ABSTRACT

Title of dissertation: ESTIMATION OF EXPECTED RETURNS, TIME CONSISTENCY OF A STOCK RETURN MODEL, AND THEIR APPLICATION TO PORTFOLIO SELECTION

Huaqiang Ma, Doctor of Philosophy, 2010

Dissertation directed by: Prof. Dilip B. Madan
AMSC / Department of Finance

Longer horizon returns are modeled by two approaches, which have different impact on skewness and excess kurtosis. The Levy approach, which considers the random variable at longer horizon as the cumulants of i.i.d random variables from shorter horizons, tends to decrease skewness and excess kurtosis in a faster rate along the time horizon than the real data implies. On the other side, the scaling approach keeps skewness and excess kurtosis constant along the time horizon. The combination of these two approaches may have a better performance than each one of them. This empirical work employs the mixed approach to study the returns at five time scales, from one-hour to two-week. At all time scales, the mixed model outperforms the other two in terms of the KS test and numerous statistical distances.

Traditionally, the expected return is estimated from the historical data through the classic asset pricing models and their variations. However, because the realized

returns are so volatile, it requires decades or even longer time period of data to attain relatively accurate estimates. Furthermore, it is questionable to extrapolate the expected return from the historical data because the return is determined by future uncertainty. Therefore, instead of using the historical data, the expected return should be estimated from data representing future uncertainty, such as the option prices which are used in our method. A numeraire portfolio links the option prices to the expected return by its striking feature, which states that any contingent claim's price, if denominated by this portfolio, is the conditional expectation of its denominated future payoffs under the physical measure. It contains the information of the expected return. Therefore, in this study, the expected returns are estimated from the option calibration through the numeraire portfolio pricing method. The results are compared to the realized returns through a linear regression model, which shows that the difference of the two returns is indifferent to the major risk factors. This demonstrates that the numeraire portfolio pricing method provides a good estimator for the expected return.

The modern portfolio theory is well developed. However, various aspects are questioned in the implementation, e.g., the expected return is not properly estimated using historical data, the return distribution is assumed to be Gaussian, which does not reflect the empirical facts. The results from the first two studies can be applied to this problem. The constructed portfolio using this estimated expected return is superior to the reference portfolios with expected return estimated from historical data. Furthermore, this portfolio also outperforms the market index, SPX.

ESTIMATION OF EXPECTED RETURNS, TIME CONSISTENCY OF A STOCK RETURN MODEL, AND THEIR APPLICATION TO PORTFOLIO SELECTION

By

Huaqiang Ma

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
2010

Advisory Committee:
Professor Dilip B. Madan, Chair
Professor Benjamin Kedem
Professor Mark Loewenstein
Professor Tobias von Petersdorff
Professor Victor M. Yakovenko

Dedication

To my mother and father

献给我的父亲母亲

Acknowledgements

First and foremost, I would like to thank my advisor Professor Dilip B. Madan. Through the years of my PhD study, Dr. Madan was always there offering his help. His width and depth of knowledge opens my eyes in the world of mathematical finance; his insight lights the way when I got lost in the mist; his intolerance for ambiguity sharpens my mind; and his sense of humor soothes my nerves when I was struggling in the research. To me, Dr. Madan is also a role model. Most of people only enjoy the time after work. Very few people enjoy their life at work, and Dr. Madan is one of them. His passion for knowledge inspired and motivated me during my PhD study, and will inspire and motivate me through my life.

I want to thank Professor Mark Loewenstein for generous advices in my thesis. I would like to extend my gratitude to Professor Benjamin Kedem, Professor Tobias von Petersdorff, and Professor Victor M. Yakovenko for agreeing to serve in the PhD advisory committee. I also want to thank Professor Michael Fu for organizing our weekly financial seminar, which broaden my knowledge in the mathematical finance.

I would like to say thanks to my fellow classmates. Samvit Prakash, Bing Zhang, and Christian Silva guided me with their own experience in math finance. Thank Yun Zhou, Linyan Cao, Guoyuan Liu, and Geping Liu for numerous hours of discussion for the problems in the research.

I would like to express my grateful feelings to the staff members. Ms. Alverda McCoy was always there willing to help. Mr. William (Bill) Schildknecht was so kind to offer me teaching opportunities in and out of math department. Mr. Chuck LaHaie helped me with financial data.

My sincere thanks go to my friends, especially Wei Hu, Yiling Luo, Baozhong Mao, Hua Sheng, Anshuman Sinha, and Min Sun. Your sincere help and encouragement helped me go through the hard time in my PhD study.

Finally, I would like to thank my parents. They always stood by me with their endless love and support. No words can express my gratitude.

Contents

List of Figures

List of Tables

Chapter 1

Empirical Study of A Stock Return Model

1.1 Introduction

Stock markets are complex dynamic systems with many elements interacting with each other. The interaction of the elements can be observed as price fluctuations through time. Because of the complexity of stock markets, price fluctuations exhibit statistical properties, which can be reproduced by financial models. Sophisticated models are capable of dynamically describing these properties, that is, they can provide statistical fit and analysis for the stock returns at various time scales, from hourly to daily returns, and even to monthly and yearly returns. These models are time consistent through the time horizon. The statistical accuracy of financial models is necessary and useful in many areas. For example, in risk management, a proper model can help reduce severe losses. With the incentive both for academic and practical purposes, many financial models have been developed.

The stochastic approach has been widely implemented to model this complex dynamic system. Dating back to 1900, Louis Bachelier [2] first proposed that the stock price behaves as random walks and modeled the price at time t as $S_t = S_0 + \sigma W_t$, where W_t is Brownian motion and σ is the volatility. In this model, the price fluctuation $\Delta S_t = S_t - S_0$ follows Brownian motion, which has independent and identical increments, with increments having Gaussian distributions. To ensure positive price, log-price fluctuation $\ln(S_t/S_0) = \ln S_t - \ln S_0$, instead of the price fluctuation, is modeled to follow Brownian motion [2]. The stochastic process of S_t is then said to follow the geometric Brownian motion, and the dynamics of S_t is $dS_t = \mu S_t dt + \sigma S_t dW_t$, which has the analytic solution $S_t = S_0 \exp\{(\mu - \sigma^2/2)t + \sigma W_t\}$. This model is also employed in option pricing by Black, Scholes [7], and Merton [58], and the work earned the latter two a Nobel Prize of economics in 1997.

Sophisticated financial models can reproduce, if not all but some important statistical properties of the stock markets. The main object is to study the probability density function (pdf) of the increments at various time scales and compare the implied statistical properties with the empirical facts. The increments in the stock markets are the log-price fluctuations (or log-price returns) $\ln(S_{t+\tau}/S_t)$ at some given time scale τ. The Brownian motion model misses some of the important statistical properties of the stock markets. Compared to Brownian motion, the empirical distribution of the increments of log-price returns has more mass at the origin and in the tails, i.e., the stock markets have more events of small returns and big returns (either losses or gains) than the model predicts. This empirical fact is called heavy tails (or fat tails), which can be quantitatively measured by excess kurtosis. The excess kurtosis of Gaussian distribution is zero while the markets always show positive excess kurtosis. The other discrepancy is that Brownian motion has a continuous sample path but the path of stock prices may show discontinuity (or jumps).

Lévy processes, which were introduced by Paul Lévy in the early 20th century, are stochastic processes with independent and identical increments. It relaxes one assumption in Brownian motion, Gaussian increments. This relaxation provides the flexibility to have more choices of distributions for the increments as long as they are infinitely divisible, and these distributions are more leptokurtic (positive excess kurtosis) than the Gaussian one. Furthermore, discontinuity is also possible to appear in the path of Lévy processes. Merton [59] is one of the first to propose one Lévy process (Brownian motion compound Poisson process, or jump-diffusion process) to model asset returns. Later, numerous Lévy models were proposed, among them we cite the Variance Gamma model by Madan and Seneta [49] [48], the NIG model by Barndorff-Nielson [3], the Generalized Hyperbolic model by Eberlein and Prause [24], Prause [67], the Meixner model by Schoutens [73], and the CGMY model by Carr *et al.* [11]. Because of the flexibility of choosing from various distributions, Lévy models can capture some of the important empirical facts, such as jumps and fat tails. The statistical fit is usually performed at some fixed time scale. However, little work has been done to do statistical analysis along the time horizon. One example of work is done by Eberlein and Ozkan [23], who investigated the time consistency of Lévy models where a Lévy distribution model is employed at different time scales, from hourly to daily return. For Lévy processes, skewness decreases in the rate of the square root of the time horizon and excess kurtosis decreases in the rate of the time horizon. However, empirical studies [17] [22] show the actual data decrease more slowly than Lévy processes predict.

Besides Lévy models, self-similarity or scaling is applied in financial markets. In a stochastic process, the scaling property means the distribution of increments of various time scales can be obtained from that of other time scale by rescaling the random variable at that time scale. Mandelbrot [52] is the first to introduce this concept into financial markets, where he considered cotton price returns having the scaling property. We cite some of the other works as Mantegna and Stanley [54], Cont, Bouchaud and Potters [18], Mandelbrot [53], Peters [63], Cont [17], and Galloway and Nolder [31]. With the assumption of the scaling behavior, the distribution at the larger timescales can be derived from those at the smaller ones, which are easier to estimate because the data is sufficient. Not like Lévy processes, the fact, the stochastic processes with scaling property have constant skewness and excess kurtosis at all timescales, also does not satisfy the empirical results.

These two approaches have different impacts on skewness and excess kurtosis through the time horizon. The Lévy approach indicates a faster decay than the market while the scaling approach has constant skewness and excess kurtosis at all horizons. Thus, Eberlein and Madan [22] proposed a model mixed of the two approaches, which provides the freedom to let the term structure of the skewness and excess kurtosis have a similar pattern as the markets. One thing needs to be pointed out is that this mixed approach is not associated with any processes. Instead, it only uses the distributional properties taken from these two processes. In this model, the random variables of the increments of the stock returns at various time horizons are decomposed into two independent parts, one is from the increments of a Lévy process, and the other comes from that of a scaling process.

In this paper, we start the statistical parameter estimation from a short horizon, e.g., one hour, because of the abundant date. A base distribution is chosen, which shall be infinite divisible and have self-decomposability characteristic (SDC) because of the requirements from Lévy processes and decomposition of the random variables. The distribution is decomposed into two parts in law, one is partial of itself, and the other is a remaining component. The distributions at longer horizons are run by these two components, represented by two parameters correspondingly: one implies the proportion of the Levy composition; the other is the scaling coefficient of the remaining component. These two parameters are estimated at the longer horizons, such as two hours, three hours, one day, one week, and two weeks in this paper. For comparison, the statistical estimations are also conducted by the associated Lévy model and scaling model at the same timescales, and statistical analyses, including the Kolmogorov Smirnov (KS) test, the Kolmogorov distance, χ^2-distance, L_1 and L_2 distance, are performed. All these statistical analyses indicate that the scaling model outperforms the Lévy model at longer horizons, while the mixed model dominates these two models at all horizons. Furthermore, the averages of the two parameters of 500 individual stocks are both around 0.4 through the time horizon. Thus, it is adequate to assume value 0.4 for these two parameters at even longer time scales, such as half-year, one year or longer, where the statistical parameter estimation are not feasible due to the lack of data.

The remaining of this chapter is organized as follows. Section 1.2 introduces and discusses the two approaches of modeling financial markets, the Lévy process approach and the scaling approach. Section 1.3 presents the empirical study of skewness and excess kurtosis in stock markets and explains how the above two approaches are questionable. In section 1.4, the mixed approach is developed to model the distribution of returns through the time horizon. Section 1.5 describes the methods to do statistical estimation. Section 1.6 presents the results. Section 1.7 concludes the study and suggests further work.

1.2 Two Approaches to Model Stock Market Returns

1.2.1 Lévy Processes: Definition, Properties, and Lévy Market Models

§ Lévy Processes and Infinitely Divisible Distributions

Relax one assumption in Brownian Motion, the Gaussian increments, and we have Lévy processes, which provide more flexibility to build continuous-time stochastic models.

Definition 1 ***Lévy Process***

A stochastic process $X = \{X_t : t \geq 0\}$ *on* (Ω, F, P) *with values in* R^d *is said to be a Lévy process if,* *(1) X has independent increment: that is, for any $0 \leq t_1 < t_2 < ... < t_n < \infty$, $X_{t_2} - X_{t_1}, ..., X_{t_n} - X_{t_{n-1}}$ are independent.*

(2) X has stationary increment: that is, the law of $X_t - X_s$ is the same as X_{t-s}, where $0 \leq s < t < \infty$

(3) X_t is stochastic continuous: that is, for $\forall\ \varepsilon > 0$ and $t > 0$

$$\lim_{h \to 0} P\left(|X_{t+h} - X_t| \geq \varepsilon\right) = 0.$$

(4) $X_0 = 0$ almost surely.

(5) X_t has the cadlag property: that is, right-continuity and left limits.

The third condition, stochastic continuity, does not imply continuous sample path. It only means that at any given (or deterministic) time t, the probability that a jump occurs is zero. However, sample path may still be discontinuous at random times. Thus, the Lévy process is capable of capturing random jumps occurring in the financial markets.

The first two conditions are the main features of Lévy processes, and the third one follows from the first two (Keller [37]). Jacod and Shiryaev [36] name processes with conditions (1) and (2) "processes with stationary independent increment (PIIS)." Given a random variable Y with the probability distribution F, if we make $X_t \overset{law}{=} Y$ for any $t > 0$, then we can construct a Lévy process through the time horizon as follows:

1. For the sample path at $t, 2t, 3t, ..., nt, ..., n > 1$, $X_{nt} = X_t + (X_{2t} - X_t) +$

$... + (X_{nt} - X_{(n-1)t})$, whose distribution is the same as that of the sum of n $i.i.d$ random variables $X_t, X_{2t} - X_t, ..., X_{nt} - X_{(n-1)t}$ because of the "stationary independent increment" property;

2. Within interval $X_t - X_0$, we choose any integer $m > 1$ that $m\Delta = t$, $X_t = X_\Delta + (X_{2\Delta} - X_\Delta) + ... + (X_{m\Delta} - X_{(m-1)\Delta})$, that is, X_t is divided into m $i.i.d$ parts and it has the same law of that of the sum of m $i.i.d$ random variables, whose distribution can be derived from X_t.

In this procedure, the distribution of X_t can be infinitely "divided" as $m > 1$ can be infinitely large. This property is called "infinite divisibility."

Definition 2 ***Infinite Divisibility***

Let X be a random variable with distribution F. A probability distribution F on R^d is infinitely divisible if for any integer $n > 1$, there exists n $i.i.d$ random variables $X_1, X_{2,}..., X_n$ such that

$$X \overset{law}{=} X_1 + X_2 + ... + X_n. \tag{1.1}$$

We have the following proposition showing the one-to-one relationship between the Lévy process and infinitely divisible distribution:

Proposition 3 ***Lévy Processes and Infinitely Divisible Distributions***

For any infinitely divisible distribution F, there exists a Lévy process $\{X_t : t \geq 0\}$ such that the law of X_1 is F. Conversely, given a Lévy process $\{X_t : t \geq 0\}$, the distribution of X_t is infinitely divisible for every $t > 0$.

Compared to Brownian motion, it's quite flexible to choose distributions for the increments of Lévy processes, with only one constraint that the distributions should be infinitely divisible.

Usually the *pdf* of X_t in Lévy processes is not easy to obtain [36]. Instead we study the characteristic function of X_t. Let $\Phi_t(u)$ or $\Phi_{X_t}(u)$ be the characteristic function of X_t, $E[e^{iu \cdot X_t}]$. Define $\Psi_t(u) = \Psi_{X_t}(u) = \ln \Phi_t(u)$ as the characteristic exponent. Then, the characteristic function of a Lévy process is given by the following proposition:

Proposition 4 ***Characteristic Function of Lévy Processes***

Let $\{X_t : t \geq 0\}$ be a Lévy process on R^d and its characteristic exponent at $t = 1$ be Ψ. Then Ψ is a continuous function $\Psi : R^d \to R$, such that:

$$E[e^{iu \cdot X_t}] = e^{t\Psi(u)}, \quad u \in R^d. \tag{1.2}$$

By this proposition, we can build a Lévy process from any infinitely divisible distribution through its characteristic function. Therefore, the law of X_t is determined by the law of X_1; both are infinitely divisible.

§ Properties of Lévy Processes

Brownian motion is a well-known Lévy process with Gaussian increment. Another simple and common Lévy process is the compound Poisson process $\{X_t : t \geq 0\}$, which is defined as $X_t = \sum_{i=0}^{N_t} Y_i$ where $\{N_t : t \geq 0\}$ is a Poisson process and Y_i are $i.i.d$ random variables independent of N_t. Other Lévy processes can be decomposed by these simple bricks.

Proposition 5 ***Lévy-Itô Decomposition***

Let $\{X_t : t \geq 0\}$ be a Lévy process on R^d. It can be decomposed into three parts:

$$
\begin{aligned}
X_t &= X_t^B + X_t^C + \lim_{\varepsilon \downarrow 0} \widetilde{X_t^\varepsilon}, \quad \textit{where} \qquad (1.3)\\
X_t^B &= rt + AW_t^{(d)}\\
X_t^C &= \int\limits_{|X| \geq 1, s \in [0,t]} xN(ds, dx)\\
\widetilde{X_t^\varepsilon} &= \int\limits_{\varepsilon \leq |X| \leq 1, s \in [0,t]} x\widetilde{N}(ds, dx).
\end{aligned}
$$

X_t^B is a d-dimensional continuous Gaussian process with drift r and covariance matrix A, $W_t^{(d)}$ is a d-dimensional Brownian motion;

X_t^C is a compound Poisson process with jump size $|X| \geq 1$. $N(ds, dx)$ is a Poisson random measure on $R^+ \times (R^d \backslash \{0\})$;

$\widetilde{X_t^\varepsilon}$ is a compensated compound Poisson process $\widetilde{N}(ds, dx) = N(ds, dx) - \nu(dx)ds$, where ν is a jump intensity (or called Lévy measure) on $R^d \backslash \{0\}$ and is given by $\nu(dx) = E[N(1, dx)]$. ν also verifies $\int\limits_{R^d \backslash \{0\}} (1 \Lambda |x|^2)\nu(dx) < \infty$.

(r, A, ν) is called the Lévy characteristic triplet or Lévy triplet.

The Lévy-Itô decomposition states that the structure of the sample path of any Lévy process consists three parts: a diffusion process with drift, X_t^B, a "large jump" process with jump size greater than one, X_t^C, and a "small jump" process with jump size less than one, $\widetilde{X_t^\varepsilon}$. As there can be infinitely many small jumps around zero and their sum may not converge, we have to compensate the compound Poisson process with small jumps to make it a martingale so that it won't explode. That's how we get the third part $\widetilde{X_t^\varepsilon}$ in the decomposition (Proposition 2.16 in [19]).

With the Lévy-Itô decomposition formula, it's easy to derive the characteristic function of Lévy processes, which is given in the next theorem:

Theorem 6 ***Lévy-Khinchin Representation***

Let $\{X_t : t \geq 0\}$ be a Lévy process. Its characteristic function satisfies:

$$E[e^{iu\cdot X_t}] = e^{t\Psi(u)}, \quad u \in R^d, \tag{1.4}$$

$$where \quad \Psi(u) = ir \cdot u - \frac{1}{2}u \cdot Au + \int_{R^d\backslash\{0\}} (e^{iu\cdot x} - 1 - iu \cdot x1_{|X|\leq 1})\nu(dx). \tag{1.5}$$

The Lévy-Khinchin representation explicitly links Lévy processes to infinitely divisible distributions. Given an infinitely divisible distribution F with characteristic component (1.5), a Lévy process X_t can be generated where its law at $t = 1$ is F. Thus, we can study any Lévy process from its corresponding infinitely divisible distribution.

§ Distributional Property of Lévy Process: Tail Behavior

The Lévy-Khinchin representation enables us to study the tail behavior of the distribution of a Lévy process through its associated infinitely divisible distribution F, which is characterized by a Lévy triplet (r, A, ν). We cite the following proposition from Cont [19], Proposition 3.13.

Proposition 7 ***Moments and Cumulants of a Lévy process***

Let $\{X_t : t \geq 0\}$ be a Lévy process on R with characteristic triplet (r, A, ν). The n-th absolute moment of X_t, $E[|X_t|^n]$ is finite for some t or, equivalently, for every $t > 0$ if and only if $\int_{|x|\geq 1} |x|^n \nu(dx) < \infty$. In this case moments of X_t can be computed from its characteristic function by differentiation. In particular, the form of cumulants of X_t is:

$$E[X_t] = t(r + \int_{|x|\geq 1} x\nu(dx)) \tag{1.6}$$

$$c_2(X_t) = Var X_t = t(A + \int_R x^2\nu(dx)) \tag{1.7}$$

$$c_n(X_t) = t\int_R x^n\nu(dx) \quad for \quad n \geq 3 \tag{1.8}$$

Skewness and excess kurtosis of the increments $X_{\Delta t}$ or $X_{t+\Delta t} - X_t$ can be derived by this proposition.

$$skewness(X_{\Delta t}) = \frac{c_3(X_{\Delta t})}{c_2(X_{\Delta t})^{3/2}} = \frac{skewness(X_1)}{\sqrt{\Delta t}} \tag{1.9}$$

$$excess\, kurtosis(X_{\Delta t}) = \frac{c_4(X_{\Delta t})}{c_2(X_{\Delta t})^2} = \frac{kurtosis(X_1)}{\Delta t} \tag{1.10}$$

We can conclude that skewness decreases at the rate of $\Delta t^{1/2}$ and excess kurtosis decreases at the rate of Δt. This proposition also implies that the distributions of the increments of Lévy processes are leptokurtic (excess kurtosis is positive) as $c_4(X_t) > 0$.

1.2.2 Scaling Property and Self-similarity

Traditionally, scaling phenomena are observed and studied in physical sciences. In the 1990s, the availability of high-frequency data and computer technology made it possible to investigate scaling behavior in economic systems. Empirical studies show that the asset prices exhibit similar statistical properties at different time scales, which bring interest to implement scaling property in economics.

In mathematics, the scaling behavior is associated with stochastic processes exhibiting the self-similarity property, which is defined below:

Definition 8 ***Self-Similarity and Self-Similar Processes***

A stochastic process $\{X_t : t \geq 0\}$ is said to be self-similar if

$$X_{\lambda t} \stackrel{law}{=} \lambda^{\gamma} X_t, \tag{1.11}$$

where $\lambda > 0$ is a scaling factor. γ is called the self-similarity exponent.

This process is also called the self-similar process or γ-self-similar process.

A well-known self-similar process is the Brownian motion with self-similarity exponent $\gamma = 1/2$. Some of the other Lévy processes may also have self-similar properties, and they are named α-stable Lévy processes. However, self-similarity property does not solely appear in Lévy processes. It also exists in processes with dependent increments. For example, fractional Brownian motions, which have correlated increments, also show self-similarity.

If we take the unit time for X_t in (1.11), we have

$$X_t \stackrel{law}{=} t^{\gamma} X_1, \quad \forall\ t > 0, \tag{1.12}$$

which indicates the law for X_t at time t can be obtained from the law at the unit time, scaled by a coefficient t^{γ}.

Now let us study how the properties, including characteristic function, distribution function, tail behavior, and moments, of self-similar processes behave through time horizon.

