# Long-Memory Stochastic Volatility Models: A New Method for Inference and Applications in Option Pricing

Libo Xie

January 25, 2006

Thesis Proposal

## 1 Summary

Stochastic volatility (SV) models play an important role in finance. Under these models, the volatility of an asset follows an individual stochastic process. In contrast to the GARCH model, the volatility process in the SV model is autonomous with no need to refer to the asset price. It is often assumed that the log-volatility process follows a standard ARMA process in an SV model. However, empirical evidence indicates that the volatility of many assets has the "long-memory" property, which means that the autocorrelation of the volatility decays slower than exponentially as it does in an ARMA-type process. One way to incorporate this property into the SV model is by allowing the log-volatility to follow a fractionally integrated ARMA (ARFIMA) process. Such a model is called a long-memory stochastic volatility (LMSV) model. A large part of this research is focused on developing a new inference method via the sequential Monte Carlo (SMC) algorithm to estimate parameters in the LMSV model. In addition, we will check the "goodness-of-fit" of the model by comparing the LMSV model with other models based on the likelihood and other criteria. As an alternative method of model comparison, we can also price certain financial instruments, such as stock options, with our model and compare the results with the real market price and results based on other models.

In this research, I propose to:

1. Develop a likelihood-based estimation method for the LMSV model through the SMC algorithm;
2. Develop a parallelization scheme for the general SMC algorithm and apply it to our estimation method for the LMSV model;
3. Discuss how the LMSV model can be extended to the multivariate situation and related inference methods;
4. Price options under the LMSV model and compare the results with the real market price and results from other pricing methods; discuss possible trading strategies utilizing the long-memory property;

5. Compare the LMSV model with existing volatility models including the GARCH models and the (short-memory) SV model.

# 2 Introduction

A time series is a collection of data observed at a sequence of time points, which are usually spaced at equal intervals. Time series analysis is an important branch of statistics and involves fitting time series data to statistical models and forecasting future values based on the fitted model.

Correlations of values of a time series at different times contain important information about the time series and are one of the basic properties to check at the beginning of time series analysis. A good time series model fitted to data should be able to produce theoretical correlations close to the sample correlations calculated from the data. While correlations of a time series are expected to vanish ultimately when observations are far apart in time, the speed of the decay can be very different. For example, in an ARMA-type process, correlations decrease exponentially as time lag increases; in other time series, the decline can occur at a much slower hyperbolic rate. The latter type of time series is said to have long memory or long-range dependence and occurs frequently in social and natural phenomena. They have been empirically observed in diverse areas such as finance, economics, physics, biology, electrical engineering, hydrology, etc (Beran, 1994; Doukhan et al., 2003; Rangarajan and Ding, 2003).

Time series analysis has extensive applications in economics and finance. In particular, many time series models have been studied in order to understand the behavior of price movements of financial assets such as stocks. Besides the price itself, volatility is also of interest for a variety of reasons. Volatility of an asset (when assumed to be a constant) is the standard deviation of the log-returns of the asset within one year of time. It also serves as a measurement of uncertainty about future price changes. Volatility is an important factor in the famous Black-Scholes option pricing formula. While the formula has been widely used in the trading business, there is also obvious discrepancy between the market price and the theoretical price coming from the formula. Among the factors that may account for the imperfection is the assumption of a constant volatility used in the formula. This assumption is believed to be oversimplified because real price movements usually display apparent nonconstant volatility. To better describe the real volatility, models with time-varying or heteroskedastic volatility were developed. Among them are the well-known ARCH model by Engle (1982) and the generalized ARCH model (GARCH) by Bollerslev (1986). The GARCH model assumes an ARMA-like structure for volatility and squared returns, and therefore the model predicts the current (conditional) volatility as a deterministic function of past returns and volatility. While easy to implement and largely successful in modeling financial asset prices, GARCH models can't explain some patterns observed in prices without considerable modifications. To break through these limitations, more recently, SV models have been developed. SV models assume that volatility follows an autonomous stochastic process, the specification of which does not involve returns. Conditioning on the volatility, returns in SV models are often assumed

to be normally distributed with the volatility being the standard deviation.

Before Ding, Granger, and Engle (1993) published their finding of long memory in the volatility of the S&P 500 daily closing index, volatility models in the literature usually took the assumption of a fast-decaying correlation of volatility. Afterwards, more and more evidence of long memory in volatility was found in financial asset prices, including intraday or high-frequency stock returns (Andersen and Bollerslev, 1997; Taylor, 2005). Accordingly, a great deal of research has been done related to modeling of long memory in volatility in both academic and speculative interests. In particular, a better understanding of this phenomenon should be able to improve option pricing.

A variety of models have been proposed to incorporate long-memory volatility in volatility models. On the basis of the GARCH model, Bollerslev and Mikkelsen (1996) provided a long-memory GARCH model. Breidt et al. (1998) and Harvey (1998) proposed another type of model — the LMSV model, in which the volatility is assumed to follow a latent long-memory ARFIMA process. Although the LMSV model is attractive in several aspects, it is difficult to estimate the parameters in the model, mainly due to the difficulty in evaluating the likelihood. The purpose of this research is to develop a new method for likelihood-based estimation for both univariate and multivariate LMSV models.

This proposal document is organized as follows. The rest of section 2 is an introduction to long-memory processes, the SV model and the LMSV model. Section 3 contains our proposed work. In section 3.1, we propose a parameter estimation method for the univariate LMSV model that is based on the exact likelihood evaluated using the SMC algorithm. Section 3.2 will introduce a new parallelization scheme for the general SMC algorithm, which will be applied to the inference problem for the LMSV model. In section 3.3, we propose a multivariate LMSV model and discuss how to extend the estimation methods for the univariate model to the multivariate model. We then discuss how the LMSV model can be used in option pricing in section 3.4. We will also check the prediction power of the LMSV model. Finally, in section 3.5, we are going to discuss model comparison.

## 2.1 Long Memory

Let $\{X_t, t = 0, 1, 2, \ldots\}$ be a (weakly) stationary time series. The autocovariance function $\gamma(k, t)$ is defined as the covariance of the values of the time series with a time lag of $k$

$$\gamma(k, t) = \mathrm{Cov}(X_t, X_{t+k}), \quad \forall\, t, \quad k = 0, \pm 1, \pm 2, \ldots$$

The (weakly) stationarity condition requires $\gamma(k, t)$ to be independent of $t$, and thus we can rewrite $\gamma(k, t)$ as $\gamma(k)$. When a time series is not stationary, we usually apply certain operations on it to make it stationary before further analysis. Similarly, the autocorrelation function (ACF) $\rho(k)$ is defined as

$$\rho(k) = \mathrm{Corr}(X_t, X_{t+k}) = \gamma(k)/\gamma(0), \quad \forall\, t, \quad k = 0, \pm 1, \pm 2, \ldots$$

(note by symmetry $\rho(k) = \rho(-k)$). In practice, we replace the autocovariance with the sample covariance $\hat{\gamma}(k) = n^{-1} \sum_{t=1}^{n-k} (x_{t+k} - \bar{x})(x_t - \bar{x})$. The (sample) autocorrelation

function $\rho(k)$ can be plotted against the time lag $k$ to form a correlogram. This plot is a useful aid helping us inspect a time series quickly.

The autocorrelation will finally decay to zero as the lag increases, but different time series differ in the speed of decay. For a large family of processes including ARMA processes, the autocorrelation decays exponentially:

$$|\rho(k)| \leq Ac^k, \quad A > 0, \quad 0 < c < 1, \quad \forall\ k > 0.$$

These process are said to have short memory.

A long-memory process has persistent autocorrelation. To be precise, a long-memory process is a stationary process with a hyperbolically decaying autocorrelation function: (Brockwell and Davis, 1991),

$$|\rho(k)| \sim Ak^{2d-1}, \quad A > 0, \quad 0 < d < 0.5, \quad \text{as } k \to \infty. \tag{1}$$

$d$ is the long-memory parameter controlling the speed of decline of the autocorrelation. The time series is more persistent when $d$ is closer to 0.5. Sometimes, the Hurst parameter $H = d + 1/2$ is used in place of $d$. (Thus, for a long-memory process, $0.5 < H < 1$.)

Analogous to the autocorrelation in the time domain, there is its frequency-domain counterpart— the spectrum or the spectral density $f(\nu),\ -\pi < \nu \leq \pi$. The spectrum is defined in a relation to the autocorrelation through the Fourier transform:

$$\begin{aligned} f(\nu) &= \frac{\sigma^2}{2\pi} \sum_{k=-\infty}^{\infty} e^{-ik\nu}\rho(k), \quad -\pi < \nu \leq \pi, \\ \rho(k) &= \frac{1}{\sigma^2} \int_{-\pi}^{\pi} e^{ik\nu} f(\nu) d\nu, \quad k = 0, \pm 1, \pm 2, \ldots, \end{aligned}$$

where $\sigma^2 = \text{var}(X_t)$ is the variance of the time series. Since $\rho(k) = \rho(-k)$, the spectrum also has the symmetry: $f(\nu) = f(-\nu)$.

In practice, since we can only observe a finite number of observations of a time series, we can't determine the spectrum at all values. For this reason, we use the periodogram to approximate the spectrum. The periodogram for a series of observations $\{x_t,\ t = 1, 2, \ldots, n\}$ is defined as follows

$$I_n(\nu_j) = n^{-1} \left| \sum_{t=1}^{n} e^{-i\nu_j t} x_t \right|^2, \quad \nu_j = 2\pi j/n, \quad -\pi < \nu_j \leq \pi. \tag{2}$$

$\nu_j$'s are the Fourier frequencies. We can show that under regular conditions, the vector of $\{I_n(\nu_j),\ 0 < \nu_j < \pi\}$ converges in distribution to a vector of independent random variables with the $j$-th element being exponentially distributed with mean $2\pi f(\nu_j)$ (Brockwell and Davis, 1991). The periodogram, like the correlogram in the time domain, is very useful tool for time series analysis in the frequency domain.

When a linear filter is applied to a time series $H_t$ with spectral density $f_H(\nu)$ satisfying certain regularity conditions, it can be shown that the spectral density of the new time

series $X_t = \sum_{-\infty}^{\infty} a_s H_{t-s}$ equals

$$f_X(\nu) = \left| \sum_{s=-\infty}^{\infty} a_s e^{-is\nu} \right|^2 f_H(\nu) \tag{3}$$

Using this property, we can show that the spectrum of a long-memory process has the following asymptotic form around the origin (Brockwell and Davis, 1991):

$$f(\nu) \sim C|\nu|^{-2d}, \quad 0 < d < 0.5, \quad \text{as } \nu \to 0. \tag{4}$$

The spectrum of a long-memory process therefore diverges around zero. In contrast, the spectrum of a short-memory process has finite values.

