ABSTRACT

Title of dissertation: Discrete Time Stochastic Volatility Model

Guojing Tang, Doctor of Philosophy, 2009

Dissertation directed by: Professor Dilip B. Madan
Professor Benjamin Kedem

In this dissertation we propose a new model which captures observed features of asset prices. The model reproduces the skewness and fat tails of asset returns by introducing a discretized variance gamma process as the driving innovation process, in addition to a double gamma process to reflect the stochastic nature of volatility coefficients. The leverage effect between returns and volatilities is built in by a polynomial function describing the relationship between these two variables. One application of this model is to price volatility contracts whose payoffs depend on realized variance or volatility. Because of the scarcity of market quotes and consequent unavailability of risk neutral calibration, we propose a new scheme of pricing based on the model estimated from historical data. The estimation of the model parameters is carried out by maximizing likelihood function, which is calculated through a combination of Expectation-Maximization and Particle Filter algorithm. The resulting distribution is transformed by concave distortions, the extent of which reflects the risk aversion level of market.

Discrete Time Stochastic Volatility Model

by

Guojing Tang

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2009

Advisory Committee:
Professor Benjamin Kedem, Chair
Professor Dilip B. Madan, Co-chair
Professor Paul Smith
Professor Victor Yakovenko
Professor Abram Kagan

Acknowledgments

This dissertation would not have been possible without the support of many people. First I want to thank Professor Dilip Madan and Professor Benjamin Kedem. Professor Madan helped me pin down various points made in this work and shared with me his suggestions and attentiveness to my concerns. He has been vital in my developing a better sense of the steps necessary to do truly effective research in mathematical finance. I want to thank Professor Benjamin Kedem for giving me the great opportunity to study at Statistics Program in Mathematics Department. Through all stages of my PhD study and during numerous discussions and chalkboard sessions, he unreservedly shared with me the knowledge and enlightening insights which greatly evoked my interested in different areas of probability and statistics.

I am deeply indebted to Professor Smith for so many reasons. His careful reading of the earlier version of this dissertation yielded detailed comments, valuable suggestions and guidelines that led to substantial improvement. And more importantly, he taught me how to carry out rigorous research through an unyielded pursuit of preciseness and perfection, which has become one of the most valuable assets I earned from this journey and will continue inspire my career in the future.

Also I would like to send a special thanks to my committe members, Professor Adam Kagan and Professor Victor Yakovenko, for their kindly help and encouragement and for graciously agreeing to join my committee.

No words can express enough my gratitude towards my wife, Peng Lin, for her thoughtfulness, tenderness and love through years. And finally I would offer

my profound thanks to all my friends and colleagues who provided good company, encouragement, perspective or combinations of the above.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Stochastic Volatility

The modern asset pricing theories are dominated by the idea that higher risk should be compensated by higher return of investments. Volatility, which essentially measures the level of fluctuation in the underlying asset's return, serves as an important thermometer of risk. Stochastic volatility models deal with the time-varying volatility in financial markets and enable us to capture, for example, empirically observed departures from the Black-Scholes-Merton model for financial options and could be applied in a variety of areas such as pricing derivatives which are written on the volatility and related quantities.

The first discrete time stochastic volatility model was proposed by Taylor (1982) where the assets returns are modeled as:

$$r_i = \mu + \sigma_i \epsilon_i, \tag{1.1}$$

where r_i, $i = 1, ..., n$ are the daily returns, μ is the mean return of r_i and ϵ_i are i.i.d random variables with zero mean and unit variance. The volatility coefficients σ_i

follow a log-autoregressive process independent of ϵ_i. More specifically,

$$\begin{aligned} \sigma_i &= \exp(\zeta_i/2), \\ \zeta_{i+1} &= \mu + \phi(\zeta_i - \mu) + \eta_i, \\ \eta_i &\sim NID(0, \sigma_\eta^2). \end{aligned} \tag{1.2}$$

So this model contains a two-dimensional Markovian process and volatility is described as a latent stochastic process. It is different from the famous Autoregressive Conditional Heteroskedasticity Model (ARCH) of Engle (1978) and Generalized Autoregressive Conditional Heteroskedasticity model (GARCH) of Bollerslev (1986). Although this stochastic volatility has a simple structure, its realistic and innovative description of the volatility process has raised much academic and industrial interest and inspired a sizable amount of literature studying its econometric properties and other relevant issues. Among those are Jacquier et al. (1993), Taylor (1986), Taylor (1994), Shephard (1996) and Chernov and Ghysels (1999).

A well-documented fact about stock returns and implied/realized volatility is that volatility tends to increase when the stock price falls. A standard explanation ties this phenomenon, which is also called Leverage Effect, to the effect that a change in market valuation of a firm's equity has on the degree of leverage[1] in its capital structure, with an increase in leverage producing an increase in stock volatility. Leverage effect was studied by Black (1976) and motivated the introduction of the exponential-GARCH model of Nelson, the threshold ARCH model of Glosten et al. (1993) and Quadratic GARCH model of Sentana (1995). The economic theory of

[1]Leverage is defined as the ratio of a firm's debt to equity in its balance sheet.

these effects is discussed by Campbell and Kyle (1993).

Research on continuous time stochastic volatility can be tracked back to Clark (1973), which introduced the concept of stochastic volatility to financial economics. Johnson (1979) applied a continuous time stochastic volatility model to study the pricing of options. Heston (1993) introduced stochastic volatility model to describe returns as diffusions with stochastic volatility described by an Ornstein-Uhlenbeck process. The model accommodates a stationary and mean-reverting volatility dynamic and also it incorporates the leverage effect through the negative correlation between the return and volatility which can be expressed in the following stochastic differential equations:

$$\begin{aligned} dS_t &= \mu S dt + \sqrt{\sigma_t} S dz_{1,t}, \\ d\sigma_t &= \beta(\nu - \sigma_t)dt + \delta\sqrt{\sigma_t} dz_{2,t}, \end{aligned} \tag{1.3}$$

where S_t is the spot asset price and $z_{1,t}$ and $z_{2,t}$ are two correlated Wiener processes with instantaneous covariance $\rho dt =< dz_1(t), dz_2(t) >$.

It has been widely agreed that models built only upon continuous process cannot fully explain the extraordinary movements of asset prices and therefore it is critical to incorporate jumps into the model. That led to the development of a series of jump diffusion models. Among them, Bates (1996) added a Poisson process in addition to a diffusion process of returns. Barndorff-Nielsen and Shephard (1999) extended it by replacing the Poisson jumps with a more generic Lévy process and assumed the volatility is completely jump-driven. The jump-diffusion model

describes the return process $r(t)$ as follows,

$$dr_t = \{\mu + \beta\sigma_t^2\}dt + \sigma_t d\omega_t + \rho d\bar{z}(\lambda t),$$
$$d(\sigma_t^2) = -\lambda\sigma_t^2 dt + d\grave{z}(\lambda t), \tag{1.4}$$

where ω is the Wiener process, $\bar{z}_t = \grave{z}_t - E(\grave{z}_t)$ is the compensated Lévy process corresponding to $\grave{z}$, λ is the event rate of the Poisson process and σ_t^2 follows an Ornstein-Uhlenbeck process defined also in term of $\grave{z}$ by the stochastic differential equation (1.4).

Geman et al. (2001) suggest that price processes for financial assets must have a jump component but they need not have a diffusion component. The use of jump processes with an infinite arrival rate, such as Madan and Miline (1991) and the variance gamma process of Madan et al. (1998), can encompass the contribution of any diffusion component. Their argument is based on regarding all pricing processes as Brownian motion subordinated to a random clock and regarding the clock as a cumulative measure of economic activity, as conjectured by Clark (1973). The randomness of the clock process induces stochastic volatility and because time is increasing, the random clock can be modeled as a pure jump increasing process, as an efficient alternative to a time integral of a positive diffusion process. Therefore the diffusion part is just redundant.

Carr et al. (2003) continued to propose a more generic class of stochastic volatility Lévy processes based on three homogeneous Lévy processes (NIG, VG and CGMY models). Stochastic volatility is generated by subordinating a Lévy process to the time integral of an Ornstein-Uhlenbeck process, for example, the CIR

process of Cox et al. (1985).

1.2 Discrete time stochastic volatility model with local skewness and kurtosis

The empirical success of stochastic volatility models is not maintained when one considers pricing volatility products. The value of options written on realized variance is determined by the cumulative variance of returns and thus by higher moments of the return distribution. Most stochastic volatility models show decreasing kurtosis as term increases. For example, for a stochastic volatility Lévy process, skewness is inversely proportional to the square root of term and kurtosis is inversely proportional to the term. This implies that the value of the option on realized variance also decreases with the same speed. In contrast, the real prices observed in the market often suggest different or even opposite patterns.

This dissertation proposes a solution to this problem by turning to a discrete time model and using a new generic distribution to describe the local innovation process. In the new model the returns are driven by the product of a non-Gaussian innovation and a stochastic volatility coefficient. Therefore besides stochastic volatility, the new model introduces another source of non-Gaussian kurtosis and skewness and the overall effect is that as term increases the kurtosis accumulates rather than diminishes.

1.3 Estimation

As the model introduced in this dissertation belongs to the wide class of non-linear state-space models, classical parameter estimation is difficult due to lacking the tractable form of the likelihood. To solve the estimation problem we apply an EM algorithm of Dempster et al. (1977), enhanced by the Markov chain Monte Carlo (MCMC) and particle filter techniques for efficient computations of posterior distribution.

This dissertation also explores another modeling issue in reducing the computational cost of estimation. Unlike continuous time models, discrete time models give us the freedom to have different frequencies for the changing of return and volatility. A lower frequency of volatility change will significantly reduce the dimensions of numerical integration required for calculation of the likelihood function and therefore improve the speed of estimation. To explore the optimal frequency we can compare the fitting performance of models with different frequencies of volatility change.

1.4 Organization of this dissertation

The dissertation is organized as follows. Chapter 2 introduces the new stochastic volatility model that incorporates major features intended for financial applications. Chapter 3 describes the method of estimating in our model using the Expectation-Maximization algorithm and the Particle Filter sampling technique. We explore the issue of frequency of volatility change and study methods of detect-

ing optimal frequency using two different types of information criteria for model selection. Chapter 4 applies the new model in pricing volatility products. This is achieved by combining the distribution of volatility projected from estimated models with concave distortions based on coherent risk measure theory. Chapter 5 explores an application of the model in portfolio optimization. Chapter 6 summarizes the research, makes conclusions, and discusses possible future work.

Chapter 2

Model Structure

In this chapter we address a number of issues in building a stochastic volatility model. First, the volatility coefficient in the return process is represented by a mean-reverting and non-Gaussian process. Secondly, to bring non-zero skewness and kurtosis to the conditional distribution of returns, we propose a new distribution which is based on adapting a continuous Levy process to unit time. Thirdly, the negative correlation between price and volatility is built into the volatility process through a deterministic function of asset returns. Finally, an issue in the research on stochastic volatility models, which is about the discrepancy in the frequencies of volatility and return, is explored and we also discuss its significance in improving the computational efficiency.

The return of an underlying asset is defined as the change of the logarithm of asset price:

$$r_t = \log S_{t+1} - \log S_t, \tag{2.1}$$

where S_t is the asset's price at time t, $t = 1, ..., T$.

We assume that r_t follows a process of the form,

$$r_t = \mu h + \sigma_t x_t + g(\sigma_t),$$
$$\sigma_t^2 = \sigma_t'^2 V_t h, \tag{2.2}$$

where μ is the mean rate of the return and h is the time interval between two observations. The process x_t is the independent innovation process with a flexible marginal distribution that has non-Gaussian properties such as non-zero skewness and kurtosis. We will introduce a new distribution which is devised by adapting the variance gamma (VG) process of Madan and Seneta (1990) and Madan and Miline (1991) to unit time. The resulting distribution parsimoniously satisfies the requirements.

The volatility coefficient σ_t consists of two parts: σ_t' is a deterministic function of returns r_t through which the leverage effect is introduced. V_t is a mean-reverting stochastic process that we will introduce more details in the following sections. Overall the product $\sigma_t'^2 V_t h$ is the cumulative variance in one time period.

The function $g(z)$ in the return equation is defined as

$$g(z) = -\log \Phi_{x_t}(z) \tag{2.3}$$

where $\Phi_{x_t}(z) = E(e^{zx_t})$ is the moment generating function of x_t. By adding $g(\sigma_t)$ to the return process, the conditional martingale property of the discounted asset

price $e^{-t\mu h}S_t$ is satisfied:

$$\begin{aligned} E(S_{t+1}|S_t, \sigma_t) &= S_t e^{\mu h} E[e^{\sigma_t x_t}|\sigma_t] \\ &= S_t e^{\mu h} e^{g(\sigma_t)} \Phi_{x_t}(\sigma_t) \\ &= S_t e^{\mu h}. \end{aligned} \tag{2.4}$$

2.1 Double Gamma Stochastic volatility

The volatility component should possess a couple of properties such as non-negativity and mean-reversion. One model that satisfies these requirements was proposed by Barndorff-Nielsen and Shephard (2001), where a non-Gaussian Ornstein-Uhlenbeck process is employed in the volatility processes. However since it was built as a continuous time model and the discretization cannot guarantee the non-negativity, this method cannot be directly applied to discrete time settings.

In our model, the stochastic part of the volatility process is specified as follows:

$$V_{t+1}|V_t \sim Gamma(d, \lambda V_t + \kappa), \tag{2.5}$$

where d and $\lambda V_t + \kappa$ are, respectively, the scale parameter and the shape parameter of the conditional Gamma distribution of V_{t+1}. The conditional density function of V_{t+1} given V_t is:

$$f_{V_{t+1}}(v|V_t) = \frac{d^{\lambda V_t + \kappa}}{\Gamma(\lambda V_t + \kappa)} v^{\lambda V_t + \kappa - 1} e^{-dv}. \tag{2.6}$$

It follows that the conditional expectation and volatility of V_{t+1} given V_t are both affine functions of V_t.

Lemma 1 *Given V_t, the mean, variance, skewness and kurtosis of V_{t+1} are*

$$\begin{aligned} E(V_{t+1}|V_t) &= \frac{\lambda}{d}V_t + \frac{\kappa}{d}, \\ Var(V_{t+1}|V_t) &= \frac{\lambda}{d^2}V_t + \frac{\kappa}{d^2}, \\ Skewness(V_{t+1}|V_t) &= \frac{2}{\sqrt{\lambda V_t + \kappa}}, \\ Kurtosis(V_{t+1}|V_t) &= \frac{6}{\lambda V_t + \kappa}. \end{aligned} \tag{2.7}$$

To separate the controls of the speed of mean-reversion and the overall variance of volatility, we can refine the above structure by replacing κ in the conditional distribution of V_{t+1} by a random variable u_t, which follows another Gamma distribution.

Definition 1 *The Double Gamma Process is a process V_t defined as follows:*

$$\begin{aligned} V_{t+1}|V_t, u_{t+1} &\sim Gamma(d, \lambda V_t + u_{t+1}), \\ u_{t+1} &\sim Gamma(c, \gamma), \end{aligned} \tag{2.8}$$

where u_t are i.i.d random variables that follow a Gamma distribution whose pdf $f(u_t)$ is defined as:

$$f(u_t) = \frac{c^\gamma}{\Gamma(\gamma)} u_t^{\gamma-1} e^{-cu_t}. \tag{2.9}$$

By integrating u_{t+1} out of the conditional distribution of V_{t+1} we have the conditional expectation and variance for the new process.

Lemma 2 *The first and second moments of V_{t+1} given V_t are*

$$E(V_{t+1}|V_t) = \frac{\lambda}{d}V_t + \frac{\gamma}{cd}, \tag{2.10}$$

$$Var(V_{t+1}|V_t) = \frac{\lambda}{d^2}V_t + \frac{\gamma}{cd^2} + \frac{\gamma}{c^2 d^2}. \tag{2.11}$$

The proof is trivial using the properties of the Gamma distribution.

Lemma 3 *For $t \geq 1$, the mean and variance of V_t given V_0 are*

$$E(V_t|V_0) = a^t V_0 + b\frac{1-a^t}{1-a}, \tag{2.12}$$

$$\begin{aligned} Var(V_t|V_0) = &\left(a^{t-1}\frac{1-a^{t-1}}{1-a} + a^{2(t-1)}\right) lV_0 + q\frac{1-a^{2t}}{1-a^2} \\ &+ \frac{bl}{1-a}\left(\frac{1-a^{2(t-1)}}{1-a^2} - \frac{a^{t-1}-a^{2(t-1)}}{1-a}\right), \end{aligned} \tag{2.13}$$

where the coefficients a, b, l and q are defined as

$$a = \frac{\lambda}{d}, \quad b = \frac{\gamma}{cd}, \quad l = \frac{\lambda}{d^2}, \quad q = \frac{\gamma}{cd^2} + \frac{\gamma}{c^2d^2}.$$

Proof: for $t = 2, 3$, we have:

$$\begin{aligned} Var(V_2|V_0) &= E(Var(V_2|V_1)|V_0) + Var(E(V_2|V_1)|V_0) \\ &= E(lV_1 + q|V_0) + Var(aV_1 + b|V_0) \\ &= lE(V_1|V_0) + a^2 Var(V_1|V_0) + q \\ Var(V_3|V_0) &= E(Var(V_3|V_2)|V_0) + Var(E(V_3|V_2)|V_0) \\ &= E(lV_2 + q|V_0) + Var(aV_2 + b|V_0) \\ &= lE(V_2|V_0) + a^2 Var(V_2|V_0) + q \\ &= lE(V_2|V_0) + a^2(lE(V_1|V_0) + a^2 Var(V_1|V_0) + q) + q. \end{aligned} \tag{2.14}$$

Recursively we can generalize $Var(V_t|V_0)$ for $t \geq 1$

$$
\begin{aligned}
& Var(V_t|V_0) \\
=& l\sum_{i=1}^{t-1} a^{2(t-1-i)} E(V_i|V_0) + a^{2(t-1)} Var(V_1|V_0) + q\sum_{i=1}^{t-1} a^{2(i-1)} \\
=& l\sum_{i=1}^{t-1} a^{2(t-1-i)} \left(a^i V_0 + b\frac{1-a^i}{1-a} \right) + a^{2(t-1)}(lV_0 + q) + q\sum_{i=1}^{t-1} a^{2(i-1)} \\
=& \left(a^{t-1}\frac{1-a^{t-1}}{1-a} + a^{2(t-1)} \right) lV_0 + \frac{bl}{1-a}\left(\frac{1-a^{2(t-1)}}{1-a^2} - \frac{a^{t-1}-a^{2(t-1)}}{1-a} \right) \\
& + q\frac{1-a^{2t}}{1-a^2}.
\end{aligned}
$$

Lemma 4 *For $t \geq 1$ and an integer $p > 0$, the covariance of V_t and V_{t+p} given V_0 is*

$$cov(V_t, V_{t+p}|V_0) = AV_0 + B, \tag{2.15}$$

where the coefficients A and B are defined as

$$
\begin{aligned}
A &= a^{t+p-1} l\frac{1-a^{t-1}}{1-a} + a^{2(t-1)+p} l, \\
B &= a^p q\frac{1-a^{2t}}{1-a^2} + a^p \frac{bl}{1-a}\left(\frac{1-a^{2(t-1)}}{1-a^2} - \frac{a^{t-1}-a^{2(t-1)}}{1-a} \right).
\end{aligned}
$$

