq t) = \mathbb{P}(X_{t+h} \in B | X_t = x_t), \qquad \forall B \in B(\mathbb{R}).$$

Equivalently, the state X_{t+h} only depends on the value of X_t and not on the values $X_s, s < t$.

For discrete time processes with variables that are absolutely continuous, the above Markov property can be interpreted in terms of conditional density function as:

$$\pi(x_{t+1}|x_1, ..., x_t) = \pi(x_{t+1}|x_t), \qquad \forall t \in \mathbb{N}.$$

In the rest of this thesis we will write

$$x \sim f_\theta$$

to mean

$$\pi(x) = f_\theta(x),$$

where θ is a vector of parameters associated with the density function f.

3 Stochastic Volatility, a Filtering Problem

Volatility, in general, is a measure of the dispersiveness or variability of an asset price over time intervals. Assumptions about the underlying volatility of assets are key in many portfolio management and option valuation strategies. Volatility was classically treated as an exogenous constant. For example, the Nobel Prize winning Black-Scholes model of derivatives made the assumption of constant volatility and produced a simple closed form formula for option pricing. Such pricing models, however, are only theoretically accurate in short-term, and econometric assessments of asset volatility have produced bodies of evidence demonstrating that, far from being constant, volatilities change frequently and almost unpredictably over time [7]. Many endogenous models of volatility have been proposed to fill in the gap and to address volatility in a more appropriate manner. In this thesis, we will start with discussing, in a somewhat intuitive away, the effect of volatility on derivative pricing, thereby motivating the importance of modeling volatility in option pricing. First, we introduce the parameter volatility in the model for option pricing in Black-Scholes model of derivative pricing, and proceed to describe the Hull and White model of stochastic volatility pricing. Finally, we shall approach a discrete time approximation of the Hull-White model, found in literature, as a filtering problem in which we infer information about the underlying volatility from asset returns [12].

3.1 Volatility on Option Pricing

Options are contracts about some predetermined asset between two parties, the buyer and the seller. European call option is a specific contract that gives the buyer the option, but not the obligation, to purchase the underlying asset from the seller for a certain strike price, denoted K, at a specific time called maturity in the future. Opposite that, European put option is a contract that gives the buyer the option, but not the obligation, to sell the underlying asset to the seller of the contract for a certain strike price at maturity.

The value of a European call option at its expiration date is determined by the price of the asset at that particular future date. Let's denote the price by S_t. If the price of the asset is greater than the agreed strike price, i.e., $S_t > K$, the buyer can exercise his call and buy the asset for K from the seller, and then sell the asset in the market for S_t netting a profit of $S_t - K$ barring transaction fees. If $S_t \leq K$, it would be unwise for the buyer to exercise this option as the asset could be purchased in the market at a lower price, i.e., the option is worthless. Therefore, at the fruition date of the contract, value of the call option is max$(S_t - K, 0)$ assuming no transaction fees.

At a time before the expiration date of the contract, the price of the option must be based on the expectation of the future asset price. To introduce volatility in a very simple binomial model of asset pricing, let's assume for the sake of simplicity that the future price can assume only

High-Volatility to Strike

	S_t^-	S_t^+
Asset Price:	\$10	\$90
Option Value:	\$0	\$40

Low-Volatility to Strike

	S_t^-	S_t^+
Asset Price:	\$30	\$70
Option Value:	\$0	\$20

Table 1: Call option payoffs at strike date for two different assets. The option value is calculated as max(Asset Price - 50, 0). Volatility of the asset determines the option payoff at the strike date.

two values; a higher value S_t^+ for if the asset rises in price and a lower value S_t^- for if the asset decreases in price. Since volatility is a measure of the dispersiveness of the asset price, the distance of S_t^+ and S_t^- from the mean would represent the volatility.

Let's consider the two scenarios (modified from an example given by [3]) for an asset with European call option with a strike price of \$50 given in Table 1. We observe that high volatility increases the payoff of the option for the asset at the event of higher asset price. Additionally, since the option has a lower bound of zero for its payoff, the increased volatility does not affect its value at event of a lower asset price. From this simple example we can see that a reliable assessment/ prediction of volatility is key in determining the value of options.

3.2 Volatility in Option Pricing

We now give a more rigorous illustration of of volatility using the classic Black-Scholes equation for options pricing model.

Let's model the price of an underlying stock asset $S(t)$ by geometric Brownian motion of the form:

$$dS(t) = \alpha S(t)dt + \sigma S(t)dW(t), \tag{5}$$

where α is called the drift parameter, σ is the volatility, and $W(t)$ is a Wiener process or Brownian motion, whose definition is given in the Appendix together with the notation for Ito processes. We assume for the time being that α and σ are constant, and we want to determine the price of European call option on $S(t)$ at a time t. We denote this function, $c(t, S(t))$, where the price of the option is determined only by the time, and the process S at time t. Here, the strike price K and maturity T are fixed. If we assume that $c(t, S(t))$ is at least twice differentiable with respect to $S(t)$ and is differentiable with respect to t, Ito's integration formula prsented in the Appendix can be used to obtain a differential of its discounted value $e^{-rt}c(t, S(t))$:

$$
\begin{aligned}
d(e^{-rt}c(t,S(t))) &= e^{-rt}[-rc(t,S(t)) + c_1(t,S(t)) + \alpha S(t)c_2(t,S(t)) \\
&\quad + \frac{1}{2}\sigma^2 S^2(t)c_{22}(t,S(t))]dt + e^{-rt}\sigma S(t)c_2(t,S(t))dW(t).
\end{aligned} \tag{6}
$$

In the formula above, r is the growth rate of riskless assets.

Assume that a market portfolio of the value $X(t,S(t))$ consists of risk free assets and the assets $S(t)$, the proportion of risk free asset being $\Delta(t)$,

$$
dX(t,S(t)) = \Delta(t)dS(t) + r(X(t) - \Delta(t)S(t))dt. \tag{7}
$$

For the discounted value of the portfolio $e^{-rt}X(t,S(t))$, assuming that the $\Delta(t)$ term is not affected by the discounting (the portfolio is still self-financing). We can derive the rate of change of x using Ito's formula:

$$
d(e^{-rt}X(t,S(t))) = \Delta(t)\left[(\alpha - r)e^{-rt}S(t)dt + \sigma e^{-rt}S(t)dW(t).\right] \tag{8}
$$

If the performance of the call option can be replicated with a market porfolio, that is, $e^{-rt}c(t,S) = e^{-rt}X(t,S)$ for any $t \in [0,T]$, then

$$
d(e^{-rt}c(t,S(t))) = d(e^{-rt}X(t,S(t))),
$$

assuming $c(0, S(0)) = X(0, S(0))$. Using the expressions in (6) and (8) for the differentials and simplifying, we obtain:

$$\Delta(t) = c_2(t, S(t)) \qquad \forall t \in [0, T) \tag{9}$$

by setting the $dW(t)$ terms equal on both sides. Additionally, by equating the coefficients of dt, we obtain

$$rc(t, S(t)) = c_1(t, S(t)) + rS(t)c_2(t, S(t)) + \frac{1}{2}\sigma^2 S^2(t)c_{22}(t, S(t)) \tag{10}$$

$$\forall t \in [0, T)$$

with the boundary condition, $c(T, S(T)) = (S(T) - K)^+$. Notice that the parameter α is eliminated during our algebraic manipulation.