### 2.1.1 Modeling of Long Memory

Before looking into long-memory processes, we want to briefly review ARMA processes, since they are the most commonly used processes in time series analysis and many complicated processes are also derived from them. An ARMA$(p, q)$ process $\{X_t\}$ satisfies the following equation:

$$\phi(B)X_t = \theta(B)Z_t. \tag{5}$$

Here, $\{Z_t\}$ is white noise with mean 0 and variance $\sigma^2$. $B$ is the backward shift operator that is interpreted as $B^s X_t = X_{t-s}$ for any integer $s$. $\phi(B)$ and $\theta(B)$ are the autoregressive and moving average polynomials respectively:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p,$$

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q,$$

where $p$ and $q$ are two nonnegative integers. The roots of polynomials $\phi(z)$ and $\theta(z)$ are usually required to lie outside of the unit circle on the complex plane to satisfy the causality and invertibility conditions.

It can be shown that the ACF of an ARMA process decays exponentially, and therefore, ARMA processes are usually not used to model time series data with long memory. However, after some modifications, we are able to construct a long-memory ARFIMA process. An ARFIMA time series $\{X_t\}$ is defined as

$$\phi(B)(1 - B)^d X_t = \theta(B)Z_t. \tag{6}$$

This equation is the same as that of ARMA processes except for the extra (fractionally) differencing term $(1 - B)^d$. $d$ is the long memory parameter introduced earlier. If $d = 0$, the differencing term disappears and $X_t$ degenerates to an ARMA process. To model long memory, we are interested in $d$ in the range of $(0, 0.5)$ and define $(1 - B)^d$ as a power series of operator $B$ via the Taylor's expansion:

$$(1 - B)^d = \sum_{n=0}^{\infty} \frac{\Gamma(d + 1)}{\Gamma(n + 1)\Gamma(d - n + 1)} (-1)^n B^n. \tag{7}$$

When $p = q = 0$, equation (6) becomes $(1 - B)^d X_t = Z_t$. This process, $\{X_t\}$, is often called the fractionally integrated noise. We can write its ACF in a closed-form expression as follows (Brockwell and Davis, 1991)

$$\rho(t) = \frac{\Gamma(t + d)\Gamma(1 - d)}{\Gamma(t - d + 1)\Gamma(d)} = \prod_{0<s\leq t} \frac{s - 1 + d}{s - d}, \quad t = 1, 2, \ldots. \tag{8}$$

By applying Sterling's formula, $\Gamma(x) \sim \sqrt{2\pi} e^{-x+1}(x - 1)^{x-1/2}$ as $x \to \infty$, to the above equation, we get $\rho(k) \sim ck^{2d-1}$ for some constant $c$ as $k \to \infty$. The fact that the ACF decreases hyperbolically shows that the fractional integrated noise is actually a long-memory process. Sowell (1990) and Doornik and Ooms (1993) also developed formulae to compute the ACF of a general ARFIMA process. It can be shown that an ARFIMA process with $0 < d < 0.5$ is a long-memory process by checking the asymptotic behavior of the ACF.

We can also reach the same conclusion by computing the spectral density of an ARFIMA process. Note that the spectral density of a white noise with variance $\sigma^2$ is $\frac{\sigma^2}{2\pi}$. Applying equation (3), we can get the spectral density for an ARFIMA process:

$$f(\nu) = \left|1 - e^{i\nu}\right|^{-2d} \frac{\left|\theta(e^{i\nu})\right|^2}{\left|\phi(e^{i\nu})\right|^2} \frac{\sigma^2}{2\pi} \tag{9}$$

Since $|1 - e^{i\nu}| = 2|\sin(\nu/2)| \sim |\nu|$ as $\nu \to 0$, $f(\nu) \sim c|\nu|^{-2d}$ when $\nu$ is close to 0, which is the asymptotic behavior of long-memory processes.

### 2.1.2 Testing of Long Memory

Two methods are often used to test for long memory in time series data. The first test originates from the GPH estimator for the long-memory parameter $d$ introduced by Geweke and Porter-Hudak (1983):

$$\hat{d} = -\frac{S_{ab}}{2S_{aa}}$$

where $S_{ab} = \sum_{j=1}^{U}(a_j - \bar{a})(b_j - \bar{b})$, $S_{aa} = \sum_{j=1}^{U}(a_j - \bar{a})^2$, $a_j = \log|2\sin(\nu_j/2)|$ and $b_j = \log I(\nu_j)$. In fact, the GPH estimator is simply $-1/2$ times the slope obtained from a simple linear regression of the log-periodogram against $\log|2\sin(\nu_j/2)|$. Geweke and Porter-Hudak (1983) suggests the regression should be conducted at low frequencies to avoid the estimator being contaminated by the short-memory component of the process. Therefore, we usually choose a frequency as the upper truncation point for the regression. At low frequencies, $\log|2\sin(\nu_j/2)| \approx \log \nu_j$, so it is virtually a regression of the log-periodogram against the log-frequency. Later, Robinson (1995) suggests discarding the very first (lowest) several frequencies too in order to reduce bias. Robinson (1995) also obtained the asymptotic distribution of the GPH estimator. He found that in a long-memory process ($d \neq 0$), under some regular conditions,

$$\sqrt{U}\left(\hat{d} - d\right) \rightsquigarrow \mathrm{N}\left(0, \frac{\pi}{\sqrt{24}}\right). \tag{10}$$

However, a similar theoretical result about the asymptotic behavior of the estimator on a short-memory process ($d = 0$), which we need to construct a test for long memory, is missing. Instead, Deo and Hurvich (2003) conducted Monte-Carlo simulation study of the asymptotic distribution of $\hat{d}$ when $d = 0$. Their findings indicate that equation (10) is still valid for $d = 0$. These results are sufficient to construct confidence intervals for $d$ as well as a test for long memory.

The second test is derived from the rescaled range statistic $R/S$ (Hurst, 1951; Mandelbrot and Wallis, 1968). The $R/S$ statistic is defined as

$$R/S = \frac{1}{S(n)} \left[ \max_{0 \le i \le n} \sum_{j=1}^{i} \left( Y_j - \bar{Y}_n \right) - \min_{0 \le i \le n} \sum_{j=1}^{i} \left( Y_j - \bar{Y}_n \right) \right], \tag{11}$$

where $\bar{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i$ is the sample mean and

$$S(n) = \left[ n^{-1} \sum_{i=1}^{n} \left( Y_i - \bar{Y}_n \right)^2 \right]^{1/2}$$

is the sample standard error. The name of the $R/S$ statistic comes from the fact that the numerator in $R/S$ actually measures the range (the difference between the maximum and minimum) of the partial sums of the deviation of a time series from its sample mean. Mandelbrot and others showed that the Hurst parameter $H$ can be consistently estimated as

$$\hat{H} = \frac{\log [R/S]}{\log n}.$$

If only short memory exists in the process, $\hat{H}$ converges to 0.5. If long memory is present, $\hat{H}$ converges to a value greater than 0.5. To reduce bias, we can also regress $\log (R/S)$ against $\log n$ beyond some large $n$.

A test for long memory using the $R/S$ statistic can also be established with the knowledge of its asymptotic distribution. It is known that under the null hypothesis that the time series is IID noise, the normalized $R/S$ statistic ($\frac{1}{\sqrt{n}} R/S$) weakly converges to the range of a Brownian bridge on $[0,\ 1]$. However, if we directly use the normalized $R/S$ statistic as the test statistic, the test will reject a long-memory process as well as a simple short-memory process other than IID noise, e.g. an AR(1) process. The reason is that although $\frac{1}{\sqrt{n}} R/S$ for other short-memory processes still generally converges to the Brownian bridge on $[0,\ 1]$ up to some multiplier constant, the mean, variance and support of the density function of these limiting distributions can be very different.

To overcome this difficulty, Lo (1991) proposed a modified $R/S$ statistic which is the original $R/S$ statistic divided by some normalizing factor to account for the multiplier constant for non-IID short-memory processes. With proper normalizing factors, the modified $R/S$ statistic will have the same limiting distribution for a general class of short-memory processes, and therefore a test can be efficiently established. However, there is no clear criterion for selecting the normalizing factor and one has to take the data into consideration.

### 2.1.3 Empirical Evidence

To demonstrate long memory in volatility, we analyze the daily closing data of the Dow Jones Industrial Average Index from Jan 3, 1990 to Mar 27, 2006. Figure 1 contains the correlograms of the returns of the index and three transforms of returns that serve as proxies of volatility (see section 2.3). We use the method based on the GPH estimator to test for long memory and choose the upper truncation point of the frequency $U = n^{0.45},\ n^{0.5},\ n^{0.55}$ and the lower truncation point $L = n^{0.1}$ as Breidt et al. (1998) did. The estimation and p-value for $d$ are listed in table 1. We also show the regressions of periodograms in Figure 1. The results provide strong evidence of long memory in the volatility.

Table 1: DJIA Index data: the GPH estimator and p-value of the test for $H_0 : d = 0$

| | $r_t$ | $\lvert r_t\rvert$ | $r_t^2$ | $\log(r_t^2)$ |
|---|---|---|---|---|
| Estimation ($U = n^{0.45}$) | -0.234 | 0.684 | 0.500 | 0.574 |
| p-value | | 0.016 | 0.039 | 0.027 |
| Estimation ($U = n^{0.5}$) | -0.155 | 0.416 | 0.302 | 0.457 |
| p-value | | 0.037 | 0.076 | 0.029 |
| Estimation ($U = n^{0.55}$) | -0.163 | 0.486 | 0.363 | 0.443 |
| p-value | | 0.013 | 0.032 | 0.018 |

## 2.2 (Short-Memory) Stochastic Volatility Models

### 2.2.1 Univariate Stochastic Volatility Models

Stochastic volatility (SV) models find extensive use in modeling financial returns. In a SV model, volatility is assumed to be a latent stochastic process. This assumption is natural since volatility is neither observable nor traded.

A (short-memory) SV model can be written as follows,

$$\begin{aligned} r_t &= \mu + \sigma_t \epsilon_t, \\ h_t &= \log \sigma_t, \\ h_t - u_h &\sim \text{ARMA}(p, q). \end{aligned} \tag{12}$$

where $r_t$ are the mean-corrected returns; $\epsilon_t$ are white noise with unit variance; $\sigma_t$ is the volatility, and the log-volatility $h_t$ follows a (short-memory) ARMA process.