- Characteristic function

$$\begin{aligned} & X_t \stackrel{law}{=} t^{\gamma} X_1 && (1.13) \\ \Leftrightarrow\ & E[e^{iuX_t}] = E[e^{iut^{\gamma}X_1}] \\ \Leftrightarrow\ & \Phi_{X_t}(u) = [\Phi_{X_1}(u)]^{t^{\gamma}} \quad \text{or} \quad \Psi_{X_t}(u) = t^{\gamma}\Psi_{X_1}(u) \end{aligned}$$

- Distribution function

cdf:

$$\begin{aligned} & X_t \stackrel{law}{=} t^{\gamma} X_1 \\ \Leftrightarrow\ & P(X_t \leq x) = P(t^{\gamma} X_1 \leq x) \\ \Leftrightarrow\ & F_{X_t}(x) = F_{X_1}(\frac{x}{t^{\gamma}}) && (1.14) \end{aligned}$$

pdf: differentiate (2.13), we get $f_{X_t}(x) = \frac{1}{t^{\gamma}} f_{X_1}(\frac{x}{t^{\gamma}})$

At the center $x = 0$,

$$f_{X_t}(0) = \frac{1}{t^\gamma} f_{X_1}(0) \tag{1.15}$$

- Center and Tail Behavior

Several authors have used (1.15) to test self-similarity on stock returns and estimate γ around the center of the density function. Mantegna and Stanley [54] studied the S&P 500 Index and concluded that $\gamma \approx 0.71$ and self-similar Lévy (a α-stable process) processes describe the dynamics of the *pdf* well at zero. However, this model fails at tails. Later, power-law distributions, along with self-similar processes, are proposed by numerous authors [18] [32] [30] to model the tail behavior of stock returns. If self-similar processes have power-law tail at X_1

$$P(|X_1| > x) \sim \frac{1}{x^\alpha},$$

then at other time scales

$$P(|X_t| > x) \sim \frac{t^{\gamma\alpha}}{x^\alpha}, \quad for \quad t > 0,$$

which means the tail behavior still exhibits power-law distribution with some scaling coefficient γ.

- Moments, variance, skewness, and kurtosis

Using Eq. (1.12), it's obvious to derive the moment at $t > 0$ from $t = 1$

Moment: $\quad E[X_t] = t^\gamma E[X_1]$

$$E[X_t^n] = t^{n\gamma} E[X_1^n]$$

Variance: $Var[X_t] = t^{2\gamma} Var(X_1)$

Skewness: $skew(X_t) = \frac{E[(X_t - EX_t)^3]}{[Var(X_t)]^{3/2}} = skew(X_1)$

Kurtosis: $kurt(X_t) = \frac{E[(X_t - EX_t)^4]}{[Var(X_t)]^2} = kurt(X_1)$

We can tell that skewness and excess kurtosis of self-similar processes do not change along the time horizon.

1.3 Modeling Stock Returns with Lévy and Scaling

1.3.1 Preliminary

As discussed in the previous section, the term structures of skewness and excess kurtosis (the relationship between skewness/excess kurtosis and the time horizon) exhibit different patterns under different models. If we assume the price is moved by independent news and it is the result of the accumulation of these independent identical shocks, then the stock price is led by Lévy processes. The skewness and excess kurtosis of the price fluctuations drop at the reciprocal of $\sqrt{t}$ and t, respectively. If the stock markets, as complex dynamic systems, exhibit scaling behavior as numerous authors have indicated, then the skewness and excess kurtosis of the price fluctuations keep constant at all time scales. These two postulations have been investigated by numerous authors. The empirical studies show that the term structures of skewness and excess kurtosis behave in between of these two approaches, that is, skewness and excess kurtosis decay slower than Lévy and but faster than scaling. Thus, it is natural to propose a model that combines these two ideas: at a chosen time scale, called unit time, the random variable of log-price increment (or price fluctuation, or log return) is split into two components, one runs as the accumulation of $i.i.d$ random variables, which is Lévy, and the other behaves as a scaling random variable along the time horizon. Again, we shall point out that this construction is only for the distributions of the stock returns at various time horizons, and do not necessarily have to be associated with any stochastic processes.

1.3.2 Self-decomposable Laws

The first step in modeling is to split the random variable at the unit time, which is related to a family of limit laws and its associated property, self-decomposability.

It is known that the stock prices are moved by many pieces of information or noises. If these pieces of information are considered as a sequence of independent random variables (not necessarily identical) $\{Z_i : i = 1, 2, ...\}$, then the price fluctuation ΔP_t is the consequence of the impacts from all Z_i. Let $S_n = \sum_{i=0}^{n} Z_i$ and rewrite it as $a_n S_n + b_n$. Lévy [42] and Khinchin [40] studied the asymptotic behavior of $a_n S_n + b_n$ and defined a family of laws called "class L." It states that there exist sequences of constants a_n, the scaling coefficients, and b_n, the centering constants, such that the distribution of $a_n S_n + b_n$ converges to the law of a random variable X, which belongs to a family of laws "class L." The class L laws are limit laws. The Central

Limit Theorem, which says the distribution of the normalized sum of a large number of $i.i.d$ random variables converges to Gaussian distribution, is a special case of the class L. As ΔP_t, the price change within t time horizon, is the outcome of many independent random variables appearing in t, it can be approximated as a random variable X which has the law of class L.

Sato [72] studied another class of random variables with self-decomposable property, which is defined below.

Definition 9 ***Self-decomposable Laws***

A random variable X is self-decomposable if for $\forall\ c \in (0,1)$

$$X \overset{law}{=} cX + X^{(c)}, \tag{1.16}$$

where $X^{(c)}$ is a random variable independent of X.

This means a self-decomposable random variable X can be decomposed into a partial of itself and another independent random variable. Furthermore, Sato [72] shows that the random variable X is self-decomposable if and only if it has class L distribution. Thus, we can study the property of the price fluctuation ΔP_t through the self-decomposable laws, which is relatively easier to handle than class L.

Self-decomposable distributions belong to the family of infinitely divisible laws [41]. Their characteristic functions are given by the following proposition [72]:

Proposition 10 ***Characteristic Function of Self-decomposable Laws***

The characteristic function of a self-decomposable random variable X is

$$E[e^{iux}] = \exp\left\{iru - \frac{1}{2}\sigma^2 u^2 + \int_R \left(e^{iux} - 1 - iux1_{|x|\le 1}\frac{g(x)}{|x|}dx\right)\right\}, \tag{1.17}$$

where r, σ are constants, $\sigma^2 \ge 0$, $\int_R (|x|^2 \Lambda 1)\frac{g(x)}{|x|}dx < \infty$, and $g(x)$ is an increasing function when $x < 0$ and an decreasing function when $x > 0$.

The Lévy measure of the self-decomposable laws has the form $\frac{g(x)}{|x|}$ with some constraints for $g(x)$ as indicated above. This kind of function $g(x)$ is called the self-decomposability characteristic (SDC) of the random variable X [12].

1.3.3 Mixed Model

Let the log-price change (or the log return) $X = \ln S_t - \ln S_0$ be a self-decomposable random variable within some chosen unit time ($t = 1$), e.g., one hour, one day. By Eq. (1.16), we have $X \stackrel{law}{=} cX + X^{(c)}$. The log return Y_t at other time scales t are developed from the two components at the unit time, that is, X_t is also decomposed into $cX(t)$, which runs as a Lévy process from the cX, and $t^{\gamma}X^{(c)}$, which is scaled from $X^{(c)}$.

$$Y_t = cX(t) + t^{\gamma}X^{(c)}. \tag{1.18}$$

The characteristic function of Y_t can be derived,

$$\begin{aligned} E[e^{iuY_t}] &= E[e^{iucX(t)+t^{\gamma}X^{(c)}}] \\ &= E[e^{iucX(t)}] \cdot E[e^{iut^{\gamma}X^{(c)}}] \end{aligned}$$

as from

$$
\begin{aligned}
E[e^{iuX}] &= E[e^{iu(cX+X^{(c)})}] \\
&= E[e^{iucX}] \cdot E[e^{iuX^{(c)}}],
\end{aligned}
$$

we can get

$$
E[e^{iuX^{(c)}}] = E[e^{iuX}]/E[e^{iucX}] = \exp(\Psi(u))/\exp(\Psi(cu)) = \exp(\Psi(u) - \Psi(cu)),
$$

where $\Psi(\cdot)$ is the characteristic exponent.

so

$$
E[e^{iuY_t}] = \exp\left\{t\Psi(cu) + \Psi(ut^{\gamma}) - \Psi(cut^{\gamma})\right\}. \tag{1.19}
$$

And the following proposition [22] provides the term structure of variance, skewness, and excess kurtosis in this model:

Proposition 11 ***Variance, Skewness, and Kurtosis of the Mixed Model***

Let Var(X), Skew(X), Kurt(X) be the variance, skewness, and excess kurtosis at the unit time $t = 1$ of a self-decomposable random variable X defined by (1.16). Then the variance, skewness and excess kurtosis of Y_t defined by Eq. (1.18) are:

$$
Var(Y_t) = Var(X)(c^2 t + (1 - c^2)t^{2\gamma} \tag{1.20}
$$

$$
Skew(Y_t) = \frac{Skew(X)}{\sqrt{t}}\left[\frac{c^3 + (1 - c^3)t^{3\gamma-1}}{(c^2 + (1 - c^2)t^{2\gamma-1})^{3/2}}\right] \tag{1.21}
$$

$$
Kurt(Y_t) = \frac{Kurt(X)}{t}\left[\frac{c^4 + (1 - c^4)t^{4\gamma-1}}{(c^2 + (1 - c^2)t^{2\gamma-1})^2}\right] \tag{1.22}
$$

Remarks:

(1) By simple calculation, we can see that $\left[\frac{c^3+(1-c^3)t^{3\gamma-1}}{(c^2+(1-c^2)t^{2\gamma-1})^{3/2}}\right] < 1$ and $\left[\frac{c^4+(1-c^4)t^{4\gamma-1}}{(c^2+(1-c^2)t^{2\gamma-1})^2}\right] < 1$ when $0 < c < 1$; thus, skewness decays at the rate between $\sqrt{t}$ and 0, and excess kurtosis decay at the rate between t and 0.

(2) It can be seen from this proposition that Y_t follows Lévy process when $c = 1$ and scaling process when $c = 0$.

1.4 Related Methods and Techniques

In the experimental procedure, a couple of methods and techniques are needed, including the law at the unit time, the maximum likelihood estimation, the fast Fourier transform (FFT), and the simulation. They are used in the statistical parameter estimation and analysis both in this chapter and the other two chapters. In this section, a brief review of these methods is provided.

1.4.1 Variance Gamma Process and the Associated Law

Variance Gamma Process as a Time-changed Brownian Motion

Brownian motion captures the essentiality of the stock markets but also misses some important empirical facts. The dynamics of the markets is not homogenous through time: that is, sometimes the markets are very active while other times they are relatively slow. Time, instead of being considered as a steady increasing process, can be viewed as a randomly changing time that is "economically relevant." Thus, we have a generalized version of Brownian motion with random time, which provides more flexibility to describe the log stock prices. If this random time follows a gamma process, the time-changed Brownian motion is called the Variance Gamma process [49][48].[1]

First, we define gamma process.

Definition 12 ***Gamma Process***

A gamma process $\gamma(t;\mu,\nu)$ is a Lévy process with independent gamma increments where μ is the mean rate and ν is the variance rate. The increment $g_h = \gamma(t+h;\mu,\nu) - \gamma(t;\mu,\nu)$ is a gamma random variable and has the probability density function (pdf)

$$f_h(g) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} g^{\alpha-1} e^{-g/\beta}, \tag{1.23}$$

where $\alpha = \frac{\mu^2 h}{\nu}$, $\beta = \frac{\nu}{\mu}$, $g \geq 0$.

The gamma process is a pure jump Lévy process, i.e., no diffusion part $W(t)$. The mean of g_h is μh and the variance is νh.

The characteristic function of the gamma process $\gamma(t)$ is

$$\Phi_{\gamma(t)}(u) = \left(\frac{1}{1 - iu\frac{\nu}{\mu}} \right)^{\mu^2 t/\nu}. \tag{1.24}$$

If replacing the calendar time t in Brownian motion with a random time $\gamma(t)$, the expectation of $\gamma(t)$ should equal t, $E[\gamma(t)] = t$. Thus, the gamma process must increase with a unit mean rate, which means $\mu = 1$ and $\gamma(t;1,\nu)$ are used to model the time in this time-changed Brownian motion.

[1] Most of the results in this section, if not indicated, come from Ref [48].

Given a gamma process with a unit mean rate, we can build a Variance Gamma process.

Definition 13 ***Variance Gamma Process***

Let $b(t;\theta,\sigma)$ be a Brownian motion with drift θ and standard deviation σ, $b(t;\theta,\sigma) = \theta t + \sigma W(t)$. Let $\gamma(t;1,\nu)$ be an independent gamma process with unit mean rate.

Then the Variance Gamma process (VG) is defined as

$$\begin{aligned} X(t;\sigma,\nu,\theta) &= b(\gamma(t;1,\nu);\theta,\sigma) \\ &= \theta\gamma(t;1,\nu) + \sigma W(\gamma(t;1,\nu)), \end{aligned} \tag{1.25}$$

a time-changed Brownian motion with gamma random time.

§ Properties of the VG Process

The random variable $X(t) = \theta\gamma(t) + \sigma W(\gamma(t))$ of the VG process in the time interval t contains two independent random parts: a Gaussian random variable $W(t)$ and a gamma random variable $\gamma(t)$. Thus, the independence can let us conveniently use the conditional expectation method to derive its pdf and characteristic function.

The pdf of $X(t)$ can be obtained from a normal density function conditioned on a gamma random variable. So we can integrate the gamma part and get the density function of $X(t)$, which is given below:

$$f_{X(t)}(x) = \int_0^\infty \frac{1}{\sigma\sqrt{2\pi g}} \exp\left\{-\frac{(x-\theta g)^2}{2\sigma^2 g}\right\} \cdot \frac{g^{\frac{t}{\nu}-1}\exp(-\frac{g}{\nu})}{\nu^{\frac{t}{\nu}}\Gamma(\frac{t}{\nu})} dg. \tag{1.26}$$

Similarly, the characteristic function of $X(t)$ can be attained by conditional expectation on the gamma random variable $g(t)$:

$$\Phi_{X(t)}(u) = \left(\frac{1}{1 - i\theta\nu u + u^2 \cdot (\sigma^2\nu/2)} \right)^{\frac{t}{\nu}}. \tag{1.27}$$

Because the characteristic function has a much simpler expression than the density function, it is used most of the time.

Another representation for the VG process is to interpret it as the difference of two independent gamma processes,

$$X(t) = \gamma_p(t) - \gamma_n(t), \tag{1.28}$$

where $\gamma_p(t)$ is a gamma process representing positive jumps and $\gamma_n(t)$ is an independent gamma process with negative jumps. The VG process is the result of the effect of these two processes, which implies that the VG process is also a pure jump Lévy process.

The Lévy measure (or Lévy density) can be determined from the characteristic function,

$$k_X(dx) = \begin{cases} c\exp(Gx)\left|x\right|^{-1} dx, & x < 0 \\ c\exp(-Mx)x^{-1}dx, & x > 0 \end{cases}, \tag{1.29}$$

where

$$c = \frac{1}{\nu} > 0 \tag{1.30}$$

$$G = \left(\sqrt{\frac{\theta^2\nu^2}{4} + \frac{\sigma^2\nu}{2}} - \frac{\theta\nu}{2} \right)^{-1} > 0 \tag{1.31}$$

$$M = \left(\sqrt{\frac{\theta^2\nu^2}{4} + \frac{\sigma^2\nu}{2}} + \frac{\theta\nu}{2} \right)^{-1} > 0. \tag{1.32}$$

From (1.29) we can see that the first part ($x < 0$) is the Lévy measure of the negative jumps (γ_n) with parameter C, G, and the second part ($x > 0$) is the Lévy measure of the positive jumps (γ_p) with parameter C, M.

The central moments, skewness, and kurtosis can also be derived. The results are listed below:

$$mean = \theta \tag{1.33}$$

$$variance = \sigma^2 + \nu\theta^2 \tag{1.34}$$

$$skewness = \theta\nu(3\sigma^2 + 2\nu\theta^2)/(\sigma^2 + \nu\theta^2)^{3/2} \tag{1.35}$$

$$kurtosis = 3(1 + 2\gamma - \nu\sigma^4(\sigma^2 + \nu\theta^2)^{-2}), \tag{1.36}$$

which show that the VG process has the flexibility to control skewness and excess kurtosis, unlike Brownian motion which has fixed values.

The Lévy measure of the VG process (1.29) indicates the VG random variable has a self-decomposable distribution (Proposition 10). Thus, the VG random variable is a candidate to be the building blocks at the unit time in the mixed model.

1.4.2 Stock Price Dynamics with the VG Process

§ Preliminary - Stock Market Models

Stock price can be modeled as a stochastic process

$$S_t = S_0 \exp(\mu t + X_t), \tag{1.37}$$

where S_t, S_0 are stock prices at time t, 0, respectively, X_t represents a stochastic process, and μ is the mean rate of returns of the stock. However, the stock price S_t in (1.37) is not a martingale under the statistical probability measure (or usually called "physical measure"). The following proposition provides a way to make S_t a martingale [Theorems 2.5.1 and 2.5.3 in [72]].

Proposition 14 *Let $\{X_t : t \geq 0\}$ be a real-valued process with independent increment. If $E[e^{aX_t}] < \infty$ for some real-valued a, then $\left(\frac{e^{aX_t}}{E[e^{aX_t}]}\right)_{t\geq 0}$ is a martingale at all $t \geq 0$.*

Let $a = 1$ and assume $E[e^{X_t}] < \infty$, Eq (1.37) can be rewritten as

$$S_t = S_0 \frac{\exp(\mu t + X_t)}{E[e^{X_t}]}. \tag{1.38}$$

Take the expectation of S_t, we get $E[S_t] = S_0 e^{\mu t} E\left[\frac{e^{X_t}}{E[e^{X_t}]}\right] = S_0 e^{\mu t}$, or $S_0 = \frac{E[S_t]}{e^{\mu t}}$, which means $\frac{S_t}{e^{\mu t}}$, the stock price discounted by its drift term $e^{\mu t}$, is a martingale under the physical measure.

§ **VG Stock Market Model**

Let X_t in (1.38) be the VG random variable and $\{X_t : t \geq 0\}$ the associated VG process, we have the VG stock market model.

By the VG characteristic function (1.27),

$$
\begin{aligned}
E[\exp(X_{VG}(t))] &= \Phi_{X_{VG}(t)}(-i) \\
&= \left(\frac{1}{1 - i\theta\nu(-i) + (-i)^2 \cdot (\sigma^2\nu/2))} \right)^{\frac{t}{\nu}} \\
&= \exp\left(-\frac{t}{\nu} \ln\left(1 - \theta\nu - \frac{\sigma^2\nu}{2}\right)\right),
\end{aligned}
$$

denotes $\omega = \frac{1}{\nu} \ln\left(1 - \theta\nu - \frac{\sigma^2\nu}{2}\right)$

Eq (1.38) becomes

$$S_t = S_0 \exp\left\{\mu t + X_{VG}(t) + \omega t\right\}, \tag{1.39}$$

which is the stochastic process for the stock price. The log-price return $\ln\left(\frac{S_t}{S_0}\right)$ follows $\mu t + X_{VG}(t) + \omega t$.

Use the density function (1.29) of the VG random variable $X_{VG}(t)$ and integrate the gamma random variable, we can derive the density function (pdf) of the log-price return $\ln\left(\frac{S_t}{S_0}\right)$.

Proposition 15 *Density Function (pdf) of VG Log-price Returns*

Let $r(t) = \ln\left(\frac{S(t)}{S(0)}\right)$ *be the log-price return, and* S_t *follow Eq (1.39). Under the* *physical measure, the pdf of $r(t)$ is given by*

$$f(r(t)) = \frac{2\exp\left(\frac{\theta z}{\sigma^2}\right)}{\sqrt{2\pi}\nu^{\frac{t}{\nu}}\sigma\Gamma(\frac{t}{\nu})} \cdot \left(\frac{z^2}{\frac{2\sigma^2}{\nu}+\theta^2}\right)^{\frac{t}{2\nu}-\frac{1}{4}} \cdot K_{\frac{t}{\nu}-\frac{1}{2}}\left(\frac{1}{\sigma^2}\cdot\sqrt{z^2\left(\frac{2\sigma^2}{\nu}+\theta^2\right)}\right), \quad (1.40)$$

where K is the modified Bessel function of the second kind, and $z = r(t) - \mu t - \frac{t}{\nu}\ln\left(1-\theta\nu-\frac{\sigma^2\nu}{2}\right)$.

As $\ln\left(\frac{S_t}{S_0}\right) = \mu t + X_{VG}(t) + \omega t$, the characteristic function of the log-price return $r(t)$ is easier to attain

$$\begin{aligned} E[e^{iur(t)}] &= E\left[e^{iu(\mu t + X_{VG}(t)+\omega t)}\right] \\ &= e^{iu(\mu+\omega)t}\cdot\Phi_{VG}(u) \\ &= \exp\{iu(\mu+\omega)t\}\cdot\left(1-i\theta\nu u-\frac{\sigma^2\nu}{2}u^2\right)^{-\frac{t}{\nu}}. \end{aligned} \quad (1.41)$$

The characteristic function has a much simpler form than the density function. As these two functions have a one-on-one relationship connected by Fourier transform, people use the characteristic function most of the time.

1.4.3 Maximum Likelihood Estimation (MLE)

Probability distribution can model the log-stock returns with a fixed interval, e.g., daily returns, hourly returns, etc. Given a set of sample points, the maximum likelihood estimation (MLE) can be employed to estimate the model parameters by fitting the data to the statistical model.

Let $x_1, x_2, ..., x_n$ be n *i.i.d* sample data points collected from a population. The pdf of a proposed distribution is $f(x; \overrightarrow{\theta})$, where $\left\{\overrightarrow{\theta} : \theta_1, \theta_2, ..., \theta_k\right\}$ is the parameter set. The likelihood function is defined as

$$L(\overrightarrow{\theta}|x) = \prod_{i=1}^{n} f(x_i; \overrightarrow{\theta}). \tag{1.42}$$

The method of maximum likelihood function is to estimate $\overrightarrow{\theta}$ by finding values of a parameter set $\widehat{\theta}$ which maximize $L(\overrightarrow{\theta}|x)$. Eq (1.42) can be rewritten in the logarithm version,

$$\log L(\overrightarrow{\theta}|x) = \sum_{i=1}^{n} f(x_i; \overrightarrow{\theta}),$$

which is called log likelihood function. Maximization is then conducted on this function instead.

Due to the complexity of pdf formulas, only numerical procedure is feasible to perform MLE in most cases. When the closed-form of pdf is not known or it's too complicated to use in MLE, characteristic function is employed. At this situation, the conversion from characteristic function to pdf is performed by Fourier transform. A useful technique in the Fourier transform will be discussed in the next section.

1.4.4 Fast Fourier Transform (FFT)

The numerical MLE method is realized through optimization procedure, which requires the calculation of the likelihood function at each iteration. If the pdf is not known or computationally feasible, then the values of pdf are attained from characteristic function through Fourier transform, $f(x; \overrightarrow{\theta}) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iux}\Phi_X(u)du$.