Proof:

$$
\begin{aligned}
Cov(V_{t+p}, V_t|V_0) &= E(V_{t+p}V_t|V_0) - E(V_{t+p}|V_0)E(V_t|V_0) \\
&= E(E(V_{t+p}|V_t)V_t|V_0) - E(V_{t+p}|V_0)E(V_t|V_0) \\
&= E(a^p V_t^2 + b\frac{1-a^p}{1-a}V_t) - \left(a^{t+p}V_0 + b\frac{1-a^{t+p}}{1-a}\right)\left(a^t V_0 + b\frac{1-a^t}{1-a}\right) \\
&= a^p\left[Var(V_t|V_0) + E^2(V_t|V_0)\right] + b\frac{1-a^p}{1-a}E(V_t|V_0) \\
&\quad - \left(a^{t+p}V_0 + b\frac{1-a^{t+p}}{1-a}\right)\left(a^t V_0 + b\frac{1-a^t}{1-a}\right) \\
&= a^p\left(a^{t-1}\frac{1-a^{t-1}}{1-a} + a^{2(t-1)}\right)lV_0 + a^p\frac{bl}{1-a}\left(\frac{1-a^{2(t-1)}}{1-a^2} - \frac{a^{t-1}-a^{2(t-1)}}{1-a}\right) \\
&\quad + a^p q\frac{1-a^{2t}}{1-a^2} + a^p\left(a^t V_0 + b\frac{1-a^t}{1-a}\right)^2 + b\frac{1-a^p}{1-a}\left(a^t V_0 + b\frac{1-a^t}{1-a}\right) \\
&\quad - \left(a^{t+p}V_0 + b\frac{1-a^{t+p}}{1-a}\right)\left(a^t V_0 + b\frac{1-a^t}{1-a}\right) \\
&= \left(a^{t+p-1}l\frac{1-a^{t-1}}{1-a} + a^{2(t-1)+p}l\right)V_0 + a^p q\frac{1-a^{2t}}{1-a^2} \\
&\quad + a^p\frac{bl}{1-a}\left(\frac{1-a^{2(t-1)}}{1-a^2} - \frac{a^{t-1}-a^{2(t-1)}}{1-a}\right).
\end{aligned}
$$

Lemma 5 *(i) Assume that the initial variance V_0 follows a distribution with mean μ_0 and variance σ_0^2. Then the unconditional mean and variance of V_t are*

$$
E(V_t) = a^t\mu_0 + b\frac{1-a^t}{1-a}, \tag{2.16}
$$

$$
\begin{aligned}
Var(V_t) = &\left[a^{t-1}\frac{1-a^{t-1}}{1-a} + a^{2(t-1)}\right]l\mu_0 + a^{2t}\sigma_0^2 \\
&+ \frac{bl}{1-a}\left(\frac{1-a^{2(t-1)}}{1-a^2} - \frac{a^{t-1}-a^{2(t-1)}}{1-a}\right) + q\frac{1-a^{2t}}{1-a^2}
\end{aligned} \tag{2.17}
$$

and the covariance between V_t and V_{t+p} is

$$Cov(V_t, V_{t+p}) = \left(a^{t+p-1} l \frac{1-a^{t-1}}{1-a} + a^{2(t-1)+p} l \right) \mu_0 + a^{2t} \sigma_0^2 + a^p q \frac{1-a^{2t}}{1-a^2} + a^p \frac{bl}{1-a} \left(\frac{1-a^{2(t-1)}}{1-a^2} - \frac{a^{t-1}-a^{2(t-1)}}{1-a} \right). \quad (2.18)$$

(ii) Given $\lambda/d < 1$, as $t \to \infty$, we have the following

$$\begin{aligned} \lim_{t\to\infty} E(V_t) &= \frac{b}{1-a}, \\ \lim_{t\to\infty} Var(V_t) &= \frac{q}{1-a^2} + \frac{bl}{(1-a)(1-a^2)}, \\ \lim_{t\to\infty} Cov(V_t, V_{t+p}) &= a^p \left[\frac{q}{1-a^2} + \frac{bl}{(1-a)(1-a^2)} \right]. \end{aligned} \quad (2.19)$$

Using the original parameter set, we have

$$\begin{aligned} \lim_{t\to\infty} E(V_t) &= \frac{\gamma}{c(d-\lambda)}, \\ \lim_{t\to\infty} Var(V_t) &= \frac{\gamma(d-\gamma+cd)}{c^2(d-\gamma)(d^2-\gamma^2)}, \\ \lim_{t\to\infty} Cov(V_t, V_{t+p}) &= \left(\frac{\lambda}{d}\right)^p \frac{\gamma(d-\gamma+cd)}{c^2(d-\gamma)(d^2-\gamma^2)}. \end{aligned} \quad (2.20)$$

Also it follows that the asymptotic autocorrelation function is

$$r(p) = \lim_{t\to\infty} Corr(V_t, V_{t+p}) = \left(\frac{\lambda}{d}\right)^p. \quad (2.21)$$

Thus as $t \to \infty$, the process V_t exhibits properties of a stationary mean-reverting process. For process V_t to have asymptotic mean one, we can make $d = \lambda + \gamma/c$. As exhibited by Figure 2.1, the faster the process reverts to the asymptotic mean level, the lower the value of autocorrelation (for a fixed lag p) is. Thus the speed of mean-reverting of this process can be reflected by the autocorrelation function, and thus the value of λ/d.

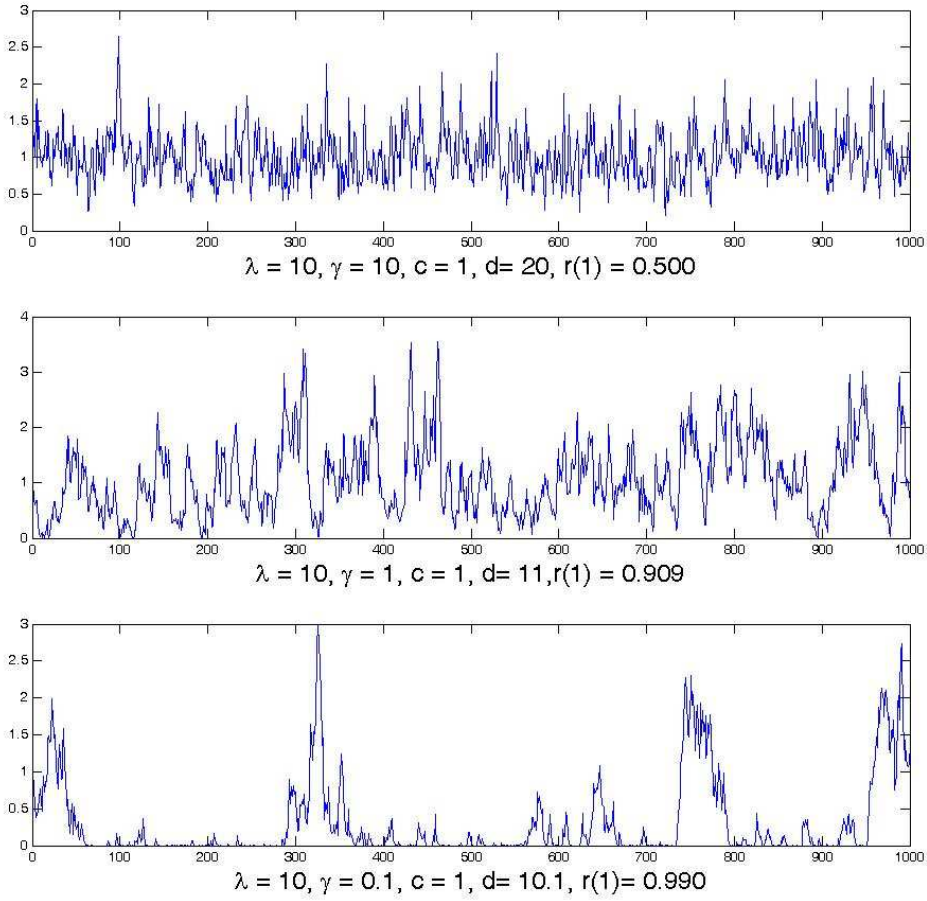

Figure 2.1: Simulated paths for double gamma process.

Figure(2.1) gives three simulated paths of double gamma processes with different set of parameters. As illustrated, c determines the overall volatility level of the Double Gamma process, that is, a greater c corresponds to smaller variances of u_t and a less volatile volatility process, and vice versa. Meanwhile, c also sets the level of persistence, which can be defined as the tendency of a mean-reverting process reverts slowly to its equilibrium (Dias and Marques (2005)). The slower the underlying process converges to its equilibrium, the more persistent the process is. Since a high autocorrelation corresponds to a slow mean-reverting speed, persistence is positively related to the asymptotic autocorrelation $r(p) = (c\lambda/(c\lambda + \gamma))^p$. With a greater value of c, the process has higher autocorrelation, slower speed of mean-reverting and thus is more persistent and vice versa. Reducing γ will offset the impact of lower c and make it possible for a process to have both high persistence and high volatility and vice versa.

Also we can calculate the moment generating function of V_{t+p} for an arbitrary step number p.

Lemma 6 *The moment generating function of V_{t+p} conditioned on V_t is:*

$$\phi_{V_{t+1}|V_t}(\omega) = \left(\frac{d}{d-\omega}\right)^{\lambda V_t} \phi_u\left(\log\frac{d}{d-\omega}\right), \tag{2.22}$$

$$\phi_{V_{t+2}|V_t}(\omega) = \phi_{V_{t+1}|V_t}\left(\lambda \log\frac{d}{d-\omega}\right) \phi_u\left(\log\frac{d}{d-\omega}\right), \tag{2.23}$$

$$\phi_{V_{t+p}|V_t}(\omega) = \phi_{V_{t+1}|V_t}(\lambda g^{(p)}(\omega)) \prod_{j=1}^{p-1} \phi_u(g^{(j)}(\omega)) \tag{2.24}$$

where $\phi_u(\omega)$ is the moment generating function of a gamma random variable u:

$$\phi_u(\omega) = \left(\frac{c}{c-\omega}\right)^{\gamma}, \tag{2.25}$$

and $g^{(t)}(\omega)$ is a sequence defined sequentially as

$$g^{(n)}(\omega) = \log\left(\frac{d}{d - g^{(n-1)}(\omega)}\right),$$

$$g^{(0)}(\omega) = \omega.$$

Proof of equation(2.23):

$$\begin{aligned}
\phi_{V_{t+2}|V_t}(\omega) &= E(e^{\omega V_{t+2}}|V_t) \\
&= E\left(E(e^{\omega V_{t+2}}|V_{t+1})|V_t\right) \\
&= E\left(\left(\frac{d}{d-\omega}\right)^{\lambda V_{t+1}+u_{t+2}}|V_t\right) \\
&= E\left(\exp\left(\lambda V_{t+1}\log\frac{d}{d-\omega}\right)|V_t\right)\phi_u\left(\log\frac{d}{d-\omega}\right) \\
&= \phi_{V_{t+1}|V_t}\left(\lambda\log\frac{d}{d-\omega}\right)\phi_u\left(\log\frac{d}{d-\omega}\right).
\end{aligned}$$

Proof of equation(2.22) is trivial. And given (2.23), it is trivial to prove equation(2.24).

Similarly we can also use characteristic function to explore the distributional properties of the Double Gamma process. Similar to Formulas (2.22-2.25), the recursive calculation of characteristic functions can be written as,

$$\widehat{\phi}_{V_{t+1}|V_t}(\omega) = \left(\frac{d}{d-i\omega}\right)^{\lambda V_t}\widehat{\phi}_u\left(-i\log\frac{d}{d-i\omega}\right), \tag{2.26}$$

$$\widehat{\phi}_{V_{t+2}|V_t}(\omega) = \widehat{\phi}_{V_{t+1}|V_t}\left(-i\lambda\log\frac{d}{d-i\omega}\right)\widehat{\phi}_u\left(-i\log\frac{d}{d-i\omega}\right), \tag{2.27}$$

$$\widehat{\phi}_{V_{t+p}|V_t}(\omega) = \widehat{\phi}_{V_{t+1}|V_t}(\lambda\widehat{g}^{(p)}(\omega))\prod_{j=1}^{p-1}\widehat{\phi}_u(\widehat{g}^{(j)}(\omega)) \tag{2.28}$$

where $\widehat{\phi}(u)$ is the characteristic function of a gamma random variable u:

$$\widehat{\phi}_u(w) = \left(\frac{c}{c - i\omega}\right)^{\gamma},$$

and $\widehat{g}^{(t)}(\omega)$ is a sequence defined recursively as

$$\widehat{g}^{(n)}(\omega) = -i \log\left(\frac{d}{d - i\widehat{g}^{(n-1)}(\omega)}\right), \widehat{g}^{0}(\omega) = \omega.$$

For ω near 0, we can use the principal branch Log() of the logarithm of the complex values to avoid the multivalued problem for complex logarithm.

2.2 Local innovation driven by adapted Lévy process

In this section we introduce the new distribution for the local innovation x_t which provides another source of non-zero skewness and kurtosis, besides stochastic volatility. Such a random variable x_t can be generated by discretizing the variance gamma process of Madan and Miline (1991).

The variance gamma process $G(t; \sigma, \nu, \theta)$ is defined by a Brownian motion with drift, $b(t; \theta, \sigma)$, and an independent random gamma time with mean rate one, $\gamma(t; 1, \nu)$. More specifically,

$$G(t; \sigma, \nu, \theta) = b(\gamma(t; 1, \nu); \theta, \sigma). \tag{2.29}$$

By definition we can obtain a variance gamma process by evaluating Brownian motion at a random time which follows a gamma process. The marginal distribution of the variance gamma process are determined by three parameters: (i) σ, the volatility of the Brownian motion; (ii) ν, the variance rate of the gamma time change

and (iii) θ, the drift in the subordinated Brownian Motion with drift. The process incorporates richer distributional features, particularly a versatile specification of skewness and kurtosis that are jointly determined by θ and ν.

Definition 2 *A random variable X has an adapted variance gamma distribution if it can be specified as follows,*

$$X|Y \sim N\left(\theta(Y-1), \sigma^2 Y\right),$$
$$Y \sim Gamma(1/\nu, 1/\nu). \tag{2.30}$$

According to the definition we can produce an adapted Variance Gamma random variable X by firstly generating a Gamma random variable Y with scale parameter $1/\nu$ and shape parameter $1/\nu$, and then generating a random variable that follows a conditional normal distribution with mean $\theta(Y-1)$ and variance $\sigma^2 Y$.

The new distribution defined above inherits the moments of a variance gamma process at unit time.

Lemma 7 *(i). The density function of the adapted variance gamma variable X is*

$$f_X(x|\theta,\sigma,\nu) = \frac{2\exp(\theta(x+\theta)/\sigma^2)}{\nu^{1/\nu}\sqrt{2\pi}\Gamma(1/\nu)}\left(\frac{(x+\theta)^2}{2\sigma^2+\theta^2}\right)^{\frac{1}{2\nu}-\frac{1}{4}}$$
$$K_{\frac{1}{\nu}-\frac{1}{2}}\left(\frac{1}{\sigma^2}\sqrt{(x+\theta)^2(2\sigma^2/\nu+\theta^2)}\right), \tag{2.31}$$

where K is the modified Bessel function of the second kind.

(ii). The unconditional central moments of X are:

$$\begin{aligned} E(X) &= 0, \\ E[X - E(X)]^2 &= \theta^2\nu + \sigma^2, \\ E[X - E(X)]^3 &= 2\theta^3\nu^2 + 3\sigma^2\theta\nu, \\ E[X - E(X)]^4 &= 3\sigma^4\nu + 12\sigma^2\theta^2\nu^2 + 6\theta^4\nu^3 + 3\sigma^4 \\ &\quad + 6\sigma^2\theta^2\nu + 3\theta^4\nu^2. \end{aligned} \tag{2.32}$$

Proof of (i).The density function of X conditional on Y is:

$$f_X(x|Y) = \frac{1}{\sqrt{2\pi Y}\sigma} \exp\left(-\frac{(x - \theta(Y-1))^2}{2\sigma^2 Y}\right) \tag{2.33}$$

and the density function of Y is:

$$f_Y(y) = \frac{(1/\nu)^{1/\nu}}{\Gamma(1/\nu)} y^{1/\nu - 1} e^{-y/\nu}. \tag{2.34}$$

Thus the density function of X can be calculated by integrating Y out of the joint density function:

$$f_X(x) = \int_0^\infty \frac{1}{\sqrt{2\pi Y}\sigma} \exp\left(-\frac{(x - \theta(Y-1))^2}{2\sigma^2 Y}\right) \frac{(1/\nu)^{1/\nu}}{\Gamma(1/\nu)} y^{1/\nu - 1} e^{-y/\nu} dy. \tag{2.35}$$

By Gradshetyn and Ryzhik (1980) 3.471.9 this form is integrable with the result given by equation (2.31).

Proof of (ii). Since $E(X|Y) = \theta(Y-1)$ and $E(Y) = 1$, $E(X) = 0$, the expectation of X^2 can be calculated as follows,

$$\begin{aligned} E(X^2) &= E[E(X^2|Y)] = E[\mathrm{Var}(X|Y)] + E(E(X|Y))^2 \\ &= \sigma^2 E(Y) + \theta^2 E(Y-1)^2. \end{aligned}$$

Since $E(Y-1)^2 = \mathrm{Var}(Y) = \nu$, we have $E(X^2) = \sigma^2 + \theta^2\nu$.

Expressing $X = \theta(Y-1) + \sigma\sqrt{Y}z$ with z being a standard normal variable, we have

$$\begin{aligned} E(X^3) =& E(E(X^3|Y)) \\ =& E[\theta^3(Y-1)^3 + 3\theta^2(Y-1)^2\sigma\sqrt{Y}z + 3\theta(Y-1)\sigma^2 Y z^2 + \sigma^3 Y^{3/2} z^3] \\ =& E[\theta^3(Y-1)^3)] + 3\theta\sigma^2\nu. \end{aligned}$$

On explicit integration we have that

$$E(Y^3) = \nu^2(2 + \frac{1}{\nu})(1 + \frac{1}{\nu})\frac{1}{\nu}.$$

It follows that

$$E[(Y-1)^3] = 3 + 3\nu + 2\nu^2 - 3(\nu+1) = 2\nu^2.$$

For the fourth moment we note on expanding X^4 and taking expectations that

$$E(X^4) = \theta^4 E[(Y-1)^4] + 6\sigma^2\theta^2 E[(Y-1)^2 Y] + 3\sigma^4 E(Y^2).$$

The expectation of Y^4 can be explicitly computed by integration and is

$$E(Y^4) = (3\nu+1)(2\nu+1)(\nu+1).$$

By substitution and collecting terms we have

$$\begin{aligned} E[X^4] =& 3\sigma^4\nu + 12\sigma^2\theta^2\nu^2 + 6\theta^4\nu^3 + 3\sigma^4 \\ &+ 6\sigma^2\theta^2\nu + 3\theta^4\nu^2. \end{aligned}$$

For this new distribution, skewness and kurtosis are determined jointly by θ and ν. For X to have a unit variance, we can specify $\theta = \sqrt{(1-\sigma^2)/\nu}$.