The model in equation (12) is a discrete-time model and is ready for time series analysis. There are also continuous-time SV models. In the continuous-time form, stochastic calculus and probability theories provide convenient tools to analyze properties of SV models, and it can be shown that the (continuous-time) SV model fits well into financial theories (see, e.g., Hull and White, 1987; Heston, 1993; Jiang, 1998). A discrete-time SV model

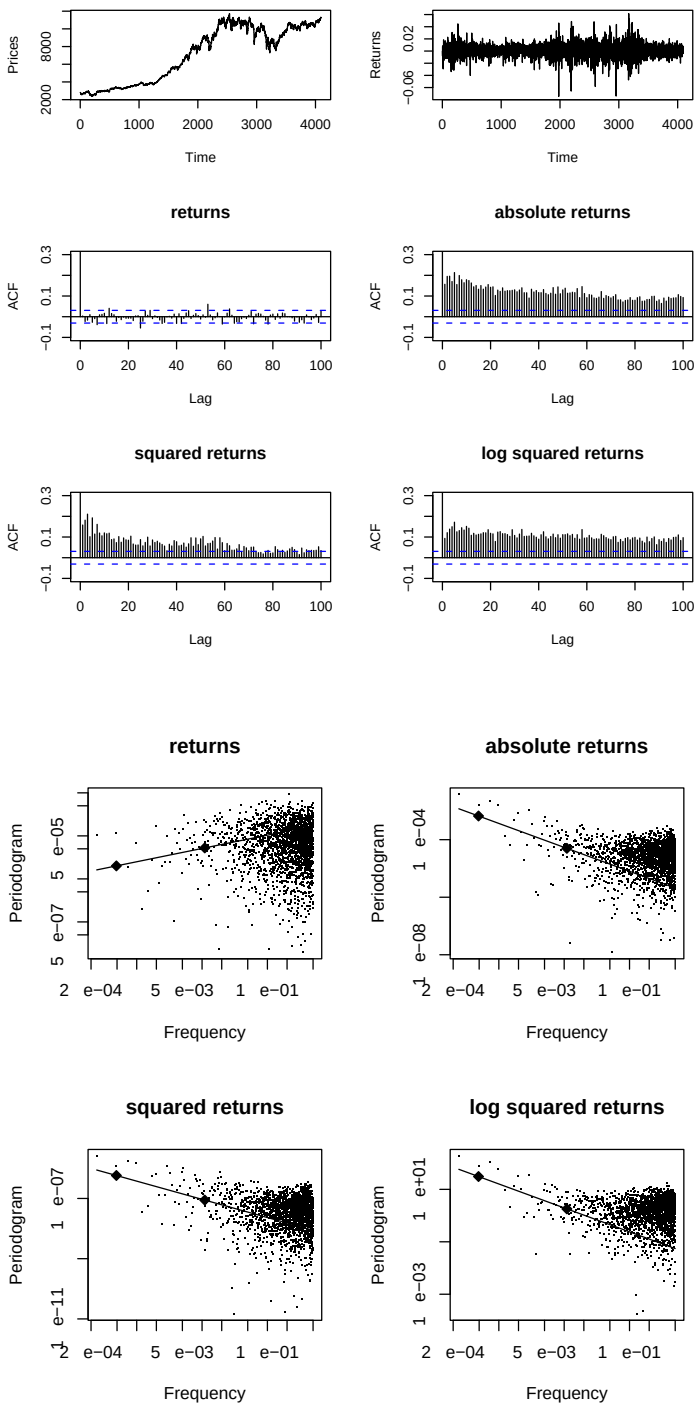



Figure 1: Upper: Correlograms of daily returns of the Dow Jones Industrial Average Index from Jan 3, 1990 to Mar 27, 2006 and three transforms of the returns (volatility proxies). Lower: Regression of the periodograms against the frequencies on the log scale. The lower/upper truncation point in frequency is $n^{0.1}/n^{0.5}$, denoted by two large spots.

is often the result of discretization of a corresponding continuous-time model. However, it is possible that some continuous-time models can be discretized into several different discrete-time models through different discretization methods and some discrete-time models don't have a continuous-time counterpart. In this research, we will mainly focus on discrete-time models.

One of the reasons that SV models become popular is that they can successfully explain many stylized facts observed in financial data. These include volatility clustering — grouping of high and low volatility periods; fat tails in the probability distribution of returns; the leverage effect — negative correlation between returns and volatilities; and the long-range dependence in the absolute or squared returns despite non-significant autocorrelation in the returns (see §2.3). Furthermore, statistical properties of (Gaussian) SV models are easy to derive using the well-known facts of the log-normal distribution. And, it is also relatively straightforward to extend a univariate SV model to a multivariate model.

### 2.2.2 Multivariate Stochastic Volatility Models

The financial market consists of a large number of assets and often we need to consider a portfolio of them at the same time. In this situation, multivariate models will be very useful. In comparison to a univariate model, a multivariate SV model needs not only model the volatility of each asset but also address how the correlation between different assets evolves with time. Because of this extra component, multivariate models can be very complicated. Here, we will present a general structure of the multivariate SV model and several specific models with the structure moving from simple to difficult.

Let $\mathbf{r}_t = (r_{1,t}, \ldots, r_{k,t})'$ be the (log-)returns of $k$ different assets with the mean $\mathbf{u} = (u_1, \ldots, u_k)'$ and the conditional variance matrix $\mathbf{V}_t$. We assume that there exists a scalar or vector stochastic process $\mathbf{h}_t$, such that $\mathbf{V}_t$ is a function of it and

$$\mathbf{r}_t | \mathbf{h}_t \sim \mathrm{N}(\mathbf{u}, \mathbf{V}_t).$$

Different multivariate SV models specify the function $\mathbf{V}(\mathbf{h}_t)$ in a different way.

A very natural extension of the univariate SV model to the multivariate model is the stochastic scalar factor model (Quintana and West, 1987; Shephard, 1994)

$$\begin{aligned} \mathbf{r}_t &= \mathbf{u} + \exp(h_t)\mathbf{e}_t, \\ \mathbf{e}_t &\sim \mathrm{IIDN}(\mathbf{0}, \mathbf{\Sigma}_e), \\ h_t - u_h &\sim \mathrm{ARMA}(p, q). \end{aligned} \tag{13}$$

$h_t$ here is a stochastic scaling factor. This model is equivalent to setting $\mathbf{V}_t = \exp(2h_t)\mathbf{\Sigma}_e$. Although the conditional variance matrix is changing with time, the conditional correlation remains the same, which is solely controlled by $\mathbf{\Sigma}_e$. To identify the model, a diagonal element of $\mathbf{\Sigma}_e$ is usually set to unity.

Instead of having a common scaling factor, we can introduce a scaling vector $\mathbf{h}_t = (h_1, \ldots, h_k)'$ in the multivariate model (Harvey et al., 1994)

$$\begin{aligned} \mathbf{r}_t &= \mathbf{u} + \exp\{\mathbf{A}_t\}\mathbf{e}_t, \\ \mathbf{A}_t &= \text{diag}(\mathbf{h}_t), \\ \mathbf{e}_t &\sim \text{IIDN}(\mathbf{0}, \mathbf{\Sigma}_e), \\ \mathbf{h}_t - \mathbf{u_h} &\sim \text{Multivariate ARMA}. \end{aligned} \tag{14}$$

$\mathbf{A}_t = \text{diag}(\mathbf{h}_t)$ is a diagonal matrix with diagonal elements being $\mathbf{h}_t$. The exponential of matrix $\mathbf{A}_t$, $\exp\{\mathbf{A}_t\} = \text{diag}(e^{h_{t,1}}, \ldots, e^{h_{t,k}})$. The $k$-dimensional $\mathbf{h}_t$ is assumed to follow a multivariate ARMA process. The conditional variance matrix in this model $\mathbf{V}_t = \exp\{\mathbf{A}_t\}\mathbf{\Sigma}_e \exp\{\mathbf{A}_t\}$. Since there are $k$ factors in $\mathbf{h}_t$, the diagonal elements of $\mathbf{\Sigma}_e$ are usually set to be 1 for identification purpose. The conditional variance of a single asset is therefore $e^{2h_{t,i}}$ and $\mathbf{h}_t$ is actually a vector of the log-volatilities. Although there are more stochastic factors in this model, the conditional correlation matrix, $\mathbf{\Sigma}_e$ does not change over time.

A more general multivariate SV model allows additional idiosyncratic errors in the prices. This model is studied in Shephard (1996); Aguilar and West (2000); Chib et al. (2005)

$$\begin{aligned} \mathbf{r}_t &= \mathbf{u} + \mathbf{L}\exp\{\mathbf{A}_t\}\mathbf{e}_t + \mathbf{w}_t, \\ \mathbf{A}_t &= \text{diag}(\mathbf{h}_t), \\ \mathbf{e}_t &\sim \text{IIDN}(\mathbf{0}, \mathbf{I_m}), \\ \mathbf{w}_t &\sim \text{IIDN}(\mathbf{0}, \text{diag}(w_1^2, \ldots, w_k^2)), \\ \mathbf{h}_t - \mathbf{u_h} &\sim \text{Multivariate ARMA}. \end{aligned} \tag{15}$$

$\mathbf{h}_t$ is an $m$-dimensional ($m \leq k$) vector of volatility factors. $\mathbf{e}_t$ is iid standard multivariate Gaussian noise. $\mathbf{L}$ is a $k \times m$ loading matrix. $\mathbf{w}_t$ are independent idiosyncratic errors. Furthermore, $\mathbf{e}_t$, $\mathbf{w}_t$ and $\mathbf{h}_t$ are often assumed to be mutually independent. The conditional variance for $\mathbf{r}_t$ is $\mathbf{V}_t = \mathbf{L}\exp\{2\mathbf{A}_t\}\mathbf{L}' + \text{diag}(w_1^2, \ldots, w_k^2)$. Both variance and correlation are time-varying in this model.