The fact that $f(x;\overrightarrow{\theta})$ is a real-valued function implies that $\Phi_X(u)$ has even real part and odd imaginary part, which derives the transform formula to $f(x;\overrightarrow{\theta}) = \frac{1}{\pi}\int_0^\infty e^{-iux}\Phi_X(u)du$.The latter can be numerically calculated by

$$f(x_k;\overrightarrow{\theta}) = \frac{1}{\pi}\sum_{j=1}^{N} e^{-iu_jx_k}\Phi_X(u_j)\cdot\eta, \tag{1.43}$$

where $\eta = \Delta u$, $k = 0, 1, ..., N-1$. Eq (1.43) is the discrete Fourier transform (DFT) and requires $O(N^2)$ operations: N computations are needed for each x_k, and there are N number of x_k. The fast Fourier transform (FFT) is an efficient algorithm to compute the DFT. It produces exactly the same result as DFT and only requires $O(N\log N)$ operations. There are several FFT algorithms. What we use in the Matlab code is based on FFTW [81], and the details of this algorithm can be found in Ref [78].

The ready-to-use formula for the numerical computation of pdf can be derived from Eq (1.43)[2]. In this formula, N usually takes the value of the power of 2. Given the step size η, the upper limit of u is $a = N\eta$, the grid point $u_j = (j-1)\eta$. Let λ be the length of grid of x, then x ranges from $-b$ to b, where $b = \frac{\lambda N}{2}$, and the grid point $x_k = -b + \lambda(k-1)$ for $k = 1, 2, ..., N$.

With the above setting, Eq (1.43) becomes

$$\begin{aligned} f(x_k;\overrightarrow{\theta}) &= \frac{1}{\pi}\sum_{j=1}^{N} e^{-i(j-1)\eta\cdot(-b+\lambda(k-1))}\Phi_X(u_j)\cdot\eta \\ &= \frac{1}{\pi}\sum_{j=1}^{N} e^{-i(j-1)(k-1)\lambda\eta}\cdot e^{ibu_j}\Phi_X(u_j)\cdot\eta. \end{aligned} \tag{1.44}$$

[2]The procedure is based on the work by Carr and Madan [13].

The formula of the standard discrete Fourier transform is

$$Z(k) = \sum_{j=1}^{N} e^{-\frac{2\pi i}{N}(j-1)(k-1)} \cdot z(j). \tag{1.45}$$

Comparing Eq (1.44) with Eq (1.45), we get

$$\lambda\eta = \frac{2\pi}{N}.$$

With properly chosen values of η and N, Eq (1.44) can be immediately used in the numerical procedure. It can be further revised by incorporating Simpon's rule,

$$f(x_k; \overrightarrow{\theta}) = \frac{1}{\pi} \sum_{j=1}^{N} e^{-\frac{2\pi i}{N}(j-1)(k-1)} \cdot e^{ibu_j} \Phi_X(u_j) \cdot \frac{\eta}{3}(3 + (-1)^j - \delta_{j-1}), \tag{1.46}$$

where δ_j is the Kronecker delta function with value one at $j = 1$ and zero elsewhere.

Comparing Eq (1.45) and Eq (1.46), we have

$$z(j) = \frac{1}{\pi} e^{ibu_j} \Phi_X(u_j) \cdot \frac{\eta}{3}(3 + (-1)^j - \delta_{j-1}), \tag{1.47}$$

which is the input in the FFT calculation.

1.4.5 VG Random Number Simulation

Simulation of VG random number is needed in the goodness of fit of the VG model in our work. The VG random number can be simulated directly from the definition of the VG process. Recall the VG process $X(t)$ is a drifted Brownian motion with its time following the gamma process, which is independent of the Brownian motion (Definition 13),

$$X(t) = \theta\gamma(t) + \sigma W(\gamma(t)). \tag{1.48}$$

To generate the random number $X(t)$ at time t, we can simulate two independent random variables separately: a standard normal random variable $W(1)$ and a gamma random variable $\gamma(t)$. $W(\gamma(t))$ in (1.48) is a Gaussian random variable, so it can be written as $W(\gamma(t)) = \sqrt{\gamma(t)}W(1)$. As $W(1)$ and $\gamma(t)$ are independent, the product of these two simulated numbers gives us $W(\gamma(t))$. Thus, the random number $X(t)$ in (1.48) is attained.

The above procedure is summarized below:

Algorithm 16 ***VG Random Number Simulation***

Step 1. Generate a standard normal random variable $z \sim N(0,1)$

Step 2. Generate a gamma random number $g \sim Gamma(\frac{t}{\nu}, \nu)$

Step 3. The VG random number $X(t)$ *at time* t *is given by* $X(t) = \theta \cdot g + \sigma\sqrt{g}z$

1.5 Numerical Implementation and Results

1.5.1 Data, Sketch of the Procedure, and Brief Introduction of the Statistical Analysis

The first largest 495 stocks are chosen from S&P 500 and S&P MidCap400 [83]. The data are the log-price returns ranging from January 2, 2003 to December 29, 2006, which include nonoverlapping one-hour, two-hour, three-hour, daily, weekly, and biweekly returns.

Sketch of the procedure is listed below:

- Step 1. Estimate the log return data at the unit time, which is chosen to be one hour, for each stock using the Variance Gamma distribution.

- Step 2. Estimate the log return data at longer horizons for each stock using three different models, which are the VG iid model (because the random variables at longer horizons are the cumulants of i.i.d random variables from the unit time), the VG scaling model (with VG law at the unit time), and the VG mixed model. The longer horizons are two hours, three hours, one day, one week, and two weeks.

- Step 3. Conduct statistical goodness of fit for each model. The statistical analyses include the Kolmogorov-Smirnov test, the Kolmogorov distance, the modified Kolmogorov distance, χ^2-distance, L_1 and L_2 distances, which are briefly described below.

○ **Kolmogorov-Smirnov test (KS-test)**

The KS-test is a statistical hypothesis test to determine whether the two data sets differ significantly. The null hypothesis is that the two data sets are from the same continuous distribution, and the alternative hypothesis is that they belong to different distributions. The significance level is usually taken as 5%.

The advantage of the KS-test is that it does not have any assumptions about the distributions of the data. On the other side, the cost or the disadvantage is it is less sensitive or accurate than other tests.

○ **Various statistical distances to measure how close the two probability distributions are**

We have two probability distributions: one is the empirical distribution obtained from the data; the other is the fitted probability distribution. If the two distributions are similar, then graphically their cdf curves are close. The distances between these two curves can be used to measure how good the fit is.

Let F_1 and F_2 be the cdf of two distributions. A couple of distances have been defined.

· *Kolmogorov distance*

$$dist_K(F_1, F_2) = \sup_{x \in R} |F_1(x) - F_2(x)|$$

measures the largest distance between the two cdfs

· *Modified Kolmogorov distance*

$$dist_{\widetilde{K}}(F_1, F_2) = \sup_{|x| > \varepsilon} |F_1(x) - F_2(x)|$$

The empirical cdf has a large jump at $x = 0$, which is called the 0-return effect. To eliminate this effect, the distance is not measured in a small area around 0.

· χ^2-*distance*

Denote n_1, n_2 the m-dimensional frequency vectors from samples of two distributions. The χ^2-distance is calculated by:

$$dist_{\chi^2}(F_1, F_2) = \sum_{i=1}^{m} \left(\frac{n_{1i}}{\Sigma n_{1i}} - \frac{n_{2i}}{\Sigma n_{2i}} \right)^2 / (n_{1i} + n_{2i})$$

where n_{1i}, n_{2i} are the elements in vectors n_1, n_2.

· L_p-*distance*

$$dist_{L_p}(F_1, F_2) = \left(\int_{\mathbb{R}} |F_1(x) - F_2(x)|^p \, dx \right)^{1/p}, \;\; p = 1, 2, ...$$

1.5.2 Statistical Estimation at the Unit Time

We first estimate the VG parameters (σ, ν, θ) for the hourly demeaned returns using MLE. The KS-test is then performed on the observed data and the simulated data using the fitted VG model. The significance level is $\alpha = 0.05$. Among 495 stocks, the VG model only has a goodness fit for one quarter (126) of the stocks, with p-value greater than 5%. Further estimations at longer horizons will be conducted only on these 126 stocks.

The statistics of the estimated parameters are presented in Table 1.1, including mean, standard deviation, minimum, maximum, quantile 1/4, quantile 3/4, and median. To graphically illustrate the statistical fit, the fitted pdf and its empirical counterpart are shown in Figure 1-1 for WMT (Walmart).

	σ	ν	θ
mean	0.4473	1.32E-04	0.0397
std.	0.0936	2.27E-05	0.3282
min.	0.2788	8.73E-05	-0.5000
quantile 1/4	0.3694	1.13E-04	-0.1886
median	0.4436	1.31E-04	0.0763
quantile 3/4	0.5124	1.48E-04	0.2940
max	0.6713	2.11E-04	0.5000

Table 1.1: Statistics of the Estimated VG Parameters (at unit time = 1 hour)

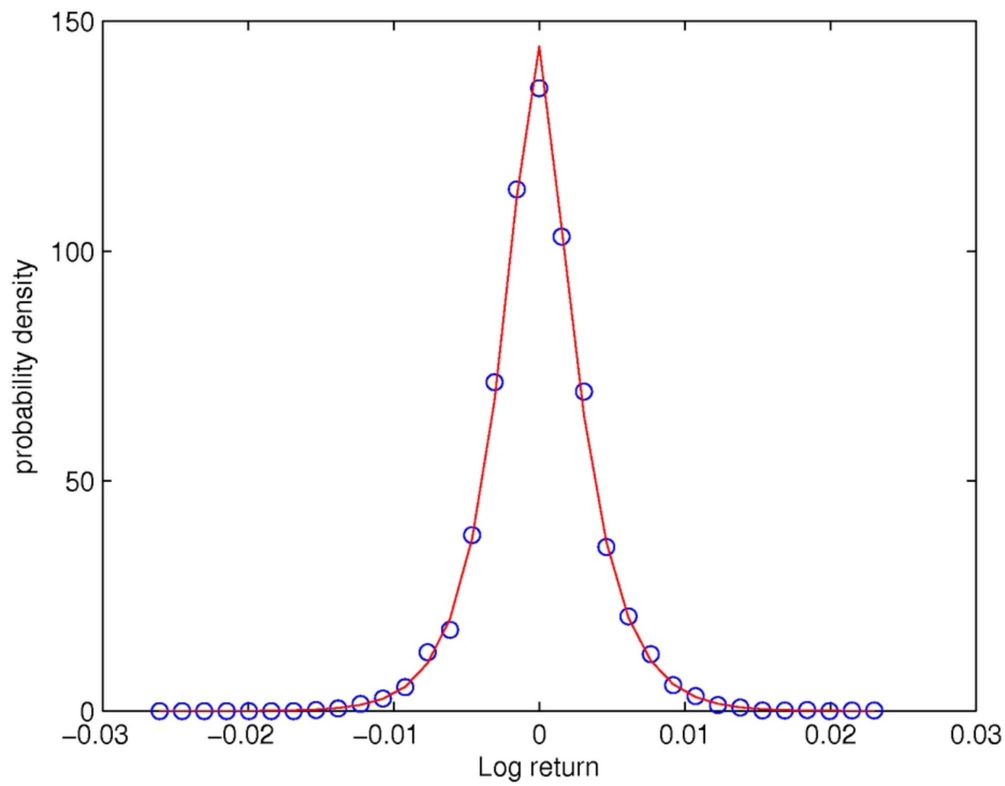


Figure 1-1: VG fit to WMT at 1-hour time scale

1.5.3 Statistical Estimation at Longer Horizons

The distribution of the mixed model is based on the distribution at the unit time, which means the estimated parameters of the base distribution $(\widehat{\sigma}, \widehat{\nu}, \widehat{\theta})$ is used at longer horizons. The remaining two parameters, c and γ, are to be estimated, where c represents the proportion of the random variable behaving like Lévy processes, and γ represents the scaling coefficient of the remaining component of the random variable. The maximum likelihood estimation is performed on the nonoverlapping demeaned log return data at each longer horizon, including two hours, three hours, one day, one week, and two weeks.

The summary statistics of the estimated $(\widehat{c}, \widehat{\gamma})$ is presented in Table 1.2 $(\widehat{c})$ and Table 1.3 $(\widehat{\gamma})$. The parameters of the VG mixed model are hard to estimate at even longer horizons, such as half year because of the lack of data. The average values of c and γ are around 0.4 in our estimation. Thus, c and γ may be properly assumed to have value of 0.4 at horizons where estimation is not feasible due to lack of data.

The performance of the VG mixed model is compared to other two models, the VG iid model and the VG scaling model. The distribution of the VG iid model at longer horizon t is the accumulated $i.i.d$ VG variables to time t, so it is known if the distribution at the unit time is given. In the VG scaling model, the random variable X_t is $t^{\gamma\prime} X_1$, a scaled version of X_1 from the unit time $t = 1$. Thus, we need to estimate $\gamma\prime$, the scaling coefficient at time t. Sample graphs of the fitted and empirical pdf of these three models at various time horizons for WMT are shown in Figure 1-2 to Figure 1-6.

	2 hours	*3 hours*	*1 day*	*1 week*	*2 weeks*
mean	0.6404	0.7204	0.3390	0.4274	0.4713
std.	0.1712	0.1122	0.1116	0.1249	0.1421
min.	0.0064	0.3985	0.0002	0.0504	0.0002
quantile 1/4	0.5882	0.6518	0.2716	0.3436	0.3959
median	0.6591	0.7420	0.3334	0.4310	0.5025
quantile 3/4	0.7665	0.7955	0.4198	0.5165	0.5764
max	0.9144	0.9792	0.5759	0.7142	0.6927

Table 1.2: Statistics of the Estimated VGMixed Parameter c

	2 hours	*3 hours*	*1 day*	*1 week*	*2 weeks*
mean	0.4966	0.5117	0.3277	0.3344	0.2671
std.	0.0617	0.0968	0.0536	0.111	0.1648
min.	0.2074	0.3043	0.0456	4.3e-06	1.9e-08
quantile 1/4	0.4674	0.4666	0.3074	0.3258	0.0536
median	0.5005	0.5028	0.3413	0.3634	0.3451
quantile 3/4	0.5363	0.5439	0.3621	0.3998	0.3864
max	0.6359	1.0000	0.4239	0.4681	0.4782