Equation (10) is the Black-Scholes partial differential equation for the price of a European call security, whose solution is of the form:

$$BS(S(t), T - t) =$$
$$S(0)\phi\left(\frac{\log(S(t)/K) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}\right) - e^{-r(T-t)}K\phi\left(\frac{\log(S(t)/K) + (r - \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}\right), \tag{11}$$

where ϕ is the cumulative standard normal distribution function,

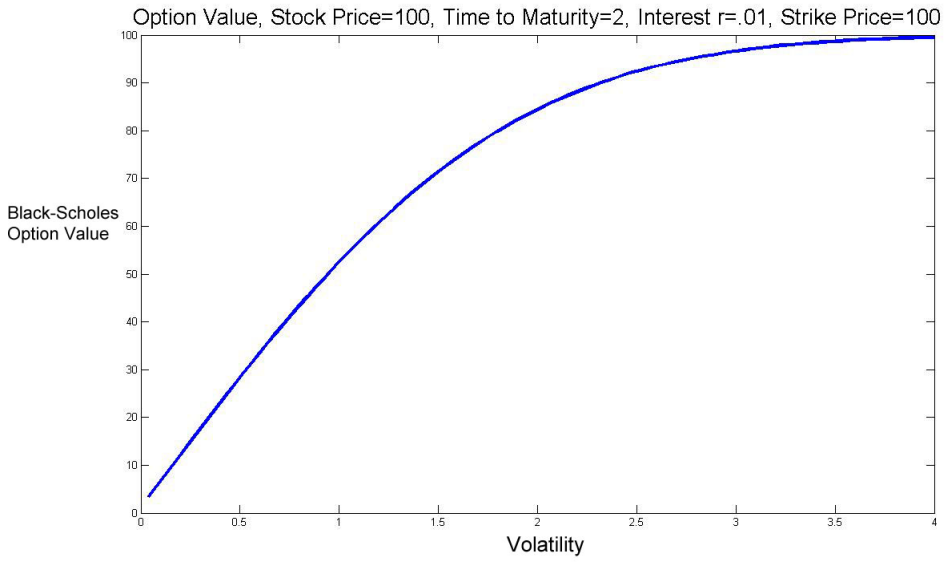

Figure 1: The pricing of derivative assets depends on the volatility of the underlying asset.

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{\infty}^{\infty} e^{-s^2/2} ds.$$

For more detailed derivations, see [4, 13, 6, 14].

The Black-Scholes function (11) shows a clear dependency between the volatility parameter and the value of the call option, as illustrated in Figure 1.

As volatility increases, there is a larger potential payoff for the option due to the greater dispersiveness of asset price. Notice that the price of the option is still bounded by the instant price $S(t)$, because if the option price were higher than the current price, it would be cheaper simply to purchase the underlying stock and hold on to it till the expiration date rather than

to purchase the options.

3.3 Hull and White Stochastic Volatility Model

The Black-Scholes equation for option pricing, while useful and powerful, is constrained by the assumption of a constant volatility. Following the observation that σ often changes over time in a non-deterministic manner, Hull and White proposed the following stochastic volatility model [8]:

$$dS = \phi S\,dt + \sigma S\,dW, \tag{12}$$

$$dV = \mu V\,dt + \xi V\,dZ, \tag{13}$$

where W and Z are Wiener processes with possible correlation $\rho, |\rho| < 1$, and the volatility V satisfies $V = \sigma^2$. Here, μ and ϕ are the drifts of the two processes. The parameter ξ can be thought of as the standard deviation of volatility. As the volatility σ is the measure of the dispersiveness of the stock prices, V must be strictly non-negative [8].

3.4 Studying Stochastic Volatility

In their original paper, Hull and White analyzed their model and its predictions for option value using numerical methods to approximate the stochastic differential equation. However, there may be significant interest

in obtaining historic volatility through data on stock market returns. As noted by Campbell et al [7], there is tremendous interest in obtaining historic volatility as implied by these models, because it gives insight on the underlying economic mechanics. However, there is an inherent difficulty in using continuous stochastic models to study volatility- since volatility itself cannot be directly measured. Research in this field had mostly abandoned approaches using continous stochastic volatility models even in their discretized form. Instead, studies have looked at volatility using sequential autoregressive models such as the ARCH and GARCH, see [7].

We will follow the article of Bordignon and Raggi [12] and consider stochastic volatility models by using a specific auxiliary particle filters technique proposed by Liu and West [11] to study the inherent volatility process of a discretized approximation of the Hull and White.

4 State Estimation Via Particle Filter

In order to apply the technique proposed by Bordignon and Raggi, some general observations are in order. One can be motivated to describe many natural, industrial, and social phenomena using discrete time stochastic processes $\{X_k\}_{k=0}^{\infty}$ and $\{Y_k\}_{k=0}^{\infty}$ called evolution and observation models. Here $\{X_k\}$ represents a phenomenon of interest, which is not directly observable, while $\{Y_k\}$ represents observations that contain indirect information on $\{X_k\}$.

Definition 3.1: Evolution and Observation Model:

A pair of stochastic processes $\{X_k\}_{k=0}^{\infty}$ and $\{Y_k\}_{k=1}^{\infty}$ is called evolution observation model if the probability distributions of the random variables X_k and $Y_k, k \geq 1$ are absolutely continuous with respect to the Lebesgue measure, the initial state X_0 is known, and the probability density functions satisfy [9]:

$$\pi(x_{k+1}|x_0, ..., x_k) = \pi(x_{k+1}|x_k), \tag{14}$$

$$\pi(y_k|x_0, ..., x_k) = \pi(y_k|x_k), \tag{15}$$

$$\pi(x_{k+1}|x_k, y_1, ..., y_k) = \pi(x_{k+1}|x_k). \tag{16}$$

Equation (14) states that $\{X_k\}_{k=0}^{\infty}$ is a discrete time Markov process, that is, the probability density of its future value depends only on its most recent

state. Equation (15) states that $\{Y_k\}_{k=1}^{\infty}$ is a Markov process with respect to X_k. Finally, the last equation of the group states that $\{X_k\}_{k=0}^{\infty}$ is dependent only on its own history.

Additionally, beyond the three equations stated above, we will assume:

$$\pi(y_k|x_k, y_{k-1}, ...) = \pi(y_k|x_k). \tag{17}$$

In such a model, $\{X_k\}_{k=0}^{\infty}$ can be thought of as the governing process and $\{Y_k\}_{k=1}^{\infty}$ as a dependent process with respect to the history of $\{X_k\}_{k=0}^{\infty}$.

Often only the dependent process $\{Y_k\}_{k=0}^{\infty}$ is observable and we are required to estimate $\{X_k\}_{k=0}^{\infty}$ based on the observations. Equation (17) states that the current observation has no memory, i.e., Y_k depends only on the current state X_k. In such a case, the requirement of estimating X_k from the observed data $D_k = \{y_1, ..., y_k\}$ is called a filtering problem.

The evolution-observation model aptly describes many situations. One interesting financial example is the serial observations of stock market returns with respect to some autocorrelated and stochastic behavior of the underlying volatility. Since volatility cannot be directly observed, a practitioner would be required to indirectly estimate it through inversion of the data on stock market returns.

For given density functions, $\pi(x_{k+1}|x_k)$ and $\pi(y_{k+1}|x_{k+1})$ that specify the transition kernel between states, particle filters can be used to estimate the distributions of the random process $\{X_k\}_{k=0}^{\infty}$ via a sequential evolution

of simulated points [9].

4.1 Introduction to Particle Filters

Particle filters, also called sequential Monte Carlo, are a class of computational methods to estimate the conditional probability density distribution $\pi(x_k|D_k)$, where $D_k = \{y_1, ...y_k\}$ is the observed data, by generating a collection of N samples $\{x_k^1, x_k^2, ..., x_k^N\}$ distributed according to the density $\pi(x_k|D_k)$. These samples are called particles, which together with the fact that the problem of estimating $\pi(x_k|D_k)$ is often referred to as a filtering problem, give the method its name.