## 2.3 Univariate Long-Memory Stochastic Volatility Models

Due to the structure of the SV model, long memory in the volatility can be easily incorporated into the SV model. In equation (12), we can assume the log-volatility follows an ARFIMA process instead of an ARMA process:

$$\begin{aligned} r_t &= \mu + \sigma_t \epsilon_t, \\ h_t &= \log \sigma_t \\ h_t - \mu_h &\sim \text{ARFIMA}(p, d, q), \quad 0 < d < 0.5. \end{aligned} \tag{16}$$

It can be shown for any positive power $s$, the absolute returns raised to power $s$ possess long memory. In terms of the autocorrelation function, this means

$$\rho(k) \sim Ak^{2d-1}, \quad \text{as } k \to \infty, \quad \text{for } |r_t|^s, \quad \forall\ s > 0.$$

Moreover, $\log(\tilde{r}_t^2)$ for mean-adjusted returns $\tilde{r}_t \equiv r_t - \mu$ can be written as

$$\begin{aligned} \log(\tilde{r}_t^2) &= h_t + \xi_t \\ \xi_t &= \log(\epsilon_t^2). \end{aligned} \tag{17}$$

Since $\log(\tilde{r}_t^2)$ is the sum of a long-memory process $h_t$ and an independent process $\xi_t$ (due to independence of $\epsilon_t$), $\log(\tilde{r}_t^2)$ is also a long-memory process.

As a result, absolute returns, squared returns and log-squared returns are often taken as volatility proxies and used in tests of long memory in the volatility. One issue with log-squared returns is that returns can occasionally be zero when the return on some point of time equals the average return. When this happens, a small number, for example, one percent of the standard error of the squared returns $0.01\ \sigma_{r^2}$, is usually added to the "problematic" squared returns before taking logarithm.

## 2.4 Option Pricing

Options are one of the most actively traded financial instruments in today's financial markets. Options, as a type of derivatives, have their value derived from an underlying asset. A call/put option on a stock is a contract that gives its holder the right but not the obligation to buy/sell the underlying stock (i.e., exercise the option) at a pre-specified strike price at or until an expiration date (maturity). European options and American options are two types of commonly traded options. European options only allow exercise at maturity, while exercise of an American option can happen at any time at or before maturity. For a European call/put option, the holder will exercise the option only if the stock price at maturity is higher/lower than the strike price. Therefore, the European call or put option has a payoff $g(x) = (x-K)^+$ or $(K-x)^+$, where $K$ is the strike price and $x$ is the stock price at maturity.

Option pricing method are intended to predict (theoretical) prices of options at some time before expiration using the information available prior to the time of interest. A lot of research has been done on option pricing, including the Nobel prize-winning Black-Scholes formula. The Black-Scholes formula works for a European option on a non-dividend paying stock in a constant interest rate environment. Moreover, the formula assumes that the underlying stock follows a geometric Brownian motion:

$$\begin{aligned} dS(t) &= \mu S(t)dt + \sigma S(t)dW(t) \\ \text{or} \quad S(t) &= S(0)\exp\left((\mu - \frac{1}{2}\sigma^2)t + \sigma W(t)\right), \end{aligned} \tag{18}$$

where rate of return $\mu$ and volatility $\sigma$ are assumed to be constant and $W(t)$ is a standard Brownian motion. With these assumptions, the Black-Scholes formula gives the price of a European call option with strike price $K$ at time $t$ before maturity $T$ when the stock price at $t$ is $x$ as follows:

$$P_c(t,x) = x\Phi\left(d_+(T-t,x)\right) - Ke^{-r(T-t)}\Phi\left(d_-(T-t,x)\right), \quad t \le T, \tag{19}$$

where

$$\begin{aligned} d_-(T-t,x) &= \frac{1}{\sigma\sqrt{T-t}}\left[\log\frac{x}{K}+\left(r-\frac{\sigma^2}{2}\right)(T-t)\right] \\ d_+(T-t,x) &= d_+(s,x)+\sigma\sqrt{T-t}, \end{aligned}$$

and

$$\Phi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$$

is the cumulative distribution function of a standard normal variable.

There are two approaches to derive this formula (see, for example, Shreve (2004)). In one approach, we set up a portfolio consisting of the underlying stock and a savings account to replicate or hedge the option. A stochastic differential equation can then be established from this strategy, and we obtain the Black-Scholes formula as the solution to this equation. The other approach takes the risk-neutral pricing principle. The principle argues that there should exist a risk-neutral probability measure, under which the discounted price ($e^{-rt}$ times price for a constant interest rate) of any financial instrument, including stocks and options, forms a martingale. The principle further claims that there will be arbitrage opportunities otherwise, which make possible strategies leading to always non-negative yet not always zero (always in a probability sense) profit with zero net investment at the beginning. The risk-neutral probability measure is equivalent (in the measure theory sense) to the objective probability measure that governs the random price movements in the real world. That is, any possible events (with non-zero probability) in the real world are also possible under the risk-neutral measure, and vice versa. Change of measure only changes the probability of those possible events in the real world.

If the stock price follows equation (19) in the real world, we can construct the risk-neutral measure and show that the stock price still follows a geometric Brownian motion under the new measure,

$$\begin{aligned} dS(t) &= \mu S(t)dt+\sigma S(t)dW(t) \\ \Theta(t) &= \frac{\mu-r}{\sigma} \\ d\widetilde{W}(t) &= dW(t)+\Theta(t)dt \\ dS(t) &= rS(t)dt+\sigma S(t)d\widetilde{W}(t) \\ \frac{d\widetilde{P}}{dP} &= \exp\left\{-\int_0^T\Theta(t)dW(t)-\frac{1}{2}\int_0^T\Theta^2(t)dt\right\}. \end{aligned}$$

The last equation defines the risk-neutral measure $\widetilde{P}$ by giving the Radon-Nikodým derivative of $\widetilde{P}$ with respect to the objective measure $P$. It turns out that this derivative is the exponential of a stochastic integral involving $\Theta(t)$, which is a constant in this simple situation but is a stochastic process in general. $\Theta(t)$ is called the market price of risk and can be interpreted as the amount of excess return (return minus risk-free interest rate) required to compensate for one unit of risk (volatility of stocks). The Girsanov's theorem guarantees that $\widetilde{W}(t)$ defined above is a standard Brownian motion under $\widetilde{P}$.

The Black-Scholes formula has been widely used in the financial industry. The price given by the formula serves as a reference price by many traders. However, there are always discrepancies between the Black-Scholes price and the actual price. To test the validity of the model, we can use the historical stock and option prices and calculate the value of volatility that makes the Black-Scholes price equal the real price. This value is called implied volatility. If the Black-Scholes model is perfect, the implied volatility of options on the same stock with the same expiration but with different strike prices should be same. However, it is often observed that the implied volatility is lowest when the strike price is close to the current stock price (at-the-money) and increases when the strike price deviates from the stock price (in/out-the-money). This phenomenon is referred to as volatility smile, as implied by the shape of the plot of implied volatility against strike price. In some cases, the implied volatility simply decreases as the strike price increases, which is often called volatility skew. Both volatility smile and volatility skew demonstrate inaccuracy of the Black-Scholes formula.

To better describe the behavior of stock prices and thus more accurately predict option prices, alternative models and methods are built. In particular, a lot of work has been done on option pricing under stochastic volatility models. Heston (1993) gave a closed-form solution for prices of European options on stocks that are assumed to follow a stochastic volatility model, in which the squared volatility follows a mean-reverting square-root process:

$$\begin{aligned} dS(t) &= \mu S(t)dt + \sqrt{v(t)}S(t)dW_1(t) \\ dv(t) &= k[\theta - v(t)]dt + \beta\sqrt{v(t)}dW_2(t), \end{aligned} \tag{20}$$

where $\theta$ is the long-run mean of the variance process $v(t)$, $k$ is the speed of mean-reversion, and the two Brownian motions $W_1(t)$ and $W_2(t)$ are allowed to be correlated. Heston's model is a special case of a general class of (continuous-time) stochastic volatility model

$$\begin{aligned} dS(t) &= \mu S(t)dt + \sigma(t)S(t)dW_1(t) \\ \sigma(t) &= f(v(t)) \\ dv(t) &= \alpha(t)dt + \beta(t)dW_2(t), \end{aligned} \tag{21}$$

where the volatility $\sigma(t)$ is determined by a stochastic process $v(t)$ through a deterministic function $f(\cdot)$, and $\alpha(t)$ and $\beta(t)$ are two arbitrary adapted processes.

The existence of two Brownian motions in these models means that the price movements are not the only source of randomness that affects value of options. The extra randomness comes from the volatility process. However, unlike stocks, volatility is neither a traded asset nor observable. This makes option pricing with stochastic volatility models more difficult. In particular, use of only the stock and a savings account is insufficient to hedge an option under such a model. Instead, we have to include another option with longer maturity in the hedging portfolio. To use the risk-neutral pricing principle, we then have to construct the risk-neutral measure in a two-dimensional space. Using the multi-dimensional Girsanov's theorem, the change of measure is done as follows,

$$d\widetilde{W}_1(t) = dW_1(t) + \Theta_1(t)dt$$

$$\Theta_1(t) = \frac{\mu - r}{\sigma}$$

$$d\widetilde{W_2}(t) = dW_2(t) + \Theta_2(t)dt$$

$$\frac{d\widetilde{P}}{dP} = \exp\left\{-\int_0^T \Theta_1(t)dW_1(t) - \int_0^T \Theta_2(t)dW_2(t) - \frac{1}{2}\int_0^T \left(\Theta_1^2(t) + \Theta_2^2(t)\right)dt\right\}.$$

If $W_1(t)$ and $W_2(t)$ are two independent Brownian motions under the objective measure $P$, the two new Brownian motions $\widetilde{W_1}(t)$ and $\widetilde{W_2}(t)$ are still independent under the risk-neutral measure $\widetilde{P}$. The market price of risk $\Theta_1(t)$ is the same as before, chosen to make the rate of return of the stock under the risk-neutral measure be the risk-free interest rate. There is a second process $\Theta_2(t)$ associated with the volatility process. Similarly, it is called the market price of volatility risk. However, $\Theta_2(t)$ can't be determined directly as $\Theta_1(t)$ since volatility is not traded. Researchers have been working on a good choice of the market price of volatility risk. For example, Heston (1993) chose $\Theta_2(t) = c\sqrt{v(t)}$ for some constant $c$ and estimate value of $c$ from historical option prices. However, there are many other choices as well. Each choice of $\Theta_2(t)$ defines a new risk-neutral measure and leads to different pricing results.