Table 1.3: Statistics of the Estimated VGMixed Parameter γ

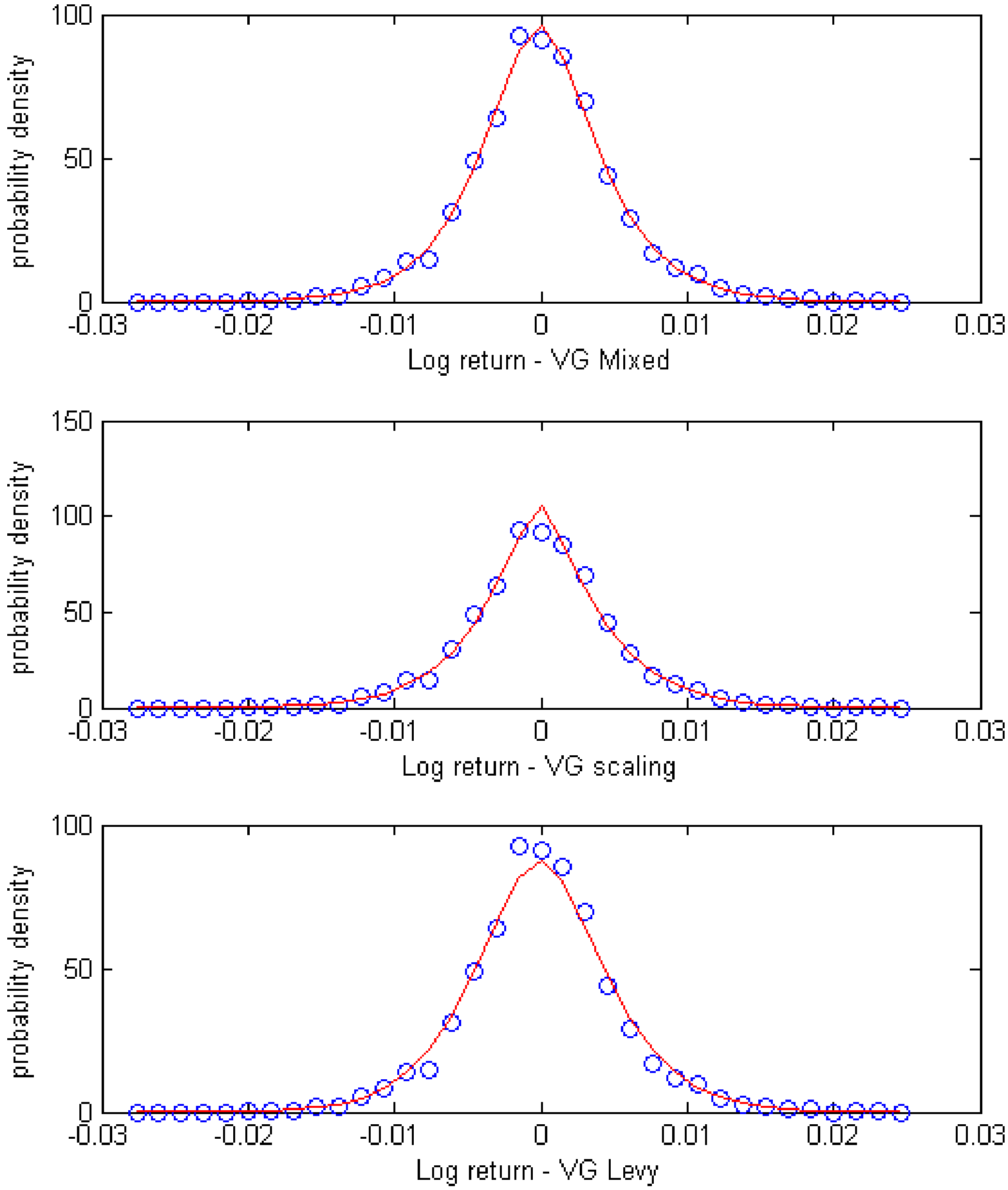


Figure 1-2: Statistical fit to WMT at 2hr timescale

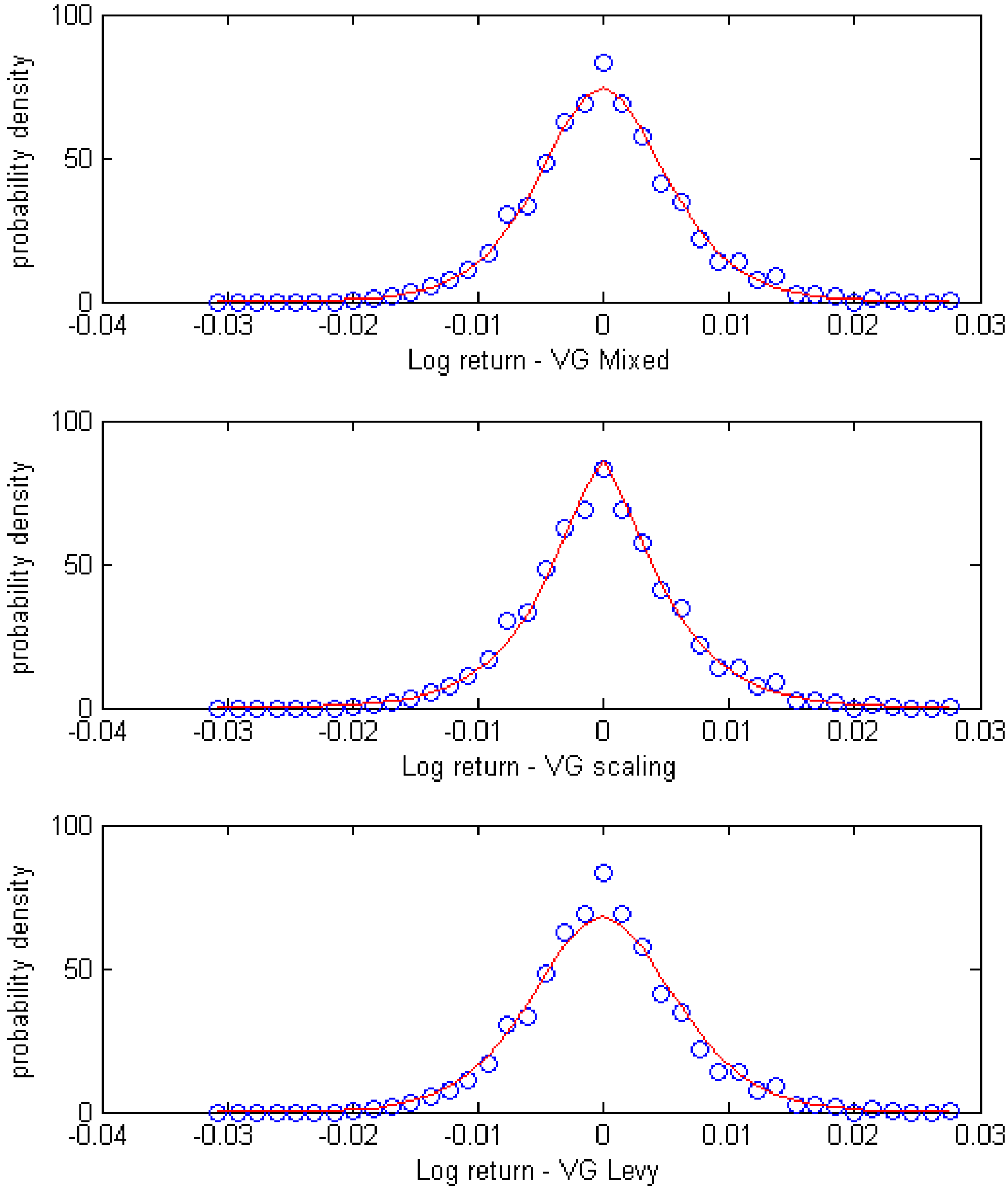


Figure 1-3: Statistical fit to WMT at 3hr timescale

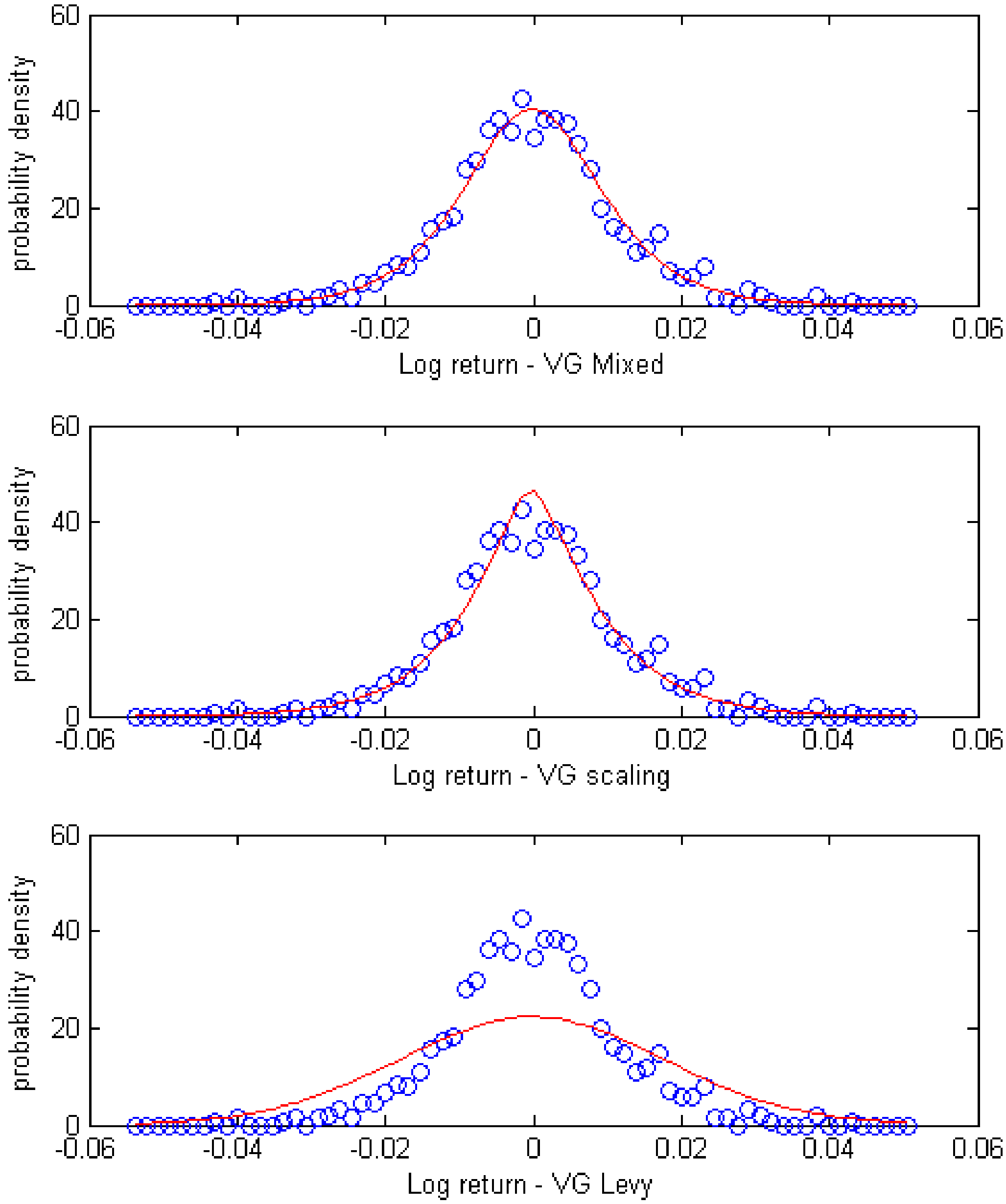


Figure 1-4: Statistical fit to WMT at 1d timescale

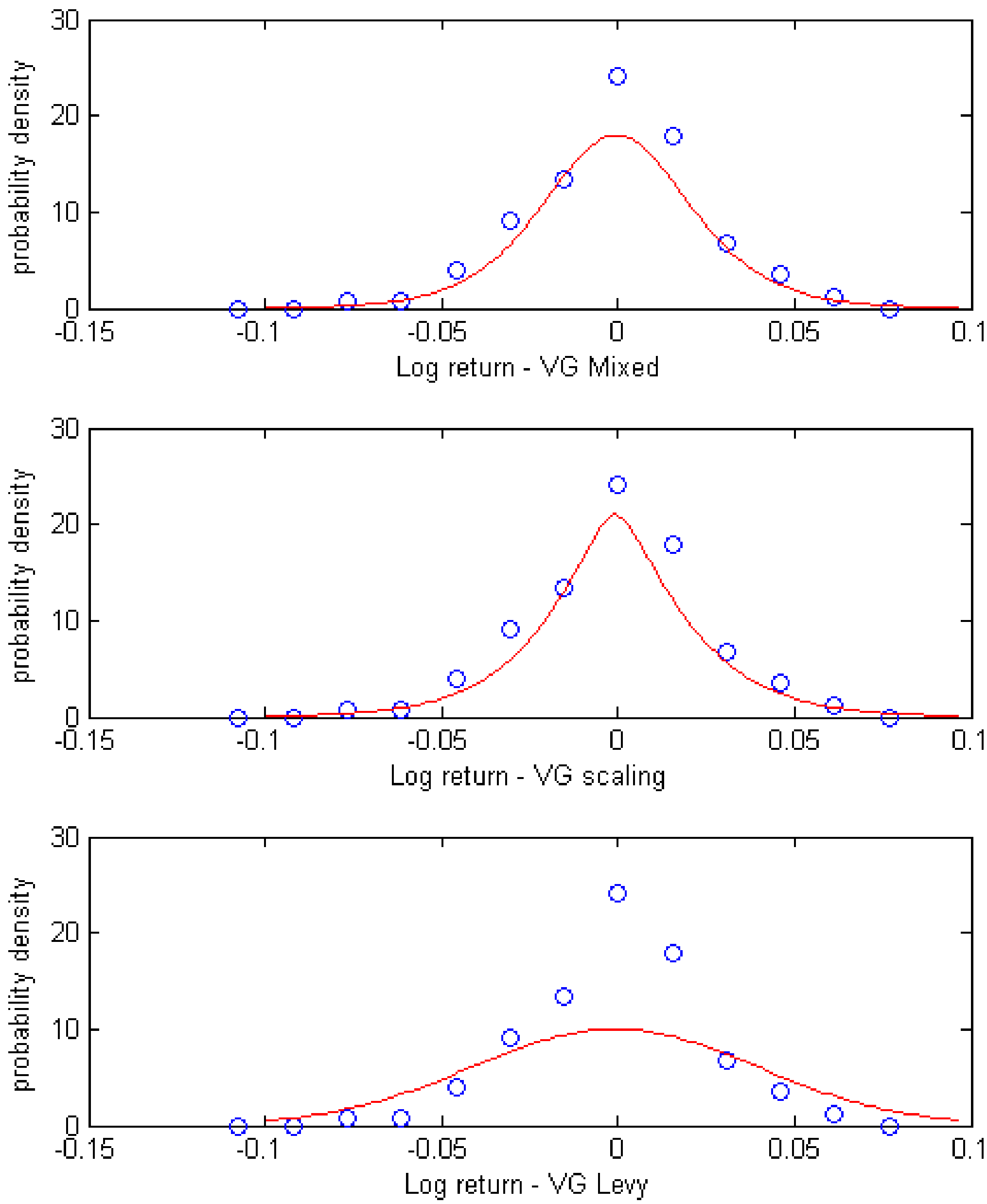


Figure 1-5: Statistical fit to WMT at 1w timescale

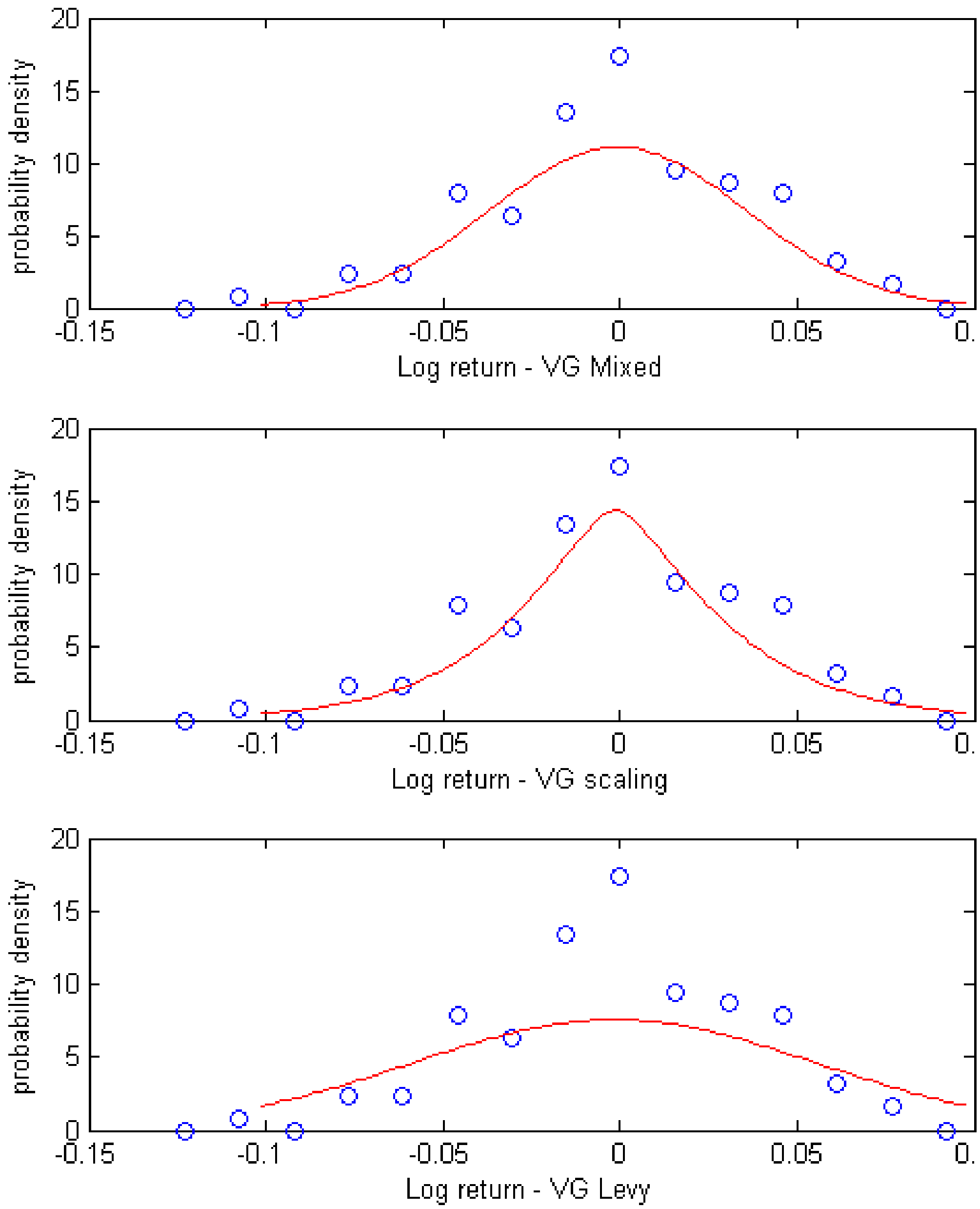


Figure 1-6: Statistical fit to WMT at 2w timescale

1.5.4 Statistical Analysis for Model Performance Comparison

A couple of statistical analyses, described in Section 1.5.1, are conducted to compare the performances of the three models, namely VG mixed, VG scaling, VG iid. The statistics of five distances are presented in Table 1.4 (mean) and Table 1.5 (std.).

The KS-test is also employed at the longer horizons to test and compare the three models. It examines whether the observed data and the simulated data from the fitted model belong to the same distribution. the p-value is attained and graphed in figures (Figure 1-7 to Figure 1-11) whose $x - axis$ is the p-value and $y - axis$ is the proportion of stocks whose p-value exceeds the corresponding p-value on the $x - axis$. The graph shows the VG scaling model has better performance than the VG iid model, and the VG mixed model outperforms both models at all horizons. At longer horizon (2-week), the VG iid model has better performance than the VG scaling model, which also confirms the empirical fact that the return distribution asymptotically approaches Gaussian, which has i.i.d increment along the time horizon.

1.6 Conclusion

This chapter investigates the performance of three stock return models, the VG iid model, the VG scaling model, and a mixed version of the two models, the VG mixed model. The first two approaches have different effects on skewness and excess kurtosis along the time horizon. The empirical study shows that skewness and excess kurtosis

Mean	$dist_K$	$dist_{\widetilde{K}}$	$dist_{\kappa^2}$	$dist_{L_1}$	$dist_{L_2}$
2h-VG Mixed	0.0064	0.0047	0.0001	0.0008	4.00E-06
2h-VG Scaling	0.0122	0.0060	0.0002	0.0013	7.00E-06
2h-VG iid	0.0160	0.0137	0.0003	0.0018	9.00E-06
3h-VG Mixed	0.0096	0.0065	0.0002	0.0011	6.00E-06
3h-VG Scaling	0.0170	0.0101	0.0004	0.0019	1.40E-05
3h-VG iid	0.0204	0.0182	0.0005	0.0024	1.20E-05
1d-VG Mixed	0.0144	0.0118	0.0007	0.0024	7.00E-06
1d-VG Scaling	0.0247	0.0224	0.0011	0.0038	1.40E-05
1d-VG iid	0.1157	0.1157	0.0083	0.0265	3.04E-04
1w-VG Mixed	0.0250	0.0239	0.0027	0.0070	1.60E-05
1w-VG Scaling	0.0346	0.0342	0.0036	0.0094	1.90E-05
1w-VG iid	0.1143	0.1143	0.0164	0.0369	2.73E-04
2w-VG Mixed	0.0388	0.0381	0.0053	0.0121	2.80E-05
2w-VG Scaling	0.0503	0.0502	0.0065	0.0155	3.30E-05
2w-VG iid	0.1140	0.1140	0.0204	0.0411	2.31E-04

Table 1.4: Mean of the statistical distances of the three models at different timescales

Std.	$dist_K$	$dist_{\widetilde{K}}$	$dist_{\kappa^2}$	$dist_{L_1}$	$dist_{L_2}$
2h-VG Mixed	0.0024	0.0021	0.0001	2.00E-04	2.00E-06
2h-VG Scaling	0.0052	0.0022	0.0001	4.00E-04	5.00E-06
2h-VG iid	0.0056	0.0054	0.0001	6.00E-04	5.00E-06
3h-VG Mixed	0.0037	0.0028	0.0001	4.00E-04	4.00E-06
3h-VG Scaling	0.0060	0.0041	0.0001	4.00E-04	1.00E-05
3h-VG iid	0.0064	0.0063	0.0002	7.00E-04	7.00E-06
1d-VG Mixed	0.0062	0.0046	0.0003	8.00E-04	4.00E-06
1d-VG Scaling	0.0094	0.0089	0.0004	1.00E-03	1.10E-05
1d-VG iid	0.0129	0.0129	0.0017	3.20E-03	2.29E-04
1w-VG Mixed	0.0103	0.0095	0.0012	2.30E-03	9.00E-06
1w-VG Scaling	0.0131	0.0129	0.0013	2.60E-03	1.30E-05
1w-VG iid	0.0211	0.0211	0.0041	6.40E-03	1.66E-04
2w-VG Mixed	0.0124	0.0124	0.0025	4.30E-03	1.60E-05
2w-VG Scaling	0.0175	0.0177	0.0026	4.80E-03	2.00E-05
2w-VG iid	0.0228	0.0228	0.0063	8.90E-03	1.34E-04

Table 1.5: Std. of the statistical distances of the three models at different timescales

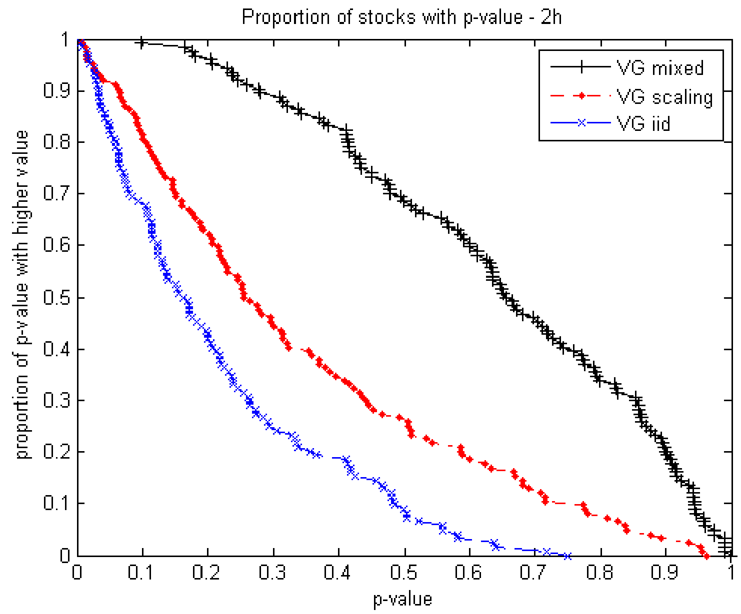


Figure 1-7: Proportion of stocks with p-value greater than certain level (2-hour timescale)

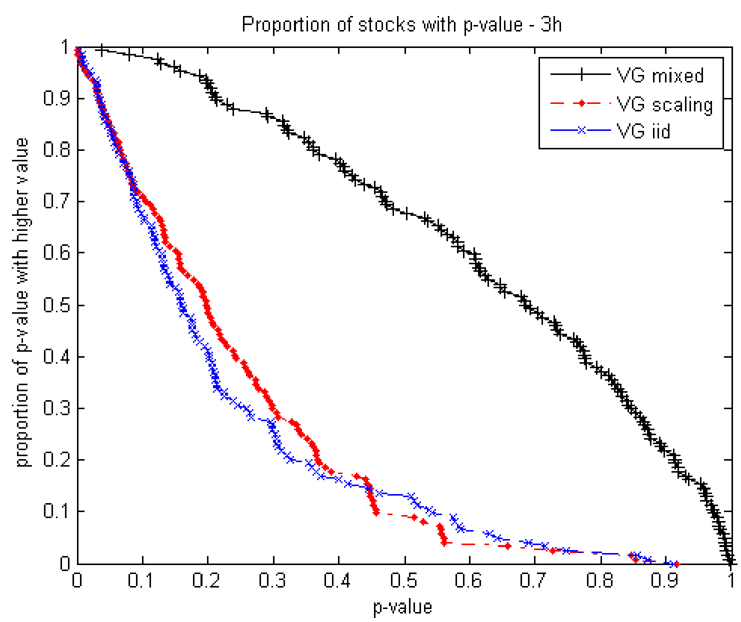


Figure 1-8: Proportion of stocks with p-value greater than certain level (3-hour timescale)

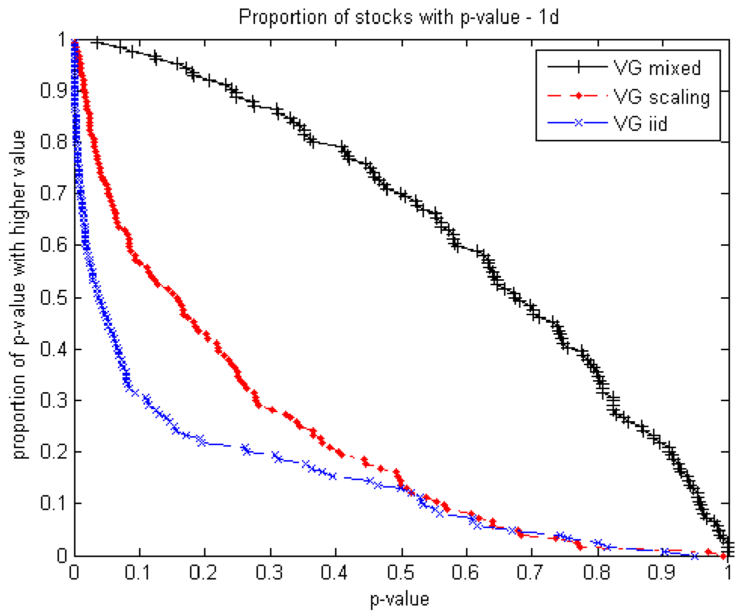


Figure 1-9: Proportion of stocks with p-value greater than certain level (1-day timescale)

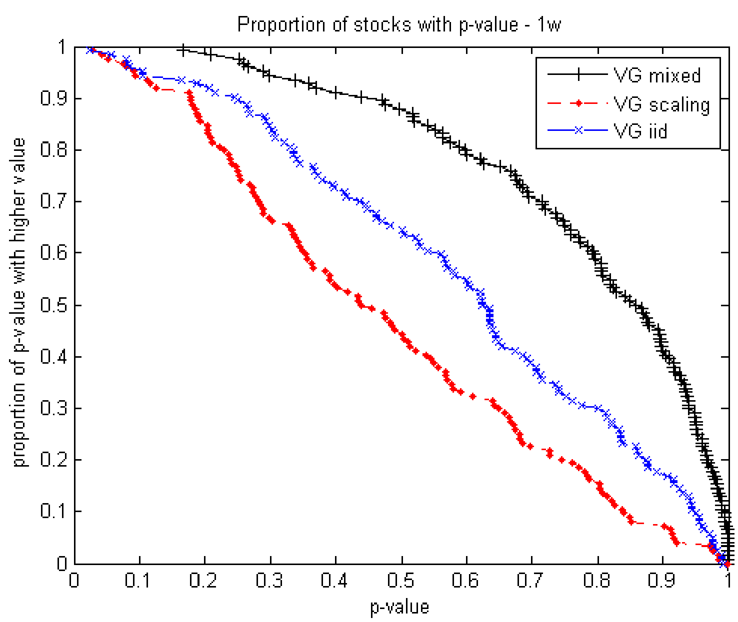


Figure 1-10: Proportion of stocks with p-value greater than certain level (1-week timescale)

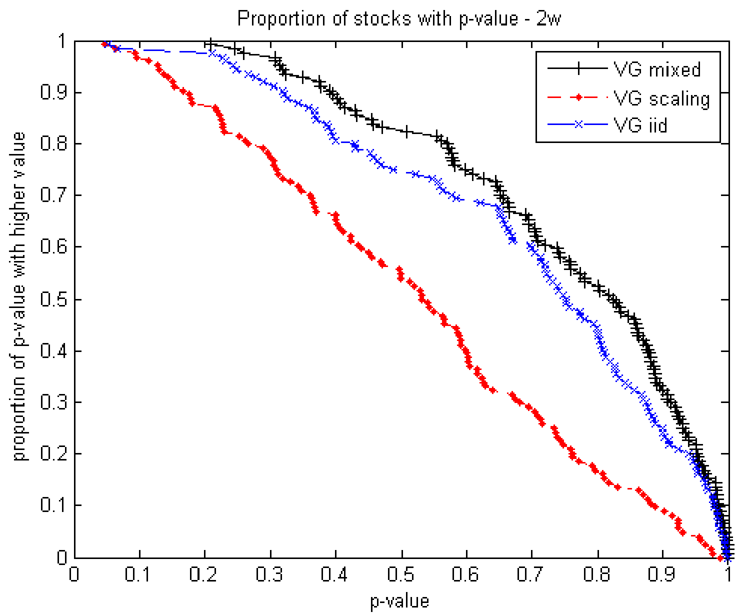


Figure 1-11: Proportion of stocks with p-value greater than certain level (2-week timescale)

in Lévy models decline much faster than the observed data when time increases while they stay constant in the scaling models. A strategy of combining the two approaches is proposed [22]: at a short horizon named unit time, e.g., one hour, we split the random variable of log return into two components, one is a fraction of itself and the other is the remaining part. The first component follows the VG iid model along the time horizon, and the second one follows the VG scaling model.

Statistical estimation and analysis are conducted. All the statistical analyses show the VG mixed model outperforms the other two models at all longer horizons. Furthermore, both the estimated coefficient, c and γ, in the VG mixed model have an average value of 0.4 at all horizons. The mixed model provides a practical method to construct return distributions at longer horizons, which has many applications in the financial industry.

There are a couple of things to investigate in the future study.

1. In this study, the VG distribution only fits one-fourth of the stocks at the unit time, one hour. Other possible self-decomposable distributions should be explored to have a better statistical fit at this unit time.

2. The value 0.4 for are assumed for c and γ when estimation is not feasible due to the lack of data. However, other values should be sought to have a better performance than 0.4. One possible way is to combine data from different horizons to estimate c and γ, which may be more accurate than 0.4.

3. The mixed model provides a strategy to build return distributions at longer horizons. The probability measure obtained from the time series stock return data is called the physical measure P. Option surface contains the information to construct another return distribution called risk neutral measure Q. It has interested researchers for a long time regarding how the ratio P/Q behaves along the time horizon. This question is not easy to answer due to the difficulty in obtaining the physical measure at longer horizons. The mixed model provides a possible way to study this topic.

Chapter 2

Estimating Expected Return By Numeraire-Portfolio Method

2.1 Introduction

The expected return is one of the most important numbers in finance, which predicts risky asset's future performance. The estimation of expected returns is crucial to many investment decisions, e.g., portfolio selection. Much research has been done to analyze and model expected returns by various risk factors. However, few studies have been done to estimate this important number. Furthermore, there is no universally accepted agreement on the estimation method.