Particle filter algorithms are based on two observations about the evolution-observation model of stochastic processes [9].

The conditional probabilities satisfy the following identities:

1. Time Evolution Updating (Chapman-Kolmogorov equation for Markov Process):

$$\pi(x_{k+1}|D_k) = \int \pi(x_{k+1}|x_k)\pi(x_k|D_k)dx_k. \tag{18}$$

2. Observation Updating:

$$\pi(x_{k+1}|D_{k+1}) \propto \pi(y_{k+1}|x_{k+1})\pi(x_{k+1}|D_k) \tag{19}$$

Proof:

The Chapman-Kolmogorov equation (18) is a general result for Markov processes. To derive it, we start with Bayes' formula for probability densities from Definition 1.13:

$$\begin{aligned}\pi(x_{k+1}|D_k) &= \frac{\pi(x_{k+1}, D_k)}{\pi(D_k)} = \frac{\int \pi(x_{k+1}, x_k, D_k)dx_k}{\pi(D_k)} \\ &= \frac{\int \pi(x_{k+1}|x_k, D_k)\pi(x_k, D_k)dx_k}{\pi(D_k)}.\end{aligned}$$

Since $\pi(x_{k+1}|x_k, D_k) = \pi(x_{k+1}|x_k)$ by (16), it follows that

$$\frac{\int \pi(x_{k+1}|x_k, D_k)\pi(x_k, D_k)dx_k}{\pi(D_k)} = \frac{\int \pi(x_{k+1}|x_k)\pi(x_k|D_k)\pi(D_k)dx_k}{\pi(D_k)}.$$

Since $\pi(D_k)$ is independent of x_k, it can be factored out of the integral and we obtain equation (18).

For equation (19), we again start with Bayes' formula from Definition 1.13:

$$\pi(x_{k+1}|D_{k+1}) = \frac{\pi(x_{k+1}, D_{k+1})}{\pi(D_{k+1})}.$$

Since $\pi(D_{k+1})$ is independent of x_{k+1}, we have:

$$\begin{aligned}\pi(x_{k+1}|D_{k+1}) &\propto \pi(x_{k+1}, D_{k+1}) = \pi(y_{k+1}, x_{k+1}, D_k) \\ &\propto \pi(y_{k+1}|x_{k+1}, D_k)\pi(x_{k+1}, D_k) \\ &\propto \pi(y_{k+1}|x_{k+1}, D_k)\pi(x_{k+1}|D_k)\pi(D_k).\end{aligned}$$

By (17), $\pi(y_{k+1}|x_{k+1}, D_k) = \pi(y_{k+1}|x_{k+1})$, and since $\pi(D_k)$ is constant, we have:

$$\pi(y_{k+1}|x_{k+1}, D_k)\pi(x_{k+1}|D_k)\pi(D_k) \propto \pi(y_{k+1}|x_{k+1})\pi(x_{k+1}|D_k).$$

Therefore we obtain $\pi(x_{k+1}|D_{k+1}) \propto \pi(y_{k+1}|x_{k+1})\pi(x_{k+1}|D_k)$, thus proving equation (19).

4.2 Sampling Importance Resampling Algorithm (SIR)

A straightforward implementation of particle filtering is the Sampling Importance Resampling (SIR) algorithm. Aside for the initial generation of a set of particles $\{x_0^n\}_{n=1}^N$ distributed approximately by the presumably known distribution $\pi(x_0)$, this algorithm consists of two steps. The first one consists of approximating the distribution of x_{k+1} given $\pi(x_k|D_k)$. For this purpose, the integral $\pi(x_{k+1}|D_k) = \int \pi(x_{k+1}|x_k)\pi(x_k|D_k)dx_k$ is approximated using the previously generated particles $\{x_k^n\}_{n=1}^N$ distributed according to $\pi(x_k|D_k)$. This approximated $\pi(x_{k+1}|D_k)$ is used to generate a new sample, $\{\widetilde{x}_{k+1}^n\}_{n=1}^N$ distributed according to $\pi(x_{k+1}|D_k)$. However, because $\pi(x_{k+1}|D_k) \neq \pi(x_{k+1}|D_{k+1})$, the weights of the new particles needs to be reevaluated in light of the new observations (19), effectively resampling from this set. The new set of resampled particles is used to approximate the distribution of x_{k+1} [9, 10]. The following is a description of how the algorithm is organized:

1. Initilize the procedure by generating random sample $\{x_0^n\}_{n=1}^N$ from an initial prior distribution $\pi(x_0)$.

Set $k = 0$.

2. Approximate the probability density, $\pi(x_{k+1}|D_k)$, using $\{x_k^n\}_{n=1}^N$ as follows:

$$\pi(x_{k+1}|D_k) = \int \pi(x_{k+1}|x_k)\pi(x_k|D_k)dx_k \approx \frac{1}{N}\sum_{n=1}^{N} \pi(x_{k+1}|x_k^n).$$

3. Sample from the density distribution in step 2, by drawing new particle $\widetilde{x}_{k+1}^n$ from $\pi(x_{k+1}|x_k^n)$, $1 \leq n \leq N$.

4. Reweigh the points in the sample $\{\widetilde{x}_{k+1}^n\}_{n=1}^N$ with relative likelihood:

$$w_{k+1}^n \propto \pi(y_{k+1}|\widetilde{x}_{k+1}^n).$$

5. Draw x_{k+1}^m, $1 \leq m \leq M$ from the set $\{\widetilde{x}_{k+1}^n\}_{n=1}^N$ with the sample weights w_{k+1}^n. Set $N = M$. Increase k by 1 and repeat from Step 2.

It should be noted that, in general, the number of the sampling particles N used to approximate $\pi(x_{k+1}|D_k)$ does not have to be equal to the number M of resampling particles for $\pi(x_{k+1}|D_{k+1})$.

4.3 Auxiliary Particle Filters

As noted by Pitt and Shephard in their 1997 paper, the simple Sampling Importance Resampling scheme has two basic weaknesses. The first is that, in practice, parametric densities may not perfectly model the data, and moreover there can be observation points with very low likelihood corresponding to the states that the model predicts. These observational outliers effectively result in a very uneven distribution and a significantly large number of particles would be required in order for the sample to approach the approximated density as compared to theoretical computation of paramters. This is of particular concern when $\pi(y_{k+1}|x_{k+1})$ is very sensitive to x_{k+1}. Pitt and Shephard in [10] demonstrate the sensitivity of the computational method to outliers using the following experiment, which follows very closely to [10] except for a slight change in notations.

Example: Consider an autoregressive model with noise:

$$y_k = x_k + \epsilon_k, \qquad \epsilon_k \sim N(0,1)$$

$$x_{k+1} = 0.9x_k + \eta_k, \qquad \eta_k \sim N(0, \sigma^2 = 0.01)$$

Assume that the first five observations are given by

$$(y_1, y_2, ..., y_5) = (-0.65201, -0.34482, -0.67626, 1.1423, 0.72085),$$

that the value 20.000 is assigned to y_6 and that an SIR filter algorithm is used to evaluate the expected value of x_6 given $y_1, .., y_6$. The simulation based SIR algorithm underestimates x_6 even for implementations using huge number of particles when compared to values given by calculating the estimate theoretically.

The second weakness of the method is that the tails of the distribution $\pi(x_{k+1}|D_k)$ are poorly approximated by the particles $\{x_k^n\}_{n=1}^N$ due to the limitation of the simulation process. For example, if $\pi(x_{k+1}|x_k)$ is a fat-tailed process with non-existing conditional first moment $E[x_{k+1}|x_k]$, by sampling from the mixture $\pi(x_{k+1}|x_k^n)$, we are artificially creating a finite estimate of the expectation. Therefore, there is an unavoidble loss of information while estimating fat tail distributions.