# 3 Proposed Thesis Research

In my thesis research, I am going to work on the following problems:

1. Being able to evaluate the likelihood is often central to an inference problem. Hence, we first propose a method to estimate the exact likelihood for an arbitrary set of parameters in the LMSV model. Embedded in the method is the SMC algorithm, which can simulate from the filtering and predictive densities of the state variables in a broad class of state-space models. To compute the total likelihood, we first write it as the product of a series of probabilities of each observation conditioning on its past observations. We then evaluate the conditional probabilities with simulations from the filtering densities generated by the SMC algorithm. Equipped with such a method, it seems natural to follow the maximum likelihood estimation approach. However, this won't work well because the random errors associated with simulations will deteriorate the precision of the result. Instead, we propose another approach that can be nicely integrated into the SMC algorithm. In this approach, we mix the parameters with the states and apply the SMC algorithm to the augmented states. When the algorithm ends at the final point in time, we obtain a posterior distribution of parameters, the mode or mean of which can then be taken as an estimate of the parameters. Besides this method, we also briefly discuss another likelihood-based method that uses the Whittle's approximate likelihood in the frequency domain.

2. Since our approach relies on the SMC algorithm, the accuracy of the result depends on the number of simulations. However, more simulations require more computing

time. The SMC algorithm is therefore often computationally intensive. It is desirable to put the algorithm into some parallel scheme so that we can run multiple processors at the same time to increase both accuracy and speed. Due to the nature of the SMC algorithm, a straightforward ("naive") parallelization scheme — simply let each processor finish a part of the whole SMC algorithm without interactions with other processors, does not work. Therefore, we propose a new method that can efficiently utilize multiple processors.

3. In a real situation, an investor's portfolio usually consists of multiple assets simultaneously (consider the stocks composing the S&P 500 or DJIA index). As in the short-memory case, it is desirable to have a multivariate LMSV model to address long-memory properties of several assets at the same time. For this propose, we discuss how to establish a multivariate LMSV model and related estimation methods.

4. A stock option is a financial instrument with its value relying on the (future) prices of the underlying stock. Options are becoming more and more important in today's financial market. They can be used to hedge the risk of stocks as well as for speculative purposes. Methods that can determine option prices accurately would be highly desirable for anyone trading options, and the SV model has proven useful for pricing options. With evidence of long memory in volatility, it is interesting to consider the use of the LMSV model in option pricing. We will compare the prices generated from the model with prices observed in the market. This will show us whether or not the market implicitly takes into account long memory, and provide additional information about model "goodness-of-fit".

5. Besides the LMSV model, there are other models of stochastic volatility. It is interesting to compare the LMSV model with other models including the GARCH model and the (short-memory) SV model. Criteria including the likelihood, AICC and BIC will be used in model comparison.

## 3.1 Likelihood-based Estimation for Univariate LMSV Models

We first develop a method to evaluate the likelihood of the LMSV model via the sequential Monte Carlo algorithm. Although we do not choose a maximum likelihood estimation approach, the method of likelihood evaluation still provides an important framework for a new approach to the inference problem that is embedded in the SMC algorithm. In the following, we first show how we can use the SMC algorithm to evaluate the likelihood of a short-memory AR1-SV model. Then we adapt the method for a LMSV model before we move to the estimation method. Finally, we will briefly talk about a pseudo-likelihood-based estimation method.

### 3.1.1 Likelihood Evaluation for an AR1-SV Model

The following is an AR1-SV model,

$$\begin{aligned} r_t &= \mu + \sigma_t \epsilon_t, \\ h_t &= \log \sigma_t, \\ (1 - \phi B)(h_t - \mu_h) &= \eta_t. \end{aligned} \tag{22}$$

To compute the likelihood, we first write the density function for observations $\{r_1, r_2, ..., r_n\}$ as the product of a sequence of conditional density functions:

$$p(r_{1:n}) = p(r_n|r_{1:n-1}) \cdots p(r_2|r_1)p(r_1). \tag{23}$$

For simplicity, we omit parameters in the notations, e.g., we write $p(r_{1:n}|\theta)$ as $p(r_{1:n})$ and $p(h|r;\theta)$ as $p(h|r)$, etc. Each conditional density function can be written as an integral involving the predictive density of log-volatility

$$p(r_{j+1}|r_{1:j}) = \int p(r_{j+1}|h_{j+1}, r_{1:j})p(h_{j+1}|r_{1:j})dh_{j+1}.$$

Since $r_{j+1}$ is independent of past observations given $h_{j+1}$, we can neglect $r_{1:j}$ in $p(r_{j+1}|h_{j+1}, r_{1:j})$, i.e. $p(r_{j+1}|h_{j+1}, r_{1:j}) = p(r_{j+1}|h_{j+1}) = \phi(r_{j+1}; \mu, \exp(h_{j+1}))$, where $\phi(\cdot; \mu, \sigma^2)$ is the density function of a normal distribution. Therefore,

$$p(r_{j+1}|r_{1:j}) = \int p(r_{j+1}|h_{j+1})p(h_{j+1}|r_{1:j})dh_{j+1}.$$

This integral can be approximated by a Monte Carlo sum

$$p(r_{j+1}|r_{1:j}) \approx \frac{1}{N} \sum_{i=1}^{N} p(r_{j+1}|h_{j+1}^{(i)})$$

if we have a way to draw samples $\{h_{j+1}^{(i)}\}$ from the distribution $p(h_{j+1}|r_{1:j})$.

It turns out an efficient way to do the sampling is through the sequential Monte Carlo algorithm, sometimes also called particle filtering. The algorithm is based on the sequential importance sampling (SIS) and allows us to do sampling sequentially — utilizing samples for previous state $h_j$ to sample new state $h_{j+1}$. To understand how it works in our problem, we can start with writing $p(h_{j+1}|r_{1:j})$ as

$$p(h_{j+1}|r_{1:j}) = \int p(h_{j+1}|h_j, r_{1:j})p(h_j|r_{1:j})dh_j. \tag{24}$$

Because $\{h_j\}$ follows an AR(1) process, its Markovian property enables us to neglect $r_{1:j}$

in $p(h_{j+1}|h_j, r_{1:j})$. We can justify this statement by showing

$$\begin{aligned} p(h_{j+1}|h_j, r_{1:j}) &= \int p(h_{j+1}|h_{1:j}, r_{1:j}) p(h_{1:j-1}|h_j, r_{1:j}) dh_{1:j-1} \\ &= \int p(h_{j+1}|h_{1:j}) p(h_{1:j-1}|h_j, r_{1:j}) dh_{1:j-1} \\ &= \int p(h_{j+1}|h_j) p(h_{1:j-1}|h_j, r_{1:j}) dh_{1:j-1} \\ &= p(h_{j+1}|h_j) \int p(h_{1:j-1}|h_j, r_{1:j}) dh_{1:j-1} \\ &= p(h_{j+1}|h_j). \end{aligned} \tag{25}$$

The second equation holds as $r_{1:t}$ provide no extra information for $h_{j+1}$ when $h_{1:t}$ are known, and the third equation holds for the Markovian structure of AR(1). Hence, equation (24) can be written as

$$p(h_{j+1}|r_{1:j}) = \int p(h_{j+1}|h_j) p(h_j|r_{1:j}) dh_j. \tag{26}$$

Now suppose we have obtained $N$ samples of $h_j^{(i)} \sim p(h_j|r_{1:j-1}), i = 1, ..., N$. By recognition of the following equations:

$$\begin{aligned} p(h_j|r_{1:j}) &= \frac{p(h_j, r_{1:j})}{p(r_{1:j})} = \frac{p(r_j|h_j, r_{1:j-1}) p(h_j, r_{1:j-1})}{p(r_{1:j})} \\ &= \frac{p(r_j|h_j) p(h_j, r_{1:j-1})}{p(r_{1:j})} = p(r_j|h_j) p(h_j|r_{1:j-1}) \frac{p(r_{1:j-1})}{p(r_{1:j})} \\ &\propto p(r_j|h_j) p(h_j|r_{1:j-1}), \end{aligned} \tag{27}$$

samples from $p(h_j|r_{1:j})$, $\hat{h}_j^{(i)}$, can be obtained by importance resampling of $\{h_j^{(i)}\}$ with weights $w_i = p(r_j|h_j^{(i)})$. In view of equation (26), if we draw samples $\{h_{j+1}^{(i)}\}$ from $p(h_{j+1}|h_j = \hat{h}_j^{(i)})$, then the samples together with the weights $\{w_i\}$ serve as a good approximation to the new filtering density $p(h_{j+1}|r_{1:j})$. This can be followed by an important resampling of $\{h_{j+1}^{(i)}\}$ to get equally weighted samples.

Finally, we replace $p(h_1|r_{1:0})$ and $p(r_1|r_{1:0})$ with $p(h_1)$ and $p(r_1) = \int p(r_1|h_1) p(h_1) dh_1$ respectively. $p(h_1)$ is the marginal distribution of the AR(1) process $h_t$, which is a normal distribution.

The algorithm can be summed up in the following diagram:

$$
\begin{aligned}
h_j^{(i)} \sim p(h_j|r_{1:j-1}) &\Longrightarrow p(r_j|r_{1:j-1}) \approx \frac{1}{N}\sum_{i=1}^{N} p(r_j|h_j^{(i)}) \\
&\Downarrow \text{resample } h_j^{(i)} \text{ with importance weights } p(r_j|h_j^{(i)}) \\
\hat{h}_j^{(i)} &\sim p(h_j|r_{1:j}) \\
&\Downarrow h_{j+1}^{(i)} \sim p(h_{j+1}|h_j = \hat{h}_j^{(i)}) \\
h_{j+1}^{(i)} \sim p(h_{j+1}|r_{1:j}) &\Longrightarrow p(r_{j+1}|r_{1:j}) \approx \frac{1}{N}\sum_{i=1}^{N} p(r_{j+1}|h_{j+1}^{(i)}).
\end{aligned}
$$

### 3.1.2 Likelihood Evaluation for LMSV Models

We now want to adapt the method above for use in LMSV models. In a LMSV model, the latent process for log-volatility is an ARFIMA process and the Markovian property that AR processes possess does not hold here. Without this property, equation (25) fails to hold. Hence, in equation (24), we add all previous $h_j$ in the condition of $p(h_{j+1}|r_{1:j})$:

$$
\begin{aligned}
p(h_{j+1}|r_{1:j}) &= \int p(h_{j+1}|h_{1:j}, r_{1:j}) p(h_{1:j}|r_{1:j}) dh_{1:j} \\
&= \int p(h_{j+1}|h_{1:j}) p(h_{1:j}|r_{1:j}) dh_{1:j}. \qquad (28)
\end{aligned}
$$

Thus, given samples $\hat{h}_{1:j}^{(i)} \sim p(h_{1:j}|r_{1:j})$, samples $h_{j+1}^{(i)} \sim p(h_{j+1}|r_{1:j})$ can be drawn from $p(h_{j+1}|h_{1:j} = \hat{h}_{1:j}^{(i)})$. The last density can be evaluated with the methods in Doornik and Ooms (1993) and Sowell (1990).