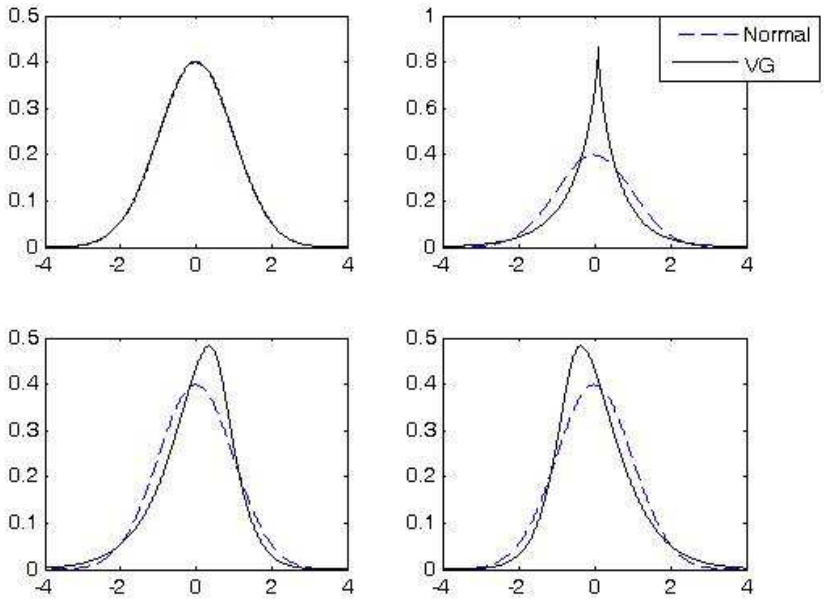

Figure 2.2: VG probability density function $f_X(x)$. Parameters (θ, σ, ν) for VG density: upper left (-4.0, 0.9985, 0.0002)– very close to Normal, upper right (-0.1, 0.994, 1.1964) –high kurtosis, lower left (-0.6, 0.92, 0.4267)–negative skewness, lower right (0.6, 0.92, 0.4267)–positive skewness.

As illustrated in Figure 2.2, this distribution parsimoniously covers the cases of non-zero skewness and excess kurtosis. The adaption can be extended to other Lévy processes, for instance, CGMY process of Carr, Geman, Madan, and Yor (2002) and the normal inverse Gaussian process of Barndorff-Nielsen and Shephard (1998).

An analytic density function, although desirable, doesn't necessarily exist for the random variables generated from the procedure described above. In such cases we can turn to the analytical characteristic functions, and apply the Fast Fourier Transform to evaluate the distribution function. Details of this methodology are given in Carr and Madan (1998).

2.3 Leverage effect

In the model description equations(2.2), the leverage effect is incorporated into the model by assuming that $\sigma_t'^2$ is a quadratic function of the lagged disturbance x_t:

$$\begin{aligned} \sigma_t^2 &= \sigma_t'^2 V_t h, \\ \sigma_t'^2 &= \sigma_0^2(1 + \alpha x_t + \beta x_t^2). \end{aligned} \tag{2.36}$$

The covariance between $\sigma_t'^2$ and x_t is as follows,

$$\begin{aligned} \operatorname{cov}(\sigma_t'^2, x_t) &= E(\sigma_t'^2 - E(\sigma_t'^2))x_t \\ &= \sigma_0^2 E[\alpha x_t + \beta(x_t^2 - E(x_t^2)]x_t \\ &= \sigma_0^2 \left(\alpha E(x_t^2) + \beta E(x_t^3)\right) \\ &= \sigma_0^2 \left(\alpha + \beta E(x_t^3)\right). \end{aligned} \tag{2.37}$$

Here we used the properties of x_t that $E(x_t) = 0$ and $E(x_t^2) = 1$. If $\alpha < 0$ and x_t

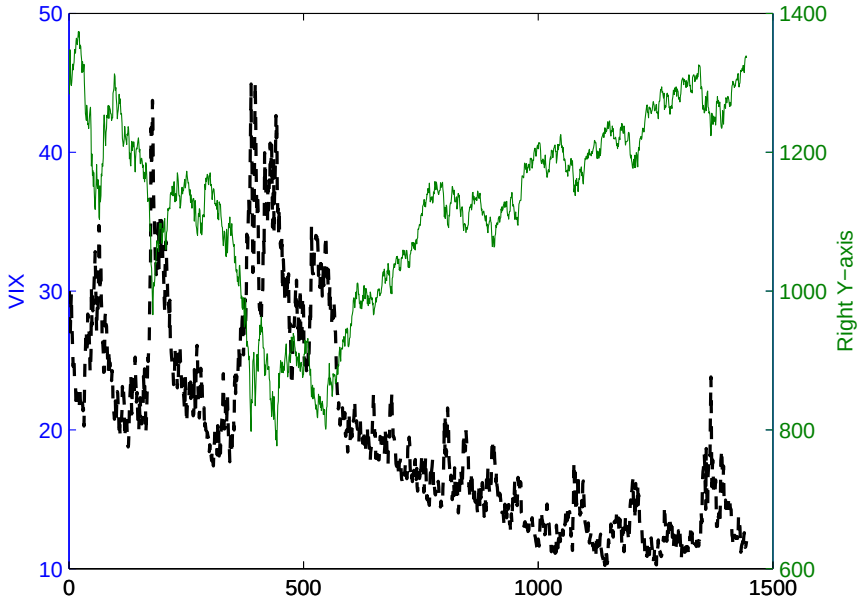

Figure 2.3: Leverage effect observed in historical data of VIX and S&P500Dashed line is the VIX index from 2001 to 2006. Solid line is S&P500 index on the same period.

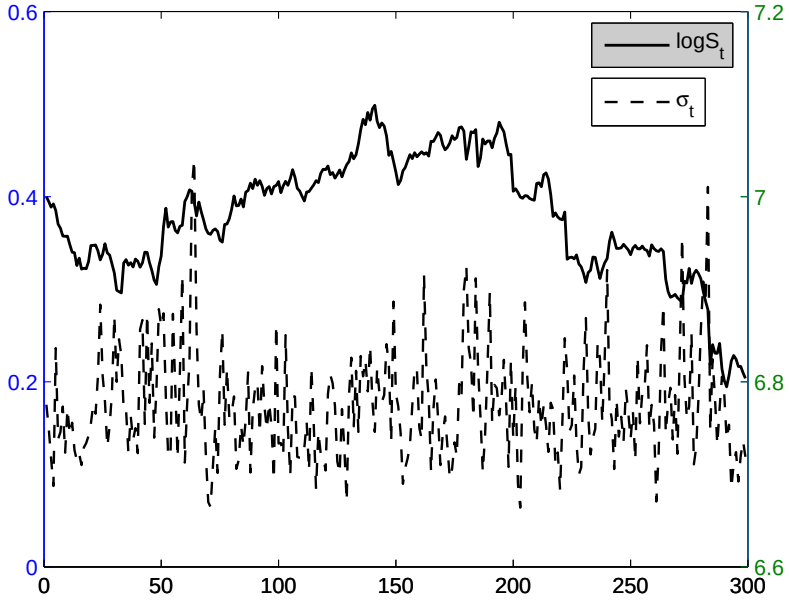

Figure 2.4: Simulated price and volatility path. $\text{corr}(\sigma_t, d\log S_t) = -0.1$, parameters used in simulation: $(\mu, \theta, \sigma) = (0.02, -0.15, 0.999)$, $(\sigma_0, \alpha, \rho) = (0.17, -0.9, 1)$ and $(\lambda, \gamma, c) = (5, 20, 2)$

has zero skewness, for example x_t follows a standard normal distribution and thus $E(x_t^3) = 0$, then $\text{cov}(\sigma_t'^2, x_t) < 0$ and σ'^2 and x_t are negatively correlated. In the case $\alpha < 0$, $\beta > 0$ and x_t has negative skewness, we also have $\text{cov}(\sigma_t'^2, x_t) < 0$.

Since this quadratic function will be non-negative if and only if $\alpha^2 - 4\beta < 0$, we set $\beta = \frac{1}{4}\alpha^2 + \eta$ with $\eta > 0$ to satisfy this condition.

2.4 Frequency of volatility change

Assuming that V_t changes every m days and hence the original time span $[0, T]$ can be cut into $N = [\frac{T}{m}]$ periods, we can model the constant level of volatility in the nth period using a new process W_n where W_n,$n = 1, ...N$ follows a double gamma

process as defined in section 2.1. More specifically,

$$\begin{aligned} V_t &= W_{[(t-1)/m]}, \\ W_{n+1}|W_n &\sim Gamma(d, \lambda W_n + U_{n+1}), \\ U_{n+1} &\sim Gamma(c, \gamma). \end{aligned} \tag{2.38}$$

This modification drops the constraint that the latent volatility process has to evolve with the same frequency as that of the price changes, and thus avoids the possible problem of overfitting. A by-product of this optimal frequency searching is that computational cost can potentionally be significantly reduced, specifically, in calculating the likelihood function during estimation. More details on this issue will be given in next chapter.

2.5 Long-term dependence

Our model and technique extends to the case where volatility follows a weighted sum of independent double-gamma processes with different levels of persistence, that is,

$$V_t = \sum_{j=1}^{M} w_j V_j(t), \tag{2.39}$$

where

$$V_{j,t}|V_{j,t-1} \sim DG(V_{j,t-1}, \gamma_i, d_j, \lambda_j, c_j, d_j) \quad \text{and} \quad \sum_{j=1}^{M} w_j = 1. \tag{2.40}$$

Since some of the components of the volatility may represent short-term variation in the process whereas others represent long-term movements, by adding together the independent double-gamma processes with different persistence levels we

can obtain a more generic volatility structure. Combining with the results in section 2.1, this implies an asymptotic autocorrelation function which is a weighted sum of exponentials:

$$r(k) = w_1' \exp(-\rho_1|k|) + ... + w_m' \exp(-\rho_m|k|) \tag{2.41}$$

with

$$\rho_i = -\log(\frac{\lambda_i}{d_i}), \quad w_i' = \frac{w_i^2 A_i}{\sum_{i=1}^{M} w_i^2 A_i} \quad \text{and}$$
$$A_i = \frac{\gamma_i(d_i - \gamma_i + c_i d_i)}{c_i^2(d_i - \gamma_i)(d_i^2 - \gamma_i^2)}.$$

Other discrete time models of this type are discussed by Engle and Lee (1993), Dacorogna, Muller, Olsen, and Pictet and Barndorff-Nielsen and Shephard (1998).

By choosing the weights and damping factors in (2.41) appropriately and letting $M \to \infty$ it is possible to construct tractable volatility models with long-range dependence. In particular, Barndorff-Nielsen and Shephard (2000) showed that there is a limiting model for which

$$r(k) = (1 + \rho|k|)^{-2(1-H)}, \tag{2.42}$$

where $\rho > 0$ and $H \in (1/2, 1)$ is the long memory parameter. The authors also developed more general models, applying the theory of independently scattered measures and Lévy random fields. Similar types of arguments have previously been used for real-valued time series models by, for example, Granger (1980) and Cox (1991). Ding and Granger (1996) have studied long memory in volatility using the addition of short memory processes, whereas Andersen and Bollerslev (1997) have applied the

theory of heterogeneous information arrivals to motivate a long memory volatility model.

Chapter 3

Estimation

The estimation of the stochastic volatility model requires evaluations of the likelihood function for a high dimensional model which contains latent volatility series. To evaluate the likelihood function we need to integrate the volatilities out of the joint density function and eventually the high dimension of the process renders the direct numerical integration impractical. For this type of problem we propose a method based on the combination of Expectation-Maximization algorithm (Dempster, Laird, and Rubin (1977)) and the particle filter method to maximize the likelihood function.

The second part of this chapter is spent on exploring the issue of determining the optimal frequency for volatility change. As different frequencies of volatility changes correspond to different models, we will apply model selection methods - Akaike Information Criterion and Likelihood Ratio Test, to search for the optimal frequency that best fits the observed return data.

3.1 Likelihood Function

Given the observed returns r_t, $t = 1, ..., T$, the likelihood function of the parameter set Ω for the stochastic volatility model is

$$L(\Omega|r_1, .., r_T) = f(r_1|\Omega) \prod_{t=2}^{T} f(r_t|r_1, ..., r_{t-1}, \Omega). \tag{3.1}$$

For the convenience of notation we denote $\{r_1, ..., r_t\}$ by $\mathbf{r}_{1:t}$ and $\{(V_1, u_1), ..., (V_t, u_t)\}$ by $\mathbf{V}_{1:t}$. To calculate the marginal conditional density function of r_t, $f(r_t|\mathbf{r}_{1:t-1}, \Omega)$, the latent variables $\mathbf{V}_{1:t}$ need to be integrated out of the joint density function of the observed and latent variables,

$$f(\mathbf{r}_{1:T}, \mathbf{V}_{1:T}|\Omega) = f(V_1, u_1|\Omega) \prod_{t=2}^{T} f(r_t|V_t, u_t) f(V_t, u_t|V_{t-1}, u_{t-1}.\Omega). \tag{3.2}$$

From a broad perspective, the Expectation-Maximization algorithm of Dempster et al. (1977) lends us the idea to solve the problem of high dimensional integration. The joint density function of $\mathbf{r}_{1:t}$ can be written as

$$f(\mathbf{r}_{1:t}|\Omega) = f(\mathbf{r}_{1:t}, \mathbf{V}_{1:t}|\Omega)/f(\mathbf{V}_{1:t}|\mathbf{r}_{1:t}, \Omega). \tag{3.3}$$

Taking logarithms on both sides we have

$$\log f(\mathbf{r}_{1:t}|\Omega) = \log f(\mathbf{r}_{1:t}, \mathbf{V}_{1:t}|\Omega) - \log f(\mathbf{V}_{1:t}|\mathbf{r}_{1:t}, \Omega). \tag{3.4}$$

Now for an arbitrary parameter set Ω', taking expectations on both sides with respect to the posterior distribution of $\mathbf{V}_{1:t}$ given $\mathbf{r}_{1:t}$ and another parameter set Ω, we see readily that

$$\log f(\mathbf{r}_{1:t}|\Omega') = E_{\mathbf{V}_{1:t}|\mathbf{r}_{1:t},\Omega}[\log f(\mathbf{r}_{1:t}, \mathbf{V}_{1:t}|\Omega')] - E_{\mathbf{V}_{1:t}|\mathbf{r}_{1:t},\Omega}[\log f(\mathbf{V}_{1:t}|\mathbf{r}_{1:t}, \Omega')]. \tag{3.5}$$

Dempster et al. (1977) proved that for a sequence of parameters $(\Omega_1, ..., \Omega_k, ...)$, which is updated sequentially in the following iteration,

$$\Omega_{k+1} = \text{argmax}_{\Omega'} E_{\mathbf{V}_{1:t}|\mathbf{r}_{1:t},\Omega_k}[\log f(\mathbf{r}_{1:t}, \mathbf{V}_{1:t}|\Omega')]. \tag{3.6}$$

Each iteration of the EM algorithm contains two steps. The expectation (E) step is to calculate the expectation of function $\log f(\mathbf{r}_{1:t}, \mathbf{V}_{1:t}|\Omega')$, as a function of the new parameter set Ω'. The posterior density for calculating expectation is a function of the former parameter set Ω_k. The maximization (M) step maximizes the expectation as indicated in equation (3.6).

The convergence of LM algorithm was studied by Wu (1983). Denote $\mathcal{M}$ as the set of local maxima in the interior of parameter space $\mathbf{\Omega}$ and $\mathcal{J}$ as the set of stationary points in the interior of $\mathbf{\Omega}$. If $L(\Omega_{k+1}|\Omega_k)$ is continuous in both Ω_k and Ω' and

$$\sup_{\Omega' \in \mathbf{\Omega}} L(\Omega'|\Omega_k) > L(\Omega_k|\Omega_k)$$

for any $\Omega_k \in \mathcal{J}/\mathcal{M}$, then the EM algorithm converges monotonically to $L^* = L(\Omega^*)$ for some local maximum Ω^*. An important class of densities satisfy this condition is the curved exponential family

$$f(x|\Omega) = b(x) \exp\{\Omega^T t(x)/a(\Omega)\}. \tag{3.7}$$

Since the expectation of the log likelihood function, designated by $Q(\Omega_k, \Omega')$, is

$$\begin{aligned} Q(\Omega_k, \Omega') &= E_{\mathbf{V}_{1:t}|\mathbf{r}_{1:t},\Omega_k}[\log f(\mathbf{r}_{1:t}, \mathbf{V}_{1:t}|\Omega')] \\ &= \int \log f(\mathbf{r}_{1:t}, \mathbf{V}_{1:t}|\Omega') f(\mathbf{V}_{1:t}|\mathbf{r}_{1:t}, \Omega_k) d(\mathbf{V}_{1:t}) \end{aligned} \tag{3.8}$$

and does not yield a closed form expression, we turn to Monte Carlo integration $Q_M(\Omega_k, \Omega')$ in the E Step to approximate $Q(\Omega_k, \Omega')$:

$$\begin{aligned} Q(\Omega_k, \Omega') &\approx Q_M(\Omega_k, \Omega') \\ &= \sum_{j=1}^{M} \log f(\mathbf{r}_{1:t}, \mathbf{V}_{1:t}^{(j)} | \Omega') p(\mathbf{V}_{1:t}^{(j)} | \mathbf{r}_{1:t}, \Omega_k) \end{aligned} \tag{3.9}$$

where $\mathbf{V}_{1:t}^{(j)}$ is jth of the M paths sampled from the distribution $f(\mathbf{V}_{1:t}|\mathbf{r}_{1:t}, \Omega_k)$. The weights $p(\mathbf{V}_{1:t}^{(j)}|\mathbf{r}_{1:t}, \Omega_k)$ are values proportional to $f(\mathbf{V}_{1:t}^{(j)}|\mathbf{r}_{1:t}, \Omega_k)$ and are subjected to the constraint that $\sum_{j=1}^{M} p(\mathbf{V}_{1:t}^{(j)}|\mathbf{r}_{1:t}, \Omega_k) = 1$.

To apply the EM algorithm to the estimation of the stochastic volatility model, there are two issues to be resolved. The first is to evaluate the posterior distribution $f(\mathbf{V}_{1:t}|\mathbf{r}_{1:t}, \Omega_k)$ and the second is how to generate samples from resulting distribution.

3.2 Calculation of Posterior Distribution

According to Bayes' theorem, the conditional density function of $\mathbf{V}_{1:t}$ given the observations $\mathbf{r}_{1:t}$ can be written as

$$f(\mathbf{V}_{1:t}|\mathbf{r}_{1:t}, \Omega) = \frac{f(\mathbf{r}_{1:t}|\mathbf{V}_{1:t}, \Omega) f(\mathbf{V}_{1:t}|\Omega)}{\int f(\mathbf{r}_{1:t}|\mathbf{V}_{1:t}, \Omega) f(\mathbf{V}_{1:t}|\Omega) d\mathbf{V}_{1:t}}. \tag{3.10}$$

The following formula can be used to recursively calculate $f(\mathbf{V}_{1:t}|\mathbf{r}_{1:t}, \Omega)$:

$$f(\mathbf{V}_{1:t}|\mathbf{r}_{1:t}, \Omega) = f(\mathbf{V}_{1:t-1}|\mathbf{r}_{1:t-1}, \Omega) \frac{f(r_t|V_t, \Omega) f(V_t|V_{t-1}, \Omega)}{f(r_t|\mathbf{r}_{1:t-1}, \Omega)}. \tag{3.11}$$

Here we apply the Markovian property of both price and volatility processes.