In order to alleviate these problems, Pitt and Shephard proposed using an index n to label the particles as an auxiliary tool to aid in the task of simulation. Here we define a joint density distribution that includes the new index:

$$\pi(x_{k+1}, n|D_{k+1}) \propto \pi(y_{k+1}|x_{k+1})\pi(x_{k+1}|x_k^n)\pi(n), \qquad n = 1, ..., N.$$

This distribution is approximated by

$$\pi(y_{k+1}|\mu_{k+1}^n)\pi(x_{k+1}|x_k^n)\pi(n), \qquad n = 1, ..., N, \tag{20}$$

where μ_{k+1}^n is some reference point (mean, mode, a sample), related to the evolution step $\pi(x_{k+1}|x_k)$, which could serve as an initial approximate of x_{k+1}.

This density is designed so that if we marginalize out x_{k+1} by integration, we obtain:

$$\pi(n|D_{k+1}) \propto \pi(n) \int \pi(y_{k+1}|\mu_{k+1}^n)\pi(x_{k+1}|x_k^n)dx_{k+1} = \pi(n)\pi(y_{k+1}|\mu_{k+1}^n),$$

since $\int \pi(x_{k+1}|x_k^n)dx_{k+1} = 1$.

The strategy is to sample from $\pi(x_{k+1}, n|D_{k+1})$ by simulating the index, n of the collection $\{x_k^n\}_{n=1}^N$ with probabilities λ^n, proportional to $\pi(n|D_{k+1})$. Then, using indices α drawn from $\pi(n|D_{k+1}) \propto \lambda^n$, one continues to sample from the evolution step given by the mixture $\pi(x_{k+1}|x_k^\alpha)$. The λ^n are called the first stage weights of the auxiliary particle algorithm.

In the second stage, the particles in the sample $\{x_{k+1}^n\}_{n=1}^N$ from the joint density $\pi(x_{k+1}, n|D_{k+1})$ are assigned new weights according to the formula:

$$\pi(n) \approx w_{k+1}^n \propto \frac{\pi(y_{k+1}|x_{k+1}^n)}{\pi(y_{k+1}|\mu_{k+1}^{\alpha_n})}, \qquad n = 1, ..., N$$

with the hope that these second stage weights are less variable than those for the normal SIR. The use of λ^n as a first stage weight instead of the uniform $1/N$ weight in the SIR algorithm is motivated by the hope that

the algorithm would spend more time sampling from particles believed to be more important given our reference point.

Here we present an outline of a structured algorithm for auxiliary particle filters proposed in [8, 10]:

1. Initialize the procedure by drawing particles $\{x_0^n\}_{n=0}^N$ and weights, $\{w_0^n\}_{n=0}^N$ from some predetermined prior, $\pi(x_0^n, w_0^n)$.

Set $k = 0$.

2. Generate auxiliary particles $\{\mu_{k+1}^n\}_{n=0}^N$, where μ_{k+1}^n is a reference point so that (20) is an effective approximation of x_{k+1}. A typical choice for μ_{k+1}^n is the conditional mean:

$$\mu_{k+1}^n = E[x_{k+1}|x_k^n], \qquad n = 1, ..., N.$$

3. Generate weights to approximate $\pi(n|D_{k+1})$

$$\lambda^n \propto w_k^n \pi(y_{k+1}|\mu_{k+1}^n), \qquad n = 1, ..., N.$$

4. Draw an index set $\{\alpha_n\}_{n=1}^N$ from the discrete weights λ^n that approximate $\pi(n|D_{k+1})$ and simulate x_{k+1} so that

$$x_{k+1}^n \sim \pi(x_{k+1}|x_k^{\alpha_n}), \qquad n = 1, ..., N.$$

5. Reweigh the samples for $\pi(n)$:

$$w_{k+1}^m \propto \frac{\pi(y_{k+1}|x_{k+1}^n)}{\pi(y_{k+1}|\mu_{k+1}^{\alpha_n})}, \qquad n = 1, ..., N.$$

Increase k by 1 and repeat from Step 2.

5 Combined Parameter and State Estimation

As a consequence of our previous constructs, it is natural to wish to use Sequential Monte Carlo (SMC) methods to estimate the unobservable states for a given evolution-observation problem. However, some relevant problems that are based on stochastic models involve not only estimating the unobservable state but also poorly known parameters θ that governs the mechanism of such a model. In such cases, we must modify our methods. The parameter θ may be scalar or multi-dimensional.

For the problems addressed in this section, we are no longer estimating the density $\pi(x_{k+1}|D_{k+1})$, but rather

$$\pi(x_{k+1}, \theta|D_{k+1}),$$

which, by Bayes' identity, can be written as

$$\pi(x_{k+1}, \theta|D_{k+1}) = \pi(x_{k+1}|D_{k+1}, \theta)\pi(\theta|D_{k+1}) \tag{21}$$

The conditions for a discrete time Markov process (Definition 3.1) still applies to the density $\pi(x_{k+1}|D_{k+1}, \theta)$. Here, we are explicitly stating that this density is conditional on the parameter, θ.

In the general case where the parameter is unknown, we need to extend the sequential updating of the simulated samples of normal particle filter to include also the parameter of interest. This is done by using an

artificial time dependency of the parameter θ, i.e., by introducing a sequence $\{\theta_1, ..., \theta_k, ...\}$. The time evolution updating and observation updating equations for combined parameter estimation are based on (18) and (19) :

1. Time Evolution Updating:

We have the following identity.

$$\pi(x_{k+1}, \theta_{k+1}|D_k) = \pi(x_{k+1}|\theta_{k+1}, D_k)\pi(\theta_{k+1}|D_k), \tag{22}$$

where, if we assume

$$\pi(x_{k+1}|\theta_{k+1}, x_k, D_k) = \pi(x_{k+1}|\theta_{k+1}, x_k),$$

and

$$\pi(x_k|\theta_{k+1}, D_k) = \pi(x_k|D_k),$$

then the right side of (22) can be written as

$$\pi(x_{k+1}|\theta_{k+1}, D_k) = \int \pi(x_{k+1}|\theta_{k+1}, x_k)\pi(x_k|D_k)dx_k,$$

which can be derived in the same way as the Chapman-Kolmogorov equation (18).

2. Observation Updating:

We have

$$\pi(x_{k+1}, \theta_{k+1}|D_{k+1}) \propto \pi(y_{k+1}|x_{k+1}, \theta_{k+1})\pi(x_{k+1}, \theta_{k+1}|D_k), \tag{23}$$

and the marginal densities $\pi(\theta_{k+1}|D_{k+1})$ and $\pi(x_{k+1}|D_{k+1})$ can be obtained by integrating out x_{k+1} and θ_{k+1}, respectively. Equation (23) can be derived by the same method as before if we assume that $\pi(y_{k+1}|x_{k+1}, \theta_{k+1}, D_k) = \pi(y_{k+1}|x_{k+1}, \theta_{k+1})$.