As in the short memory stochastic volatility model, we hope samples of previous states can be used in sampling of a new state. Analogous to equation (27), we have the following equation:

$$
\begin{aligned}
p(h_{1:j}|r_{1:j}) &= \frac{p(h_{1:j}, r_{1:j})}{p(r_{1:j})} = \frac{p(r_j|h_{1:j}, r_{1:j-1}) p(h_{1:j}, r_{1:j-1})}{p(r_{1:j})} \\
&= \frac{p(r_j|h_j) p(h_{1:j}, r_{1:j-1})}{p(r_{1:j})} = p(r_j|h_j) p(h_{1:j}|r_{1:j-1}) \frac{p(r_{1:j-1})}{p(r_{1:j})} \\
&\propto p(r_j|h_j) p(h_{1:j}|r_{1:j-1}). \qquad (29)
\end{aligned}
$$

Therefore, if we have obtained samples $h_{1:j}^{(i)} \sim p(h_{1:j}|r_{1:j-1}), i = 1, ..., N$, samples from $p(h_{1:j}|r_{1:j})$, $\hat{h}_{1:j}^{(i)}$, can be drawn by importance sampling with weights $w_i = p(r_j|h_j^{(i)})$. We have shown previously $h_{j+1}^{(i)}$ can be drawn from $p(h_{j+1}|h_{1:j} = \hat{h}_{1:j}^{(i)})$. Adding $h_{j+1}^{(i)}$ to $\hat{h}_{1:j}^{(i)}$ forms new samples $h_{1:j+1}^{(i)}$.

Now we have a sequential algorithm to move from $h_{1:j}^{(i)}$ to $h_{1:j+1}^{(i)}$ for each $j$, and we sum

it up as follows:

$$
\begin{aligned}
h_{1:j}^{(i)} \sim p(h_{1:j}|r_{1:j-1}) &\Longrightarrow p(r_j|r_{1:j-1}) \approx \frac{1}{N}\sum_{i=1}^{N} p(r_j|h_j^{(i)}) \\
&\Downarrow \text{resample } h_{1:j}^{(i)} \text{ with importance weights } p(r_j|h_j^{(i)}) \\
\hat{h}_{1:j}^{(i)} &\sim p(h_{1:j}|r_{1:j}) \\
&\Downarrow h_{1:j}^{(i)} = \hat{h}_{1:j}^{(i)},\ h_{j+1}^{(i)} \sim p(h_{j+1}|h_{1:j} = \hat{h}_{1:j}^{(i)}) \\
h_{1:j+1}^{(i)} \sim p(h_{1:j+1}|r_{1:j}) &\Longrightarrow p(r_{j+1}|r_{1:j}) \approx \frac{1}{N}\sum_{i=1}^{N} p(r_{j+1}|h_{j+1}^{(i)}).
\end{aligned}
$$

### 3.1.3 Estimation Method for LMSV Models

The idea behind our approach is that we augment the state variable (log-volatility in the LMSV model) to include the parameters in the model. As in the method above, the SMC algorithm simulates from the filtering densities of log-volatilities, now it will simulate from the filtering densities of the augmented state variables. If we give a prior distribution to the parameters at the beginning, we will obtain a posterior distribution at the end of the SMC algorithm. We then take the model or mean of this posterior distribution as the estimates for the model.

A main difficulty of this approach is that while there is an equation that governs the evolution of the log-volatility, there is no such equation for the parameters — which are simply fixed. Hence, researchers developed methods to artificially evolute the parameters. However, we have to choose the "dynamics" for the parameters appropriately so that the information contained in their filtering densities won't lose after the evolution. In this research, we are going to adapt the method in Kitagawa and Sato (2001) and Liu and West (2001) for use on the LMSV model.

### 3.1.4 Pseudo-likelihood-based Estimation

Besides the method proposed above, there exist other estimation methods. Jacquier et al. (1994) developed a Bayesian method based on Markov Chain Monte Carlo techniques for the ARSV model. This method depends on the Markovian structure of the AR process, which is however lacking in an ARFIMA process. Another method is derived from the equation set (17). There, $\xi_t = \beta + \log(\epsilon_t^2)$ is not normally distributed. Nonetheless, if we assume a Gaussian distribution for $\xi_t$, equation set ( 17) represents a Gaussian state-space model, and the maximum likelihood estimator is relatively easy to obtain. Harvey et al. (1994) discussed the method for short-memory SV models. It can also be extended to LMSV models, but there are convergence and computing problems (Breidt et al., 1998).

A more feasible method is to use the Whittle's estimator, which maximizes the frequency-domain Gaussian-likelihood for the log squared returns $\{\log r_t^2\}$. For a Gaussian time

series, the Whittle's estimator is asymptotically equivalent to the traditional maximum likelihood estimator (MLE), its time-domain counterpart. In a long-memory context, the advantage of the Whittle's method is its much faster speed and better numerical stability (Fox and Taqqu, 1986; Dahlhaus, 1989). Although $\{\log r_t^2\}$ is not Gaussian, Breidt, Crato, and de Lima (1998) have verified the consistency of the estimator. Their simulation study shows that the Whittle's estimator is a relatively effective and easy-to-implement method for LMSV models.

The Whittle's estimator is achieved by maximizing the spectral likelihood function:

$$L_n(\theta) = \frac{2\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \left\{ \log f(\lambda_j; \theta) + \frac{I(\lambda_j)}{f(\lambda_j; \theta)} \right\}, \tag{30}$$

where

$$\lambda_j = \frac{2\pi j}{n}$$

is the $j$th Fourier frequency, and

$$f(\lambda_j) = \left|1 - e^{i\lambda_j}\right|^{-2d} \frac{\left|\theta(e^{i\lambda_j})\right|^2}{\left|\phi(e^{i\lambda_j})\right|^2} \frac{\sigma^2}{2\pi} + \frac{\sigma_\xi^2}{2\pi}$$

$$I(\lambda_j) = \frac{1}{2\pi n} \left( \sum_{i=1}^{n} r_t \cos \lambda_j i \right)^2 + \frac{1}{2\pi n} \left( \sum_{i=1}^{n} r_t \sin \lambda_j i \right)^2.$$

$f(\lambda_j)$ is the spectral density of $\log(r_t^2)$ at $\lambda_j$. It is actually the sum of the spectral density of an ARFIMA process and a constant because the log squared returns are the sum of the log volatility and an independent series. $I(\lambda_j)$ is the $j$-th periodogram ordinate.

## 3.2 Parallelization of The Sequential Monte Carlo Algorithm

We have shown in earlier sections how we use the sequential Monte Carlo method to obtain the posterior distribution of the latent volatility given the observed returns in order to calculate the likelihood of the model. Actually, an accurate evaluation of likelihood is very important. This is because many of the proposed models in the finance literature perform well, and they all have log-likelihoods which are close together. Thus the Monte Carlo error in SMC evaluation of the likelihood can be critically important in model selection. Standard results tell us that SMC-based likelihood estimators converge to the true likelihoods as the number of particles $M$ approaches infinity. However, this obviously increases computational burden. Therefore we are interested in finding ways to "parallelize" sequential Monte Carlo schemes.

Sequential Monte Carlo methods can be applied not only to stochastic volatility models but also to the extremely rich class of generalized state-space models. These models consist of two components, a latent process, and an observed process. The latent process

is assumed to be Markovian, while elements of the observation process are conditionally independent given the latent process. Formally, we have

$$\begin{aligned} X_t &\sim f(\cdot; X_{t-1}, \theta) \\ Y_t &\sim g(\cdot; X_t, \theta), \end{aligned}$$

where $\{X_t\}$ is the latent process, $\{Y_t\}$ is the observable process, and $f(\cdot;\cdot)$ and $g(\cdot;\cdot)$ are some (arbitrary) conditional densities in their first arguments, which may depend on a parameter vector $\theta$. In particular, using the SMC approach, one can estimate $\theta$ by maximizing the likelihood. Computation of the likelihood requires evaluating of the filtering densities, i.e. $p(x_t|y_1, \ldots, y_t)$ for each $t$, along with the one-step predictive densities $p(x_{t+1}|y_1, \ldots, y_t)$. We can also handle the case where $\{X_t\}$ is non-Markovian by modifying the method appropriately just like what we do on the long-memory stochastic volatility model .

The SMC algorithm works by representing each $\pi_t = p(x_t|y_1, \ldots, y_t)$ by an approximating collection of "particles",

$$x_t^{(i)} \sim p(x_t|y_1, \ldots, y_t), \quad i = 1, \ldots, M.$$

Here $M$ is the number of particles used in the representation. When the particles are indeed realizations of the distribution $\pi_t$, properties of $\pi_t$ can be derived from the sample $\{x_t^{(i)},\ i = 1, \ldots, M\}$. The basic form of the SMC algorithm works by recursively obtaining particle collections using the following steps.

1. **Simulation.** Draw $\tilde{x}_{t+1}^{(i)}$ from $f(\cdot; x_t^{(i)}, \theta)$.
2. **Importance Resampling.** Assign weights $w_{t+1}^{(i)} = g(y_{t+1}|\tilde{x}_{t+1}^{(i)}, \theta)$. Then construct a sample of size $M$, which will become $\{x_{t+1}^{(i)},\ i = 1, \ldots, M\}$ by sampling with replacement from $\{\tilde{x}_{t+1}^{(i)},\ i = 1, \ldots, M\}$, with probabilities proportional to $\{w_{t+1}^{(i)}\}$.

When we have observations $y_1, \ldots, y_T$, Steps 1 and 2 are repeated for $t = 1, 2, 3, \ldots, T$ to obtain approximations to the required filtering and predictive distributions. Because of its non-parametric representation of the required distributions, the SMC algorithm is not bound by typical constraints placed on model structure by previously existing methods such as the Kalman filter.

To use parallel computing on SMC, we need to design a suitable parallel algorithm. Suppose we have $Q$ processors available. A seemingly obvious algorithm is to let each processor independently sample $M/Q$ number of particles from time 1 to the final time $T$. However, due to the error introduced in the resampling step (Step 2 above) of the SMC algorithm, this method is worse than using one processor to produce $M$ particles. (In this paper, we demonstrate this in empirical studies.)

We have proposed a different approach to utilize parallel computing. Suppose we still have $Q$ processors and need to generate $M$ particles. We start with our collection of $M$ particles distributed across all $Q$ processors (as close to evenly as possible).