More specifically,

$$f(r_t|\mathbf{r}_{1:t-1}, \mathbf{V}_{1:t}, \Omega) = f(r_t|V_t, \Omega),$$

$$f(V_t|\mathbf{r}_{1:t-1}, \mathbf{V}_{1:t-1}, \Omega) = f(V_t|V_{t-1}, \Omega). \tag{3.12}$$

3.2.1 Sequential Importance Sampling

Importance sampling is applied to generate samples from the posterior distribution of $\mathbf{V}_{1:t}$, $f(\mathbf{V}_{1:t}|\mathbf{r}_{1:t}, \Omega)$, which cannot be directly simulated. The idea is to simulate the paths of $\mathbf{V}_{1:t}$ from the proposal distribution $\pi(\mathbf{V}_{1:t}|\mathbf{r}_{1:t})$ and then to modify equation (3.9) as follows:

$$\begin{aligned} Q(\Omega_k, \Omega') &\approx Q_M(\Omega_k, \Omega') \\ &= \sum_{j=1}^{M} \log f(\mathbf{r}_{1:t}, \mathbf{V}_{1:t}^{(j)}|\Omega') p(\mathbf{V}_{1:t}^{(j)}|\mathbf{r}_{1:t}, \Omega_k) \\ &= \sum_{j=1}^{M} \log f(\mathbf{r}_{1:t}, \mathbf{V}_{1:t}^{(j)}|\Omega') w(\mathbf{V}_{1:t}^{(j)}|\mathbf{r}_{1:t}, \Omega_k) \pi(\mathbf{V}_{1:t}^{(j)}|\mathbf{r}_{1:t}, , \Omega_k), \end{aligned} \tag{3.13}$$

where $w(\mathbf{V}_{1:t}^{(j)}|\mathbf{r}_{1:t}, \Omega_k)$ is known as the importance weight,

$$w(\mathbf{V}_{1:t}^{(j)}|\mathbf{r}_{1:t}, \Omega_k) = \frac{p(\mathbf{V}_{1:t}^{(j)}|\mathbf{r}_{1:t}, \Omega_k)}{\pi(\mathbf{V}_{1:t}^{(j)}|\mathbf{r}_{1:t}, \Omega_k)}. \tag{3.14}$$

Following equation (3.13) we can simulate M i.i.d paths $\{V_{1:t}^{(j)}, j = 1, ..., M\}$ from $\pi(\mathbf{V}_{1:t}^{(j)}|\mathbf{r}_{1:t}, \Omega_k)$, and then calculate a Monte Carlo estimate of $Q(\Omega_k, \Omega')$ through equation (3.13).

Similar to equation (3.11) the approximation process for the conditional distribution $\pi(\mathbf{V}_{1:t}^{(j)}|\mathbf{r}_{1:t}, \Omega_k)$ can also be carried out recursively applying the Markovian

property of both the r_t and V_t processes, so that

$$\pi(\mathbf{V}_{1:t}|\mathbf{r}_{1:t},\Omega_k) = \frac{\pi(\mathbf{r}_{1:t}|\mathbf{V}_{1:t},\Omega_k)\pi(\mathbf{V}_{1:t},\Omega_k)}{\int \pi(\mathbf{r}_{1:t}|\mathbf{V}_{1:t},\Omega_k)\pi(\mathbf{V}_{1:t}|\Omega_k)d\mathbf{V}_{1:t}}. \tag{3.15}$$

Combining equations (3.11) and (3.15) we can calculate the importance weight function $\tilde{w}_t^{(j)} = w(V_{1:t}^{(j)}|\mathbf{r}_{1:t},\Omega_k)$ as:

$$\begin{aligned}\tilde{w}_{t+1}^{(j)} &= \tilde{w}_t^{(j)} \frac{f(V_{t+1}|V_{1:t}^{(j)},\mathbf{r}_{1:t+1},\Omega_k)}{\pi(V_{t+1}|V_{1:t}^{(j)},\mathbf{r}_{1:t+1},\Omega_k)} \\ &\propto \tilde{w}_t^{(j)} \frac{f(\mathbf{r}_{t+1}|V_{t+1}^{(j)},\Omega_k) f(V_{t+1}^{(j)}|V_t^{(j)},\Omega_k)}{\pi(V_{t+1}|V_{1:t}^{(j)},\mathbf{r}_{1:t+1},\Omega_k)}.\end{aligned} \tag{3.16}$$

One convenient choice for the proposal distribution $\pi(\mathbf{V}_{1:t}|\mathbf{r}_{1:t},\Omega_k)$ is the prior distribution $f(\mathbf{V}_{1:t}|\Omega_k)$, by which the proposal distribution function $\pi(\mathbf{V}_{1:t}|\mathbf{r}_{1:t},\Omega_k)$ can be calculated as:

$$\pi(\mathbf{V}_{1:t}|\mathbf{r}_{1:t},\Omega_k) = f(\mathbf{V}_{1:t}|\Omega_k) = f(V_1|\Omega_k)\sum_{i=2}^{t} f(V_i|V_{i-1},\Omega_k). \tag{3.17}$$

Then the importance weights are simplified to

$$\tilde{w}_{t+1}^{(j)} \propto \tilde{w}_t^{(j)} f(\mathbf{r}_{t+1}|V_{t+1}^{(j)},\Omega_k). \tag{3.18}$$

Sequential importance sampling is an attractive method but also has significant drawbacks when applied to problems with high dimension, as pointed by Gilks, Richardson, and Spiegelhalter (1996). As t increases, the distribution of the importance weight $w_t^{(j)}$ becomes increasingly skewed and therefore after a few time steps, only one path has non-zero importance weight. Consequently the algorithm fails to represent the posterior distributions. To avoid the degeneracy problem, one more step needs to be added to resample the sampled paths.

3.2.2 Particle filter

Particle filtering is applied to eliminate the simulated paths (particles) with low importance weights $w_t^{(j)}$ and to multiply paths with greater weights (Gordon). In implementation we replace the weighted empirical distribution

$$\widehat{P}_M(\mathbf{V}_{1:t}|\mathbf{r}_{1:t}) = \sum_{j=1}^{M} w_t^{(j)} \delta_{\mathbf{V}_{1:t}^{(j)}}(d\mathbf{V}_{1:t}) \tag{3.19}$$

by

$$P_M(\mathbf{V}_{1:t}|\mathbf{r}_{1:t}) = \frac{1}{N} \sum_{j=1}^{N} N_t^{(j)} \delta_{\mathbf{V}_{1:t}^{(j)}}(d\mathbf{V}_{1:t}), \tag{3.20}$$

where $\delta_{\mathbf{V}_{1:t}^{(j)}}$ denotes the Dirac mass located at $\mathbf{V}_{1:t}^{(j)}$ and $N_t^{(j)}$ is the number of offspring associated with particle $\mathbf{V}_{1:t}^{(j)}$, which satisfies $\sum_{j=1}^{M} N_t^{(j)} = M$. If $N_t^{(j)} = 0$ then the particle $\mathbf{V}_{1:t}^{(j)}$ is filtered out. After the selection step the surviving particles $\mathbf{V}_{1:t}^{(j)}$ with $N_t^{(j)} > 0$, are distributed as $f(\mathbf{V}_{1:t}|\mathbf{r}_{1:t})$. We obtain the surviving particles by sampling N times from the distribution $\widehat{P}_M(\mathbf{V}_{1:t}|\mathbf{r}_{1:t})$, which is equivalent to sampling the number of offspring $N_t^{(j)}$ according to a multinomial distribution with parameters $w_t^{(j)}$.

The next sections will summarize the implementation for two cases. We begin with the case where volatility is assumed to change with the same frequency as prices, then advance to a more generic setting where different frequencies of volatility change are allowed.

3.2.3 Algorithm for daily-changing volatility

Assuming volatility changes daily, we have the frequency parameter m equal to 1. Then $\mathbf{V}$ could be simulated by carrying out the following steps.

Step 1. At $t = 1$, set $V_1^{(j)} = 1$ and $w_1^{(j)} = 1/M$ for $j = 1, ..., M$.

Step 2. At $t = i \geq 2$, generate $V_i^{(j)}$ from the double gamma process conditional on $V_{i-1}^{(j)}$, for $j = 1, ..., M$. Reset the weight of each particle in the sample by

$$w_i^{(j)} \propto w_{i-1}^{(j)} f(r_i | V_i^{(j)}) \tag{3.21}$$

where $f(r_i|V_i^{(j)})$ is the conditional density function of r_i given $V_i^{(j)}$.

Normalize $w_i^{(j)}$ such that $\sum_{j=1}^{M} w_i^{(j)} = 1$.

Add $V_i^{(j)}$ to the sequence $\mathbf{V}_{1:i-1}^{(j)}$ and set $\tilde{\mathbf{V}}_{1:i}^{(j)} = (\mathbf{V}_{1:i-1}^{(j)}, V_i^{(j)})$.

Step 3. Resample with replacement M particles $\{\mathbf{V}_{1:i}^{(j)}, j = 1, .., M\}$ from $\{\tilde{\mathbf{V}}_{1:i}^{(j)}, j = 1, .., M\}$ with weights $\{w_i^{(j)}, j = 1, .., M\}$.

Set $i = i + 1$ and go to step 2.

3.2.4 Algorithm for volatility with arbitrary frequency

In this case m is an integer greater than or equal to one. With some modification to the algorithm described in the last section, we sample $\mathbf{V}$ as follows.

Step 1. At $n = 1$, set $W_1^{(j)} = 1$ and $w_1^{(j)} = 1/M$ for $j = 1, ..., M$. Set $V_k^{(j)} = W_1^{(j)}$ for $k = 1, ..., m$.

Step 2. At $n = i$, generate $W_i^{(j)}$ from the double gamma process conditional on $W_{i-1}^{(j)}$, for $j = 1, ..., M$. Set $V_{k+(i-1)m}^{(j)} = W_i^{(j)}$ for $k = 1, ..., m$. Reset the weight of each particle in the sample by

$$w_i^{(j)} \propto w_{i-1}^{(j)} \prod_{k=1}^{m} f(r_{k+(i-1)m}^{(j)} | V_{k+(i-1)m}^{(j)}) \tag{3.22}$$

where $f(r^{(j)}_{k+(i-1)m}|V^{(j)}_{k+(i-1)m})$ is the conditional density function of $r^{(j)}_{k+(i-1)m}$ given $V^{(j)}_{k+(i-1)m}$.

Normalize $w_i^{(j)}$ such that $\sum_{j=1}^{M} w_i^{(j)} = 1$.

Add $W_i^{(j)}$ to the sequence $\mathbf{W}^{(j)}_{1:i-1}$, set $\tilde{\mathbf{W}}^{(j)}_{1:i} = (\mathbf{W}^{(j)}_{1:i-1}, W_i^{(j)})$.

Step 3. Resample with replacement M particles $\{\mathbf{W}^{(j)}_{1:i}, j = 1, .., M\}$ from $\{\tilde{\mathbf{W}}^{(j)}_{1:i}, j = 1, .., M\}$ with weights $\{w_i^{(j)}, j = 1, .., M\}$.

Set $V_t^{(j)} = W^{(j)}_{[(t-1)/m]}$ for $t = 1, .., im$, $j = 1, ..., M$.

Set $i = i + 1$ and go to step 2.

3.3 Data and results

In this section we implement the method explicated above for estimation of the new model. Historical data of eight indices cross different countries, which include the Standard and Poor 500 index (SPX), Dow-Jones Industrial Average Index (DJIA), National Association of Securities Dealers Automated Quotations Index (NASDAQ), Nikkei 225 Stock market index (NIKKI), Financial Time Security Exchange Index (FTSE), Deutscher Aktien IndeX (DAX 30), Hang Seng Index (HSI) and Cotation Assistee en Continu Index (CAC 40), are taken as inputs from the period between January 1, 2001 and September 30, 2006, inclusive.

The estimation results for models with and without leverage effects are, respectively, shown in Table 3.1 and Table 3.2. We note that the leverage effect component significantly improves the fitting performance, which is shown by the fact that the maximized likelihoods increase by 71.9, 90.8, 31.3, 67.1, 69.1,72.8, 69.3 and 134.6,

respectively, for each index after adding the leverage effect component. Secondly, the skew parameter θ remains negative for all indices with and without the leverage effect. This finding implies that the leverage effect cannot explain by itself the skewness in return data and therefore the non-zero skewness in the distribution of innovation is necessary. Finally, the stochastic volatilities show different levels of persistence for each index. Table 3.3 summarizes the decay rate as a measure of persistence for each index.

Another observation is that the latent volatility processes exhibit strong similarity across different indices, as illustrated in Figure 3.1. Carrying out a principal component analysis on the covariance matrix of the filtered volatility processes to measure the similarity, we find out the first three principal components explain 94% of overall variation. This finding indicates the possibility of using a multifactor model to link the dynamics of different assets' volatility processes, which is further explored in the next chapter.

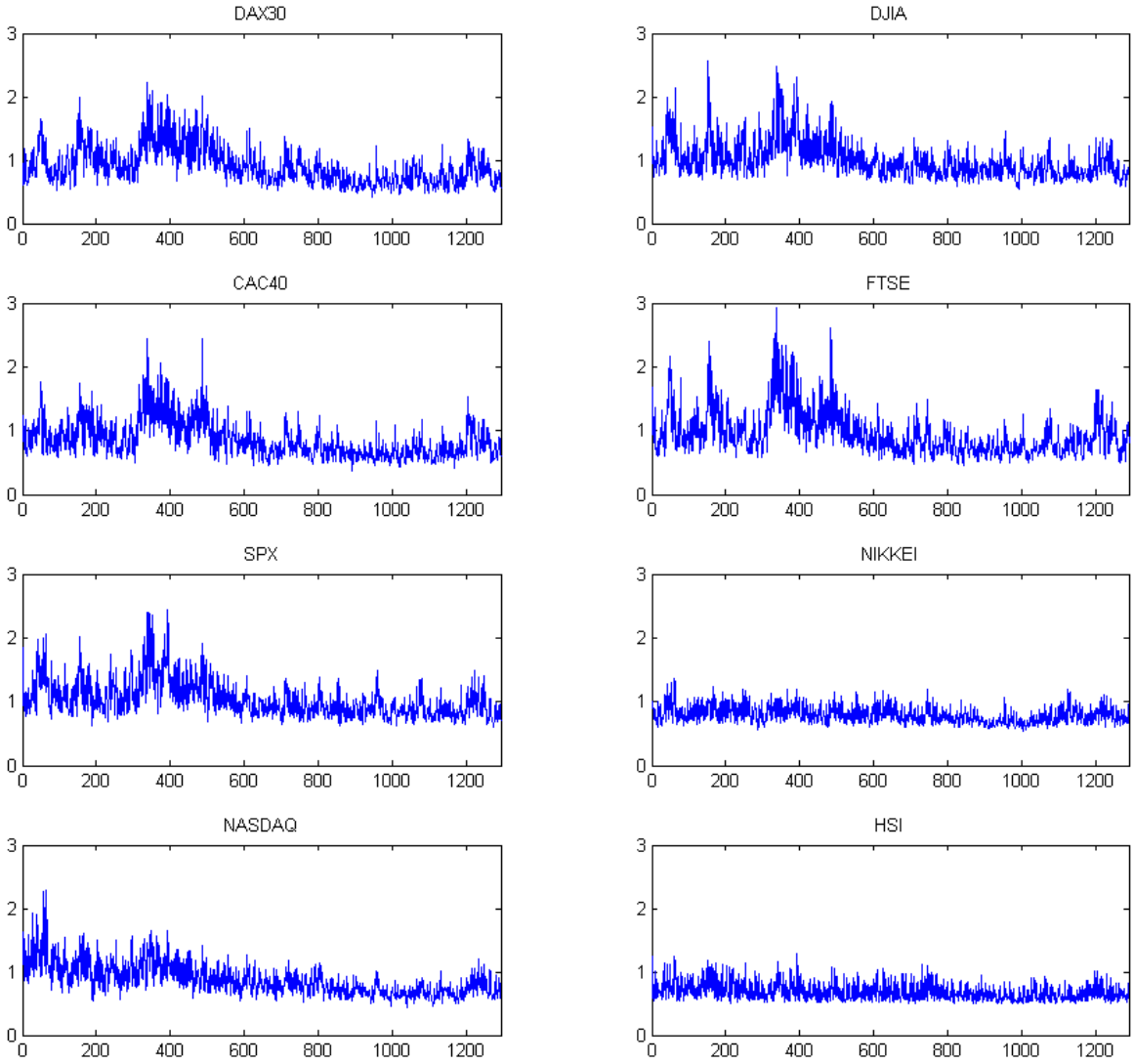

Figure 3.1: Mean of posterior distribution of latent volatility $f(V_t|r_t, \Theta)$ for eight global indices, plotted against time.

Table 3.1: Estimated parameters of SV model without leverage effect.

	DAX30	DJIA	CAC40	FTSE	SPX	NIKKEI	NASDAQ	HSI
μ	0.300	0.266	0.300	0.297	0.294	0.300	0.300	0.279
θ	-0.155	-0.136	-0.178	-0.154	-0.176	-0.158	-0.17	-0.146
θ	-0.155	-0.136	-0.178	-0.154	-0.176	-0.158	-0.170	-0.146
σ	0.998	0.999	0.999	0.999	0.999	0.999	0.999	0.999
ν	0.166	0.108	0.063	0.084	0.065	0.080	0.069	0.094
σ_0	0.235	0.172	0.205	0.159	0.160	0.210	0.238	0.176
λ	8.048	2.114	2.000	6.533	7.813	2.000	5.052	2.000
c	1.262	1.000	1.000	1.000	4.900	7.314	1.685	1.197
γ	0.100	1.736	0.100	0.457	0.754	8.894	9.692	1.021
$-\log L$	4,492.0	4,955.5	4,706.1	5,011.7	4,917.5	4,340.0	4,347.2	4,723.0

*Note:*Estimation is based on assumption of no leverage effect,i.e., $\alpha = \rho = 0$. Here μ is drift parameter, θ, σ and ν are VG parameters, σ_0 is average level of volatility and λ, d, γ and c are double-gamma stochastic volatility parameters with average level one. When sampling volatility from the posterior distribution we use the particle filter with $M = 2000$ paths.

3.4 Testing frequencies of volatility change

In this section we study the impact of different frequencies of changing volatility on the computational cost of model estimation. Different updating frequencies corresponds to different values of m in our model. For instance, $m = 1$ is for daily change and $m = 20$ is for monthly change. Following a similar procedure to that of the last section, we estimate the model for different frequencies of volatility change, and also that of constant volatility (CV). The results are shown in Table 3.4:

Table (3.4) shows that higher frequency of volatility change always leads to higher likelihoods as the granularity and finesse of volatility process increase. The cost of overfitting the model, however, is apparently also expensive. As the model estimation intensively depends on simulation, overestimating the frequency of volatil-

Table 3.2: Estimated parameters of SV model with leverage effect.