Equation (22) raises the question how to simulate the density $\pi(\theta_{k+1}|D_k)$ from $\pi(\theta_k|D_k)$. Gordon et al (1993) (referenced from [11]) recommend that we complete the auxiliary filtering process by adding Gaussian noise to the prior parameter samples at each step. We introduce a stochastic process $\{\theta_k\}_{k=1}^{\infty}$ with an artificial evolution model:

$$\theta_{k+1} = \theta_k + \zeta_{k+1}, \qquad \zeta_{k+1} \sim N(0, W_{k+1}), \tag{24}$$

for some pre-specified covariance matrix W_{k+1}. In this case, θ_k is a random variable, and its realization defines the sample for the parameter vector θ at the index k used to estimate $\pi(\theta_k|D_k)$ and θ_{k+1} is the "evolved" step used to estimate $\pi(\theta_{k+1}|D_k)$. In this scenario, the observation updating step of the particle filtering algorithm then re-establishes the respective probability weights via the approximation for the density $\pi(x_{k+1}, \theta_{k+1}|D_{k+1})$. The artificial noise evolution ideally provides a mechanism to the auxiliary filter algorithm for generating new parameters at each time step in the sim-

ulation.

However, this can be more problematic than anticipated. If we allow θ to vary sequentially, we are imposing an artificial condition that θ is "time" dependent. For fixed model parameters, this method generates an artificial dispersion of what should be a static θ between time points. The resultant estimates may become too diffuse to be useful. To see this, letting $V_k = V[\theta_k]$ be the variance of θ_k, and choosing ζ_{k+1} conditionally independent of θ_k given D_k, we have the variance:

$$V[\theta_{k+1}] = V[\theta_k + \zeta_{k+1}] = V[\theta_k] + W_{k+1} = V_k + W_{k+1}, \tag{25}$$

showing that we have made our estimate density $\pi(\theta_{k+1}|D_k)$ artificially overly dispersed relative to $\pi(\theta_k|D_k)$.

In order to correct this, we will follow Liu and West's strategy [11] of modifying (24) to deal with fixed parameter and stochastic state estimation.

5.1 Artificial Parameter Evolution

In (24), if ζ_{k+1} is independent of θ_k as was originally proposed, the variance V_k of the implied evolution step density $\pi(\theta_{k+1}|D_k)$ increases by the variance matrix W_{k+1} as demonstrated by equation (25). However, if we allow θ_k and ζ_{k+1} to be correlated, then

$$V[\theta_{k+1}] = V[\theta_k + \zeta_{k+1}] = V[\theta_k] + W_{k+1} + 2C[\theta_k, \zeta_{k+1}]. \tag{26}$$

To avoid artificial dispersion, we require that $V[\theta_{k+1}] = V[\theta_k]$, leading to

$$W_{k+1} + 2C[\theta_k, \zeta_{k+1}] = 0,$$

or, equivalently,

$$C[\theta_k, \zeta_{k+1}] = -W_{k+1}/2.$$

From the evolution noise model, we have $E[\theta_{k+1}] = E[\theta_k] + E[\zeta_{k+1}] = \bar{\theta}_k + 0$. We need to characterize the joint distribution of θ_k and θ_{k+1} assuming that $\zeta \sim N(0, W_{k+1})$. First assume that

$$E\left[\begin{pmatrix} \theta_{k+1} \\ \theta_k \end{pmatrix}\right] = \begin{pmatrix} \bar{\theta}_k \\ \bar{\theta}_k \end{pmatrix},$$

and the covariance is given by

$$V\left[\begin{pmatrix} \theta_{k+1} \\ \theta_k \end{pmatrix}\right] = E\left[\begin{pmatrix} \theta_{k+1} - \bar{\theta}_k \\ \theta_k - \bar{\theta}_k \end{pmatrix}\begin{pmatrix} \theta_{k+1} - \bar{\theta}_k \\ \theta_k - \bar{\theta}_k \end{pmatrix}^T\right]$$

$$= \begin{bmatrix} V_k & E[(\theta_k + \zeta_{k+1} - \bar{\theta}_k)(\theta_k - \bar{\theta}_k)^T] \\ E[(\theta_k - \bar{\theta}_k)(\theta_k + \zeta_{k+1} - \bar{\theta}_k)^T] & V_k \end{bmatrix},$$

due to the assumption that $V[\theta_{k+1}] = V[\theta_k] = V_k$.

Since

$$E[\zeta_{k+1}(\theta_k - \bar{\theta}_k)] = C[\theta_k, \zeta_{k+1}] = -\frac{W_{k+1}}{2}$$

and

$$E[(\theta_k - \bar{\theta}_k)(\theta_k + \zeta_{k+1} - \bar{\theta}_k)^T] = E[(\theta_k - \bar{\theta}_k)(\theta_k - \bar{\theta}_k)^T + (\theta_k - \bar{\theta}_k)\zeta_{k+1}^T],$$

we have:

$$V\left[\begin{pmatrix} \theta_{k+1} \\ \theta_k \end{pmatrix}\right] = \begin{bmatrix} V_k & V_k - W_{k+1}/2 \\ (V_k - W_{k+1}/2)^T & V_k \end{bmatrix}.$$

Let's denote this covariance matrix by Γ, partitioned as

$$\Gamma = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix}.$$

with this notation, the joint probability density $\pi(\theta_k, \theta_{k+1})$ is then

$$\pi(\theta_{k+1}, \theta_k) \propto \exp\left(-\frac{1}{2}\begin{bmatrix} \theta_{k+1} - \bar{\theta}_k \\ \theta_k - \bar{\theta}_k \end{bmatrix}^T \Gamma^{-1} \begin{bmatrix} \theta_{k+1} - \bar{\theta}_k \\ \theta_k - \bar{\theta}_t \end{bmatrix}\right).$$

Using Theorem A.1 in the appendix, we have:

$$\pi(\theta_{k+1}|\theta_k) = N(\bar{\theta}_k + (V_k - W_{k+1}/2)^T V_k^{-1}(\theta_k - \bar{\theta}_k), \tilde{\Gamma}_{22}^{-1})$$

where

$$\tilde{\Gamma}_{22}^{-1} = (V_k - (V_k - W_{k+1}/2)V_k^{-1}(V_t - W_{t+1}/2)^T)^{-1}$$

$$= (I - (I - W_{t+1}V_k^{-1}/2)(I - W_{t+1}V_k^{-1}/2)^T)V_k^{-1}$$

Now, if we denote $A_{k+1} = I - W_{k+1}V_k^{-1}/2$, we can express the covariance of the conditional density as

$$\tilde{\Gamma}_{22}^{-1} = (I - A_{k+1}^2)V_k^{-1}$$

and the mean as

$$\mu = A_{k+1}\theta_k + (1 - A_{k+1})\bar{\theta}_k,$$

therefore

$$\theta_{k+1} \sim N(A_{k+1}\theta_k + (1 - A_{k+1})\bar{\theta}_k, (I - A_{k+1}^2)V_k^{-1}).$$

Liu and West suggest the use of a so called discount factor technique on W_{k+1} by choosing

$$W_{k+1} = V_k(\frac{1}{\delta} - 1),$$

where δ is a scalar factor in (0,1], whose value is normally chosen to be between .95 and .99. With this choice, we have $A = I - \frac{1}{2}(\frac{1}{\delta} - 1)I = \frac{3\delta-1}{2\delta}I = aI$ and $(I - A_{k+1}^2)V_k^{-1} = (1 - a^2)V_k = h^2V_k$. We approximate the conditional evolution density, $\pi(\theta_{k+1}|D_k)$ by generating particles θ_{k+1}^n, which are drawn from

$$N(a\theta_k^\alpha + (1 - a)\bar{\theta}_k, h^2V_k) \tag{27}$$

for a series of $\alpha = \alpha_n$ selected from the weight of the respective particles.

5.2 Fixed Parameter Auxiliary Particle Filter Algorithm

We shall take the density given in equation (27) as a method to implement a modified Gordon artificial evolution of the parameters, and use this to complete the mechanism of auxiliary particle filters. An algorithmic scheme for applying auxiliary particle filters to fixed parameter estimation

models is as follows.