At each time $t$, each processor carries out the simulation step for each of the particles it is storing. (This is the same as in the obvious approach.) But we introduce a procedure which gets around the problem of the obvious approach by genuinely resampling from the entire collective group of $M$ particles spread across $Q$ processor. The idea is to perform weighting separately on each processor, then communicate sums of weights from each processor to a central "master" process. At this point, the master process can draw from a multinomial distribution to decide how many samples will be drawn from each processor. Then processors can perform individual resampling.

While the proposed procedure is functionally equivalent to an $M$-particle SMC algorithm, it can become inefficient as the number of particles stored locally on a processor varies significantly from the "balanced" number $M/Q$. To get around this problem, we introduce a further "rebalancing" step, where particles with high weights are redistributed across different processors. The rebalancing needs to be done carefully in order to prevent communication overhead from eliminating gains in speed from parallelization.

Figure 2 is preliminary results on efficiency of the proposed algorithm and comparison of it in terms of speed to single-processor and the aforementioned parallelized method without rebalancing, by applying the algorithms to analysis of the following linear Gaussian model:

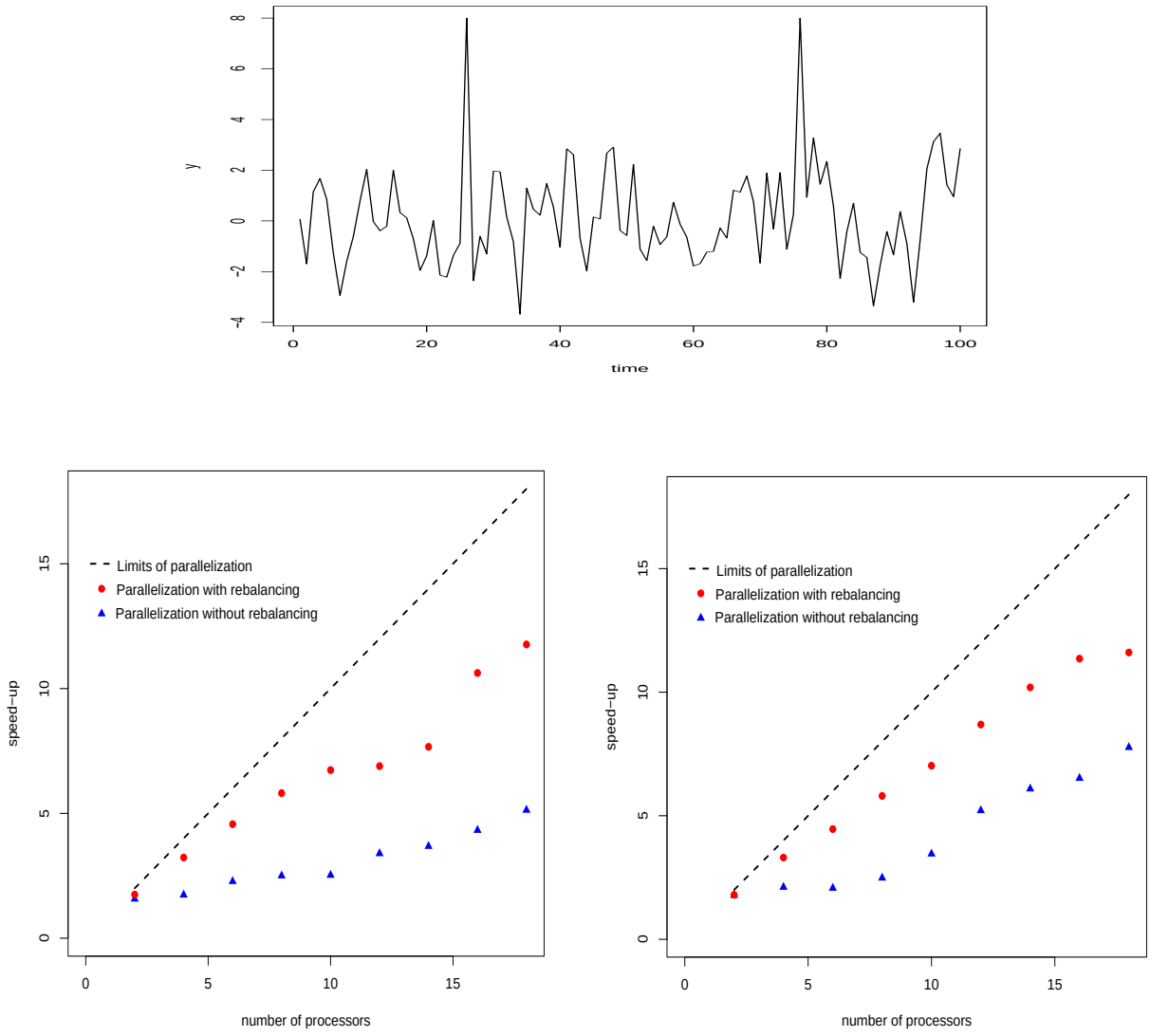
$$\begin{aligned} Y_t &= X_t + Z_t, \quad Z_t \sim \mathrm{N}(0,1) \\ X_t &= 0.5X_{t-1} + E_t, \quad E_t \sim \mathrm{N}(0,1) \end{aligned}$$

## 3.3 Extension to Multivariate LMSV Models

We can modify the (short-memory) multivariate SV models to make them be able to model long-memory volatility. For example, we can assume that $h_t$ in the one-factor model in equation (13) follows an ARFIMA process. That is,

$$\begin{aligned} \mathbf{r}_t &= \mathbf{u} + \exp(h_t)\mathbf{e}_t, \\ \mathbf{e}_t &\sim \mathrm{IIDN}(\mathbf{0}, \mathbf{\Sigma}_e), \\ h_t - u_h &\sim \mathrm{ARFIMA}(p,d,q). \end{aligned} \tag{31}$$

In this model, although asset returns are correlated, there is only one (long-memory) process that controls the magnitude of the variances and covariances while the conditional correlation is constant. We expect that the likelihood-based estimation method for the




$5 \times 10^5$ particles $10^6$ particles

Figure 2: Simulated observations and speed-up of the parallelization scheme of the sequential Monte Carlo method with different number of simulations

univariate model also applies to this model with modifications:

$$h_{1:j}^{(i)} \sim p(h_{1:j}|\mathbf{r}_{1:j-1}) \Longrightarrow p(\mathbf{r}_j|\mathbf{r}_{1:j-1}) \approx \frac{1}{N}\sum_{i=1}^{N} p(\mathbf{r}_j|h_j^{(i)})$$

$$\Downarrow \text{ resample } h_{1:j}^{(i)} \text{ with importance weights } p(\mathbf{r}_j|h_j^{(i)})$$

$$\hat{h}_{1:j}^{(i)} \sim p(h_{1:j}|\mathbf{r}_{1:j})$$

$$\Downarrow h_{1:j}^{(i)} = \hat{h}_{1:j}^{(i)},\ h_{j+1}^{(i)} \sim p(h_{j+1}|h_{1:j} = \hat{h}_{1:j}^{(i)})$$

$$h_{1:j+1}^{(i)} \sim p(h_{1:j+1}|\mathbf{r}_{1:j}) \Longrightarrow p(\mathbf{r}_{j+1}|\mathbf{r}_{1:j}) \approx \frac{1}{N}\sum_{i=1}^{N} p(\mathbf{r}_{j+1}|h_{j+1}^{(i)}).$$

In the multivariate model, $p(\mathbf{r}_j|h_j) = (2\pi|\mathbf{V}_t|)^{-k/2}e^{-(\mathbf{r}_t-\mathbf{u})'\mathbf{V}_t^{-1}(\mathbf{r}_t-\mathbf{u})/2}$, where $\mathbf{V}_t = \exp(2h_t)\mathbf{\Sigma}_e$.

## 3.4 Option Pricing under The LMSV Model

As shown in previous parts, we can model volatility persistence in the discrete-time stochastic volatility framework with the use of ARFIMA processes for modeling log-volatility. We are also interested in continuous-time LMSV models, since continuous-time models are often useful in option pricing. Comte and Renault (1998) presented a continuous-time model, which is analogous to the discrete-time LMSV model with an ARFIMA$(1, d, 0)$ $(0 < d < 1/2)$ log-volatility process,

$$\begin{aligned} dS(t) &= \mu S(t) + \sigma(t)S(t)dW_1(t) \\ v(t) &= \ln\sigma(t) \\ d(v(t)) &= -k(\Theta - v(t)) + \beta dW_{2,d}(t). \end{aligned} \tag{32}$$

The log-volatility in this model satisfies an Ornstein-Ulenbech type of stochastic differential equation, which leads to the AR order of 1. $W_{2,d}(t)$ is a fractional Brownian motion, which is defined as an integration with respect to a standard Brownian motion

$$W_d(t) = \int_0^t \frac{(t-s)^d}{\Gamma(d+1)} dW(s).$$

Comte and Renault (1996) and Tsai and Chan (2005) also discussed the more general continuous-time long-memory model.

To use the LMSV model in option pricing, we have to handle the issue of non-uniqueness of risk-neutral measure. As in short-memory models, one can either choose a specific parametric form for the market price of volatility risk and estimate the parameters from historical data, or just assume it is zero for the sake of simplicity. The latter strategy was assumed in Comte and Renault (1998) and in papers cited in it. Due to the much more complicated structure of the ARFIMA process in the discrete-time LMSV model, we also choose a zero market price of volatility risk in this paper. Therefore, we assume that the volatility process is the same under both the objective measure and the risk-neutral measure.

Using the risk-neutral pricing principle, the price at the current time (time 0) of a European option with maturity $T$ and payoff function $g(S_T)$ is

$$V(0) = \widetilde{E}_{RN}\left[e^{-rT}g(S_T)|\mathcal{F}(0)\right],$$

where RN means the expectation is conducted under the risk-neutral measure, and $\sigma$-field $\mathcal{F}(0)$ contains all the price information up to and including time 0. To evaluate this expectation, we have to work with the model under the risk-neutral measure, under which the stock earns a rate of return equal to the risk-free interest $r$,

$$\begin{aligned} dS(t) &= rS(t)dt + \sigma(t)S(t)d\widetilde{W}(t) \\ \text{or } S(t) &= S(0)\exp\left(\left[r - \frac{1}{2}\sigma(t)^2\right]t + \sigma(t)\widetilde{W}(t)\right). \end{aligned}$$

Discretization of $S(t)$ yields

$$r_t = (r - \sigma_t^2/2) + \sigma_t \epsilon_t,$$

where $r_t = \log(S(t)/S(t-1))$ is the log-return on the $t$-th day and $\epsilon_t$ is Gaussian random noise. Along with the log-volatility following an ARFIMA process, this leads us to the model in the set of equations (16), except that the rate of return $\mu$ is replaced by $r$.