One widely used method implements classic asset pricing models, mainly the Capital Asset Pricing Model (CAPM) and the Fama-French three-factor model, to estimate expected returns from historical data. In these models (and their variations), the expected return is affected by one or more than one of the risk factor(s), named beta(s). In the estimation procedure, Beta(s) is first estimated using a simple OLS regression on historical data. Then the expected return is obtained by the product of the estimated beta(s) and the associated risk premium. However, realized returns are so volatile that a huge amount of data is required to obtain relatively precise estimates. Detailed discussion can be found in [6]. An empirical study by Bartholdy and Peare [4] also indicates that none of the two popular models provides an accurate good fit, where both regressions in the method can only explain an average 5% of differences in returns.

Numerous authors, including Breeden [8], Lucas [46], Mehra and Prescott [57], and Rubenstein [71], demonstrate that expected returns are determined by future uncertainty and investors' preferences, instead of implied by realized returns. Therefore, a discount cash flow model which links expected returns to future cash flow (uncertainty) is proposed. The estimator is originally derived from the Dividend Discount Model by Preinreich [68], which says an asset's current price is the future payoffs discounted by the expected return. Edwards and Bell [25] and Ohlson [62] improved the model and derived the Edwards-Bell-Ohlson equation, which, along with its modified versions, is implemented by numerous authors. We cite Claus and Thamas [15] and Philips [64]. However, this method is not robust for assets with dividends or earning growth rates.

For other estimation approaches, we refer to Welch's paper [79], which provides a review of the existing estimates of the expected returns. In this paper, an interesting survey is conducted among a group of academic financial economists, and their forecasts of equity premium are reported.

In this chapter, we propose a novel estimation approach, which also tries to extrapolate the expected return from the future uncertainty that is represented by option prices. Unlike the classic risk-neutral pricing, the option can be priced by an alternative method, which is related to a so-called numeraire portfolio [45] and the associated pricing method. The numeraire portfolio is a self-financing, positive portfolio, which maximizes the expected log utility at the terminal time. It exists if, and only if, there is no arbitrage opportunity. A striking feature of this portfolio is that the price process of any asset in the same market, if denominated by this portfolio, is a martingale under the physical measure. Therefore, the numeraire portfolio provides a pricing method for contingent claims, which is proposed by Platen [9] [65] [66]. More explicitly speaking, an option's price, denominated by the numeraire portfolio, is the expectation of its numeraire-denominated terminal payoff under the physical measure. The physical measure implies the expected return through stochastic stock price models. Therefore, the numeraire pricing method links the expected return and future uncertainty, and it leads to a new method to estimate the expected return.

The numeraire portfolio is required in this method. However, its composition is not as easy to determine as its existence. Long [45] demonstrates that it is a levered position in the market portfolio. Furthermore, empirical studies [29] [69] [70] suggest the market portfolio and the numeraire portfolio can be proxied by value-weighted or equal-weighted portfolios, such as S&P 500, NYSE.

Option calibration is conducted to estimate the parameters, which include the desired expected return. A simulation technique is employed in the calibration procedure. As stock and the numeraire portfolio (or its proxy) are correlated, the bivariate random variable is simulated through the full-rank Gaussian copula (FGC) [39] [50] [51], which transforms the marginal samples to a standard normal random variable, constructs the dependence structure from the binormal random variable, and then transforms the simulated standard normal back to the desired bivariate random numbers. In the calibration procedure, a stock price model for long-horizons (one month in this study) is required, which is the VG mixed model by Eberlein and Madan [22].

The expected returns of the first 50 stocks in the S&P 500 are estimated once every month from January 1999 to October 2009. Unlike the realized return or its sample mean, nearly 95% of the estimated expected returns are positive. The statistics of these estimates are more stable than the realized returns. A simple linear regression model further shows that the estimated returns and the realized returns have the same mean for nearly 80% of the stocks. The results indicate the estimated return can be served as an estimator for the expected return, and it is superior to the estimators from the historical data.

The rest of the chapter is organized as follows. Section 2.2 introduces the numeraire portfolio and its pricing method. Section 2.3 describes the estimation procedure using the numeraire-portfolio method. The numerical implementation and results are presented in Section 2.4. Section 2.5 concludes.

2.2 Pricing with Physical Measure

2.2.1 A Simple Example

In this section, we present a simple example to illustrate how to price an asset.[1] Consider a single-period binomial model in Figure 2-1. We are interested in valuing a stock A whose current price is $S_0 = \$100$. At time $t = 1$, its price will either be \$100 or \$95, each with 50% possibility. To simplify the situation, the interested rate is assumed to be zero.

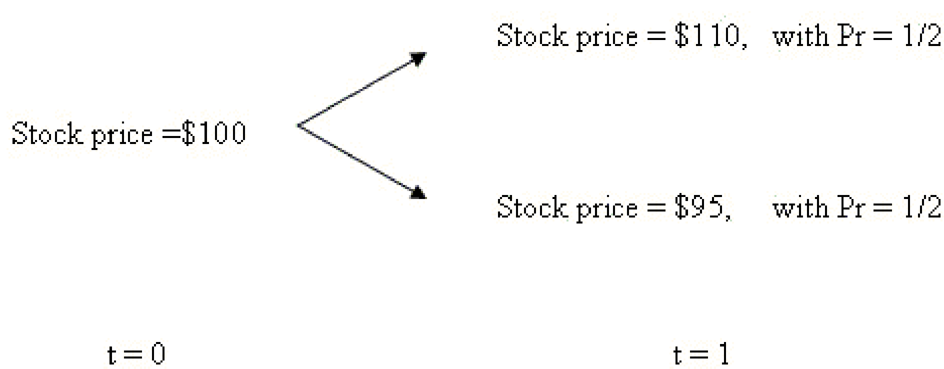


Figure 2-1: A single-period binormial model.

In this example, the probability at $t = 1$ is the real world probability which is called the "physical measure." Simply taking the expectation using the physical measure, $E[S_1|t_0] = 110 \cdot \frac{1}{2} + 95 \cdot \frac{1}{2} = 102.5$, does not give us $S_0 = \$100$, the stock price at $t = 0$. If the price is \$102.50, nobody will buy this stock as people can invest this amount at time 0 in the money market, which is risk free, and get back \$102.50 at time 1 guaranteed with no worry to lose. So the actual price is lower than the

[1] Please note that it is a rough example, only for illustration purpose.

expected price using the physical measure, and the extra amount \$2.50 is the risk premium, a compensation for the uncertainty that people take in this risky asset.

One approach to obtaining the price is to take the expectation under another probability measure called the "risk-neutral measure." The idea was first proposed by Cox and Ross [20] in 1976 and it is now the widely used method in pricing derivatives. Let $P^Q(S_1 = 110) = \frac{1}{3}$ and $P^Q(S_1 = 95) = \frac{2}{3}$, and then take the expectation under this measure, we get $E^Q[S_1|t_0] = 110 \cdot \frac{1}{3} + 95 \cdot \frac{2}{3} = 100$, which is the actually price at time 0. This new measure is not the actual probability measure but equivalent to the physical measure. Harrison and Kreps [34] name this risk-neutral measure an "equivalent martingale measure" as under this formulated measure, the asset price processes are martingales.

How can we still obtain the actual price if taking the expectation under the physical measure? In other words, under what condition can the price process still be martingale under the physical measure? Let us assume a portfolio with value \$100 at time 0, \$105 in the up state and \$99.75 in the down state at time 1. Now divide the stock A's prices by the values of this portfolio and then take the expectation under the physical measure, and we get expected dominated price at time 1 $E[S_1^d|t_0] = \frac{110}{105} \cdot \frac{1}{2} + \frac{95}{99.75} \cdot \frac{1}{2} = 1$, which is the dominated price at time 0 $S_0^d = \frac{100}{100}$. Thus, the stock's dominated price is a martingale under the physical measure. This portfolio is found by Long [45] and named the numeraire portfolio. Again, please keep in mind that this is a rough example to illustrate the idea of asset pricing, no further information is implied. The structure and property of the numeraire portfolio is discussed in the following section.

2.2.2 The Numeraire Portfolio

§ The Setting

In this section, the basic definition and assumptions are set up. The numeraire portfolio is discussed within this context.

A single-period model of an asset market is considered. We assume no transaction costs and restrictions on short sales. *N tradable assets* exist in the market, with price S_{ti} for asset i at time t, where $i = 1, ..., N$ and $t = 0, 1$. To simplify the situation, the asset prices are adjusted values, in which the information of dividend and split is reflected. All the assets are assumed to have strictly positive prices, i.e., $S_{ti} > 0$. It is also reasonable to assume all prices are bounded, denoted as $P(S_{ti} < D$, $i = 1, ..., N$, $t = 0, 1) = 1$, which means all prices are less than a finite number D for sure. Let R_{ti} be the rate of return for asset i from time $i - 1$ to i. Thus, R_{ti} is also bounded. Now, we have the price and rate of return $N \times 1$ vector S_t and R_t for the N assets at time t.

Portfolios can be constructed using the N assets. Denote α_{ti} the number of units of asset i at time t, and α_t the associated $1 \times N$ composition vector. We also assume finite portfolios, which make α_{it} a finite number for all i and t. The market value of the portfolio A at time t is denoted as V_{α_t}, which equals $\alpha_t S_t$ at time t.

There are *some specific requirements for the portfolios* in our context: *self-financing* and always *positive value*. In a self-financing portfolio, the purchase of new assets must be funded by the sale of its own assets, expressed in the mathematical formula as $\alpha_{t-1} S_t = \alpha_t S_t$ for all $t \geq 1$. Because of self-financing, only portfolios with positive values can survive in the market, as when one portfolio's value is below zero, and there is no exogenous infusion and the portfolio is valueless. We assume there exists at least one self-financing portfolio with positive value all the time. In that case, we have at least one portfolio with good performance to serve as a numeraire portfolio, which will be defined later.

Last, we define arbitrage, or "profit opportunities" termed in Long's paper [45]. Roughly speaking, arbitrage is the opportunity to get something from nothing, or a "free lunch." In our case, a portfolio with arbitrage opportunity has initial zero cost ($t = 0$) but probability one to have nonnegative terminal value ($t = 1$ or in more general case $t = T$ where $T \geq 1$), and a positive probability to have strictly positive gain terminal value.

Mathematically it is defined as follows:

(1) $V_{\alpha_0} = 0$;

(2) $P\{V_{\alpha_1} \geq 0\} = 1$;

(3) $P\{V_{\alpha_1} > 0\} > 0$.

§ The Numeraire Portfolio: Definition

Within the above scenario, let us find a portfolio with the maximal expected log return at the terminal time $t = 1$. The initial value of all portfolios is set to 1, i.e.. $V_{\alpha_0} = 1$. The composition of portfolio at time t is α_t, a $1 \times N$ vector. In a single-period model, we select the portfolio at $t = 0$ and hold the position at $t = 1$. Therefore, α_t is the same at $t = 0$ and 1 and can be simplified as α. Correspondingly, α_i is the shares of asset i. The portfolio value is still denoted as V_{α_t} as it equals αS_t

which is still related with t. Under the physical measure, this maximization problem can be formulated as below:

$$\max_{\alpha} E_0 \left[\ln \alpha S_1\right] \quad \text{or} \quad \max_{\alpha} E_0 \left[\ln V_{\alpha_1}\right], \qquad \text{st.} \quad \alpha S_0 = 1.$$

Using the Lagrangian method,

$$\frac{\partial \left(E_0 \left[\ln \alpha S_1\right] - \lambda(\alpha S_0 - 1)\right)}{\partial \alpha_i} = 0$$

for each α_i, $i = 1, ..., N$. Then

$$E_0 \left[\frac{S_{1i}}{V_{\alpha_1^*}}\right] - \lambda S_{0i} = 0, \text{ where } V_{\alpha_1^*} = \alpha^* S_1, \;\; i = 1, ..., N. \tag{2.1}$$

Multiply α^* on both sides of these N equations,

$$E_0 \left[\frac{\alpha^* S_1}{V_{\alpha_1^*}}\right] = \lambda \alpha^* S_0.$$

Because $\alpha^* S_0 = 1$ and $V_{\alpha_1^*} = \alpha^* S_1$, we have $\lambda = 1$. Eq (2.1) then becomes

$$E_0 \left[\frac{S_{1i}}{V_{\alpha_1^*}}\right] = S_{0i} = \frac{S_{0i}}{V_{\alpha_0^*}}.$$

Maximum is achieved by the second order condition if the portfolio $V_{\alpha_1^*}$ has positive value.

The above result means each asset denominated by a portfolio with maximized expected log return is a martingale under the physical measure. Furthermore, this conclusion can be extended to multi-period discrete-time case by the following theorem[2].

Theorem 17 *In a discrete-time market with N assets, $\alpha = \{\alpha_t : t \geq 0\}$ is a positive self-financing portfolio, where α_t is a $1 \times N$ vector representing the portfolio's composition at time t, $t = 0, ..., T$. If this portfolio maximizes $E_0[\ln V_T]$ at T, then for any asset i in this market*

$$\frac{S_{it}}{V_{\alpha_t}} = E_t\left[\frac{S_{iT}}{V_{\alpha_T}}\right], \quad for\ 0 \leq t < T, \tag{2.2}$$

where S_{it} is the price of asset i at t, and V_{α_t} is the value of portfolio α at t.

Long [45] defines the kind of denominating portfolio the "*numeraire portfolio*".

Definition 18 ***Numeraire Portfolio***

In a discrete-time market, a numeraire portfolio V^* is a self-financing portfolio with maximized terminal expected log return. When each asset in the market is denominated by this portfolio, it is a martingale under the physical measure.

$$\frac{S_{it}}{V_t^*} = E_t\left[\frac{S_{iT}}{V_T^*}\right], \quad for\ 0 \leq t < T, \ \ i = 1, ...N.$$

The numeraire portfolio is obtained by maximizing the expected log return, which is also called the expected growth (or expected growth rate in the continuous-time

[2] The idea of the proof can be found on page 54 in [45].

model). Kelly [38] proposed an investment portfolio, which is named the growth-optimal portfolio (GOP), by maximizing the expected growth rate of portfolios. Thus, the numeraire portfolio is the growth-optimal portfolio, and both reach optimal for log utility investors when the initial portfolio value is set to one.

Let R^*_{ti} be the rate of return for the denominated asset i from $t-1$ to t, $R^*_{ti} = \frac{S^*_{ti}}{S^*_{t-1,i}} - 1 = \frac{S_{ti}}{V^*_t} / \frac{S_{t-1,i}}{V^*_{t-1}} - 1$. Take the conditional expectation on both sides, we get $E_{t-1}[R^*_{ti}] = E_{t-1}\left[\frac{S_{ti}}{V^*_t}\right] / \frac{S_{t-1,i}}{V^*_{t-1}} - 1$. The numeraire portfolio implies that $E_{t-1}[R^*_{ti}] = 0$, the best conditional forecast of the rate of return for any denominated asset is zero. This is an impressive feature that implies its relationship with the market portfolio of the Capital Asset Pricing Model (CAPM), which provides information of the composition and proxy of the numeraire portfolio. Details will be discussed on page 62.

§ The Numeraire Portfolio: Existence

The numeraire portfolio provides striking properties described in the previous session. Now comes the question of under what conditions does such kind of portfolio exist.

Theorem 19 ***Existence and Uniqueness of Numeraire Portfolio***

In a market all the portfolios are assumed to have bounded values. A numeraire portfolio exists if and only if no arbitrage opportunity exists in the market. If there are two numeraire portfolios, then, they have the same rates of return.

We will sketch the idea of the proof to have a better understanding of the conditions of the existence. Detailed proof can be found on page 53 in [45].

First, if there exists a numeraire portfolio α, then its definition implies that any other portfolio β has

$$\widehat{V}_{\beta_0} = E_0\left[\widehat{V}_{\beta_t}\right], \tag{2.3}$$

where $\widehat{V}_{\beta_0} = V_{\beta_0}/V_{\alpha_0}$, $\widehat{V}_{\beta_1} = V_{\beta_1}/V_{\alpha_1}$, the denominated values of the portfolio β.

If β contains arbitrage opportunity, by the arbitrage definition, we have

$$\widehat{V}_{\beta_0} = 0 \quad \text{and} \quad E_0\left[\widehat{V}_{\beta_t}\right] > 0,$$

which contradicts Eq (2.3). Thus, the existence of a numeraire portfolio excludes arbitrage.

Secondly, we demonstrate the "if" part. The numeraire portfolio is derived from maximization of $E_0\left[\ln V_{\beta_t}\right]$ with constraint $V_{\beta_0} = 0$. On page 56 we assumed that the prices of all assets are bounded and there exists at least one self-financing portfolio with always positive values. The solution of the maximization problem exists only under the no arbitrage condition and the above listed assumptions. Thus, there exists a numeraire portfolio. Last, the uniqueness of the rate of returns. If we have two numeraire portfolios A and B with different rates of return R_{A_t}, R_{B_t}, then they can be denominated by each other,

$$E_0\left[\frac{R_{A_t}}{R_{B_t}}\right] = E_0\left[\frac{R_{B_t}}{R_{A_t}}\right] = 0.$$

The above equations exist if and only if $R_{A_t} = R_{B_t}$. However, the uniqueness of the rate of return does not imply the unique composition of the numeraire portfolios. Vasicek [77] provides an example that with redundant assets, and there exists numeraire portfolios with different compositions.

§ The Numeraire Portfolio: Composition and Proxies

On page 60, we described an impressive property of the numeraire portfolio: zero is the best conditional forecast of the rate of return for any numeraire-denominated asset. This property connects the numeraire portfolio with the market portfolio in CAPM, which is a portfolio consisting of all assets in the market, with weights proportional to their values in the market. Denote R_i and R_α the rates of return of asset i and the numeraire portfolio α, respectively. Then, we have

$$1 + \widehat{R}_i = \frac{S_{1i}}{V_1^*} / \frac{S_{0i}}{V_0^*} = \frac{S_{1i}}{S_{0i}} / \frac{V_1^*}{V_0^*} = \frac{1 + R_i}{1 + R_\alpha}, \tag{2.4}$$

which has a conditional expectation 1 because $E_0[\widehat{R}_i] = 0$. This implies that if asset i has high rate of return, it also has high covariance with the numeraire portfolio's rate of return. This relationship is similar to the relationship between individual assets and the market portfolio in the CAPM. Thus, the numeraire portfolio is similar to the market portfolio. Furthermore, Long [45] indicates that the numeraire portfolio is a levered position in the mean-variance efficient portfolio. Roll [70] states the mean-variance efficient portfolio P can be served as the market portfolio in the CAPM equation,

$$E[R_i] = R_f + \beta_{iP}[E[R_p] - R_f],$$

where the original CAPM equation is

$$E[R_i] = R_f + \beta_{im}[E[R_m] - R_f],$$

where m represents the market portfolio, and the mean-variance efficient portfolio is the market portfolio. Thus, the numeraire portfolio is also a levered position in the market portfolio.

Eq (2.4) shows that the numeraire-denominated rate of return of asset i is

$$\widehat{R}_i = \frac{1 + R_i}{1 + R_\alpha} - 1.$$

As discussed previously, the above rate of return has mean zero. This property can be used to test and compare different proxies for the numeraire portfolio. Let $R_{iP} = \frac{1+R_i}{1+R_P} - 1$ be the proxy-denominated returns. Roll [69] uses the market portfolio proxy, the S&P 500, for the numeraire portfolio, Fama and MacBeth [29] choose the NYSE equal-weighted portfolio, and Long [45] picks the NYSE value-weighted portfolio. The Hotelling T^2 hypothesis tests are employed with the null hypothesis that the expected proxy-denominated return equals to zero. All the tests indicate zero expected returns with sufficient high p-values. Long [45] also find the proxy-denominated returns have means close to zero with small standard deviations. These empirical tests suggest value-weighted or equal-weighted portfolios, such as S&P 500, NYSE indices, can serve as proper proxies for the numeraire portfolios.

2.2.3 Pricing Under the Physical Measure

As discussed in Section 2.2.2, any asset in a market, when denominated by a numeraire portfolio, is a martingale, so does any portfolio that is the linear combination of all the assets in this market. This numeraire-denominated portfolio process is called the *fair price process* by Bühlmann and Platen [9].

Definition 20 ***Fair Price Process***

In a no-arbitrage market, $\alpha = \{\alpha_t : t \geq 0\}$ *is the numeraire portfolio with value process* $V = \{V_{\alpha_t} : t \geq 0\}$*, and* $\pi = \{\pi_t : t \geq 0\}$ *is a self-financing portfolio with price process* $V = \{V_{\pi_t} : t \geq 0\}$*. If the numeraire-denominated price of this portfolio follows:*

$$\frac{V_{\pi_t}}{V_{\alpha_t}} = E_t\left[\frac{V_{\pi_s}}{V_{\alpha_s}}\right] \qquad 0 \leq t < s < T, \quad t, s, T \in N,$$

then, V *is called a fair price process. And this market is a fair market.*

In a fair market, denote a contingent claim $H = \{H_t : 0 \leq t \leq T\}$, which is F_t-measurable and has nonnegative payoff H_t on or before maturity T. Given a numeraire portfolio in this market, a fair price for this contingent claim can be defined [9].

Definition 21 ***Numeraire Pricing***

Given a numeraire portfolio $\alpha = \{\alpha_t : t \geq 0\}$ *in a no-arbitrage market, the fair price at time* t *of a contingent claim* $H = \{H_t : 0 \leq t \leq T\}$ *is defined by*

$$V_{H_t} = V_{\alpha_t} \cdot E_t^P\left[\frac{H_T}{V_{\alpha_T}}\right] \qquad 0 \leq t < T, \tag{2.5}$$

where H_T is the contingent claim's terminal payoff. P is the physical measure.

The numeraire-denominated price $\widehat{V}_{H_t} = \frac{V_{H_t}}{V_{\alpha_t}}$ is called the numeraire-portfolio price (or "*benchmarked price*" in [9]), which is a martingale under the physical measure P.

The numeraire pricing requires the existence of the physical measure and a numeraire portfolio. As discussed in Theorem 19, the numeraire portfolio exists if there is no arbitrage opportunity, and the market indices are proper proxies. If the physical measure is easy to attain, then, this method is feasible and easy to implement. Another advantage is assets' expected return is contained in the pricing formula. Thus, the numeraire-portfolio pricing method provides a possible way to estimate expected returns, which is not easy to obtain. Details will be discussed in the next section.

This pricing method is also named the "*real word pricing*" in [9] because the pricing formula (2.5) uses the physical measure (or the real world probability measure). According to Cochrane [16], the price in Eq (2.5) is the conditional expectation of final payoff discounted by a stochastic factor V_{α_t}. Alternatively, we have the widely used risk-neutral pricing method, which vales financial asset by the expected future payoff, discounted by a risk-free asset B_t, under the risk-neutral measure. Platen [66] claims the risk-neutral pricing is a special case of the numeraire pricing.

Let $\{B_t : t \geq 0\}$ be the riskless saving account (or a risk-free bond). Starting from Eq (2.5),

$$\begin{aligned} V_{H_0} &= V_{\alpha_t} \cdot E_0^P \left[\frac{H_T}{V_{\alpha_T}} \cdot \frac{B_T}{B_0} \cdot \frac{B_0}{B_T} \right] \\ &= E_0^P \left[H_T \frac{B_T}{V_{\alpha_T}} \cdot \frac{V_{\alpha_0}}{B_0} \cdot \frac{B_0}{B_T} \right] \\ &= B_0 \cdot E_0^P \left[\frac{H_T}{B_T} \cdot \Lambda_{T,0} \right], \end{aligned}$$

where $\Lambda_{T,0} = \frac{B_T}{V_{\alpha_T}} / \frac{B_0}{V_{\alpha_0}} = \frac{\widehat{B}_T}{\widehat{B}_0}$. $\widehat{B}_0$ and $\widehat{B}_T$ are the numeraire-portfolio risk-free bond. B_0 and V_{α_0} equal 1 as initially set.