	DAX30	DJIA	CAC40	FTSE	SPX	NIKKEI	NASDAQ	HSI
μ	0.102	0.118	0.150	0.135	0.095	0.096	0.137	0.076
θ	-0.153	-0.177	-0.184	-0.183	-0.168	-0.164	-0.164	-0.168
σ	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999
ν	0.128	0.096	0.089	0.089	0.1063	0.112	0.112	0.106
σ_0	0.175	0.133	0.225	0.151	0.1273	0.168	0.180	0.140
α	-0.100	-0.080	-0.700	-0.100	-0.135	-0.599	-0.014	-1.305
ρ	119.1	113.1	133.8	113.1	118.5	680.0	394.8	650.0
λ	2.005	3.469	2.000	4.604	10.66	6.625	2.000	10.51
d	7.117	5.613	6.000	4.969	13.71	7.806	2.101	13.04
γ	6.301	1.053	1.000	1.006	3.895	3.625	1.010	3.50
c	1.233	0.491	0.250	2.754	1.277	3.070	10.00	1.380
$-\log L$	4,420.1	4,864.7	4,614.8	4,944.6	4,848.4	4,267.2	4,277.9	4,588.4

Note: Here α and ρ are additional leverage effect parameters and $d = \lambda + \gamma/c$.

Table 3.3: Persistence of stochastic volatility.

	DAX30	DJIA	CAC40	FTSE	SPX	NIKKEI	NASDAQ	HSI
λ/d	0.770	0.785	0.461	0.962	0.952	0.718	0.333	0.477
τ	0.113	0.105	0.336	0.017	0.021	0.144	0.477	0.322
$K_{0.5}$	2.655	2.867	0.895	17.990	14.207	2.091	0.631	0.936

Note: The decay rate $\tau = -\log(\lambda/d)$ also reflects the level of persistence. $K_{0.5} = -1/\tau \log 0.5$ is half-life (in days) for decaying volatility.

ity change will significantly and unnecessarily increase the time spent in simulating latent volatility paths.

To choose the optimal frequency of volatility change we can employ model selection methods that balance between fitting and controlling the numbers of parameters. In the next section we will show this experiment using two representative methods: Akaike Information Criterion and Likelihood Ratio test.

Table 3.4: Negative maximized log-likelihoods for different frequencies and eight indices, with leverage effect component.

m	DAX30	DJIA	CAC40	FTSE	SPX	NIKKEI	NASDAQ	HSI
1	4,420.1	4,864.7	4,614.8	4,944.6	4,848.4	4,267.2	4,277.9	4,588.4
5	4,930.3	5,338.6	5,130.6	5,454.2	5,362.2	4,783.6	4,816.4	5,057.2
10	4,979.8	5,364.6	5,172.0	5,498.0	5,389.0	4,825.8	4,851.8	5,102.5
20	4,992.2	5,399.8	5,198.3	5,538.6	5,416.1	4,848.9	4,887.7	5,135.8
40	5,015.8	5,429.7	5,209.3	5,557.8	5,441.7	4,859.1	4,927.2	5,158.9
60	5,049.5	5,451.6	5,237.5	5,574.9	5,462.3	4,877.7	4,943.6	5,181.4
120	5,063.4	5,467.8	5,252.1	5,590.1	5,479.8	4,901.7	4,974.6	5,201.1
240	5,114.1	5,493.9	5,302.0	5,650.9	5,508.3	4,923.1	4,996.0	5,228.1
CV	5,447.7	5,668.4	5,578.1	5,897.6	5,737.2	5,047.3	5,382.3	5,346.0

3.4.1 Akaike information criterion

The Akaike information criterion (AIC) is a measure of the goodness of fit of of an estimated model. It is defined as

$$AIC = -2 \log L + 2p, \tag{3.23}$$

where L is the likelihood and p is the total number of parameters, which is determined by:

$$p = p_1 + [T/m] - p_2 \tag{3.24}$$

where p_1 is the number of parameters in the observation equation and p_2 is that of the latent equation.

We can test the performance of AIC in detecting real frequencies by looking at the Type I error and Type II errors. For each of the two models with $m = 1$ and $m = 10$, this method is tested against 1,000 simulated paths of 500-day returns. The parameters for these two models are estimated for each of the overall 2,000 return paths. Then on each path we choose the model and the corresponding m value with

Table 3.5: Testing power of AIC.

		Actual m	
		m=1	m=10
Test result	m=1	0.863	0.137
	m=10	0.101	0.899

the lowest AIC. In this way we approximate Type I and Type II error by computing the ratio of false rejection and false acceptance respectively.

The test result are shown in Figure 3.2. We find that for all eight indices the AIC changes with the same pattern. There's a noticeable drop of AIC when the model moves from constant volatility to stochastic volatility where volatility changes every 240 days. This trend continues, however with a decreasing speed, until it reaches a minimum around $m = 20$ or 40, which corresponds to monthly or bimonthly change.

3.4.2 Likelihood ratio test

The likelihood ratio is defined as

$$\lambda(\mathbf{r}) = \frac{\sup_{\Theta_0} L(\theta|\mathbf{r})}{\sup_{\Theta} L(\theta|\mathbf{r})}, \tag{3.25}$$

where Θ_0 and Θ are parameter spaces corresponding to the null hypothesis and alternative hypothesis and $\mathbf{r}$ is the observed data . The quantities $\sup_{\Theta_0} L(\theta|\mathbf{r})$ and $\sup_{\Theta} L(\theta|\mathbf{r})$ are two supreme of $L(\theta|\mathbf{r})$ over Θ_0 and Θ. The likelihood ratio test is defined by the rejection region

$$\{\mathbf{r} : -2\log\lambda(\mathbf{r}) \geq \chi^2_\alpha(p)\}, \tag{3.26}$$

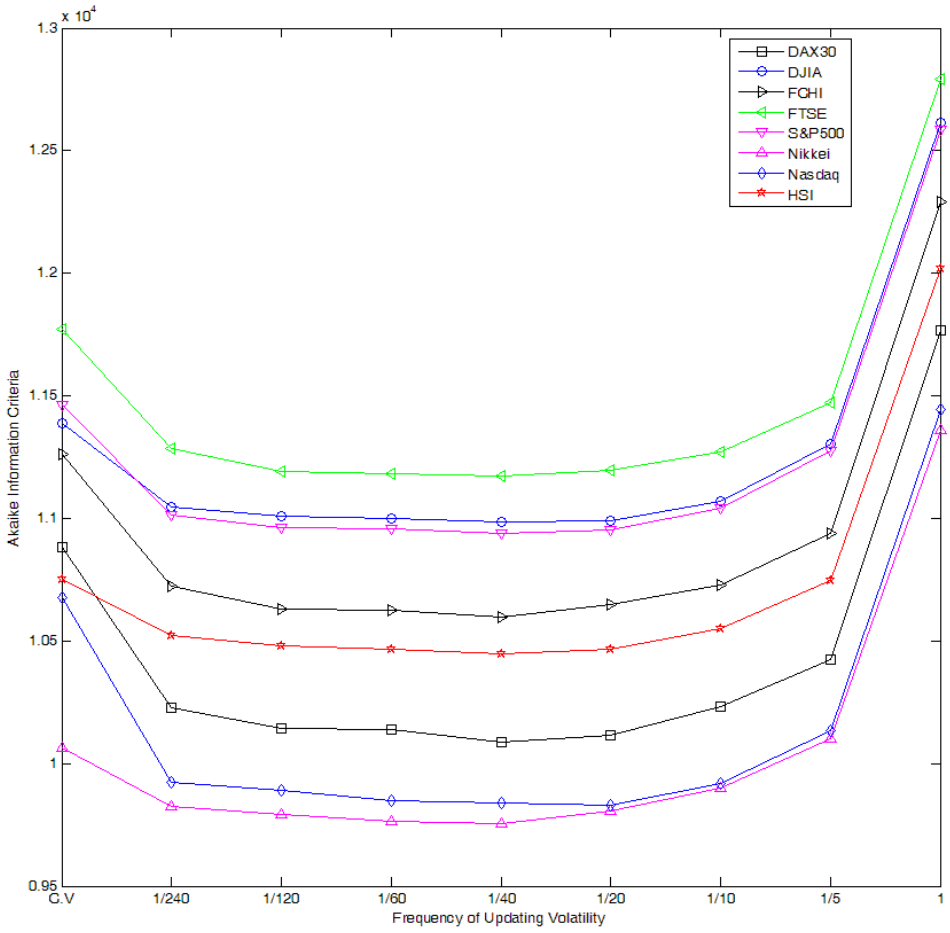

Figure 3.2: AIC for different frequencies of volatility change(times/day).

Table 3.6: Testing Power of Likelihood Ratio Tests.

m: H0 vs H1	Type I Error	Type II Error
10 vs 5	0.061	0.144
20 vs 10	0.073	0.137
40 vs 20	0.085	0.124
60 vs 40	0.091	0.110
120 vs 60	0.109	0.106
240 vs 120	0.117	0.083
CV vs 240	0.129	0.075

Note: Hypotheses are tested against 1,000 simulated paths of 500-day data. Estimations with different m values are run for each of the 2,000 returns paths. On each path we accept or reject the hypothesises according to rejection region (3.26). Type I and Type II error are approximated by computing the ratio of false rejection and false acceptance respectively.

where $\chi^2_\alpha(p)$ is the cutoff point of a χ^2 distribution with p degrees of freedom and p equals the difference between the number of free parameters specified by $\theta \in \Theta_0$ and the number of free parameters specified by $\theta \in \Theta$.

Barndorff-Nielsen and Shephard (2004) and Bos and Shephard (2006) used likelihood ratio test in selecting different granularities of volatility changes. Here to examine the testing power we carry out the same simulation based Type I and Type II calculation as in last section. The result, shown in Table 3.6, suggests that using Likelihood Ratio Test, Type I error increases and Type II error decreases as the value of m increases.

The results of log-likelihood tests are shown in Table 3.6. From the table we can see that models with lower frequencies are rejected until reach the optimal frequency. Afterward the increase in log-likelihood cannot compensate the increased

Table 3.7: Stepwise Likelihood ratio tests with asymptotic size $\alpha = 0.05$.

m: H0 vs H1	DAX30	DJIA	FCHI	FTSE	SPX	NIKKEI	NASDAQ	HSI
10 vs 5	98.9	52.1	82.7	87.7	53.6	84.3	70.7	90.5
	(175.2)	(173.0)	(176.3)	(174.1)	(173.0)	(169.7)	(173.0)	(170.8)
20 vs 10	26.4	52.4	69.7	79.2	81.4	50.7	35.3	61.8
	(93.9)	(92.8)	(95.1)	(93.9)	(92.8)	(91.7)	(92.8)	(91.7)
40 vs 20	48.7	61.9	20.9	58.4	51.0	19.3	85.7	54.5
	(52.2)	(51.0)	(52.2)	(51.0)	(51.0)	(49.8)	(51.0)	(51.0)
60 vs 40	68.3	40.7	52.2	31.9	31.4	34.2	31.4	54.5
	(21.1)	(21.1)	(21.1)	(21.1)	(21.1)	(21.1)	(21.1)	(21.1)
120 vs 60	22.9	29.8	30.8	35.4	31.6	51.5	69.4	37.8
	(21.1)	(21.1)	(21.1)	(21.1)	(21.1)	(21.1)	(21.1)	(21.1)
240 vs 120	92.2	51.6	85.3	105.2	56.1	42.6	51.1	55.7
	(12.6)	(12.6)	(12.6)	(12.6)	(12.6)	(12.6)	(12.6)	(12.6)
CV vs 240	665.4	348.6	554.1	497.6	460.8	248.5	766.3	236.1
	(12.6)	(12.6)	(12.6)	(12.6)	(12.6)	(11.1)	(12.6)	(11.1)
m^*	40	20	40	20	20	40	20	20

*Note:*The first column indicates the stepwise hypothesis test for models with different frequencies (m values). Numbers in other columns without parentheses are $-2\log\lambda(\mathbf{r})$. Numbers with parentheses are the corresponding cutoff value of chi-square distribution $\chi^2_\alpha(p)$. The last row gives the optimal m value for each index. The null hypothesis (lower frequency, greater m) is rejected if the likelihood ratio is greater than the cutoff chi-square value.

complexity of the model.

Computational cost for estimation varies significantly for models with different frequencies. For the daily-changing volatility model, it takes 12.5 seconds on a PC with Intel Core 2 Duo 2.13GHz processor and 2GB RAM to sample 2000 paths from the posterior distribution of volatility using particle filter algorithm. For a model with volatility changing every 10 days the sampling costs 1.45 seconds on the same machine. If the frequency decreases to once every 20 days, the cost in each sampling reduces to 0.87 seconds. This substantial reduction in computation cost

on each sampling, multiplied by the number of iterations in optimization, leads to a significant improvement of computational efficiency of estimation and application of the model.

This conclusion has important implication for our practice in developing and applying stochastic volatility model. The latent volatility process is unobservable and thus a pure abstract concept, and most existing stochastic volatility models including ours are built for serving specific practical purposes instead of to establish unchallengeable latent structure of the volatility processes. For instance GARCH type models are applied primarily in the field of risk management to evaluate the exposure to volatility changes. Continuous stochastic volatility models including those by Heston (1993), Bates (1996) and Carr et al. (2003) are used mainly for interpolating market traded vanilla option prices cross maturities and strikes. For our case, the purpose is to evaluate and optimize portfolio of volatility contracts that cannot be handled well by any existing volatility models. Thus although the method described above does not determine the exact dynamic of volatility process, it gives us a superior model judged jointly by fitting performance, computational efficiency and practical applicability.

Chapter 4

Application to Pricing Volatility Derivatives

In this chapter we will discuss how to price volatility derivatives by combining the estimated stochastic volatility model with the concave distortion of distribution functions introduced by Cherny and Madan (2007) and Cherny and Madan (2006).

The essential idea is that we could compute the price of a certain financial asset as the minimum value that can sufficiently compensate the taken risk to an acceptable level. The level of acceptability is measured by one coherent performance measure that can be interpreted as the expected cash flow in an extreme scenario. The payoff distribution under the extreme scenario can be derived from the model estimated from the historical data.

This chapter starts with a brief review of two volatility products: volatility swaps and options. In the second part of this chapter we present the concept of coherent risk measures and the associated concave transformations of distribution functions, and then introduce a new method of pricing volatility products by using the combination of historical distribution and the concave distortions. In the third part we will show the results of the application.

4.1 Volatility swaps and options

Volatility derivatives cover a group of financial products whose values are directly determined by the level of volatility. The pricing methodology introduced in this chapter is initialized by, but not limited to, the study on pricing of the two fundamental volatility derivatives: volatility swaps and volatility options.

Traditional volatility investment strategies include taking a delta-hedged position in options, buying a straddle or a straggle, and buying a volatility swap or option. A delta-hedged position in options removes the exposure to the stock price. However, the accuracy of delta-hedging depends on the soundness of the employed models which might be far from reality. Moreover, other issues such as transaction costs and liquidity concern often attenuate the feasibility of such investment strategies.

In contrast, since volatility or variance swaps are directly written on the volatility level, they provide much cleaner exposure to volatility than other strategies. Applications of this new category of volatility derivatives include speculation on future volatility levels, trading on the spread between realized and implied volatility, or hedging the volatility exposure of other positions. For example, if an investor foresees a rapid decline in political and financial turmoil after a forthcoming election, a short position in volatility might be appropriate. Another example is that hedge funds or risk arbitrageurs might want to short volatility since they often take positions which assume that the spread between stocks of companies planning mergers will narrow. If overall market volatility increases, the merger may become less likely

and the spread may widen and cause loss, but a short position on volatility will provide positive cash flow to smooth the total payoff.

The first volatility swap trade was done in 1997 and was written on FTSE. The participants in this new market initially were a small number of very quantitative hedge funds, but since 2005 the volume of trade increased dramatically and now the market has been extended to new users like insurance companies, asset managers, pension funds and individual investors.

The realized volatility σ_R of an asset over some period is the annualized standard deviation of the daily returns. Given observations of stock prices S_i, $i = 1, ..., T$, σ_R is defined in the following equations:

$$\begin{aligned} y_i &= \log \frac{S_i}{S_{i-1}}, \\ \sigma_R^2 &= \frac{252}{T-1} \sum_{i=1}^{T} y_i^2. \end{aligned} \tag{4.1}$$

The annualized realized variance does not coincide with the classic statistical definition of variance, but follows the usual market convention of not subtracting the mean.

4.1.1 Variance swap and volatility swap

The payoff P of a volatility swap at maturity depends on the difference between the realized volatility σ_R during the contract period and a predetermined constant value K_{vol},

$$P = N(\sigma_R - K_{vol}), \tag{4.2}$$

where K_{vol} is the annualized volatility delivery price and N is the notional amount[1] of the swap in dollars per annualized volatility point. The holder of a volatility swap at expiration receives N dollars for every unit, usually in percentage, that the stock's realized volatility σ_R exceeds the volatility delivery price K_{vol}.

The delivery price K_{vol} is typically quoted as a volatility and the notional amount is typically quoted in dollars per volatility point. One example is a swap contract with $K_{vol} = \$35\%$ and $N = \$250,000/\%$. If at the expiry of the contract, the realized volatility $\sigma_R = 36\%$, then the final payoff equals to $\$250,000$.

Variance swap is a similar forward contract written on future realized variance of return. The payoff of variance swap at maturity equal to

$$P = N(\sigma_R^2 - K_{var}), \tag{4.3}$$

where K_{var} is delivery price for variance, N is the notional amount in units of a certain currency per variance point, and σ_R^2 is the realized variance over the life of the contract. Compared to volatility swaps, variance swaps are less commonly traded, and therefore lack a general quoting convention. The fair value of variance is the delivery price that sets the expectation of the final payoff to zero:

$$K_{var} = E(\sigma_R^2). \tag{4.4}$$

Similarly we have

$$K_{vol} = E(\sigma_R). \tag{4.5}$$

[1]The notional amount (or notional principal amount or notional value) on a financial instrument is the nominal or face amount that is used to calculate payments made on that instrument. This amount generally does not change hands and is thus referred to as notional.

4.1.2 Option on realized variance

Option on realized variance, which is also called variance swaption[2] is a financial contract with payoff contingently depending on the realized variance σ^2_{vol} and a specified strike level K_{var}. For example, the payoff function of a variance call option (also named "call swaption") is

$$P = N(\sigma^2_R - K_{var})^+. \tag{4.6}$$

Assuming the current price of this variance call option is C, the cash flow to the owner when the contract expires equals to

$$X = Ce^{rT} - N(\sigma^2_R - K_{var})^+, \tag{4.7}$$

where r is the interest rate and T is the time to the expiration day.

4.1.3 Pricing

As volatility contract is still relatively illiquid, they are traded with a relatively wide bid-ask spread and the price for trading largely depends on the direction of trade. To solve this problem we apply a method proposed by Cherny and Madan (2007), which sets the price of volatility contracts to a value that locates the corresponding cash flow X inside an acceptable range $\mathcal{A}_a$. Here X is a bounded random variable defined on a standard probability space $(\Omega, \mathcal{F}, P)$, $X \in L^\infty$ and $L^\infty = L^\infty(\Omega, \mathcal{F}, P)$. The acceptable range $\mathcal{A}_a$ is then the set of cashflows whose

[2]Swaption: option on swap, a financial contract giving the buyer the right but not the obligation to buy or sell swap contract when the payoff is favorable when the option contract expires.

acceptability measure $\alpha(X)$, a real valued map from L^∞ to $[0, \infty)$, is above certain level a,

$$\mathcal{A}_a = \{X : \alpha(X) \geq a \quad a \in \mathbb{R}^+\}. \tag{4.8}$$

By investigating the levels of acceptability available in the economy across a broad set of activities, one can identify acceptability levels that are good target values. This knowledge may be combined with a historical distribution to price a contingent claim with a view to attaining a predetermined target level of acceptability.