1. Initialize $\{x_0^n\}_{n=0}^N$ and $\{\theta_0^n\}_{n=0}^N$ particles from some predetermined prior. The weight w_0 can be uniform or be derived from the prior distribution. Let $k = 0$. Choose an appropriate $\delta \in [.95, .99]$.

2. Generate auxiliary particles $\{\mu^n\}_{n=0}^N$ and $\{\theta_{aux}^n\}_{n=0}^N$ for x_{k+1} and θ, with the auxiliary weight λ^j, where

$$\mu^n = E[x_{k+1}^n | x_k^n, \theta_k^n],$$

$$\theta_{aux}^n = a\theta_k^n + (1-a)\bar{\theta}_k, \qquad \bar{\theta}_k = \frac{1}{N}\sum_{n=1}^{N}\theta_k^n, \qquad a = (3\delta - 1)/2\delta,$$

$$\lambda^n \propto w_k^n \pi(y_{k+1} | \mu^n, \theta_{aux}^n), \qquad 1 \leq n \leq N.$$

3. For each particle $n = 1, 2, ..., N$, draw $\alpha_n \in \{1, 2, ..., N\}$ with discrete weight λ^n given by the above equation. The weight λ^n approximates $\pi(n|D_{k+1})$, $1 \leq n \leq N$.

4. For each $\alpha = \alpha_n$ drawn, generate a parameter particle

$$\theta_{k+1}^n \sim N(\theta_{aux}^{\alpha}, h^2\Sigma_t),$$

where

$$\Sigma_k = \mathsf{C}(\theta_k), \qquad h^2 = 1 - a^2$$

5. For each $\alpha = \alpha_n$ drawn, generate a state particle $x_{k+1}^n \sim \pi(x_{k+1}|x_k^\alpha, \theta_{k+1}^n)$

6. Calculate new weight,

$$w_{k+1}^n \propto \frac{\pi(y_{k+1}|x_{k+1}^n, \theta_{k+1}^n)}{\pi(y_{k+1}|\mu^\alpha, \theta_{aux}^\alpha)}.$$

Increase k by 1 and repeat from Step 2.

We will demonstrate this algorithm with the following simple example taken from [11], and then apply this to our practical model of stochastic volatility.

5.3 A Simple Parameter Estimation Example

Consider a simple autoregressive model with observable data $\{Y_k\}$, and unobservable state $\{X_k\}$. Since this is a demonstration of the parameter estimation process, we will simplify the setting and take

$$Y_k = X_k$$

$$\pi(x_{k+1}|x_k) = N(\phi x_k, 1)$$

for constant ϕ.

In this example, the filtering side of the code is removed, leaving in only the sequential parameter estimation, ϕ being the only parameter to

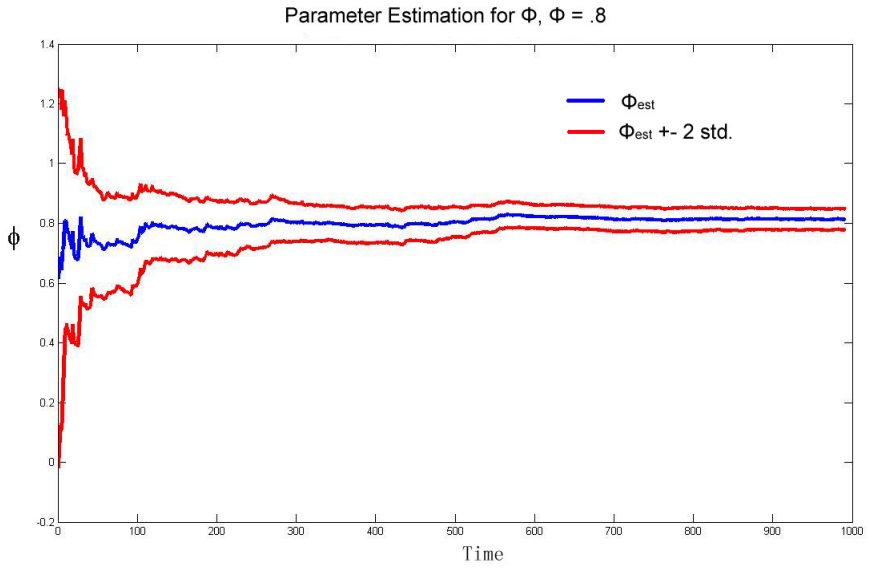

Figure 2: Parameter Estimation of ϕ in 1000 iterations. The estimate converges to the actual value fairly quickly even though we provided the algorithm with an uninformative prior.

be estimated.

The previous problem is solved using auxiliary particle filters in Matlab. In the code, true values of x_0 and ϕ are set to 0.5 and 0.8, respectively, and the model is simulated for one thousand time steps. A particle filter is then implemented. To initialize the algorithm, ϕ_0 is then drawn from the prior, which is $N(0, 10)$. Figure 2 shows a plot of ϕ_{est} plus minus two standard deviation intervals through the one thousand time steps.

Clearly the estimate of ϕ converges to the actual value within a fairly fast window.

5.4 Discrete Model and Filtering

Bordignon and Raggi [12] proposed the following discrete system as an approximation of the Hull and White stochastic model of stock pricing and volatility. We index this model with t rather than k since this model simulates the volatility of returns over time. The starting point for testing particle filters in volatility estimation is the following observation evolution model found in literature [16]:

$$y_{t+1} = e^{V_t/2}\epsilon_{t+1}$$

$$V_{t+1} = \mu + \phi V_t + \sigma_\eta \eta_{t+1}$$

$$\pi(\epsilon_{t+1}, \eta_{t+1}) = N([0,0], B)$$

where

$$B = \begin{bmatrix} 1 & \rho\sigma_\eta \\ \rho\sigma_\eta & \sigma_\eta \end{bmatrix}.$$

Here, y_t represents the logarithmic return of the asset in question, and V_t is the parameter for stochastic volatility, which is an autoregressive process. The correlation ρ can be thought of as leveraging effect [12].

This problem requires us to estimate the volatility state V_t and the four parameters $\theta = [\mu, \phi, \rho, \sigma_\eta]$. However, since y_{t+1} depends on the previous state V_t and not the current state V_{t+1}, this is slightly different from our

requirement for the observation-evolution model expressed in equation (15). Despite this, the algorithm can still be implemented based on the historic V_t values.

A more detailed description of the implementation of the method is given below.

0. Set up discount parameters, $\delta, a,$ and h, where the value of δ is held constant, $a = \frac{(3\delta-1)}{2\delta}$, and $h = \sqrt{1-a^2}$. The number of particles for each step is set to $N = 4000$, the initial weight is uniform.

1. Initialize θ_0 and V_0 using the following priors:

$$\begin{aligned}
\mu &\sim N(0, 10) \\
\phi &\sim \text{Beta}(30, 1.5) \\
\rho &\sim N(-0.5, 1) \\
\sigma_\eta &\sim |N(0.15, .3)| \\
V_0 &\sim N\left(\frac{\mu}{(1-\phi)}, \frac{\sigma_\eta^2}{(1-\phi)^2}\right).
\end{aligned}$$

for $t = 1$ to T

2. For $j = 1, 2, ..., N$, calculate the auxiliary state, the auxiliary parameter, and the auxiliary weight, using the following formulas:

$$V_{aux}^j = E[V_{t+1}|V_t^j, \theta_t^j] = \mu_t^j + \phi_t^j V_t^j,$$
$$\theta_{aux}^j = a\theta_t^j + (1-a)\bar{\theta}_t.$$

3. Determine the auxiliary likelihood:

$$\pi(y_{t+1}|V_t^j, \theta_{aux}^j)$$

Using the previous given probability density, calculate the auxiliary sampling mass, λ^j.