To price European options, our method is as follows. We first use historical returns, $\{r_0, \ldots, r_{-(T'-1)}\}$ (suppose that we use the historical daily prices observed in the past $T'+1$ days including the current day) to estimate the parameters in the LMSV model with the method proposed in this paper. By our assumption, the volatility process remains the same under both the objective measure and the risk-neutral measure, while the rate of return of the stock changes to the interest rate $r$. The risk-neutral pricing method requires simulations of the stock price at $T$, which in turn requires simulations of stock returns at all future times ($S_T = S_0 e^{r_1 + \cdots + r_T}$). Since $r_t \,|\, \mathcal{F}_{t-1} \sim \mathrm{N}(r - \sigma_t^2/2, \sigma_t^2)$, we need to simulate future volatilities as well.

In our inference method, we have already obtained sample trajectories of historical log-volatilities via the SMC algorithm, and the empirical distribution of them forms an approximation to $p(\log \sigma_0, \ldots, \log \sigma_{-(T'-1)} \,|\, r_0, \ldots, r_{-(T'-1)})$. We then simulate $\log \sigma_t$, for $t > 0$, from the conditional distribution $p(\log \sigma_t \,|\, \log \sigma_{t-1}, \ldots, \log \sigma_0, \ldots, \log \sigma_{-(T'-1)})$, conditioning on the simulated trajectory, including newly generated samples of future log-volatilities as well. The conditional distributions can be determined using properties of multivariate Gaussian distributions, since the ARFIMA process is a Gaussian process. We can show the sample trajectories into the future, $\{\log {\sigma_1}^{(j)}, \ldots, \log {\sigma_T}^{(j)}\}$, are actually samples from $p(\log \sigma_1, \ldots, \log \sigma_T \,|\, r_0, \ldots, r_{-(T'-1)})$. The inclusion of all historical prices is necessary since we are dealing with a non-Markovian ARFIMA process. For each log-volatility $\log \sigma_t$ ($t \geq 0$), we simulate the return $r_t \sim \mathrm{N}(r - \sigma_t^2/2, \sigma_t^2)$. At the end, Many samples of $S_T = S_0 e^{r_1 + \cdots + r_T}$ are collected, and for each sample, $e^{-rT} g(S_T)$ is calculated. Taking the average of these simulated discounted payoffs, we obtain a Monte Carlo simulation for the option price $V(0)$.

Pricing of American options are usually more difficult because of the early-exercise possibility. The price of an American option depends on the path of the stock price (path-dependent). A successful pricing method should contain an algorithm that can determine the optimum time for exercise. At any point of time before expiration, we should compare the value of the option at immediate exercise with its expected discounted future payoff under the risk-neutral measure if it continues to be hold. If immediate exercise is more profitable, we should exercise the option; otherwise, we should continue to hold it. Since the value of the option at a certain time depends on its value in the future, we usually take a backward induction strategy. We first calculate the value of option (payoff) at maturity, which of course just depends on the stock price at maturity, then we go one step backward, do that comparison and repeat this procedure until the current time. This algorithm works well for simulation methods, including the commonly used tree or lattice methods.

There is also a large body of research on pricing American options under SV models. Among them, Rambharat and Brockwell (2006) developed an algorithm based on the

SMC algorithm. However, to my knowledge, the current simulation methods for pricing American options are all developed for short-memory (Markovian) volatility processes. In principle, we can extend some methods, for example, the method in Rambharat and Brockwell (2006) to the long-memory situation. However, the inclusion of the whole history at each time step, due to the non-Markovian property of long-memory processes, will enormously increase the computing burden and make implementation of these methods almost impossible in practice. Therefore, in this research, we will focus on pricing European options.

## 3.5 Model Comparison

Besides the SV model the current research is focused on, there are other models of stochastic volatility. In particular, the GARCH model and its variants have proven useful in practical use. A GARCH model (Engle, 1982; Bollerslev, 1986) assumes that the conditional variance (squared volatility) of a stock is a linear combination of the historical conditional variances and squared returns:

$$\begin{aligned} r_t &= \mu + \sigma_t \epsilon_t, \quad \epsilon_t \sim \text{IIDN}(0,1) \\ \sigma_t^2 &= \alpha_0 + \alpha_1 r_{t-1}^2 + \ldots + \alpha_p r_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \ldots + \beta_q \sigma_{t-q}^2 \\ \alpha_0 &> 0, \ \alpha_i \geq 0, \ \beta_i \geq 0, \ \text{for } i > 0 \end{aligned}$$

The equations above represent a GARCH$(p, q)$ model.

Baillie et al. (1996) extended the GARCH model to FIGARCH$(p, d, q)$ model in order to model long memory in volatility:

$$\phi(B)(1-B)^d r_t^2 = \omega + [1 - \beta(B)](r_t^2 - \sigma_t^2),$$

where $\beta(B) = \beta_1 B + \ldots + \beta_q B^q$, $\alpha(B) = \alpha_1 B + \ldots + \alpha_p B^p$ and $\phi(B) = [1 - \alpha(B) - \beta(B)](1-B)^{-1}$.

We will compare the LMSV model with the (short-memory) SV model, GARCH and FIGARCH model based on the likelihood, AICC and BIC.

# References


O. Aguilar and M. West. Bayesian dynamic factor models and portfolio allocation. *Journal of Business and Economic Statistics*, 18:338–357, 2000.

T. G. Andersen and T. Bollerslev. Heterogeneous information arrivals and return volatilty dynamics: uncovering the long-run in high frequency returns. *Journal of Finance*, 52: 975–1005, 1997.

R. T. Baillie, T. Bollerslev, and H. O. Mikkelsen. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74:3–30, 1996.

J. Beran. *Statistics for Long-Memory Processes*. Chapman & Hall/CRC, 1994.

T. Bollerslev. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31:307–327, 1986.

T. Bollerslev and H. O. A. Mikkelsen. Modeling and pricing long-memory in stock market volatility. *Journal of Econometrics*, 73:151–184, 1996.

F. J. Breidt, N. Crato, and P. de Lima. Detection and estimation of long memory in stochastic volatility. *Journal of Econometrics*, 83:325–348, 1998.

P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer, 1991.

S. Chib, F. Nardari, and N. Shephard. Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics (in press, available online)*, 2005.

F. Comte and E. Renault. Long memory continuous time models. *Journal of Econometrics*, 73:101–149, 1996.

F. Comte and E. Renault. Long memory in continuous time stochastic volatility models. *Mathematical Finance*, 8:291–323, 1998.

R. Dahlhaus. Efficient parameter estimation for self-similar processes. *Annals of Statistics*, 17:1749–66, 1989.

R. S. Deo and C. M. Hurvich. Estimation of long memory in volatility. In *Theory and Applications of Long-Range Dependence*, pages 313–324. Birkhauser, 2003.

Z. Ding, C. W. J. Granger, and R. F. Engle. A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1:83–106, 1993.

J. A. Doornik and M. Ooms. Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models. *Computational Statistics and Data Analysis*, 42:333–348, 1993.

P. Doukhan, G. Oppenheim, and M. S. Taqqu, editors. *Theory and Applications of Long-Range Dependence*. Birkhauser, 2003.

R. F. Engle. Autoregressive conditional heteroscedasticity with esitmates of the variance of uk inflation. *Econometrica*, 50:987–1007, 1982.

R. Fox and M. S. Taqqu. Large-sample properties of parameter estimates for strongly dependent stationary gaussian time series. *The Annals of Statistics*, 14:517–532, 1986.

J. Geweke and S. Porter-Hudak. The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4:221–238, 1983.

A. C. Harvey. Long memory in stochastic volatility. In *Forecasting Volatility in The Financial Markets*, pages 307–320. Oxford ; Boston : Butterworth Heinemann, 1998.

A. C. Harvey, E. Ruiz, and N. Shephard. Multivariate stochastic variance models. *Review of Economic Studies*, 61:247–264, 1994.

S. L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6:327–343, 1993.

J. Hull and A. White. The pricing of options on assets with stochastic volatilities. *Journal of Finance*, 42:281–300, 1987.

H. Hurst. Long term storage capacity of reservoirs. *Transactions of The American Society of Civil Engineers*, 116:770–799, 1951.

E. Jacquier, N. G. Polson, and P. E. Rossi. Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics*, 12:69–87, 1994.

G. J. Jiang. Stochastic volatility and option pricing. In *Forecasting Volatility in The Financial Markets*, pages 47–96. Oxford ; Boston : Butterworth Heinemann, 1998.

G. Kitagawa and S. Sato. Monte carlo smoothing and self-organising state-space model. In *Sequential Monte Carlo Methods in Practice*, pages 177–196. Springer, 2001.

J. Liu and M. West. Combined parameter and state estimation in simulation-based filtering. In *Sequential Monte Carlo Methods in Practice*, pages 197–217. Springer, 2001.

A. W. Lo. Long-term memory in stock market prices. *Econometrica*, 59:1279–1313, 1991.

B. B. Mandelbrot and J. R. Wallis. N. joseph and operational hydrology. *Water Resources Reseach*, 4:909–918, 1968.

J. M. Quintana and M. West. An analysis of international exchange rates using multivariate dlms. *The Statistician*, 36:589–594, 1987.

B. R. Rambharat and A. Brockwell. Sequential monte carlo pricing of american-style options under stochastic volatility models. *Discussion Paper, ISDS, Duke University*, 2006.

J. Rangarajan and M. Ding, editors. *Processes with Long-Range Correlations: Theory and Applications*. Springer-Verlag, 2003.

P. M. Robinson. Log-periodogram regression of time series with long range dependence. *Annals of Statistics*, 23:1048–1072, 1995.

N. Shephard. Statistical aspects of arch and stochastic volatility. In *Time Series Models in Econometrics, Finance and Other Fields*, pages 1–67. Chapman & Hall, 1996.

Neil Shephard. Local scale models : State space alternative to integrated garch processes. *Journal of Econometrics*, 60:181–202, 1994.

S. E. Shreve. *Stochastic Calculus for Finance II: Continuous-Time Models.* Springer, 2004.

F. B. Sowell. Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, 53:165–188, 1990.

S. J. Taylor. *Asset Price Dynamics, Volatility, and Prediction.* Princeton University Press, 2005.

H. Tsai and K. S. Chan. Maximum likelihood estimation of linear continuous time long memory processes with discrete time data. *Journal of the Royal Statistical Society, Series B*, 67:703–716, 2005.