In the following sections we will review the concept of coherent acceptability index with which Cherny and Madan (2007) proposed a set of measures to define the cone of acceptability of investments.

4.2 Coherent Acceptability index $\alpha(X)$

4.2.1 Coherent measures

There are eight properties that the acceptability index $\alpha(a)$ should satisfy. The first four properties define a coherent acceptability index. The remaining properties are additional properties that enable us to provide a number of tractable examples of coherent acceptability indices.

Definition 3 *A function of X, $\alpha(X)$ is a coherent acceptability index if it satisfies the following four properties:*

1. **Convexity.** This condition requires the function $\alpha(X)$ to satisfy the condition that if $\alpha(X) \geq a$ and $\alpha(Y) \geq a$, then $\alpha(\lambda X + (1-\lambda)Y) \geq a$ for any $\lambda \in [0,1]$.

2. **Monotonicity.** This is the condition that if X is acceptable and Y dominates X as a random variable, then Y is acceptable. Equivalently, we require that if $X \leq Y$, then $\alpha(X) \leq \alpha(Y)$.

3. **Scale invariance.** It's required that $\alpha(\lambda X) = \alpha(X)$ for a constant $\lambda > 0$.

4. **Fatou property.** For any countable collection of random variables X_n with $|X_n| \leq 1$ such that $\alpha(X_n) \geq a$, we require that if X_n converges to X in probability, then $\alpha(X) \geq a$.

In addition there are four other properties useful in constructing practical examples of coherent acceptability indices.

1. **Law invariance** This property requires that: if $X \overset{law}{=} Y$, then $\alpha(X) = \alpha(Y)$

2. **Second order monotonicity.** For the index to be consistent with expected utility, we must have that if $X \preccurlyeq^2 Y$, then $\alpha(X) \leq \alpha(Y)$ where $X \preccurlyeq^2 Y$ means that Y dominates X in the second order[3] which can be met if $E[f(X)] \leq E[f(Y)]$ for any increasing concave function f.

3. **Arbitrage consistency.** We require the index is infinity for arbitrages or equivalently that $\alpha(X) = \infty$ if and only if $X \geq 0$.

[3]Distribution $F(x)$ dominates $G(x)$ in the second order, or $F \geq_2 G$, if $T(x) = \int_a^x [G(t) - F(t)]dt \geq 0$ for all $x \in [a,b]$.(Please refer to Bawa (1975).)

4. **Expectation consistency** We relate the positivity of the index to the positivity of the mean and require that if $E(X) \leq 0$, then $\alpha(X) = 0$; if $E(X) > 0$, then $\alpha(X) > 0$.

Denote $\mathcal{P}$ as the set of probability measures absolutely continuous with respect to P, for each positive real a we can define an acceptable set $\mathcal{A}_a$ of random variables associated with test measures $\mathcal{D}_a \subset \mathcal{P}$ such that $X \in \mathcal{A}_a$ if and only if for every measure $Q \in \mathcal{D}_a$ we have that $E^Q(X) \geq 0$. We make the cones of acceptability $\mathcal{A}_a$ smaller and converging to the positive orthant as we raise the level a by increasing the sizes of the set of test measures $(\mathcal{D}_a)_{a\in\mathbb{R}}$. The acceptability index $\alpha(X)$ is then defined by the largest level of a with $X \in \mathcal{A}_a$,

$$\alpha(X) = \inf\{a \in \mathbb{R}_+ : \inf_{Q\in\mathcal{D}_a} E^Q[X] < 0\}. \tag{4.9}$$

Two acceptability indices, Sharpe Ratio and Value at Risk, are widely used in practice to measure risk. However, neither of these two satisfy all the required properties of a coherent measure.

Specifically, it is well known that the Sharpe Ratio $S(X) = E(X)/\sigma(X)$ does not have the monotonicity property and is also not consistent with arbitrage. One example is that by using the Sharpe Ratio as benchmark to measure the acceptability of an investment we will reject an arbitrage opportunity that has a positive payoff but also a variance converging to infinity and making the Sharpe Ratio converge to zero.

Value at Risk (VAR) is defined as the negative of the quantile q_α^+ of the final net flow X,

$$VAR_\alpha(X) = -\inf\{a|\mathbb{P}(X \leq a) > \alpha\}. \tag{4.10}$$

While VAR satisfies the requirements of monotonicity, scale invariance and Fatou property, it fails to satisfy the convexity or subadditivity properties.

Other acceptability indices, based on risk measures such as Gain-Loss ratio, Tilt Coefficient and Tail VAR, provide remedies for some of these problems but still fail to qualify as coherent and economically sensible measures of risk. For instance, the Gain-Loss ratio is defined as

$$GLR(X) = \begin{cases} \frac{E(X^+)}{E(X^-)} - 1, & \text{if } E(X) > 0; \\ 0, & \text{otherwise.} \end{cases}$$

The associated acceptability index satisfies convexity, law and scale invariance, Fatou property arbitrage and expectation consistency, and it's monotonic and therefore it's a coherent measure. However Gain-Loss ratio doesn't have an explicit analytic formula for the state-price density [4], which is critical in practice. In addition, as pointed out by Cherny and Madan (2007), the state-price density treats small losses

[4]State price density is often called the risk-neutral density and is related to Arrow-Debreu securities, which is elementary securities each paying one dollar in one specific state of nature and nothing in any other state. In a continuum of states, state-price density defines the prices of Arrow-Debreu securities by giving each state x the price of a security paying one dollar if the state falls between x and $x + dx$. Ross (1976) and Cox and Ross (1976) first observed that the Black-Scholes formula can be derived by assuming that all investors are risk neutral and all assets in such a world-including options-must yield an expected return equal to the risk-free rate of interest.

and large losses symmetrically and hence exaggerates losses uniformly.

The motivation for incorporating an exaggeration of losses relative to a simultaneous deflation of gains is rooted in economic principles advocating weightings proportional to marginal utility with a similar effect. This led to the work of Artzner et al. (1997) and Carr et al. (2001) on the tilt coefficient, which conceptualizes acceptable risks in terms of a convex cone or set containing the positive orthant.

The tilt coefficient (TC) for the random cash flow X is the highest level of absolute risk aversion for exponential utility such that the cash flow is still attractive to such a utility at the margin:

$$TC(X) = \inf\{\lambda \in \mathbb{R}_+ : E(Xe^{-\lambda X}) < 0\}. \tag{4.11}$$

TC satisfies the requirement of monotonicity, law invariance and the Fatou property. It is arbitrage and expectation consistent, but it is not convex or scale invariant.

4.2.2 Acceptability Index Based on TVAR $AIT(X)$

Tail Value at risk or TVAR is a coherent risk measure defined based on a coherent utility measure $u^\lambda(X)$:

$$TVAR(X) = -u^\lambda(X), \tag{4.12}$$

$$u^\lambda(X) = \inf_{Q \in \mathcal{D}^\lambda} E^Q[X], \tag{4.13}$$

where $\mathcal{D}^\lambda$ is the set of probability measures absolutely continuous with respect to P such that $dQ/dP \leq \lambda^{-1}$, where $\lambda \in (0, 1]$ is a parameter. In particular, if X has a

continuous distribution, then the above infimum is attained at the measure $Q^*(X)$ with

$$\frac{dQ^*(X)}{dP} = \lambda^{-1}\mathbf{1}_{X\leq q^\lambda(X)},$$

where q^λ denotes the λ-quantile of X. From this it is clear that $u^\lambda(X) = E[X|X \leq q^\lambda(X)]$.

The acceptability index based on TVAR is AIT and is defined by

$$AIT(X) = (\inf\{\lambda \in (0,1] : u^\lambda(X) \geq 0\})^{-1} - 1. \tag{4.14}$$

As the map $\lambda \to u^\lambda(X)$ is continuous in λ, we have that $AIT(X) \geq a$ if and only if $u^{\frac{1}{a+1}}(X) \geq 0$, and hence, AIT has the convexity property. It is also monotone, scale invariant and has the Fatou property. The law invariance of $TVAR$ is obvious. if X has a continuous distribution. Second order monotonicity is also inherited from the same property of u^λ.

Arbitrage and expectation consistency properties follow from the relations

$$\lim_{\lambda\downarrow 0} u^\lambda(X) = \text{essinf}_\omega X(\omega) = \sup\{c \in \mathbb{R} : X \geq c \quad \text{a.s.}\},$$

$$\lim_{\lambda\uparrow 1} u^\lambda(X) = E[X].$$

for X with a continuous distribution, the extreme measures or coherent state-price densities are given by

$$\frac{dQ_a^*(X)}{dP} = (a+1)\mathbf{1}_{X\leq q^{1/(a+1)}(X)}, \quad a \in \mathbb{R}_+.$$

These measures are even more extreme than those supporting GLR. Here by employing measures that are zero for gains and uniform with respect to the size

of losses, the gains are ignored. From the perspectives of economics considerations these are unreasonable measures and probably more so than those associated with GLR. this problem can be overcome by the AIW index proposed by Cherny and Madan (2007).

4.2.3 Weighted VAR and Acceptability Indices $AIW(X)$

Weighted VAR is a coherent risk measure defined as $WVAR(X) = -U(X)$, where

$$U(X) = \int_{(0,1]} u^{\lambda}(X)\mu(d\lambda), \tag{4.15}$$

μ is a probability measure on $(0, 1]$ and u^{λ} is the utility function defined in equation (4.13). To have a tractable expression for this risk measure Cherny and Madan (2007) introduced a concave distribution function $\Psi(y)$ on $(0, 1]$.

$$\Psi^{\mu}(y) = \int_0^y \int_{(0,z]} \lambda^{-1}\mu(d\lambda)dz, \quad y \in (0, 1]. \tag{4.16}$$

We have the map $\mu \to \Psi^{\mu}$ between probability measures μ on $(0, 1]$ and concave distribution functions Ψ^{μ} on $(0, 1]$. Differentiation of Ψ^{μ} establishes that it is monotone and concave. Since it can be easily proved that $\Psi^{\mu}(1) = 1$ and $\Psi^{\mu}(0) = 0$, it's a probability measure on $(0, 1]$. The inverse map is given by $\mu(dy) = -y(\Psi(y))''(dy)$, where $\Psi(y)$ is second derivative in the sense of distributions.

Follmer and Schied (2004) established

$$U(X) = \int_{\mathbb{R}} yd(\Psi^{\mu}(F_X(y)), \tag{4.17}$$

where F_X is the distribution function of X. The right-hand side of (4.17) is the expectation of a random variable X with $\Psi^\mu(F_X)$ being the distribution function of X. The representation (4.17) provides a convenient way to construct particular examples of the class $WVAR$. The coherent utility $U(X)$ associated with weighted value at risk can be simply computed by taking expectations under a concave distortion with the concave distortion specified by the function Ψ^μ.

The function $U(X)$ is a coherent utility with all the required properties inherited from u^λ. According to Cherny (2006), its determining set is given by

$$\mathcal{D}^\mu = \{Z : Z \geq 0, E(Z) = 1, \text{and } E((Z-y)^+) \leq \Phi^\mu(y) \quad \forall y \in \mathbb{R}^+\}, \tag{4.18}$$

where Φ^μ is the convex conjugate of Ψ^μ:

$$\Phi^\mu(y) = \sup_{z\in[0,1]} (\Psi^\mu(z) - yz), \quad y \in \mathbb{R}^+. \tag{4.19}$$

For a random variable X with continuous distribution, the minimum of expectations $E^Q(X)$ over $Q \in \mathcal{D}^\mu$ is attained at the measure $Q^*(X)$ that is given by

$$\frac{dQ^*(X)}{dP} = (\Psi^\mu)'_-(F_X(X)), \tag{4.20}$$

where $(\Psi^\mu)'_-$ is the left-hand derivative.

Representation (4.17) allows us to define the $WVAR$ acceptability index AIW by

$$AIW(X) = \inf\{a \in \mathbb{R}^+ : \int_{\mathbb{R}} y d(\Psi_a(F_X(y))) < 0\}, \tag{4.21}$$

where $(\Psi_a)_{a\in\mathbb{R}^+}$ is a collection of concave distribution functions on $[0,1]$ that satisfy $\Psi_{a_1}(y) \leq \Psi_{a_2}(y)$ for all y and $a_1 \leq a_2$.

As the mapping $a \mapsto \int_{\mathbb{R}} yd(\Psi_a^-(F_X(y)))$ is left-continuous, we may then write

$$AIW(X) \geq a \Leftrightarrow \int_{\mathbb{R}} yd(\Psi_a(F_X(y))) \geq 0 \tag{4.22}$$

from which it is clear that *AIW* satisfies convexity, the Fatou property, monotonicity, scale invariance, law invariance and second order monotonicity.

The notation

$$u_a(X) = \int_{\mathbb{R}} yd\Psi_a(F_X(y)) \tag{4.23}$$

is employed for the series of coherent utilities associated with the index, which will be discussed in the next subsections.

Four special examples of $WVAR$, namely, *MINVAR, MAXVAR, MINMAXVAR, MAXMINVAR* and also the corresponding acceptability indices, *AIMINVAR, AIMAXVAR, AIMINMAXVAR, AIMAXMINVAR* are described. The extreme measures of this set form the counterpart of state-price densities in classical expected utility theory and may be viewed as coherent analogs of the tilt coefficient and thereby are more closely related to the intuitions embedded in classical economics.

4.2.3.1 *MINVAR* and Acceptability Indices *AIMIN(X)*

MINVAR is associated with a concave distortion

$$\Psi_a(y) = 1 - (1-y)^{a+1}, \quad a \in \mathbb{R}^+, \quad y \in [0,1]. \tag{4.24}$$

For integer a we have $u_a(X) = E(Y)$, where

$$Y \overset{law}{=} \min\{X_1, ..., X_{a+1}\}, \tag{4.25}$$

and $X_1, ..., X_{a+1}$ are independent draws of X. *MINVAR* refers to the risk measure $\rho_a(X) = -\mu_a(X)$ and *AIMIN* is the associated acceptability index. It is actually the largest number a such that the expectation of the minimum of $a+1$ draws from the cash flow distribution is still positive.

For X with a continuous distribution, the state-price densities are given by

$$\frac{dQ_a^*(X)}{dP} = (a+1)(1 - F_X(X))^a, \quad a \in \mathbb{R}^+. \tag{4.26}$$

As Cherny and Madan (2007) pointed out, a potential drawback of *AIMIN* is that this density converges to 0 at $+\infty$ and a finite value $a+1$ at $-\infty$, the latter indicates the density exaggerates losses uniformly. However, from a economic perspectives we are accustomed to measure changes that exaggerate large losses more than small losses.

4.2.3.2 *MAXVAR* and Acceptability Indices *AIMAX(X)*

Another concave distortion is to consider

$$\Psi_a(y) = y^{\frac{1}{a+1}}, \quad a \in \mathbb{R}^+, \quad y \in [0,1]. \tag{4.27}$$

For integer a we have $u_a(X) = E(Y)$, where Y is a random variable with property

$$\max\{Y_1, ..., Y_{a+1}\} \overset{law}{=} X, \tag{4.28}$$

where $Y_1, ..., Y_{a+1}$ are independent draws of Y. *MAXVAR* is the risk measure $\rho_a(X) = -u_a(X)$ and *AIMAX* is the associated acceptability index .

Then the state-price densities of X is given by

$$\frac{dQ_a^*(X)}{dP} = \frac{1}{a+1}(F_X(X))^{-\frac{a}{a+1}}, \quad a \in \mathbb{R}^+. \tag{4.29}$$

This density tends to $+\infty$ at $-\infty$ but has another potential drawback: it tends to a strictly positive value $1/(a+1)$ at $+\infty$. This corresponds to an asymptotically linear utility for large gains. However, according to law of diminishing marginal utility, the marginal utility of returns for typical risk-aversive investor diminishes as X increases, hence the state-price densities should tend to zero at positive infinity.

4.2.3.3 *MAXMINVAR* and Acceptability Indices *AIMAXMIN(X)*

Combining *MINVAR* and *MAXVAR*, we consider the distortion

$$\Psi_a(y) = (1-(1-y)^{a+1})^{\frac{1}{a+1}}, \quad a \in \mathbb{R}^+, \quad y \in [0,1]. \tag{4.30}$$

For an integer a we have $u_a(X) = E(Y)$, where Y is a random variable with the property:

$$\max\{Y_1, ..., Y_{a+1}\} \overset{law}{=} \min\{X_1, ..., X_{a+1}\}, \tag{4.31}$$

where $X_1, ..., X_{a+1}$ are independent draws of X, and $Y_1, ..., Y_{a+1}$ are independent draws of Y. *MAXMINVAR* is the corresponding risk measure $\rho_a(X) = -u_a(X)$ and *AIMAXMIN* is the associated acceptability index. It is in recognition of the fact that we construct the worst case scenario first using a *MINVAR* perspective followed by a *MAXVAR* perspective.

For X with a continuous distribution, the state-price densities are given by

$$\frac{dQ_a^*(X)}{dP} = (1-F_X(X))^a(1-(1-F_X(X))^{a+1})^{-\frac{a}{a+1}}, \quad a \in \mathbb{R}^+. \tag{4.32}$$

In particular, this density tends to $+\infty$ at $-\infty$ and to 0 at $+\infty$.

4.2.3.4 *MINMAXVAR* and Acceptability Indices *AIMINMAX(X)*

Another way to combine *MINVAR* and *MAXVAR* is to consider

$$\Psi_a(y) = 1 - (1 - y^{\frac{1}{a+1}})^{a+1}, \quad a \in \mathbb{R}^+, \quad y \in [0,1]. \tag{4.33}$$

For integer a we have $u_a(X) = E(Y)$, where Y is a random variable with the property

$$\begin{aligned} Y &\overset{law}{=} \min\{Z_1, ..., Z_{a+1}\}, \\ \max\{Z_1, ..., Z_{a+1}\} &\overset{law}{=} X, \end{aligned} \tag{4.34}$$

and $Z_1, ..., Z_{a+1}$ are independent draws of Z. *MINMAXVAR* is the risk measure $\rho_a(X) = -u_a(X)$ and denote the associated acceptability index as *AIMINMAX* in recognition of the fact that we construct the worst case scenario first using a *MAXVAR* perspective followed by a *MINVAR* perspective.

For X with a continuous distribution, the state-price densities are given by

$$\frac{dQ_a^*(X)}{dP} = (1 - F_X(X)^{\frac{1}{a+1}})^a F_X(X)^{-\frac{a}{a+1}}, \quad a \in \mathbb{R}^+. \tag{4.35}$$

4.3 Pricing

Our method of pricing volatility products by combining historical distribution of return and concave distortions could be implemented as follows. Firstly based on the estimated model we simulate a number of price paths and then calculate the realized variance or volatility σ_R for each path using equation (4.1). Based

on the resulted empirical distribution of σ_R we then apply the concave distortion to generate the distributions of payoff under an extreme scenario, from which we could easily calculate the expectation of the payoff and set the target price as the acceptable ask price for volatility products.