In a no-arbitrage market, any numeraire-portfolio asset is a martingale under the physical measure P. Therefore, $E_0[\Lambda_{T,0}] = E_0 \left[\frac{B_T}{V_{\alpha_T}} \right] = 1$. As the risk-free bond B_t and the numeraire portfolio α_t are nonnegative, $\frac{B_T}{V_{\alpha_T}}$ is also a nonnegative random variable. Thus, $\Lambda_{T,0}$ is a Radon–Nikodym derivative, which transforms the physical measure P to another measure Q by the following formula:

$$V_{H_0} = E_0^P \left[\frac{H_T}{B_T} \cdot \Lambda_{T,0} \right] = E_0^Q \left[\frac{H_T}{B_T} \right],$$

which is the risk-neutral pricing formula. Thus, Q is the risk-neutral measure.

2.3 Estimating Expected Returns Via Numeraire-portfolio Pricing Method

2.3.1 Idea of the Expected Return Estimation

Consider a European call option of asset i with terminal payoff $H_{Ti} = (S_{Ti} - K)^+$, where S_{Ti} is the asset price and K is the strike. Let $V_\alpha = \{V_{\alpha_t} : t \geq 0\}$ be the price of the proxy of the numeraire portfolio. Using the numeraire pricing formula (2.5), the price of this call option is

$$V_{H_{0i}} = V_{\alpha_0} \cdot E_0^P \left[\frac{(S_{Ti} - K)^+}{V_{\alpha_T}} \right], \tag{2.6}$$

where the conditional expectation is under the physical measure of the stock and the proxy. The associated probability distribution is a bivariate distribution of the random variables (S_{Ti}, V_{α_T}).

Recall Section 1.4.2, the model of the asset price is given by Eq (1.39)[3]

$$S_T = S_0 \exp \{\mu T + X(T) + \omega T\}, \tag{2.7}$$

where μ is the mean rate of return, μT is the expected return in the time interval 0 to T, ω is a "convexity correction" to make the expected rate of return be μ under the physical measure.

[3] Although Eq (1.39) is the asset price model using VG distribution, it is applicable to all other proper laws for $X(t)$.

As discussed in Theorem 19, the existence of the numeraire portfolio is under the assumption that all portfolios in the market are bounded. If the assets are modeled by Eq (2.7), we need to check the portfolios constructed by the assets with dynamics of Eq (2.7) are bounded. Let the value of such portfolio be $\sum_i \alpha_{ti} S_{ti}$ at time t, where S_{ti} is asset i's price and α_{ti} is its shares. By Eq (2.7), we have

$$\begin{aligned}
\ln\left(\sum_i \alpha_{ti} S_{ti}\right) &= \ln\left(\sum_i \alpha_{ti} S_{0i} \exp\{\mu t + X_i(t) + \omega_i t\}\right) \\
&\leq \ln\left(\sqrt{\sum_i (\alpha_{ti})^2} \cdot \sqrt{\sum_i (S_{0i} \exp(\mu t + \omega_i t))^2} \cdot \sqrt{\sum_i e^{2X_i(t)}}\right) \\
&= \frac{1}{2} \ln\left(\sum_i (\alpha_{ti})^2\right) + \frac{1}{2} \ln\left(\sum_i (S_{0i} \exp(\mu t + \omega_i t))^2\right) + \frac{1}{2} \ln\left(\sum_i e^{2X_i(t)}\right).
\end{aligned}$$

$$\begin{aligned}
E\left[\ln\left(\sum_i e^{2X_i(t)}\right)\right] &\leq \ln E\left(\sum_i e^{2X_i(t)}\right) \\
&= \ln\left(\sum_i E\left[e^{2X_i(t)}\right]\right) \\
&= \ln\left(\sum_i \Phi_i(-2i)\right) < \infty.
\end{aligned}$$

$\sum_i (S_{0i} \exp(\mu t + \omega_i t))^2$ is finite. If $\sum_i (\alpha_{ti})^2$ is bounded, which is a reasonable constraint, then all the portfolios in this model setting are bounded. Thus, the numeraire portfolio exists when the asset prices in the market follows the dynamics of Eq (2.7).

To estimate the mean rate of return μ using Eq (2.6), we need to calibrate the option prices of asset i, which is conducted by minimize the average absolute error between the market prices and the model prices $V_{H_{0t}}$. A couple of average absolute errors have been defined, which are summarized in Schoutens's book [74].

- Average Pricing Error (APE)

$$APE = \frac{1}{\overline{P}_R} \sum_{i=1}^{N} \frac{|P_R - P_M|}{N}$$

- Average Relative Percentage Error (ARPE)

$$ARPE = \frac{1}{N} \sum_{i=1}^{N} \frac{|P_R - P_M|}{P_R}$$

- Average Absolute Error (AAE)

$$AAE = \sum_{i=1}^{N} \frac{|P_R - P_M|}{N}$$

- Root-mean-square Error (RMSE)

$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{|P_R - P_M|^2}{N}} \tag{2.8}$$

where P_R is the market price, $\overline{P}_R$ is the mean of the market prices, P_M is the model price, and N is the number of options. The model parameters in Eq (2.2.2) can be estimated by minimizing one of these errors.

The asset prices (S_{Ti}, V_{α_T}) can be obtained either by analytical calculation or simulation. In our work, S_{Ti} and V_{α_T} are simulated by a technique called full-rank Gaussian copula (FGC), which will be introduced in the next section.

2.3.2 Multivariate Random Number Simulation Via FGC

Simulation of multivariate random numbers requires the information of the associated multivariate distribution function, which does not always have the closed-form. Also the correlation is complicated in non-Gaussian distributions and covariance may not be a proper measure for the correlation. Copula provides a general approach to constructing dependence structures to formulate multivariate distribution from arbitrary marginal distributions, where the dependence structure is independent from the marginal distributions. We will briefly introduce the definition and some important properties of copula below. One type of copula and the application in simulation will also be described. The details of copula can be found in [61].

Definition 22 ***Copula***

A n-dimensional copula C is a multivariate joint distribution function defined on $[0,1]^n$ with mapping $[0,1]^n \rightarrow [0,1]$, which has the following properties:

(1) C is ground[4] *and n-increasing;*

(2) $C(1,...,1,u_i,1,...,1) = u_i$, $u_i \epsilon [0,1]$, $i = 1,...,n$;

(3) $C(u_1,...,u_n) = 0$ if at least one u_i equals zero.

This definition indicates that copula C, as a multivariate distribution function, has uniform marginal distributions.

How is a multivariate distribution function related to other arbitrage multivariate distribution functions? Sklar [76] provides a solution, which is the foundation of the most applications of the copula.

[4] *A function $f(x_1,x_2...,x_n)$ is called grounded if $f(a_1,...) = f(...,a_2,...) = ... - f(...,a_n) = 0$, where a_i is the least element in the domain of x_i.*

Theorem 23 ***Sklar's Theorem***

Let F be an n-dimensional multivariate distribution function with continuous marginal distributions $F_1, F_2, ..., F_n$. Then, there exists a unique n-dimensional copula C defined on $[0,1]^n$ such that

$$F(x_1, x_2, ..., x_n) = C(F_1(x_1), F_2(x_2), ..., F_n(x_n)).$$

Sklar's theorem states that given a join law F and the corresponding marginal laws, there exists a copula C that describing the dependence structure, and that copula does not contain any information of the expression of the marginal laws. Thus, the joint law F can be constructed from the marginal laws $F_1, F_2, ..., F_n$ and the dependence structure C separately.

Instead of expressing in random variables $(x_1, x_2, ..., x_n)$, the Sklar's theorem has an equivalent form expressed by the probability distributions $F_1, F_2, ..., F_n$.

Corollary 24 ***An Equivalent Form of Sklar's Theorem***

Let H be an n-dimensional multivariate distribution function with continuous marginal distributions $F_1, F_2, ..., F_n$. Then, there exists a unique n-dimensional copula C defined on $[0,1]^n$ such that

$$C(u_1, u_2, ..., u_n) = F(F_1^{-1}(u_1), F_2^{-1}(u_2), ..., F_n^{-1}(u_n)),$$

where $u_i \epsilon [0,1]$ for $i = 1, 2, ..., n$.

The copula also has a useful invariant property given by Embrechts *et al.* [27].

Theorem 25 ***Invariant Property of Copula***

Let $(x_1, ..., x_n)$ be a continuous n-dimensional random vector with copula C and $h_1(x_1), ..., h_n(x_n)$ be strictly increasing continuous functions on the ranges of $x_1, ..., x_n$. then the n-dimensional random vector $(h_1(x_1), ..., h_n(x_n))$ also has the same copula C.

This invariant property provides a powerful way to construct the multivariate distributions. If the distribution of the random vector $\overrightarrow{X}$ is not easy to obtain, then, we can transform it to a new random vector whose dependence structure is easy to build. The only requirement for this procedure is that the transform function is strictly increasing.

Now let us introduce one copula, called Gaussian copula proposed by Li [43], which is widely used in financial modeling. The function of the Gaussian copula has the same structure as the cumulative distribution function (cdf). of the standard multivariate Gaussian random variables.

Definition 26 ***Gaussian Copula***

The Gaussian copula function is

$$C_G(u_1, ..., u_1) = \Phi(\Phi_1^{-1}(u_1), ..., \Phi_n^{-1}(u_n)) \qquad u_i \in [0,1] \quad for \ i = 1, ..., n,$$

where Φ is the multivariate Gaussian cdf with mean zero and correlation matrix A. Φ_i is the univariate standard Gaussian cdf.

Malevergne and Sornette [50] provides the method to build the joint distribution using the Gaussian copula. The idea is briefly described here. Details can be found from [50]. Let X be a n-dimensional random vector with marginal cdf $F(x_i)$ and pdf $f(x_i)$ for X_i, $\overrightarrow{Z}$ be a n-dimensional standard Gaussian random vector with the conservation of probability

$$f(x_i)dx_i = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z_i^2}{2}\right) dz_i.$$

Integrate this equation,

$$\begin{aligned} F_i(x_i) &= \Phi(z_i) \qquad \text{where } \Phi(z_i) \text{ is the cdf of } Z_i, \\ z_i &= \Phi^{-1}(F_i(x_i)). \end{aligned} \tag{2.9}$$

Eq (2.9) is strictly increasing. Thus, by Theorem 25, $\overrightarrow{X}$ and $\overrightarrow{Z}$ have the same copula C. $\overrightarrow{Z}$ has a simple and well-defined dependence structure, which is the covariance matrix A. The copula of $\overrightarrow{Z}$ is its multivariate cdf, which is the Gaussian copula. The joint distribution can then be easily obtained by combining this Gaussian copula with the marginal distribution of $\overrightarrow{X}$.

We are interested in applying the copula method in simulating multi-asset returns. A model of dependence, termed the full rank Gaussian copula (FGC), is employed. It is proposed and studied by Malevergne and Sornette [51], later summarized by Madan and Khanna [39].

FGC can be a very useful tool to simulate multivariate non-Gaussian random numbers in finance. Empirical study ([17] [54] [32], and many others) indicates that the distributions of returns have power-law tails, whose variance and covariance are either not well-defined, or only exist in principle but are hardly accurately to estimate because of the poor convergence of the sample estimators. These multivariate random variables $\overrightarrow{X}$ can first be transformed to standard Gaussian variables $\overrightarrow{Z}$, which have well-defined correlations, the covariance matrix A with possibly full rank. Then, the correlation can be estimated more accurately than the direct estimation on the original random samples of $\overrightarrow{X}$. Next, multivariate Gaussian random numbers can be simulated using the estimated covariance matrix $\widehat{A}$. Finally, the simulated Gaussian random numbers is transformed back to get the random sample of the original random vector $\overrightarrow{X}$.

The algorithm is summarized below:

Algorithm 27 ***Multivariate Simulation Using FGC***

Step 1. Calculate the values of cdf for sample of X_i with the marginal cdf $F_i(x_i)$

$$P(X_i \leq x) = F_i(x)$$

Step 2. Transform the marginal distribution $F_i(x)$ to the standard Gaussian variable

$$z_i = \Phi^{-1}(F_i(x_i)), \quad \Phi(z_i) \text{ is the cdf of the standard normal } Z_i$$

Step 3. Estimate the covariance matrix A with the transformed sample. The *estimated covariance matrix is denoted as* $\overrightarrow{A}$.

Step 4. Simulate multivariate Gaussian variables with $\overrightarrow{A}$. *The simulated random numbers are denoted as* $\widetilde{Z}$.

Step 5. Convert each $\widetilde{Z}_i$ *to* $\widetilde{X_i}$ *by*

$$\widetilde{X_i} = F_i^{-1}(\Phi(\widetilde{Z}_i))$$

$\widetilde{X}$ *are the simulated random numbers.*

Remark:

1. By Theorem 25, any multivariate random variables $\overrightarrow{Y}$, if connected with $\overrightarrow{X}$ by strictly increasing functions $Y_i = g_i(X_i)$, can be simulated by Algorithm 27. We can start from random samples of $\overrightarrow{X}$ in Step1 and obtain the random numbers of $\overrightarrow{Y}$ in Step 5;

2. Algorithm 27 also works in a special case when X is a univariate random variable.

2.4 Numerical Implementation and Results

The expected returns are estimated for the S&P 500 Index and the first 50 stocks of S&P 500 from January 1999 to October 2009. The estimation is conducted once every month, on a Wednesday of the middle of that month, for a total of 130 months. These 130 days are termed the *estimation days*. S&P 500 is chosen as the proxy of the numeraire portfolio. The stock and option data are obtained from WRDS to estimate the expected returns. The price data of eight sectors of S&P 500 are attained from Reuters to perform statistical analysis.

2.4.1 Estimating Expected Returns

Expected return is estimated from calibrating one-month options[5], which is realized by minimizing one of the average absolute errors given in Section 2.3.1. RMSE (Eq 2.8) is chosen in our study, $RMSE = \sqrt{\sum_{i=1}^{N} \frac{\left|P_R - V_{H_{0t}}\right|^2}{N}}$, where $V_{H_{0t}}$ is option's model price (Eq 2.6), $V_{H_{0i}} = V_{\alpha_0} \cdot E_0^P \left[\frac{(S_{Ti}-K)^+}{V_{\alpha_T}}\right]$, in which S_{Ti} and V_{α_T} are modeled by Eq (2.7) $S_T = S_0 \exp\{\mu T + X(T) + \omega T\}$. Simulation technique FGC is employed in the calibration. To simulate S_{Ti} (stock i) and V_{α_T} (proxy to the numeraire portfolio) in Eq (2.6), a distribution model at horizon T is required. Because of its better performance at a longer horizon, the VG mixed model is employed in this study, as T is one month. The input of this model is the VG parameters $(\sigma_i, \nu_i, \theta_i)$ of the marginal distribution for each stock i at the unit time, which is one day in this study. On each estimation day, these parameters are needed, which are estimated from four-year daily stock price data prior to this estimation day. The expected return μ_i of stock i and other parameters (c_i, γ_i) in the VG mixed model can be estimated through iterations of simulating (S_{Ti}, V_{α_T}) in the optimization. As the index V_{α_T} also appears in Eq (2.6), we first need to estimate its parameters $(\mu_\alpha, c_\alpha, \gamma_\alpha)$ with its option pricing formula $V_{H_{0\alpha}} = V_{\alpha_0} \cdot E_0^P \left[\frac{(V_{\alpha_T}-K)^+}{V_{\alpha_T}}\right]$.

[5] Actual maturity varies from four to five weeks (roughly one month), depending on the days between the estimation day to the next available maturity).

The estimation is conducted on 130 estimation days from January 1999 to October 2009. On each estimation day the following procedure is employed. S&P 500 (SPX) is used as the proxy of the numeraire portfolio.

1. Estimate the VG parameter (σ, ν, θ) for SPX and the 50 stocks using four-year daily asset price data prior to the estimation day.

2. Estimate $(\mu_\alpha, c_\alpha, \gamma_\alpha)$ for SPX from the calibratoin of SPX one-month option data. VG mixed model is employed to model the price of SPX.

3. Estimate (μ_i, c_i, γ_i) for each stock i from the calibration of its one-month option data.

Figure 2-2 and Figure 2-3 present two sample calibration results, the former is from SPX option data on July 11, 2007, with RMSE value 2.01, the latter is from HPQ option data on the same day, with RMSE 0.098. Both calibrations use one option data with single maturity of one month.

Among all the estimated returns of the 51 assets in 130 days, there are 94.52% of $\widehat{\mu}$ and 73.73% of $\widehat{\mu} - r_f$ (risk premium, where r_f is the risk-free rate) with positive value. The percentage of positive risk premium $(\widehat{\mu} - r_f)$ for each asset is presented in Table 2.1.

As an example, the estimated return $\widehat{\mu}_\alpha$ and the realized return $\widetilde{\mu}$ of SPX are compared in Figure 2-4.

From Table 2.1 we can tell the estimated risk premium $(\widehat{\mu} - r_f)$ are positive most of the time, which is consistent with the argument of positive risk premium of risky

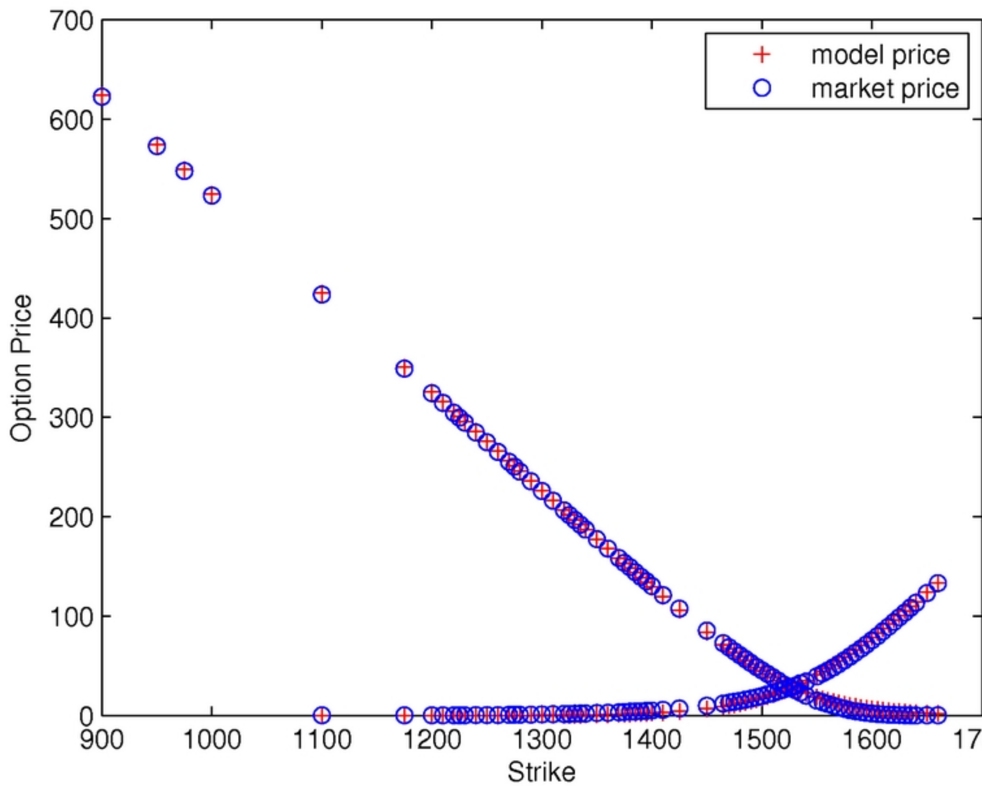


Figure 2-2: The fitted option data of SPX on July 11, 2007, with one-month maturity, RMSE=2.01

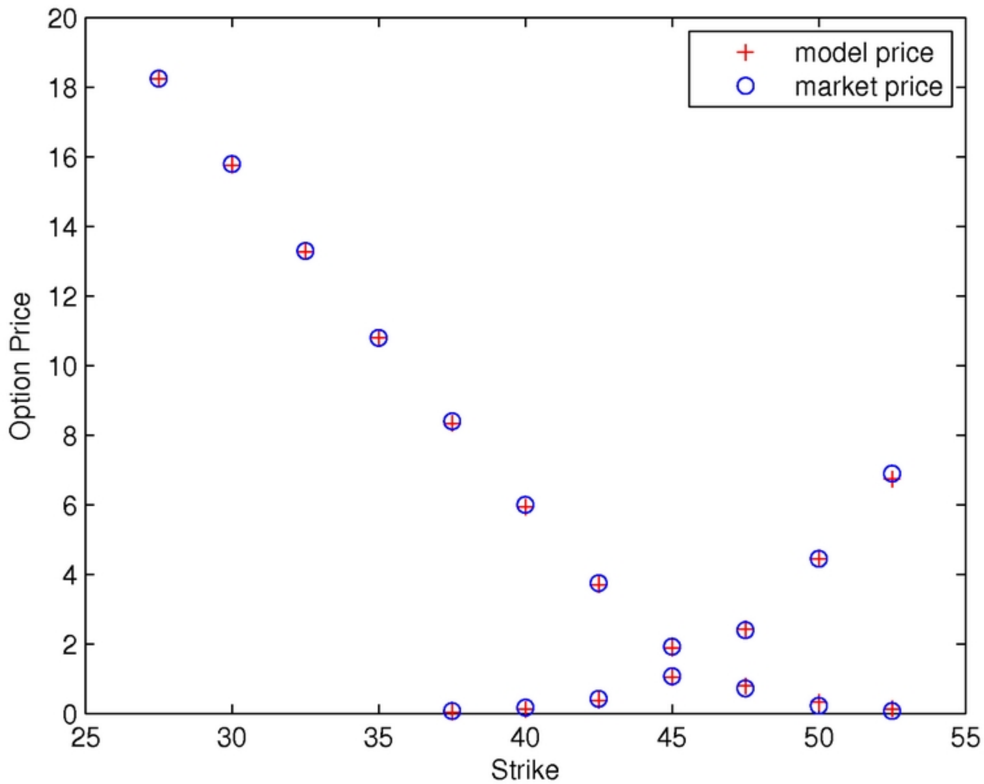


Figure 2-3: The fitted option data of HPQ on July 11, 2007, with one-month maturity, RMSE=0.0617

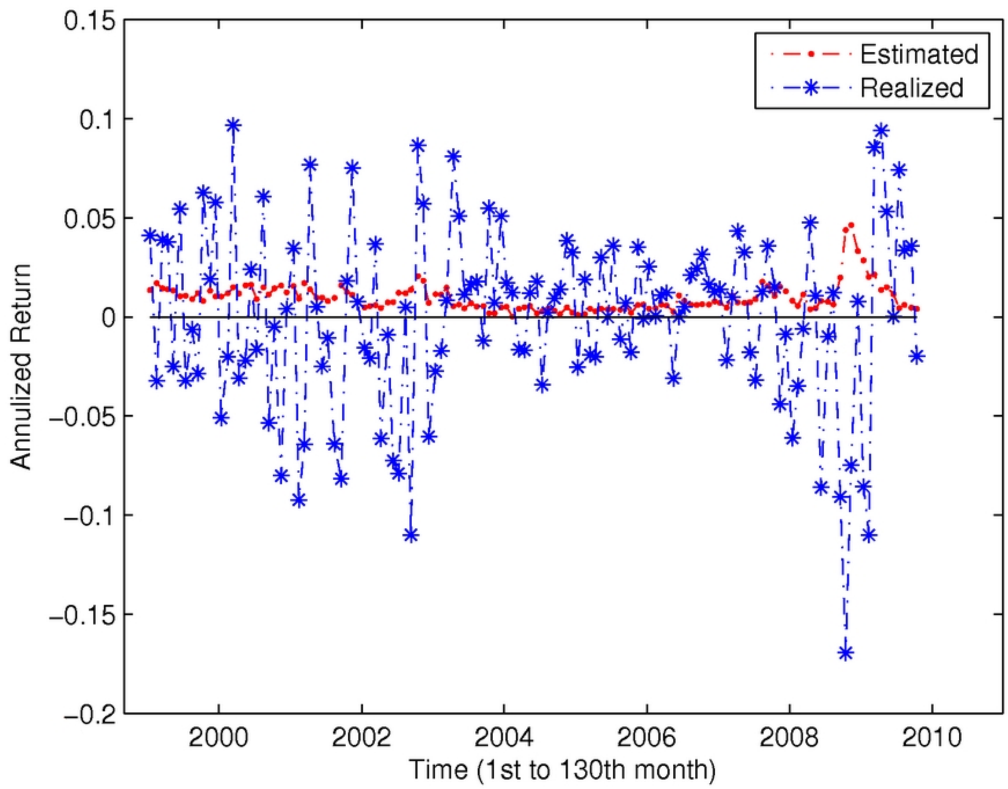


Figure 2-4: Estimated return $\widehat{\mu}_{SPX}$ vs. realized return $\widetilde{\mu}_{SPX}$ for SPX (January 1999 to October 2009)

Symbol	*positive %*	*Symbol*	*positive %*	*Symbol*	*positive %*
SPX	89.2	DOW	64.6	OXY	63.1
ABT	64.6	DD	68.5	ORCL	92.3
MO	53.8	EMR	71.5	PNC	76.9
AXP	88.5	XOM	63.1	PEP	59.2
AMGN	86.2	F	64.6	PFE	64.6
AAPL	93.1	GE	72.3	PG	54.6
BAC	67.7	HAL	80.8	SLB	78.5
BA	83.1	HPQ	90.0	TGT	89.2
CVS	82.3	HD	82.3	TXN	90.0
CAT	70.8	INTC	82.3	MMM	63.1
CVX	56.2	JNJ	47.7	UNP	79.2
CSCO	82.3	LLY	64.6	UTX	76.2
C	74.6	LOW	88.5	UNH	85.4
KO	60.0	MCD	70.0	VZ	64.6
CL	65.4	MRK	67.7	WMT	75.4
COP	59.2	MSFT	78.5	WAG	80.0
DIS	78.5	NKE	80.8	WFC	74.6

Table 2.1: Percentage of positive risk premium of each stock in 130 days

asset in all the financial models. Figure 2-4 also shows that the estimated expected returns are more stable than the observed ones for SPX. Other stocks have the similar results.