4. For each $j \in \{1, 2, ..., N\}$, draw an α_j, the auxiliary index, from $\{1, 2, ..., N\}$ with weights $\{\lambda^1, \lambda^2, ..., \lambda^N\}$.

5. For each $j \in \{1, 2, ..., N\}$, update $\theta_{t+1}^j \sim N(\theta_{aux}^\alpha, h^2\Sigma_t)$, where Σ_t is a measure of the variance of the sample θ_t^α. Moreover, we will impose the following conditions on the values of σ_η and ρ: $\sigma_\eta > 0$ and $|\rho| \leq 1$.

6. For each $j \in \{1, 2, ..., N\}$, update $V_{t+1}^j \sim \pi(V_{t+1}^j|V_t^j, \theta_{t+1}^j)$.

7. Recalculate the weight w_{t+1}^j:

$$w_{t+1}^j \propto \frac{\pi(y_{t+1}|V_t^\alpha, \theta_{t+1}^j)}{\pi(y_{t+1}|V_t^\alpha, \theta_{aux}^\alpha)}.$$

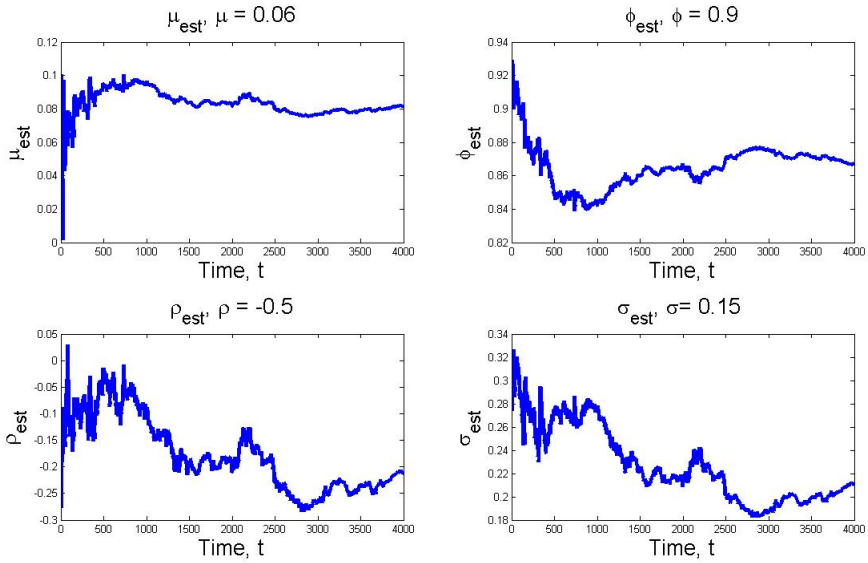

Figure 3: Parameter estimates of θ in 4000 iterations. While the estimates for μ and ϕ seem to get close to their actual values fairly quickly, the ones for ρ and σ are still imperfect.

The results displayed in Figure 3, obtained after about 4000 iterations $N = 2000$, are estimates of the correlation and variance. While not perfect, they seem to be moving in the right direction. Futhermore, the volatility state estimates and their $\pm$ 2 standard deviation range closely match the actual stochastic volatility, as shown in Figure 4.

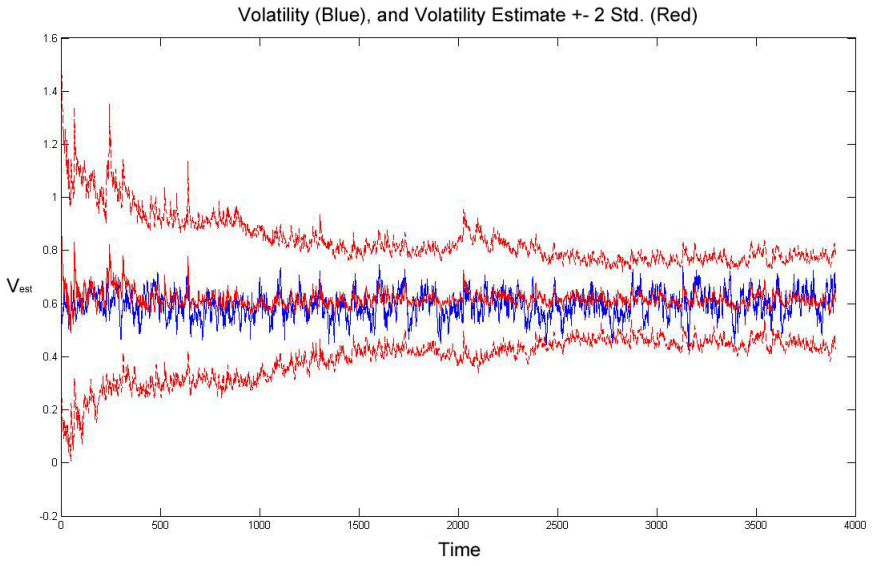

Figure 4: Volatility state estimates rendered by auxiliary particle filter in 4000 iterations. Here we have better luck as the estimates seem to envelop around the true values of the volatility.

6 Conclusion

While imperfect, the auxiliary particle filter method proposed by Liu and West gives a partial solution to the problem of estimating volatility states through historic observations of asset returns. In addition, it has also demonstrated some limited success in estimating parameters for a specific discrete approximation of the Hull-White asset model using stochastic volatility. However, further investigations are warranted. For example, this method can be applied to other assets models of stochastic volatility to determine its effectiveness beyond this specific model. Furthermore, model misspecification and outliers can be introduced to test the robustness of this algorithm.

7 Appendix

In this appendix, we present some technical details used in the derivation of the algorithms.

Lemma 1: Given a positive definite symmetric matrix

$$\Gamma = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix},$$

where $\Gamma_{11} \in \mathbb{R}^{k\times k}, \Gamma_{22} \in \mathbb{R}^{(n-k)\times(n-k)}$, with $n > k$, and $\Gamma_{21} = \Gamma_{12}^T$.
Denote by $\tilde{\Gamma}_{11}, \tilde{\Gamma}_{22}$ the Schur complements $\tilde{\Gamma}_{22} = \Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21}$, and $\tilde{\Gamma}_{11} = \Gamma_{22} - \Gamma_{21}\Gamma_{11}^{-1}\Gamma_{12}$.
Then $\tilde{\Gamma}_{11}$ and $\tilde{\Gamma}_{22}$ are invertible matrices and

$$\Gamma^{-1} = \begin{bmatrix} \tilde{\Gamma}_{22}^{-1} & -\tilde{\Gamma}_{22}^{-1}\Gamma_{12}\Gamma_{22}^{-1} \\ -\tilde{\Gamma}_{11}^{-1}\Gamma_{21}\Gamma_{11}^{-1} & \tilde{\Gamma}_{22}^{-1} \end{bmatrix}.$$

Proof: See [15].