4.3.1 Determination of Tilt Coefficient

Tilt Coefficient could be calculated by minimizing the overall difference between the observed market prices and the model prices. The exact implementation depends on the type of market contracts taken into calibrations. As an example, assume we have $N = n_T \cdot n_K$ European options across n_T different maturities T_i, $i = 1, ..., n_T$ and n_K strikes K_j, $j = 1, .., n_K$. Using the estimated model together with a tilt parameter a we have the model prices $P_a(T_i, K_j)$ for the same options. The tilt coefficient a^* is the value that minimizes sum of squared errors of pricing:

$$a^* = \text{argmin}_a g(a), \tag{4.36}$$

where $g(a) = \sum_{i,j}(P(T_i, K_j) - P_a(T_i, K_j))^2$ and $P(T_i, K_j)$ is the market quotes for the N contracts.

For different concave distortions we have different optimal Tilt Coefficients. In the following pricing section, we choose those parameters as follows.

Table 4.1: Optimal Tilt Coefficients.

Tilt Coefficient	MINVAR	MAXVAR	MAXMINVAR	MINMAXVAR
a	0.5	0.25	0.15	0.15

4.3.2 Pricing of variance and volatility swaps

Here we calculate the swap rate K_{vol} or K_{var} for which the volatility or variance swap has zero initial expected value. This is equivalent to finding the price level at which the expectation of future cash flow under the concavely distorted measure Q^*_α equals to zero. That is, for a volatility swap K_{vol} is the swap rate satisfying

$$\begin{aligned} Z &= K_{vol} - \sigma_R, \\ E^{Q^*_\alpha}(Z) &= \int_{\mathbb{R}} z d\Psi_\alpha(F_Z(z)) \geq 0, \end{aligned} \tag{4.37}$$

and for a variance swap, K_{var} is the swap rate that satisfies

$$\begin{aligned} Z &= K_{var} - \sigma_R^2, \\ E^{Q^*_\alpha}(Z) &= \int_{\mathbb{R}} z d\Psi_\alpha(F_Z(z)) \geq 0. \end{aligned} \tag{4.38}$$

The results for variance swap rates using different concave distortions are presented in Table 4.2.

For each of the four concave distortions, the calculated swap rates are all greater than the one computed without distortion. The new distributions shift weight to combinations of negative cash flow and high realized volatility and therefore reflect the fact that the sellers of the products ask for risk compensation by exaggerating the possibilities of loss.

4.3.3 Options on Realized Variance and Volatility

The ask price[5] P for variance or volatility option is determined by the expected cash flow from the short position. The expected cash flow on expiration day under

[5]Ask price is the price for a financial security quoted by a market maker for an immdediate purchase. Bid price is the price for an immediate sale.

Table 4.2: Prices of Variance Swap. (Units are percentage)

Maturity	Distortion	DAX30	DJIA	CAC40	FTSE	SPX	Nikkei	Nasdaq	HSI
3 Month	None	13.84	10.81	12.21	7.42	5.27	13.68	21.40	10.65
	MINVAR	15.98	13.09	14.24	8.87	6.11	16.16	26.80	12.70
	MAXVAR	15.27	12.56	13.66	8.57	6.05	15.42	27.99	12.13
	MAXMINVAR	15.43	12.63	13.77	8.61	6.03	15.57	27.10	12.24
	MINMAXVAR	15.50	12.72	13.85	8.66	6.06	15.66	27.39	12.31
6 Months	None	10.74	8.47	9.88	5.49	4.11	10.65	20.60	7.87
	MINVAR	11.89	9.86	11.03	6.27	4.56	12.09	25.86	9.01
	MAXVAR	11.50	9.48	10.67	6.07	4.51	11.63	27.02	8.66
	MAXMINVAR	11.59	9.55	10.75	6.11	4.50	11.73	26.17	8.73
	MINMAXVAR	11.63	9.60	10.79	6.14	4.52	11.78	26.45	8.78
1 Year	None	9.28	7.27	8.68	4.50	3.56	9.18	19.53	6.62
	MINVAR	9.94	8.15	9.38	4.94	3.81	10.06	24.27	7.28
	MAXVAR	9.69	7.89	9.13	4.81	3.76	9.76	24.61	7.05
	MAXMINVAR	9.75	7.94	9.19	4.84	3.76	9.83	24.14	7.10
	MINMAXVAR	9.77	7.97	9.21	4.86	3.77	9.86	24.37	7.13
24 Months	None	8.56	6.74	8.11	4.06	3.27	8.39	19.08	5.94
	MINVAR	8.96	7.33	8.54	4.32	3.41	8.95	23.26	6.33
	MAXVAR	8.80	7.13	8.38	4.24	3.37	8.74	22.86	6.19
	MAXMINVAR	8.84	7.18	8.42	4.26	3.38	8.80	22.76	6.23
	MINMAXVAR	8.85	7.20	8.43	4.27	3.39	8.82	22.93	6.24

Note: Stress levels: *MINVAR* $a = 0.5$, *MAXVAR* $a = 0.25$, *MAXMINVAR* $a = 0.15$, *MINMAXVAR* $a = 0.15$.

distorted measure Ψ_α is

$$Z = P - (\sigma_{var} - K_{var})^+,$$
$$E^{Q^*_\alpha}(Z) = \int_{\mathbb{R}} z d\Psi_\alpha(F_Z(z))) \tag{4.39}$$

for a variance option with strike price at K_{var}. Similarly

$$Z = P - (\sigma_{vol} - K_{vol})^+,$$
$$E^{Q^*_\alpha}(Z) = \int_{\mathbb{R}} z d\Psi_\alpha(F_Z(z)) \tag{4.40}$$

for a variance option with strike price at K_{vol}. The ask price is the value that sets the expectation equal to zero. Table 4.3 presents the pricing results.

Table 4.3: Price of 30% OTM Variance Swaption using SVPS (concave distorted).

Maturity	Distortion	DAX30	DJIA	CAC40	FTSE	SPX	Nikkei	Nasdaq	HSI
3 Month	NONE	0.52	1.24	0.72	0.35	0.10	1.07	8.08	0.61
	MINVAR	3.42	4.35	3.87	1.90	0.96	4.62	17.42	2.96
	MAXVAR	3.51	4.52	3.95	1.93	1.01	4.74	19.80	3.01
	MAXMINVAR	3.42	4.37	3.86	1.88	0.97	4.61	18.53	2.94
	MINMAXVAR	3.46	4.43	3.90	1.91	0.99	4.67	18.80	2.98
6 Months	NONE	0.24	0.82	0.38	0.17	0.04	0.61	7.58	0.34
	MINVAR	2.30	3.67	2.87	1.42	0.54	3.73	17.52	2.32
	MAXVAR	2.45	3.74	2.97	1.45	0.62	3.78	19.50	2.37
	MAXMINVAR	2.35	3.65	2.88	1.41	0.58	3.70	18.39	2.31
	MINMAXVAR	2.39	3.70	2.92	1.43	0.59	3.74	18.65	2.34
1 Year	NONE	0.09	0.46	0.15	0.08	0.01	0.31	6.49	0.15
	MINVAR	1.34	2.90	1.90	0.93	0.20	2.70	16.68	1.58
	MAXVAR	1.54	2.94	2.05	1.00	0.27	2.81	18.01	1.68
	MAXMINVAR	1.44	2.87	1.96	0.96	0.24	2.72	17.18	1.61
	MINMAXVAR	1.46	2.91	1.99	0.97	0.24	2.76	17.41	1.64
24 Months	NONE	0.02	0.23	0.04	0.02	0.00	0.11	4.75	0.06
	MINVAR	0.48	2.07	0.86	0.43	0.05	1.57	15.26	0.88
	MAXVAR	0.66	2.16	1.06	0.52	0.08	1.75	15.69	1.01
	MAXMINVAR	0.58	2.09	0.97	0.48	0.07	1.65	15.25	0.94
	MINMAXVAR	0.59	2.12	0.98	0.49	0.07	1.68	15.45	0.97

Note: Stress levels: *MINVAR* $a = 0.5$, *MAXVAR* $a = 0.25$, *MAXMINVAR* $a = 0.15$, *MINMAXVAR* $a = 0.15$.

4.3.4 Portfolio of underlying assets and variance swaps

In this section we will solve the optimization problem for a portfolio that consists of variance contracts and the underlying asset. The cash flow to the owner of this portfolio at expiration is

$$X = S_T - S_0 e^{rT} + \beta N_{var}(\sigma_{vol}^2 - K_{var}), \tag{4.41}$$

where N_{var} is the notional value for per unit of realized variance, K_{var} is the predetermined swap rate, $S_T - S_0 e^{rT}$ is the profit from investing one share of stock and $N_{var}(\sigma_{vol}^2 - K_{var})$ is the profit from one variance swap contract. The problem is to decide the weight β of the variance contract to achieve the best combination of profitability and risk taking.

The value of β is set to make the resulting cash flow fall inside the acceptable set at level a

$$\mathcal{A}_a = \{X : \alpha(X) \geq a\}. \tag{4.42}$$

For two different strike prices, $K = E(\sigma_{vol}^2)$ and $K = 1.2E(\sigma_{vol}^2)$, the resulting β are shown in Table 4.4 and Table 4.5.

Table 4.4: Weight β of variance swap in portfolio with underlying assets. $K = E(\sigma^2_{vol})$.

Maturity	Distortion	DAX30	DJIA	CAC40	FTSE	SPX	Nikkei	Nasdaq	HSI
3 Month	MINVAR	0.07	-0.30	0.19	2.34	-0.89	-0.45	-0.31	-0.44
	MAXVAR	2.03	2.76	-1.92	-2.70	-1.38	3.77	-0.61	5.46
	MAXMINVAR	-1.34	-2.93	-1.54	2.03	0.94	2.43	0.03	3.53
	MINMAXVAR	-1.96	1.58	1.13	2.12	-1.19	2.25	-0.48	3.26
6 Months	MINVAR	3.84	4.65	2.35	6.88	2.23	-9.68	-0.76	12.69
	MAXVAR	-9.76	-12.97	7.72	14.40	6.22	-18.51	-1.13	23.13
	MAXMINVAR	-7.88	-10.80	6.02	7.27	3.44	-15.52	-1.26	17.57
	MINMAXVAR	-7.44	-10.24	-5.62	-11.33	3.34	-14.67	-1.13	16.67
1 Year	MINVAR	18.66	-22.61	15.68	27.65	-15.30	52.21	-1.75	-55.84
	MAXVAR	-38.70	-37.00	-32.13	-52.46	-22.93	63.73	-2.69	90.132
	MAXMINVAR	29.12	30.78	24.21	16.99	-20.71	51.75	-2.45	-81.63
	MINMAXVAR	28.16	-36.18	23.32	40.89	-19.68	48.39	1.19	68.50

Note: Stress levels: *MINVAR* $a = 0.5$, *MAXVAR* $a = 0.25$, *MAXMINVAR* $a = 0.15$, *MINMAXVAR* $a = 0.15$.

Table 4.5: β of variance swap in portfolio with underlying assets. $K = 1.2E(\sigma^2_{vol})$

Maturity	Distortion	DAX30	DJIA	CAC40	FTSE	SPX	Nikkei	Nasdaq	HSI
3 Month	MINVAR	-0.64	-0.25	-0.56	0.35	0.09	-4.53	-0.46	-2.71
	MAXVAR	0.17	0.85	0.52	2.35	0.33	1.03	-0.23	1.61
	MAXMINVAR	0.19	-1.97	0.36	0.74	-2.63	0.61	0.00	0.99
	MINMAXVAR	0.18	0.53	0.32	0.79	-1.65	0.55	-6.29	0.03
6 Months	MINVAR	0.91	1.62	0.79	2.71	-0.22	2.35	-5.34	3.51
	MAXVAR	1.96	3.49	1.76	4.14	1.23	2.36	-0.33	6.52
	MAXMINVAR	1.63	2.87	1.47	3.64	1.06	2.36	0.02	5.52
	MINMAXVAR	1.58	2.77	1.42	3.57	1.12	3.66	-3.76	5.36
1 Year	MINVAR	4.21	7.09	4.07	9.64	2.93	9.85	-8.96	14.50
	MAXVAR	6.23	10.63	5.83	12.74	3.70	14.34	0.82	20.39
	MAXMINVAR	5.58	9.44	5.27	11.68	3.43	12.87	-0.23	18.49
	MINMAXVAR	5.47	9.26	5.18	11.53	3.09	12.64	-7.65	18.17

Note: Stress levels: *MINVAR* $a = 0.5$, *MAXVAR* $a = 0.25$, *MAXMINVAR* $a = 0.15$, *MINMAXVAR* $a = 0.15$.

Chapter 5

Markowitz portfolio theory for variance contracts

In this chapter we will discuss how to extend the portfolio optimization theory to the application of stochastic volatility models to set the optimal strategy for variance contracts.

5.1 Markowitz theory

Markowitz theory describes how rational investors will use diversification to optimize their portfolio. If we have n assets with returns $\mathbf{r} = \{r_1, r_2, ..., r_n\}$ and the weights of assets are $\mathbf{w} = \{w_1, w_2, .., w_n\}$, then the overall return of the portfolio is the weighted average of the returns of each asset

$$r_P = \mathbf{wr} = \sum_{i=1}^{n} w_i r_i. \tag{5.1}$$

Markowitz theory uses an exponential utility function to measure the degree of satisfaction that an investor gets from the investments:

$$U(r_P) = 1 - \exp(-\eta r_P), \tag{5.2}$$

where η is a positive scalar parameter which measures the degree of risk aversion of investor. It further assumes that a typical investor determines the optimal strategy

by maximizing the expected utility, under the restriction that the overall risk can't exceed a certain level. The risk is measured by the variance of the portfolio, $V_{r_P} = \mathbf{w}'V_{\mathbf{r}}\mathbf{w}$, where $V_{\mathbf{r}} = cov(\mathbf{r})$, the covariance matrix of $\mathbf{r}$. The problem of finding the optimal investment strategy can be represented as an optimization problem, that is, to find the weights of contracts in the portfolio that maximizes the expected utility.

$$\mathbf{w}^* = \text{argmax}E[U(\mathbf{wr})] \tag{5.3}$$

subject to

$$\mathbf{w}'V_{\mathbf{r}}\mathbf{w} \leq V_{max}. \tag{5.4}$$

If all the returns follow normal distributions,

$$\mathbf{r} \sim N(\mu_{\mathbf{r}}, \mathbf{V_r}), \tag{5.5}$$

then so does the portfolio return,

$$r_P \sim N(\mu_{r_P}, V_{r_P}), \tag{5.6}$$

where $\mu_{r_P} = \mathbf{w}\mu_{\mathbf{r}}$ and $\mu_{\mathbf{br}} = \mathbf{E}(\mathbf{r})$. Then the expected utility is

$$\begin{aligned} E[U(r_P)] &= E[1 - \exp(-\eta r_P)] = 1 - \exp(-\eta\mu_{r_P} + \frac{1}{2}\eta^2 V_{r_P}) \\ &= 1 - \exp(-\eta\mathbf{w}\mu_{\mathbf{r}} + \frac{1}{2}\eta^2\mathbf{w}'V_{\mathbf{r}}\mathbf{w}). \end{aligned} \tag{5.7}$$

So the expected utility is an increasing function of expected return and a decreasing function of variance. The solution of the portfolio weight vector $\mathbf{w}$ is

$$\mathbf{w}^* = V_{\mathbf{r}}^{-1}\frac{\mu_{\mathbf{r}}}{\eta}. \tag{5.8}$$

The implications of this result for portfolio management are as follows. First, the investors are risk aversive such that given portfolios of same risk level, a rational investor will choose the one with highest expected return. Conversely, given

portfolios with same level of expected return, the investor will chose the one least risky.

Secondly, the benefit of diversification comes from the fact that if the assets are not perfectly correlated, then investors can reduce their exposure to individual risk by holding a diversified portfolio of assets since the variance of the portfolio is strictly less than the weighted average of the variance of individual assets. That is, $V_R \leq \sum_{i=1}^{n} \sigma_i^2$. Diversification allows for the same portfolio return with reduced risk.

We can apply the essential ideas of Markovian theory in solving the problem of allocating volatility contracts. However since the normality assumption doesn't hold for stochastic volatility processes, the above result can't be applied directly. Based on our volatility model estimated from the historical data, we can still find out how to maximize the expected utility function while keeping the overall risk level under a certain level. In the next sections we will describe how to factorize the volatility process and introduce correlation among volatility processes of different assets. This will lead to our final solution to optimization of portfolio of volatility contracts.

5.2 Multidimensional volatility process

In chapter 3 we observed that the volatility processes for eight different indexes seem to follow a similar pattern. This suggests that a multifactor model can be employed to correlate those processes. We assume the stochastic volatility process for each

of the M indices, $W_{i,n}, i = 1, \ldots, M,\ n = 1, \ldots, N$, is determined by two global factors (F_1 and F_2) and one local idiosyncratic factor ϵ_i. The weights $\{\alpha_{i,1}, \alpha_{i,2}\}_{i=1}^{M}$ indicate how significant the impact of the factors are to each volatility process. The structure of the model for multi-dimensional case is specified as follows

$$V_{i,t} = F_{1,t}^{\alpha_{i,1}} F_{2,t}^{\alpha_{i,2}} \vartheta_{i,t}^{\zeta_i} \quad i = 1, ..., 8, \quad t = 1, ..., T \tag{5.9}$$

while F_1 and F_2 are two independent double-gamma processes, $\vartheta_{i,t} = \exp(\epsilon_i)$ and ϵ_i is normally distributed disturbance in the volatility:

$$\begin{aligned} F_{1,t} &\sim Gamma(d_1, \lambda_1 F_{1,t-1} + u_{1,t}), \quad u_{1,t} \sim Gamma(c_1, \gamma_1), \\ F_{2,t} &\sim Gamma(d_2, \lambda_2 F_{2,t-1} + u_{2,t}), \quad u_{2,t} \sim Gamma(c_2, \gamma_2). \end{aligned} \tag{5.10}$$

To estimate this multidimensional version of the stochastic volatility model we need to adjust the particle filter used in the univariate case.

Table 5.1: Parameters for two Double Gamma factors.