The mean and the standard deviation of the estimated return $\widehat{\mu}$ and observed return $\widetilde{\mu}$ for each asset are displayed in Table 2.2, which shows the estimated return $\widehat{\mu}$ have lower standard deviation than the observed ones.

2.4.2 Statistical Analysis

The previous section shows the estimated expected return using the numeraire-portfolio method is better than the realized return and its sample mean. Further investigation is required to test how good it is.

Name	*mean* ($\widehat{\mu}$)	*std.* ($\widehat{\mu}$)	*mean* ($\widetilde{\mu}$)	*std.* ($\widetilde{\mu}$)	*Name*	*mean* ($\widehat{\mu}$)	*std.* ($\widehat{\mu}$)	*mean* ($\widetilde{\mu}$)	*std.* ($\widetilde{\mu}$)
AAPL	0.088	0.055	0.451	1.655	LLY	0.055	0.049	-0.02	0.869
ABT	0.048	0.039	0.076	0.772	LOW	0.085	0.070	0.132	1.113
AMGN	0.079	0.053	0.130	1.250	MCD	0.053	0.045	0.075	0.876
AXP	0.114	0.134	0.121	1.295	MMM	0.052	0.046	0.092	0.780
BA	0.065	0.048	0.129	1.051	MO	0.033	0.053	0.157	0.917
BAC	0.066	0.120	0.053	1.670	MRK	0.053	0.052	0.041	0.953
C	0.096	0.200	-0.009	1.894	MSFT	0.073	0.054	0.092	1.040
CAT	0.060	0.058	0.125	1.224	NKE	0.067	0.051	0.262	1.085
CL	0.049	0.044	0.057	0.664	ORCL	0.102	0.073	0.217	1.436
COP	0.046	0.052	0.165	0.919	OXY	0.052	0.067	0.298	0.971
CSCO	0.097	0.068	0.073	1.286	PEP	0.045	0.037	0.089	0.667
CVS	0.062	0.047	0.066	0.995	PFE	0.055	0.047	-0.017	0.832
CVX	0.042	0.048	0.109	0.735	PG	0.035	0.047	0.073	0.802
DD	0.059	0.054	0.016	1.015	PNC	0.080	0.077	0.024	1.191
DIS	0.076	0.065	0.072	0.991	SLB	0.060	0.053	0.194	1.115
DOW	0.061	0.080	0.103	1.503	TGT	0.086	0.062	0.120	1.066
EMR	0.065	0.057	0.070	0.783	TXN	0.089	0.054	0.075	1.294
F	0.060	0.131	-0.055	1.819	UNH	0.070	0.052	0.240	1.227
GE	0.076	0.082	0.019	1.052	UNP	0.065	0.058	0.110	0.898
HAL	0.070	0.064	0.188	1.518	UTX	0.068	0.055	0.148	0.874
HD	0.077	0.062	0.045	1.116	VZ	0.054	0.062	-0.008	0.749
HPQ	0.081	0.053	0.199	1.190	WAG	0.061	0.043	0.086	0.837
INTC	0.088	0.062	0.028	1.300	WFC	0.075	0.096	0.089	1.213
JNJ	0.034	0.037	0.095	0.601	WMT	0.066	0.048	0.051	0.728
KO	0.045	0.040	0.059	0.730	XOM	0.051	0.05	0.117	0.677

Table 2.2: Mean and std. of the estimated return $\widehat{\mu}$ and realized return $\widetilde{\mu}$ (annualized)

Let $R_{i,t+1}$ be asset i's return from t to $t+1$ and $\widehat{\mu}_{i,t}$ be the associated estimated expected return from t to $t+1$ using the numeraire-portfolio method. $R_{i,t+1}$ is F_{t+1}-measurable and $\widehat{\mu}_{i,t}$ is F_t-measurable as it is obtained at time t from the option prices, which are F_t-measurable. If $\widehat{\mu}_{i,t}$ properly estimates $R_{i,t+1}$, then the conditional expectation of the difference of these two returns is zero, i.e., $E_t[R_{i,t+1} - \widehat{\mu}_{i,t}] = 0$, or furthermore, $E[R_{i,t+1} - \widehat{\mu}_{i,t}] = E[E_t(R_{i,t+1} - \widehat{\mu}_{i,t})] = 0$. The generalized method of moments (GMM) [33] can be employed to test this hypothesis. In the hypothesis test of GMM, the null hypothesis is $H_0 : E[u] = 0$, where u is the orthogonality condition. In our study, let z_t, which is F_t-measurable, be the instrumental variables (vector) that may affect $R_{i,t+1}$ or $R_{i,t+1} - \widehat{\mu}_{i,t}$. Thus, the null hypothesis becomes $E[(R_{i,t+1} - \widehat{\mu}_{i,t}) \cdot z_t] = 0$.

The asset i's return $R_{i,t+1}$ can be similarly modeled by the classical asset return models, such as the Capital Asset Pricing Model (CAPM) [44] [60] [75], the Fama-French three-factor model [28], which establish the relation between assets' expected returns and their risk attributes. Among these models, one asset's expected return and return itself are linearly determined by various factors which represent different risks this asset is exposed to. The CAPM measures the asset's return with its sensitivity to one factor, systematic risk or market risk.

$$E[R_{i,t}] = r_{f,t} + \beta_{iM}(E(R_{M,t}) - r_{f,t}),$$

where:

- $R_{M,t}$ is the market return from $t-1$ to t with expectation $E(R_{M,t})$

- $r_{f,t}$ is the risk-free rate
- β_{iM} measures the sensitivity of asset i's expected return to the market return, and $\beta_{iM} = \frac{Cov(R_i, R_M)}{Var(R_M)}$

In CAPM, only a single factor β_{iM} is used to measure asset i's expected return, which oversimplifies the complicate situations in the market. Fama-French three-factor model introduces two more risk factors, firm size and book-to-market ratio which are represented by two classes of stocks, small cap stocks and value stocks (stocks with a high book-to-market ratio, BTM). These stocks tend to outperform the market. Fama and French include these two factors in their model to adjust assets' outperformance tendency:

$$R_{i,t} = r_{f,t} + \beta_{iM}(R_{M,t} - r_{f,t}) + \beta_{iS} \cdot SMB_t + \beta_{iH} \cdot HML_t + \varepsilon_{i,t},$$

where additionally to CAPM:

- SMB represents "Small Minus Big stocks," which is the excess return of small stocks over big ones
- HML represents "High BTM Minus Low BTM," which is the excess return of high BTM stocks over small ones
- β_{iS} and β_{iH} measure the sensitivities to firm size and book-to-market ratio

Besides the risk factor in these two traditional models, other factors are proposed by many people in academia and industry. MSCI Barra [82] suggests numerous factors in their industrial models. We choose two factors in our analysis, asset's own performance and the influence from asset's sector[6]. The third one is the market risk in CAPM and the Fama-French model. These returns associated with the three factors are also considered as the instrumental variables z_t.

$R_{i,t+1}$ is F_{t+1}-measurable. $\widehat{\mu}_{i,t}$ and the three returns, namely the market return $R_{M,t}$, asset i's return $R_{i,t}$, and the sector return $R_{S,t}$ are F_t-measurable. At time t, if the null hypothesis is chosen to be

$$E_t[(R_{i,t+1} - \widehat{\mu}_{i,t}) \cdot z_t] = 0, \tag{2.10}$$

then the linear regression model assumed in our study is

$$R_{i,t+1} - \widehat{\mu}_{i,t} = \beta_0 + \beta_{iM}(R_{M,t} - r_{f,t}) + \beta_{ii}(R_{i,t} - r_{f,t}) + \beta_{iS}(R_{S,t} - r_{f,t}) + \varepsilon_{i,t+1}, \tag{2.11}$$

where:

- $R_{M,t}$ is the market return at t
- $R_{S,t}$ is the sector's return at t
- $r_{f,t}$ is the risk-free rate at t
- β_{ii} reflects asset i's exposure to its own performance
- β_{iS} measures asset i's sensitivity to its sector
- β_{iM} and $\varepsilon_{i,t+1}$ are the same as those in the CAMP and the Fama-French's model

[6] The category an asset belongs to, such as Exxon in energy sector.

- β_0 is the intercept.

If $\widehat{\mu}_{i,t}$ properly measures asset i's return, then, all the betas in Eq (2.11) should be zero. Under this scenario, the null hypothesis (2.10) can be rewritten as

$$E_t[z_t(y_{t+1} - x'_t\beta)] = 0,$$

where:

- z_t is the column return vector $(1, R_{M,t} - r_{f,t}, R_{i,t} - r_{f,t}, R_{S,t} - r_{f,t})'$
- $y_{t+1} = R_{i,t+1} - \widehat{\mu}_{i,t}$
- β is the column risk factor vector $(\beta_0, \beta_{iM}, \beta_{ii}, \beta_{iS})'$

This is equivalent to the ordinary least square (OLS) model [80] $y = X\beta + \varepsilon$. The estimator $\widehat{\beta} = (X'X)^{-1}X'y$ is the same as the OLS estimator. After β is estimated through OLS, the F-test in OLS can be performed, where the null hypothesis is $H_0 : \beta = 0$. The acceptance of this null hypothesis implies the acceptance of the GMM null hypothesis (2.10).

Linear regressions are employed on two pairs of return using Eq (2.11): one pair is the estimated returns $\widehat{\mu}_{i,t}$ and the associated realized returns $R_{i,t+1}$; the other is the estimated returns $\widehat{\mu}_{i,t}$ and the sample mean $\overline{R}_{i,t+1}$. Both regressions are conducted for 50 stocks using 130 data points from 130 estimation days from January 1999 to October 2009. Let Δt_i be the number of days to the next available maturity from the estimation day i on which $\mu_{i,t}$ is estimated. $R_{i,t+1}$ is the associated realized return during this time period, and $R_{i,t}$, $r_{f,t}$, $R_{S,t}$ and $R_{M,t}$ are the realized returns during the time period back Δt_i days from the estimation day i. $\overline{R}_{i,t+1}$ is the sample mean calculated from the daily returns during the time period forward Δt_i days from the estimation day i. All these returns are annualized.

Two hypothesis tests are conducted to test the beta values:

t test is to determine the significance of each individual beat, with the null hypothesis $H_0 : \beta_i = 0$, and the alternative hypothesis $H_1 : \beta_i \neq 0$, where β_i is β_0, β_{iM}, β_{ii}, or β_{iS}.

F test is for the overall significance of all the betas with the null Hypothesis tests $H_0 : \beta_0 = \beta_{iM} = \beta_{ii} = \beta_{iS} = 0$, and the alternative hypothesis H_1 : one or more than one beta is not equal to zero.

Both tests are performed with 95% confidence level. The test results are displayed in Table 2.3 and 2.4. The results are also summarized below:

t test:

The hypothesis tests for both pairs indicates large portion of stocks have $\beta_i = 0$:

$R_{i,t+1} - \widehat{\mu}_{i,t}$: 76% for $\widetilde{\beta}_{ii}$, 90% for $\widetilde{\beta}_{iM}$, 72% for $\widetilde{\beta}_{iS}$, and 80% for β_0.

$\overline{R}_{i,t+1} - \widehat{\mu}_{i,t}$: 76% for $\widetilde{\beta}_{ii}$, 84% for $\widetilde{\beta}_{iM}$, 76% for $\widetilde{\beta}_{iS}$, and 66% for β_0.

F test:

The F test shows that 34 out of 50 stocks' p-value is greater than 0.05, which means all the betas of these stocks have zero value with 95% confidence level.

Similar results are also attained using the Fama-French three-factor model. Thus, the numeraire-portfolio method provides a good approach to estimate expected returns.

Name	*indi.* β_{ii}	*p-value* β_{ii}	*indi.* β_{iM}	*p-value* β_{iM}	*indi.* β_{iS}	*p-value* β_{iS}	*indi.* β_0	*p-value* β_0
AAPL	0	0.138	0	0.799	0	0.854	1	0.004
ABT	0	0.612	0	0.328	0	0.513	0	0.348
AMGN	0	0.534	0	0.410	0	0.269	0	0.240
AXP	0	0.360	0	0.338	1	0.039	0	0.423
BA	1	0.005	0	0.422	1	0.033	0	0.371
BAC	0	0.052	0	0.887	0	0.309	0	0.804
C	0	0.770	0	0.867	0	0.707	0	0.366
CAT	0	0.195	0	0.157	0	0.067	0	0.260
CL	0	0.949	1	0.004	1	0.001	0	0.152
COP	0	0.543	0	0.632	0	0.678	0	0.111
CSCO	1	0.049	0	0.151	1	0.012	0	0.692
CVS	0	0.579	0	0.697	0	0.975	0	0.355
CVX	0	0.471	0	0.373	0	0.069	0	0.067
DD	1	0.006	0	0.122	1	0.001	0	0.992
DIS	1	0.042	0	0.300	0	0.881	0	0.405
DOW	0	0.506	0	0.333	0	0.367	0	0.963
EMR	0	0.944	0	0.334	0	0.811	0	0.633
F	0	0.875	0	0.573	0	0.471	0	0.117
GE	1	0.040	0	0.976	0	0.151	0	0.937
HAL	0	0.791	0	0.080	0	0.562	1	0.027
HD	1	0.018	0	0.290	1	0.006	0	0.966
HPQ	0	0.292	0	0.524	1	0.039	0	0.359
INTC	0	0.629	1	0.009	0	0.058	0	0.675
JNJ	0	0.488	0	0.099	1	0.001	0	0.137
KO	0	0.281	0	0.543	0	0.210	0	0.288
LLY	0	0.984	0	0.533	1	0.032	0	0.075
LOW	0	0.211	0	0.594	0	0.211	0	0.241
MCD	0	0.769	0	0.518	0	0.271	0	0.102
MMM	0	0.268	0	0.693	0	0.588	0	0.183
MO	1	0.010	0	0.845	1	0.014	1	0.001
MRK	0	0.665	0	0.576	0	0.725	0	0.703
MSFT	0	0.152	0	0.055	0	0.580	0	0.511
NKE	1	0.001	0	0.383	0	0.900	1	0.000
ORCL	1	0.000	1	0.033	0	0.385	0	0.314
OXY	0	0.085	0	0.626	0	0.779	1	0.000
PEP	1	0.003	0	0.735	0	0.189	1	0.029
PFE	1	0.025	0	0.935	1	0.025	0	0.958
PG	0	0.859	0	0.075	1	0.014	1	0.012
PNC	0	0.187	0	0.828	0	0.260	0	0.223
SLB	0	0.401	0	0.067	0	0.177	0	0.059
TGT	0	0.640	0	0.730	0	0.747	0	0.273
TXN	0	0.707	0	0.547	0	0.225	0	0.806
UNH	0	0.331	0	0.906	0	0.698	1	0.015
UNP	0	0.635	1	0.004	1	0.007	1	0.048
UTX	0	0.052	0	0.842	0	0.386	1	0.007
VZ	1	0.016	1	0.012	1	0.014	0	0.907
WAG	0	0.123	0	0.671	0	0.618	0	0.704
WFC	0	0.343	0	0.081	0	0.075	0	0.301
WMT	0	0.530	0	0.061	0	0.441	0	0.495
XOM	0	0.140	0	0.891	0	0.352	0	0.304

Table 2.3: t-test of each β_i ($indi. = 0$ represents $\beta_i = 0$, $indi. = 1$ represents $\beta_i \neq 0$)

Name	*indi.*	*p-value*	*Name*	*inid.*	*p-value*	*Name*	*inid.*	*p-value*
AAPL	0	0.292	F	0	0.912	OXY	0	0.171
ABT	0	0.791	GE	0	0.146	PEP	1	0.028
AMGN	0	0.497	HAL	0	0.345	PFE	0	0.056
AXP	0	0.169	HD	1	0.008	PG	1	0.027
BA	1	0.020	HPQ	0	0.102	PNC	0	0.227
BAC	0	0.182	INTC	0	0.079	SLB	0	0.171
C	0	0.820	JNJ	1	0.002	TGT	0	0.714
CAT	0	0.320	KO	0	0.633	TXN	0	0.589
CL	1	0.001	LLY	1	0.009	UNH	0	0.796
COP	0	0.876	LOW	1	0.006	UNP	1	0.044
CSCO	0	0.097	MCD	0	0.417	UTX	0	0.238
CVS	0	0.930	MMM	0	0.213	VZ	1	0.013
CVX	0	0.105	MO	1	0.022	WAG	1	0.050
DD	1	0.014	MRK	0	0.792	WFC	0	0.339
DIS	0	0.169	MSFT	0	0.115	WMT	1	0.003
DOW	0	0.695	NKE	1	0.005	XOM	0	0.473
EMR	0	0.313	ORCL	1	0.003			

Table 2.4: F test of β (*indi.* $= 0$ represents $\beta = 0$, *indi.* $= 1$ represents $\beta \neq 0$)

2.5 Conclusion

Expected returns are determined by future uncertainty, which can be represented by option prices. The numeraire-portfolio pricing method links expected returns to option prices. This method states that the numeraire-denominated option price is the conditional expectation of the numeraire-denominated terminal payoff under the physical measure, which contains the information of the expected return. The expected return is estimated by the option calibration and a statistical analysis on the results is performed.

A couple of advantages of this method are summarized below:

1. Stocks are riskier than riskless assets such as the money market account. There-

fore, their expected returns should be higher than the risk-free rate. Otherwise, it does not make sense to invest in them. Realized returns representing the history are so volatile that they may not accurately reveal expected returns. For example, they could be outperformed by the risk-free assets for a relatively long period: the stock market's return was on average less than the risk-free asset for eleven years, from 1973 to 1984 [35]. Furthermore, conditions in markets may change overtime in the long run. Therefore, past average returns may not represent the current situation [6].

Expected returns estimated by the numeraire-portfolio method use option prices with a short period maturity, e.g., one month. Thus, the price information revealed is "local" and it represents future uncertainty. The results in this study show the risk premiums are positive most of the time and more stable than the realized returns. Furthermore, the OLS regressions indicate that the difference of the estimated returns and the realized returns is indifferent to two major risk factors for a large portion of assets. This result indicates that the estimated returns properly estimate the expected returns.

2. The traditional estimation using the CAMP and the Fama-French model requires the input of betas, the risk factors. However, there is no uniformly accepted agreement what betas should be chosen. Academicians and practitioners try to "fine gold" by data mining [6]. The uncertain input in the numeraire-portfolio method is the proxy of the numeraire portfolio. Numerous empirical studies show that market indices or equal-weighted/value-weighted portfolios

can serve as good proxies. Thus, there is less uncertain input in the numeraire-portfolio approach than the traditional methods.

Therefore, our study demonstrates that the numeraire-portfolio approach provides a good estimation for expected returns.

Future study may include the generalization of estimating the expected return. Option is one example that represents future uncertainty. The idea to estimate the expected return has two steps: first, we need to find any equity that can reveal future uncertainty; second, this equity contains the information of the expected return. Futures may be one of the candidates that satisfy the above criteria.

Chapter 3

A New Approach to Portfolio Selection

3.1 Introduction

In financial market, investors often face the question of how to allocate their wealth among various assets, and in what sense. Modern portfolio theory (MPT), first articulated by Markowitz [55] [56], provides selection principles for maximizing a portfolio's expected return when fixing its variance, or minimizing the variance for a fixed level of expected return. These two principles formulate the efficient frontier from which investors can choose their preferred portfolio with the optimal combination of gain (the expected return) and risk (defined as the standard deviation of return), where investors' preference is the trade-off between gain and risk. Another important concept is the diversification. Because every asset is correlated with other assets, a properly constructed portfolio's variance can be smaller than the sum of all assets' variances.

Thus, investors can reduce the risk with a diversified portfolio instead of investing in individual asset. In the modern portfolio theory, asset returns are assumed to be multivariate Gaussian random variables. To optimize a portfolio, investors first need to estimate each asset's expected return, its variance, and the correlation to other assets. Portfolio selection is well developed both in theory and implementation. We refer to Elton and Gruber's review paper [26], which provides literatures on each topic in the modern portfolio theory.

However, various aspects are questioned in the modern portfolio theory. As discussed in Chapter 1, empirical studies indicate individual asset's return is not normally distributed, which makes correlation complicated. Under this situation, covariance may not properly measure correlation. Another issue is in the implementation procedure. The return in the model input is the expected return, which is the prediction to asset's future return. In practice, the expected return is estimated from the historical data, which does not necessarily provide a good prediction. Furthermore, as discussed in Chapter 2, the historical data is very volatile and it cannot give a relatively precise estimation.