Theorem A.1: (Adapted from [9]) Let $X : \Omega \to \mathbb{R}^n$ and $Y : \Omega \to \mathbb{R}^k$ be two Gaussian random variables. Let the representation of their joint probability density function $\pi : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}_+$ be of the form:

$$\pi(x,y) \propto \exp\left(-\frac{1}{2}\begin{bmatrix} x-x_0 \\ y-y_0 \end{bmatrix}^T \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix}^{-1} \begin{bmatrix} x-x_0 \\ y-y_0 \end{bmatrix}\right). \tag{28}$$

Then the probability density of X conditional on $Y = y$ is of the form

$$\pi(x|y) \propto \exp\left(-\frac{1}{2}(x-\bar{x})^T\tilde{\Gamma}_{22}^{-1}(x-\bar{x})\right) \tag{29}$$

where $\tilde{\Gamma}_{22}$ is the Schur complement of Γ_{22}, and $\bar{x} = x_0 + \Gamma_{12}\Gamma_{22}^{-1}(y-y_0)$

Proof: Without loss of generality, we may assume that $\{x_0, y_0\} = \{0, 0\}$ with lossing generality. It follows from Bayes' formula that $\pi(x|y) \propto \pi(x,y)$. Expand the quantities in (29) and using Lemma 1, we have

$$\pi(x|y) \propto \exp\left(-\frac{1}{2}(x^T\tilde{\Gamma}_{22}^{-1}x - 2x^T\tilde{\Gamma}_{22}^{-1}\Gamma_{12}\Gamma_{22}^{-1}y + y^T\tilde{\Gamma}_{11}^{-1}y)\right),$$

which upon completing the square gives

$$\pi(x|y) \propto \exp\left(-\frac{1}{2}(x-\Gamma_{12}\Gamma_{22}^{-1}y)^T\tilde{\Gamma}_{22}^{-1}(x-\Gamma_{12}\Gamma_{22}^{-1}y)\right) \times \exp\left(-\frac{1}{2}y^T(\tilde{\Gamma}_{11}^{-1} - \Gamma_{22}^{-1}\Gamma_{21}\tilde{\Gamma}_{22}^{-1}\Gamma_{12}\Gamma_{22}^{-1})y\right).$$

The expression $\exp\left(y^T(\tilde{\Gamma}_{11}^{-1} - \Gamma_{22}^{-1}\Gamma_{21}\tilde{\Gamma}_{22}^{-1}\Gamma_{12}\Gamma_{22}^{-1})y\right)$ is independent of x and

therefore can be proportioned out, giving:

$$\pi(x|y) \propto \exp\left(-\frac{1}{2}(x - \Gamma_{12}\Gamma_{22}^{-}1y)^T\tilde{\Gamma}_{22}^{-1}(x - \Gamma_{12}\Gamma_{22}^{-}1y)\right).$$

7.1 Brownian Motion

This is adapted from [4, 6, 5].

Definition: Let $(\Omega, F, \mathbb{P})$ be a probability space. For each $\omega \in \Omega$, suppose there is a continuous realisation $W(t)$, i.e., $t \to W(t)$ is continuous almost certainly, such that $W(0) = 0$. Then $W(t)$ is a Brownian motion if it statisfies the following properties:

1. Stationary Independent Increments: for any sequence $\{t_1, t_2, ..., t_n\}$ with $t_1 < ... < t_n$ and $n \geq 1$,

$$W(t_2) - W(t_1),$$
$$W(t_3) - W(t_2),$$
$$...$$
$$W(t_n) - W(t_{n-1})$$

are independent random variables.

2. Gaussian Increments:

$$W(t_{i+1}) - W(t_i) \sim N(\mu, \sigma)$$

with

$$\mu = E[W(t_{i+1}) - W(t_i)] = 0,$$

and

$$\sigma = V[W(t_{i+1}) - W(t_i)] = t_{i+1} - t_i.$$

7.2 Ito Calculus

Definition: Ito Integral

Let $W(t), t \geq 0$, be a Brownian motion. An Ito integral $\int_0^t \Delta(t) dW(s)$ is defined as:

$$\int_0^t \Delta(t) dW(s) = \lim_{n \to \infty} \sum_{h_{i-1}, h_i \in \pi_n} \Delta(t_{i-1})(W(t_i) - W(t_{i-1})),$$

where π_n is a sequence of partitions of $[0, t]$ with the mesh going to 0 and the limit understood in the L^2 sense [4].

Definition: Ito Process

Let $W(t), t \geq 0$, be a Brownian motion. As defined in [4], an Ito process, X(t), is a stochastic process of the form:

$$X(t) = X(0) + \int_0^t \Delta(u)dW(u) + \int_0^t \Theta(u)du$$

or in differential notation:

$$dX(t) = \Delta(t)dW(t) + \Theta(t)dt.$$

Theorem A.2: Ito's Formula

Ito's formula will be stated and not proved. For a detailed derivation, check [13, 14]. Let $X(t)$ be an Ito Process as defined previously. For any function, $f(t, X(t))$ that has derivatives $\frac{\delta f}{\delta t} = f_1(t, X(t)) = f_1$, $\frac{\delta f}{\delta X(t)} = f_2(t, X(t)) = f_2$, and $\frac{\delta^2 f}{\delta X(t)^2} = f_{22}(t, X(t)) = f_{22}$, we have:

$$df(t, X(t)) = \left(f_1 + f_2\Theta(t) + \frac{1}{2} f_{22}\Delta^2(t) \right) dt + f_2\Delta(t)dW(t),$$

which is the differential form of the equation

$$f(t, X(t)) = f(0, X(0)) + \int_0^t \left(f_1 + f_2\Theta(u) + \frac{1}{2} f_{22}\Delta^2(u) \right) du + \int_0^t f_2\Delta(u)dW(u),$$

where the last integral is understood to as an Ito integral.

References

[1] Robert B. Ash, Catherine A. Dolans-Dade, *Probability & Measure Theory*. Academic Press, San Diego, 2nd Edition, 2000.

[2] Daniela Calvetti, Erkki Somersalo, *Introduction to Bayesian Scientific Computing*. Springer Science + Business Media, New York, 2007.

[3] Peter A. Abken, Saikat Nandi, "Options and Volatility." *Economic Review* 81, no. 3-6 (1996): 25-35 .

[4] Steven E. Shreve, *Stochastic Calculus for Finance II*. Springer Science + Business Media, New York, 2nd Edition, 2008.

[5] Tomas Björk, *Arbitrage Theory in Continuous Time*. Oxford University Press, Oxford, 2nd Edition, 2004.

[6] Thomas Mikosch, *Elementary Stochastic Calculus*. World Scientific, New Jersey, 1998.

[7] John Y. Campbell, Andrew W. Lo, A. Craig MacKinlay, *The Econometrics of Financial Markets*. Princeton University Press, New Jersey, 2nd Edition, 1997.

[8] John Hull, Alan White, "The Pricing of Options on Assets with Stochastic Volatilities." *The Journal of Finance*. 2, (1987): 281-300.

[9] Jari Kaipio, Erkki Somersalo, *Statistical and Computational Inverse Problems*. Springer Science + Business Media, New York, 2004.

[10] Michael K. Pitt, Neil Shephard, "Filtering Via Simulation: Auxiliary Particle Filters." *Journal of the American Statistical Association*. 94, no. 446 (1999): 590-599.

[11] Jane Liu, Mike West, "Combined Parameter and State Estimation in Simulation-Based Filtering." *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, New York, 2000.

[12] Silvano Bordignon, Davide Raggi, "Inference for Stochastic Volatility Models: a Sequential Approach." *Convegno MAF 2006*. Metodi Matematici e Statistici per le Assicurazioni e la Finanza.

[13] Ioannis Karatzas, Steven E. Shreve, *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York, 2nd Edition, 2008.

[14] Bernt Oksendal, *Stochastic Differential Equations: An Introduction with Applications*. Springer, New York, 6th Edition, 2007.

[15] Lloyd N. Trefethen, Bavid Bau III, *Numerical Linear Algebra*. Society for Industrial and Applied Mathematics, Philadelphia, 1997.

[16] S.J. Taylor, "Financial returns modelled by the product of two stochastic preocesses - a study of daily sugar prices 1961-79." In *Time Series Analysis: Theory and Practice*. Anderson (Ed.), Amsterdam: North-Holland, 1982.