	λ	d	γ	c	Persistence=λ/d
F1	11.921	21.508	0.479	0.050	0.554
F2	10.521	43.756	3.414	0.103	0.241

Table 5.1 shows the parameters of the two factors estimated from the historical data of eight indexes. Table 5.2 gives the coefficients of each factor in the volatility processes. Two important facts can be discovered from the results. First, the two factors differ in the degree the autocorrelation, persistence and mean-reverting speed, as can be observed in Figure 5.1. The first factor has a damping rate 0.5542, compared to that of the second factors, 0.2405, and is more autocorrelated, less persistent and converges to mean level with a slower speed. This finding is in line

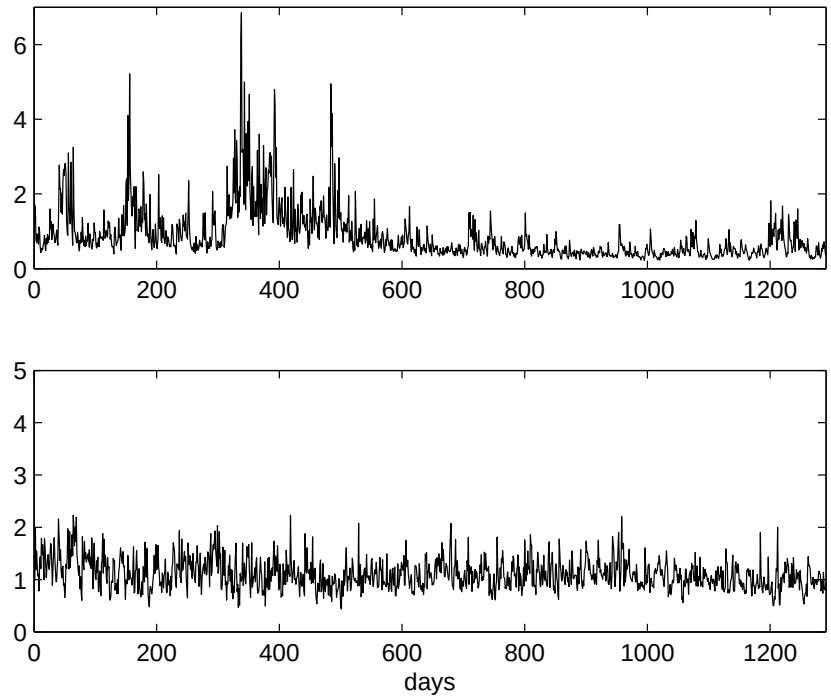

Figure 5.1: Two factors for stochastic volatility process.

with Engle and Lee (1993), who conjected that the volatility can be decomposed into a long-term component and a short-term component. The two factors are both mean-reverting processes, with different mean-reverting rates which lead to different persistence rates. This structure can also serve as a preliminary attempt to explain the long-term dependence property of volatility process which is crucial for pricing volatility options, whose value largely depends on the long-term level of volatility.

Table 5.2: Coefficients for two Double Gamma factors and the idiosyncratic factors.

	DAX30	DJIA	CAC40	FTSE	SPX	NIKKEI	NASDAQ	HSI
F1	0.448	0.349	0.451	0.466	0.347	0.106	0.336	0.110
F2	-0.245	0.413	-0.355	-0.419	0.455	-0.042	0.508	-0.031
ζ_i	0.121	0.084	0.095	0.113	0.059	0.114	0.101	0.071

Another observation is that the difference in the coefficients of the two factors reveals their nature. For the first factor all eight indexes have positive coefficients. For the second factor the three American indexes (DJIA, SPX, Nasdaq) have positive loadings while the five foreign indexes (DAX30, CAC40, FTSE, NIKKEI, HSI) have negative loadings. Therefore the first factor can be regarded as a global trend, while the second one can be deemed as a domestic influence of United States. Finally the idiosyncratic factors are more regional and can explain how local economic, political and many other local factors affect the behavior of stock markets.

This multidimensional stochastic volatility model is of particular importance because it holds together different volatility processes, builds up a codependency structure and meanwhile maintains the local idiosyncratic features. The joint law gives the expectation of functions of the volatility values at arbitrary time points

and thus expected utility can be calculated and optimization of the portfolio can be achieved.

5.3 Optimal strategy for portfolio of volatility contracts

Suppose we have n volatility swap contracts, written on the realized volatilities $\mathbf{V} = \{V_1, V_2, ..., V_n\}$ of indexes with swap prices $\mathbf{K} = \{K_1, K_2, ...K_n\}$. The payoff of each contract at expiration date is $\mathbf{r} = \mathbf{V} - \mathbf{K}$. Thus given the weights $\mathbf{w}$ of these contracts in the portfolio, the overall payoff equals

$$r_P = \mathbf{w}(\mathbf{V} - \mathbf{K}) = \sum_{i=1}^{n} w_i(V_i - K_i). \tag{5.11}$$

Our goal is to find the optimal strategy which maximizes the expected utility, with restriction that the variance cannot exceed a certain level:

$$\mathbf{w}^* = \text{argmax}E[U(\mathbf{wV})], \quad s.t. \quad V_{r_P} \leq V_{max}. \tag{5.12}$$

Unlike the case where asset returns follows a normal distribution, there is no closed-form solution for $\mathbf{w}^*$ due to the complex nature of the distribution of $\mathbf{V}$. Nevertheless, since the processes can be easily simulated, we can calculate the expectation by Monte Carlo simulation based on the multidimensional volatility model. This can be done through four steps.

In step one we simulate a number of paths for the two independent double gamma factors and eight idiosyncratic factors using the parameters estimated from historical data. In the second step we construct paths for each of the eight individual volatility processes using the two common factors and eight idiosyncratic factors. In the third step we calculate the realized volatility on each path using Equation (4.1). Finally we can approximate the expectation of the utility function through the simulated realized volatilities and the weights of each volatility swap contract

using Equation (5.13) .

Although the expected utility is calculated based on simulation, the total computational cost for finding the optimal strategy is fairly low. The reason is because during the whole procedure, the simulation is needed only once. Once the empirical probability space of the realized volatility is constructed, the expected utility can be calculated fairly quickly as a simple function of the simulated realized volatilities $\mathbf{V}_m, m = 1, ..., N$:

$$E[U(r_P)] \approx 1 - \frac{1}{N} \sum_{m=1}^{N} \exp[-\eta \mathbf{w}(\mathbf{V}_m - \mathbf{K}_m)] = f(\mathbf{w}). \tag{5.13}$$

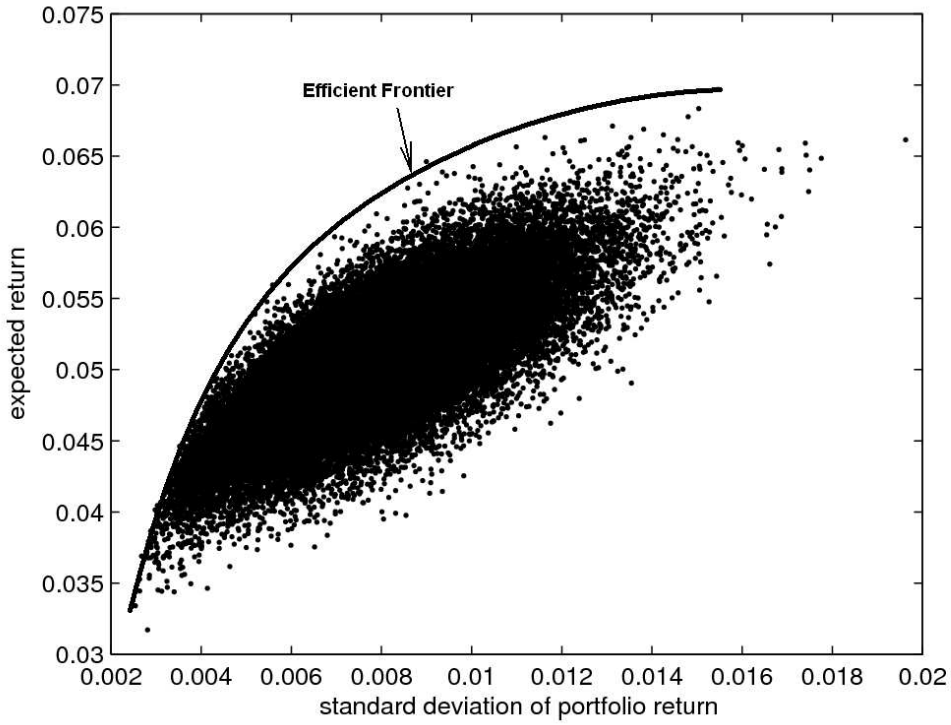

Figure 5.2: Efficient frontier for portfolio of volatility contracts.

The efficient frontier of a portfolio is the set of all optimal combinations of

expected return and risk of the portfolio. We can build up the efficient frontier for eight volatility contracts as shown in Figure 5.2.

Table 5.3 shows an example of the result of optimal portfolio of volatility contracts for different risk tolerance levels V_{max}. The expected return of the portfolio r_P increases as the risk tolernce level increases and the weights on each index's volatility contract also vary in the optimal portfolioes as results of different balancing of expected returns and risk. For investor with high level of risk-tolerence, the allocation is more skewed to contracts with higher expected return and higher uncertainty (volatility). For investor with lower risk-tolerence level, the optimal portfolio will be more diversified to reduce overal risk. Among the candidate contracts in the portfolio, volatility swap written on FTSE has low expected return, high risk and meanwhile has great correlation with some other contracts (CAC40 and DAX30), thus is not chosen in the porfolio even at the lowest risk tolerence.

Table 5.3: Optimal portfolios of volatility swaps for different risk tolerance level (V_{max}).

V_{max}	0.0025	0.003	0.006	0.008	0.01	0.012	0.016
DAX30	0.027	0.055	0.214	0.343	0.510	0.651	0.890
DJIA	0.237	0.180	0.000	0.000	0.000	0.000	0.000
CAC40	0.005	0.005	0.022	0.037	0.032	0.023	0.007
FTSE	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SPX	0.269	0.148	0.000	0.000	0.000	0.000	0.000
NIKKEI	0.179	0.271	0.330	0.064	0.000	0.000	0.000
NASDAQ	0.039	0.135	0.435	0.556	0.458	0.326	0.104
HSI	0.244	0.206	0.000	0.000	0.000	0.000	0.000
r_P	0.035	0.041	0.059	0.065	0.069	0.072	0.077
U_{r_P}	0.051	0.060	0.085	0.093	0.098	0.102	0.109

Chapter 6

Conclusions

In this dissertation we assumed the stochasticity of volatility is determined by a double-gamma process. The benefit from this unique specification is twofold. First, it accommodates the non-Gaussian properties of volatility as documents in Chernov et al. (2003). Secondly it has an Affine structure for conditional expectation and variance of volatility, which is convenient for the pricing of derivative products.

To solve the problem of fast decaying skewness and excess kurtosis we construct an innovations part of the return process based on Lévy processes at unit time. Although we only used a Variance-Gamma process to derive the density function of innovation, this method can be easily expanded to other Lévy processes.

Another stylized fact addressed is the leverage effect. In our model, it is incorporated by one component which is a quadratic function of past innovations in return. Its main distinctive feature is that it allows an asymmetric effect of positive and negative lagged values of innovations on volatility.

To estimate the model we applied the Particle filter EM algorithm to evaluate the likelihood function, which is used for calculate the MLE of the models parameters. A particle filter is applied to sample paths from a posterior distribution of

volatility and it significantly improves the efficiency and accuracy of the estimation.

To lessen the burden of estimation, we also explored an untapped issue in stochastic volatility modeling, that is, volatility might change with different frequencies to return. The likelihood ratio tests on models with different frequencies of volatility not only dismisses the restriction that the volatility changes with the same pace to return, but also provides a way to search for optimal updating frequencies. With a lower-valued optimal frequency the time spent on sampling latent process is significant and the computational burden is greatly eased.

An important application of the stochastic volatility model is to price volatility contracts including variance/volatility swap and variance/volatility option. Since the market is not liquid enough to provide sufficient price data to do the calibration, we rely on the physical measure which is ready after we solve the estimation problem based on historical data. To compensate the risk we distort the physical measure under a coherent utility function and the corresponding acceptability index. This method overcomes the calibration problem in pricing derivatives and surpasses the non-arbitrage pricing methodology and therefore provides a robust and efficient way to price the exotic derivative contracts combining historical price information, market risk aversion level and direction of trade.

Future researches will include study on calibration issues in capturing the implied market information instead of just historical information, further theoretical study on hypothesis test among different volatility change frequencies, and issues in evaluation of more exotic volatility products using this model.

Bibliography

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

Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. "Coherent measures of risk". *Mathematical Finance*, 9:203–228, 1997.

Ole E. Barndorff-Nielsen and Neil Shephard. "Processes of Normal Inverse Gaussian type". *Financial Stochasticity*, 2:41–68, 1998.

Ole E. Barndorff-Nielsen and Neil Shephard. "Incorporation of a leverage effect in a stochastic volatility model".Unpublished Paper. 1999.

Ole E. Barndorff-Nielsen and Neil Shephard. "Superposition of Ornstein-Uhlenbeck type processes". *Theor. Probab. Applic.*, 2000.

Ole E. Barndorff-Nielsen and Neil Shephard. "Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics". *Journal of Royal Statistical Society, B*, 63:167–241, 2001.

Ole E. Barndorff-Nielsen and Neil Shephard. "Power and bipower variation with stochastic volatility and jumps ". *Journal of Financial Econometrics*, 2:1–48, 2004.

David Bates. "Jumps and stochastic volatility: Exchange rate processes implicit in Deutsches Mark options". *The Review of Financial Studies*, 9(1):69–107, 1996.

Vijay S. Bawa. "Optimal rules for ordering uncertain prospects". *Journal of Financial Economics*, 2:95–121, 1975.

F. Black. "Studies of stock price volatility changes". *Proceedings of the Business and Ecnomic Statistics Section, American Statistical Association*, pages 177–181, 1976.

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

C. Bos and Neil Shephard. "Iinference for adaptive time series models:stochastic volatility and conditionally gaussian state space form". *Econometric Reviews*, 25: 219–244, 2006.

John Y. Campbell and Albert S. Kyle. "Smart money, noise trading and stock price behaviour". *Review of Ecnomic Studies*, 60:1–34, 1993.

Peter Carr and Dilip B. Madan. "Option valuation using the Fast Fourier Transform". *Journal of Computational Finance*, 2:61–73, 1998.

Peter Carr, Helyette Geman, and Dilip B. Madan. "Pricing and hedging in incomplete markets". *Journal of Financial Economics*, 62:131–167, 2001.

Peter Carr, Helyette Geman, Dilip B. Madan, and Marc Yor. "The fine structure of asset returns: An empirical investigation". *Journal of Business*, 75(2):305–332, 2002.

Peter Carr, Helyette Geman, Dilip B. Madan, and Marc Yor. "Stochastic volatility for Levy processes". *Mathematical Finance*, 13:345–382, 2003.

Mikhail Chernov and Eric Ghysels. "Estimation of the stochastic volatility models for the purpose of options valuation". In Andrew W.Lo Andreas S.Weigend Yaser S.Abu-Mostafa, Blake LeBaron, editor, *Computational Finance - Proceedings of the Sixth International Conference*, 1999.

Mikhail Chernov, Ronald Gallant, Eric Ghysels, and George Tauchen. "Alternative models for stock price dynamics". *Journal of Econometrics*, 116:225–257, 2003.

Alexander Cherny and Dilip B. Madan. "From required returns to required Sharpe ratios". *Available at Social Science Research Network*, January 12 2006.

Alexander Cherny and Dilip B. Madan. "On measuring the degree of market efficiency". *Available at Social Science Research Network*, January 1 2007.

C. Cherny. "Weighted VAR and its properties". *Finance and Stochastics*, 10:367–393, 2006.

P.K. Clark. "A suboridinated stochastic process model with fixed variance for speculative prices". *Econometrica*, 41:135–156, 1973.

D.R. Cox. "Long-range dependence,non-linearity and time irreversibility". *Journal of Time Series Analysis*, 12:329–335, 1991.

John Cox and Stephen Ross. "The valuation of options for alternative stochastic processes". *Journal of Financial Economics*, 166:145, 1976.

John Cox, Jonathan Ingersoll, and Stephen Ross. "A theory of the term structure of interest rates". *Econometrica*, 53(2):385–407, 1985.

M.M. Dacorogna, U.A. Muller, R.B. Olsen, and V. Pictet. "Modeling short term volatility with GARCH and HARCH". In Christian Dunis and Bin Zhou, editors, *Nonlinear Modeling of High Frequency Financial Time Series*. Wiley, Chichester.

A.P. Dempster, N.M. Laird, and D.B. Rubin. "Maximum likelihood from incomplete data via the EM algorithm". *Journal of the Royal Statistical Society. Series B*, (1), 1977.

Daniel Dias and C.R. Marques. "Using mean reversion as a measure of persistence". *European Central Bank. Working Paper Series*, 450:5–48, 2005.

Z. Ding and C.W.J. Granger. "Modeling volatility persistence of speculative returns: A new approach". *Journal of Econometrics*, 73:185–215, 1996.

R.F. Engle. "Estimating structural models of seasonality". *Seasonal Analysis of Economic Time Series. Washington D.C: Bureau of the Census*, pages 281–308, 1978.

Robert F. Engle and G.J. Lee. "A permanent and transitory component model of stock return volatility". *University of California at San Diego, Economics Working Paper Series*, 1993.

H. Follmer and A. Schied. "Convex measures of risk and trading constraints". *Finance and stochastics*, 6:429–447, 2004.

Helyette Geman, Dilip B. Madan, and Marc Yor. "Time changes for Levy processes". *Mathematical Finance*, 11:131–156, 2001.

W.R. Gilks, S. Richardson, and D.J. Spiegelhalter. *"Markov Chain Monte Carlo in Practice"*. Chapmean and Hall, 1996.

L. Glosten, R. Jagannathan, and D. Runkle. "On the relation between the expected value and the volatility of the nominal excess return on stocks". *Journal of Finance*, 48:1779–1801, 1993.

N. Gordon. "Bayesian Methods for Tracking",Phd thesis, University of London.

I.S. Gradshetyn and I.M. Ryzhik. *"Table of Integrals, Series, and Products"*. New York:Academic Press, 1980.

C.W.J. Granger. "Long memory relationships and the aggregation of dynamic models". *Journal of Econometrics*, 14:227–238, 1980.

S.L. Heston. "A closed-form solution for options with stochastic volatility with applications to bond and currency options". *The Reviews of Financial Studies*, 6 (2):327–343, 1993.

E. Jacquier, N.G. Ploson, and P.E. Rossi. "Bayesian analysis of stochastic volatility models (with discussion)". *Journal of Business and Econometric Statistics*, 12: 371–417, 1993.

H. Johnson. "Option pricing when the variance rate is changing". *Working paper, University of California, Los Angeles*, 1979.

Dilip B. Madan and Frank Miline. "Option pricing with VG martingale components". *Mathematical Finance*, 1(4):39–55, 1991.

Dilip B. Madan and Eugene Seneta. "The variance gamma model for share market returns". *Journal of Business*, 63:511–524, 1990.

Dilip B. Madan, Peter Carr, and Eric Chang. "The variance gamma process and option pricing". *European Finance Review*, 2:79–105, 1998.

Stephen Ross. "Options and efficiency". *Quarterly Journal of Economics*, 90:75–89, 1976.

Enrique Sentana. "Quadratic ARCH models". *The Review of Economic Studies*, 62:639–661, 1995.

Neil Shephard. "Statistical aspects of ARCH and stochastic volatility". In *Time Series Models in Econometrics, Finance and Other Fields*, pages 1–67. London:Chapman & Hall, 1996.

S.J. Taylor. "Financial returns modeled by the product of two stochastic processes - a study of daily sugar prices 1961-1979". *Time Series Analysis: Theory and Practice*, 1:203–226, 1982.

S.J. Taylor. *"Modeling Financial Time Series"*. Wiley, Chichester,UK, 1986.

S.J. Taylor. "Modeling stochastic volatility: A review and comparative study". *Mathematical Finance*, 4:183–204, 1994.

C.F.Jeff Wu. "On the convergence properties of the em algorithm". *Ann.Statist.*, 11:95–103, 1983.