In this chapter, we propose some alternative approaches for these questions. A non-Gaussian law, the VG mixed distribution described in Chapter 1, is employed to model the marginal distributions of asset returns. This model well captures the skewness and excess kurtosis patterns exhibited in the data. The joint law is formulated by FGC, a simulation technique proposed by Malevergne and Sornette [50] [51], and summarized by Madan and Khanna [39]. The FGC transforms all marginal random numbers to standard normal random numbers, and then constructs the dependence structure by the covariance matrix of the multivariate normal distribution, which is well defined to measure the correlation. Last, the expected return estimated by the numeraire-portfolio method introduced in Chapter 2 is employed. The estimation is conducted on option prices, which can be viewed as the prediction to asset's future values. It is also demonstrated in Chapter 2 that this estimator is more stable than that from the historical data. Thus, the estimated expected return by the numeraire-portfolio method is expected to serve as a better and more precise estimator.

Criteria in portfolio optimization are other issues to consider, which include the mathematical formulation for the optimization and the measures for portfolio evaluation. Traditionally, the utility function is the objective function and to be maximized in the optimization [56]. A variety of measures have been discussed to evaluate portfolios. The paper by Biglova *et al.* [5] provides a good review and also compares various measures (or risk estimations) that are all in the form of ratios between the expected return and certain risk measures.

New criteria are proposed in this study. We construct a portfolio from the buyer's side. To be a competitive player, the buyer should charge a price as minimal as possible, which is called the bid price [47]. This price is defined based on the acceptability index developed by Cherny and Madan [14], and is the negative of the distorted expectation of the terminal payoff. Different risk level leads to different bid price. Given a risk level, the buyer can reach his or her highest profit by maximizing the bid price which also depends on the composition of the portfolio. Thus, the bid price, or the distorted expectation, serves as the objective function and the associated acceptability index evaluates portfolio's performance.

The outline of the rest of the chapter is as follows. Section 3.2 introduce the criteria of the portfolio optimization, the bid price, and the acceptability index. In Section 3.3, the optimization problem is formulated and the numerical implementation and results are presented. Section 3.4 concludes with a summary.

3.2 Portfolio Evaluation - Acceptability Indices

In this section, we start with the traditional utility function that leads to the concept of an acceptance set and the associated coherent risk measure, which measures the risk level of the acceptance set. The acceptability indices, in the sense of the coherent risk measure, is introduced, along with examples. Finally, given a fixed acceptable level, the bid and ask price are described, which will be employed as the objective function in the portfolio optimization problem.

3.2.1 Acceptance Sets and Coherent Risk Measure

In the classical portfolio optimization problem, an investor allocates wealth by maximizing portfolio's utility function. If he or she starts from a position with a zero-cost portfolio, any positions that will increase the portfolio's terminal expected utility are acceptable by this investor. These positions form a convex set that contains nonnegative terminal cash flows. Every investor has his or her acceptable set depending on the preference, or the utility function. The set that is accepted by all investors is the intersection of all these sets, which is a convex cone. It is called the acceptance set [1], which also includes the nonnegative terminal cash flows. The acceptance sets are studied by Artzner *et al.* [1] and Carr, Geman, and Madan [10]. The model is set up for the random variable X, the terminal cash flow of zero-cost trades, on a probability space (Ω, F, P). The risk measure $\rho(X)$ for the acceptance sets, defined as a mapping from the set of risks to the real-line R, is connected with the acceptance set by a nonempty set of probability measures, denoted as D, which are equivalent to P [1] [10]. This risk measure is called coherent risk measure [1]. Delbaen [21] further indicates the coherent risk measure has the form

$$\rho(X) = -\inf_{Q \in D} E^{Q}[X], \tag{3.1}$$

and a trade X is acceptable when $\rho(X) \leq 0$.

3.2.2 Acceptability Indices

Based on the axioms of the coherent risk measure in [1], Cherny and Madan [14] define the "index of acceptability," a mapping α from the class of bounded random variables X to the positive real line $R^{+} = [0, \infty]$, which has the following four properties:

1. Monotonicity:

   If X is dominated by random variable Y, $X \leq Y$, then, $\alpha(X) \leq \alpha(Y)$

2. Scale Invariance:

   $\alpha(X)$ stays the same when X is scaled by a positive number, $\alpha(cX) = \alpha(X)$ for $c > 0$.

3. Quasi-concavity:

If $\alpha(X) \geq \gamma$ and $\alpha(Y) \geq \gamma$, then, $\alpha(X+Y) \geq \gamma$

4. Fatou Property (Convergence)

   Let $\{X_n\}$ be a sequence of random variables. $|X_n| \leq 1$ and X_n converges in probability to a random variable X. If $\alpha(X_n) \geq x$, then $\alpha(X) \geq x$.

$\alpha(X)$ can be considered as the degree to measure the quality of terminal cash flow X, where bigger the value of $\alpha(X)$, closer is X to the arbitrage. $\alpha(X) = +\infty$ represents arbitrage and all random variables in the acceptable cone are nonnegative.

Under the above four conditions, a basic representation theorem is derived by Cherny and Madan [14], which connects acceptability indices $\alpha(X)$ to family of probability measures.

Theorem 28 ***Representation Theorem of Acceptability Indices***

Let $L^\infty = L^\infty(\Omega, F, \widetilde{P})$ be the probability space of the bounded random variables X. $\alpha(X)$ is an acceptability index which is a map $\alpha : L^\infty \to [0, \infty]$ and satisfies the condition 1-4 if and only if there exists a family of subset $\{D_\gamma : \gamma \geq 0\}$ of $\widetilde{P}$ such that

$$\alpha(X) = \sup\{\gamma \in R^+ : \inf_{Q \in D_\gamma} E^Q[X] \geq 0\}, \tag{3.2}$$

and $\{D_\gamma : \gamma \geq 0\}$ is an increasing family of sets of probability measures, i.e., $D_\gamma \subseteq D_{\acute{\gamma}}$ for $\gamma \leq \acute{\gamma}$.

Remark:

(1) The probability measures in D_γ are absolutely continuous with respect to the original probability measure P for X and each $Q \in D_\gamma$ is equivalent to P.

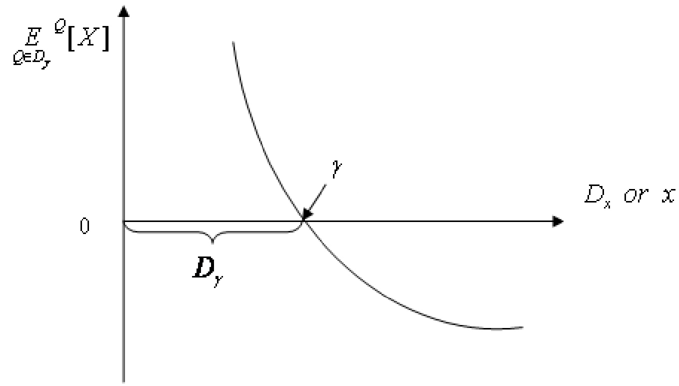


Figure 3-1: Graphic illustration of the Representation Theorem

(2) Eq (3.2) indicates that $\alpha(X) = \gamma$ is the largest value that makes the expectation of X positive under all probability measures in D_γ. This can be roughly illustrated in Figure 3-1.

(3) Recall the coherent risk measure $\rho(X)$ has the relationship in Eq (3.1) with acceptable sets if $\rho(X) \leq 0$. Then, the acceptability index $\alpha(X)$ is linked with $\rho(X)$ by

$$\alpha(X) = \sup\{\gamma \in R^+ : \rho_\gamma(X) \leq 0\}.$$

Thus, $\alpha(X)$ is the largest risk level that the cash flow X is acceptable, and the risk level is γ.

(4) The sets of probability measures $\{D_\gamma : \gamma \geq 0\}$ can be considered as pricing kernels, which will be discussed later.

3.2.3 WVAR Acceptability Indices

There are many acceptability indices that satisfy the four conditions on page 96. The weighted VAR (WVAR) acceptability indices [14] are used in this study because of their computational feasibility. The WVAR has the following form:

$$WVAR_{\gamma}(X) = -\int_{R} x \cdot d\Psi_{\gamma}(F_X(x)), \tag{3.3}$$

where $F_X(x)$ is the cdf of random variable X. $\{\Psi_{\gamma} : \gamma \geq 0\}$ is a set of increasing concave continuous functions with mapping $\Psi : [0,1] \rightarrow [0,1]$, where $\Psi(0) = 0$ and $\Psi(1) = 1$. Furthermore, $\Psi_{\gamma}(y)$ increases in γ with fixed y value. Thus, $\Psi_{\gamma}(y)$ can be viewed as a function to distort the cdf $y = F_X(x)$, adding more weights to the losses, which are the area when $F_X(x)$ is close to x or X decreases in negative values. $\Psi_{\gamma}(F_X(x))$ again is sees as a probability distribution function, and WVAR can be viewed as a distorted expectation of cash flow X.

Apply Eq (3.3) to Eq (3.2), the WVAR acceptability index $\alpha(X)$ is defined as

$$\alpha(X) = \sup\{\gamma \in R^{+} : \int_{R} x \cdot d\Psi_{\gamma}(F_X(x)) \geq 0\}, \tag{3.4}$$

where $\alpha(X)$ is the biggest γ such that the distorted expectation is still positive.

The expectation of X is taken under a new probability measure $Q_{\gamma} \in D_{\gamma}$ by a measure change $\frac{dQ_{\gamma}}{dP} = \Psi_{\gamma}'(F_X(x))$ where P is the original probability measure of X. D_{γ} is the set of probability measures discussed in Section 3.2.2. When γ increases, Ψ_{γ} also increases, which distorts cdf F_X more and more to the left, or in another word, gives more weight to the losses. Under this situation, if the distorted expectation of X still remains positive, it means the trading strategy X is more acceptable as it can survive the worse situations, and thus, a better performance. Therefore, $\alpha(X) = \gamma$ is a performance measure for trade X. Higher γ, better the performance of trading strategy X. γ is also seen as a "stress level" for the cash flow X, higher γ, more stressed of X.

The computation of the distorted expectation is relatively simple. Given a sample $x_1, x_2, ..., x_N$ of cash flow X, the numerical formula is

$$\int_R x \cdot d\Psi_\gamma(F_X(x)) = \sum_{i=1}^{N} x_{(i)} \left(\Psi_\gamma(\frac{i}{N}) - \Psi_\gamma(\frac{i-1}{N}) \right), \tag{3.5}$$

where $\{x_{(i)}\}$ are ordered values sorted increasingly, $\Psi_\gamma(\frac{i}{N})$ is the empirical distribution function with $\Psi_\gamma(\frac{i}{N}) - \Psi_\gamma(\frac{i-1}{N}) = \frac{1}{N}$ for all $i = 1, 2, ...N$.

Four WVAR indices are provided in [14], namely MINVAR, MAXVAR, MINMAXVAR, and MAXMINVAR.

We first look at a simple case. Let $Y \overset{law}{=} \min\{x_1, x_2, ..., x_{\gamma+1}\}$, where $x_1, x_2, ..., x_{\gamma+1}$ are $\gamma + 1$ independent draws from X. If the cdf of X is $z = F_X(x)$, then by the order statistics, the probability function of Y is given by

$$\Psi_\gamma(z) = 1 - (1 - z)^{1+\gamma} \quad \text{where} \;\; z \in [0, 1], \;\gamma \geq 0. \tag{3.6}$$

In this case, γ is the largest number of draws such that the expected value of the minimum of these γ draws still remains positive. Larger value of γ represents better performance of trade X. This acceptability index is termed *MINVAR* as it's related with the minimum from a number of independent draws.

Now let us check how the distortion in MINVAR reweight the loss and gain. Differentiate the distortion function (3.6)

$$\frac{d\Psi_\gamma(F_X(x))}{dx} = (\gamma+1)(1-F_X(x))^\gamma \cdot f_X(x),$$

where $f_X(x)$ is the pdf of X. This derivative shows MINVAR distortion adds more weight to large losses (when $F_X(x)$ is close to 0) and reduces more weight to large gains (when $F_X(x)$ is close to 1). However, large losses can not be reweighted to infinitely large levels. Thus, a modified MINVAR is considered: $\Psi_\gamma(z) = 1-(1-z^{\frac{1}{1+\gamma}})^{1+\gamma}$ with its differentiation

$$\frac{d\Psi_\gamma(F_X(x))}{dx} = (\gamma+1)(1-F_X(x)^{\frac{1}{1+\gamma}})^\gamma \cdot F_X(x)^{-\frac{\gamma}{1+\gamma}} \cdot f_X(x).$$

Under this situation, the large losses can be reweighted infinitely large and the large gains can be reweighted down to zero. This index is called $MINMAXVAR$, which will be implemented in this study.

Details of the other two indices, MAXVAR and MAXMINVAR, can be found in [14].

3.2.4 Bid and Ask Prices

Given a fixed acceptability index and level, we study the corresponding acceptable price for the terminal cash flow X, either from seller's or buyer's view, namely ask price and bid price [47], respectively. Let us derive a trading strategy from buyer's side. If b is the bid price, then the buyer's residual cash flow is $X - b$. To be a competitive buyer in the market, he or she should offer as much as possible. If the residual cash flow $X - b$ is $\gamma - acceptable$ (acceptable level is γ) with certain acceptability index α, then the competitive bid price is taken maximum and is defined as

$$b_\gamma(X) = \sup\{b : \alpha(X - b) \geq \gamma\}. \tag{3.7}$$

For the bid price, by Theorem 28, we have

$$\inf_{Q \in D_\gamma} E^Q[X - b] \geq 0 \Leftrightarrow b \leq \inf_{Q \in D_\gamma} E^Q[X].$$

By Eq (3.7), $b = \inf_{Q \in D_\gamma} E^Q[X]$. Therefore, the bid price is the minimum of the distorted expectation of X among all $Q \in D_\gamma$ with a fixed level γ.

If WVAR is chosen to be the acceptability index, by Eq (3.4), we have

$$\int_R x \cdot d\Psi_\gamma(F_{X-b}(x)) \geq 0 \Leftrightarrow -b + \int_R x \cdot d\Psi_\gamma(F_X(x)) \geq 0.$$

By Eq (3.7), the competitive bid price is

$$b_\gamma(X) = \int_R x \cdot d\Psi_\gamma(F_X(x)), \tag{3.8}$$

the distorted expectation of the terminal cash flow X.

Similarly, if the seller has the residual cash flow $a - X$ with acceptable level γ under the acceptability index α, where a is the ask price. The competitive ask price is the minimal price defined by

$$a_\gamma(X) = \inf\{a : \alpha(a - X) \geq \gamma\}.$$

With the similar procedure, given the acceptable level γ, we can derive the competitive ask price

$$a = \sup_{Q \in D_\gamma} E^Q[X],$$

and

$$a_\gamma(X) = -\int_R x \cdot d\Psi_\gamma(F_{-X}(x)),$$

if using WVAR as the acceptability index.

3.3 Numerical Implementation and Results

3.3.1 Trading Strategy

A trading strategy over a single period is implemented. As a buyer, we construct a stock-only portfolio at time 0 with an bid price b and the maturity t. All the payoff or cash flow X is delivered at the maturity. The resulting residual cash flow $X - b$ is set to be at $\gamma-$*acceptable* level with WVAR the acceptability index α. A buyer maximize the distorted expectation, or the bid price b, which turns out to be Eq (3.8) derived in Section 3.2.4. Different weights of assets result in different trading strategies or the final cash flows X. Thus, this maximized distorted expectation depends on the weights of assets. Optimal weight leads to a maximal distorted expectation for the buyer.

Let $w = (w_1, ..., w_i, ..., w_n)$ and $R = (R_1, ..., R_i, ..., R_n)$ be the weight and return vectors of the portfolio, where n is the total number of stocks. The cash flow x is defined as $x = w\acute{} \cdot R$, which is the portfolio's return. The portfolio optimization problem is formulated as follows:

$$\max_{w} \int_{R} x \cdot d\Psi_{\gamma}(F_X(x)), \tag{3.9}$$

$$s.t. \quad \sum_{i=1}^{n} w_i = 1, \qquad -1 \leq w_i \leq 1.$$

Objective function in (3.9) can be numerically computed by Eq (3.5), in which N samples of x are obtained by simulation. $x = w\acute{} \cdot R = \sum_{i=1}^{n} w_i R_i$, where R_i has the VG mixed distribution [22]. The stock's log return is given by Eq (1.39) reformulated as

$$r_i = \ln \frac{Sti}{S_{0i}} = \mu_i t + X^{(i)}_{VGMixed}(t) + \omega_i t, \tag{3.10}$$

where μ_i is the expected return which can be estimated by the numeraire-portfolio method. The stock's return R_i then equals to $\exp(r_i) - 1$.

3.3.2 Procedure and Results

The results and data obtained from Chapter 2 are employed in this study. The trading strategy is implemented on the 130 estimation days in Chapter 2, ranging from January 1999 to October 2009. On each day, a portfolio containing the first 50 stocks from the S&P 500 (SPX) is constructed, and the holding period is the same as the time-to-maturity of the options used in the calibration. The FGC in Algorithm 27 is implemented to simulate the multivariate random variables $r = (r_1, ..., r_i, ..., r_n)$ with VG mixed distribution as the marginal law for each r_i. The input of FGC, the VG parameters $(\sigma_i, \nu_i, \theta_i)$ from the marginal daily return, the expected return μ_i estimated from the numeraire-portfolio method $(\widehat{\mu}_i)$, and the parameters (c_i, γ_i) in the VG mixed distribution at time horizon t, are all obtained from the results in Chapter 2. Thus, the only parameters left in the objective function (3.9) are the weights, which are attained from optimization. The optimal portfolio is named the estimated-return portfolio (ERP).

Two reference portfolios are constructed to compare to the estimated-return portfolio. They use the same input parameters as the estimated-return portfolio except the one for the expected return μ_i: one using the realized return R_i is named the realized-return portfolio (RRP); the other using the sample mean $\overline{R}_i$ is named the mean-return portfolio (MRP). On each of the 130 estimation days, these two portfolios are attained by the objective function (3.9).

The actual return of the optimized portfolio at the maturity can be calculated by $\sum_{i=1}^{n} \widehat{w}_i f_i$, where $\widehat{w}_i$ is stock i's weight in the optimized portfolio, and f_i is stock i's actual return during the period from the estimation day to maturity. The actual return for each of the three optimized portfolios is calculated.

Besides comparing to the two reference portfolios, we are also interested in comparing the performance of the estimated-return portfolio to the market index, which is SPX in this study. The adjustment for SPX's return is required before comparison due to the leverage effect. In the estimated-return portfolio, the weight ranges from -1 to 1 for each stock and there are 50 stocks in the portfolio. The weight in SPX can be considered as 1. Thus, the estimated-return portfolio is leveraged compared to SPX. To set them to the same leverage level, the return of SPX is multiplied by some ratios, which are defined as follow:

$$l_{pos} = \sum_{i=1}^{n} \widehat{w}_i^{+} \quad \text{and} \quad l_{neg} = \sum_{i=1}^{n} \widehat{w}_i^{-},$$

where

$$\widehat{w}_i^{+} = \begin{cases} \widehat{w}_i, & \text{if } \widehat{w}_i > 0 \\ 0, & \text{otherwise} \end{cases}$$

$$\widehat{w}_i^{-} = \begin{cases} \widehat{w}_i, & \text{if } \widehat{w}_i < 0 \\ 0, & \text{otherwise} \end{cases}.$$

We have two leveraged returns for SPX

$$R^+_{SPX} = l_{pos} \cdot R_{SPX}, \quad R^-_{SPX} = |l_{neg}| \cdot R_{SPX},$$

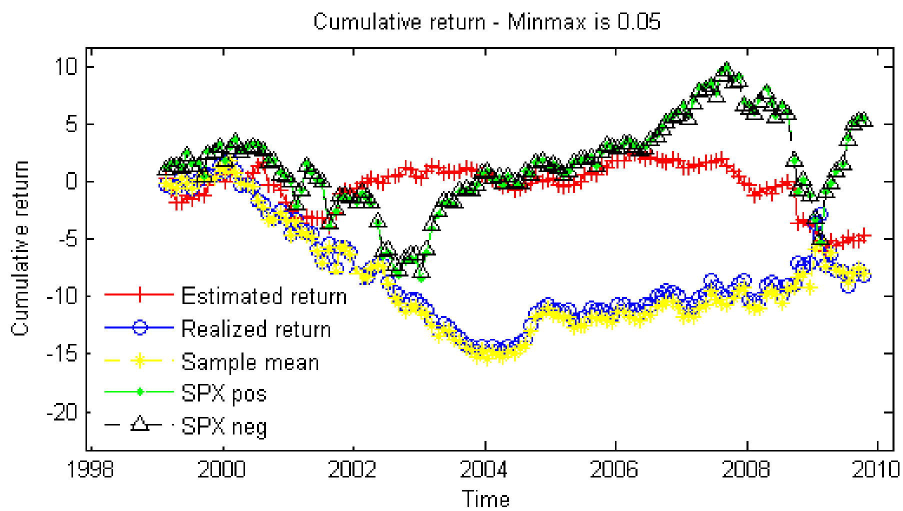
where R_{SPX} is SPX's actual return during the period from the estimation day to maturity. These two leveraged SPX are named positive-leveraged SPX and negative-leveraged SPX.

The three optimized portfolios are constructed at different risk levels, namely $\gamma = 0.05$, $0.10, 0.15$, 0.20, and 0.25. MINMAXVAR is employed as the acceptability index.

At each risk level, cumulative returns are calculated to compare the performances of the five portfolios, namely the estimated-return portfolio, the realized-return portfolio, the mean-return portfolio, the positive-leveraged SPX, and the negative-leveraged SPX. The results of the cumulative returns are graphed in Figure 3-2 to Figure 3-6 for each risk level. The cumulative returns of the estimated-return portfolio at different risk levels is shown in Figure 3-7 to check the effect of risk level on its performance.

The statistics of the returns for each portfolio at every risk level are also displayed in Table 3.1 and 3.2. All the returns are annualized before the analysis.

It is observed from the figures and also confirmed from the tables that the estimated-return portfolio is superior to the other two reference portfolios, and its performance is even better than SPX at all risk level except $\gamma = 0.05$. Furthermore, Table 3.2 also shows the return of the estimated-return portfolio is less volatile than SPX, which means it is mean-variance optimal than SPX.



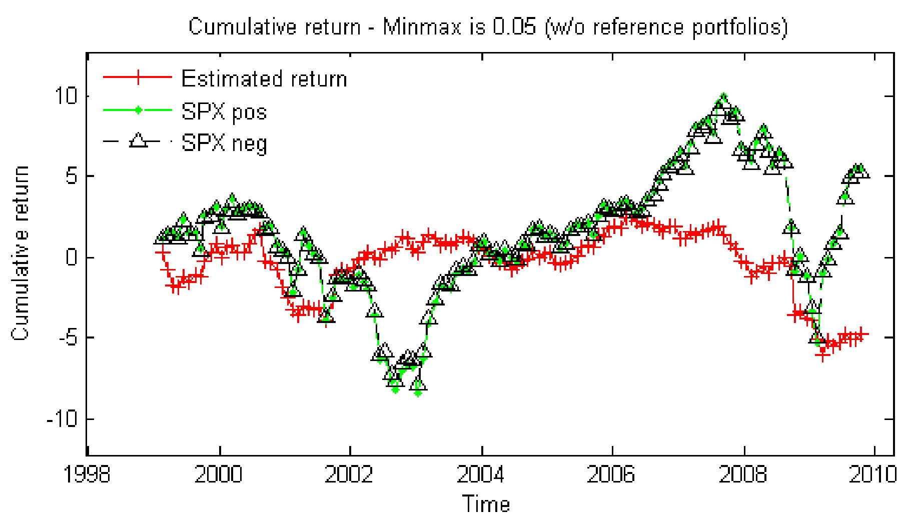
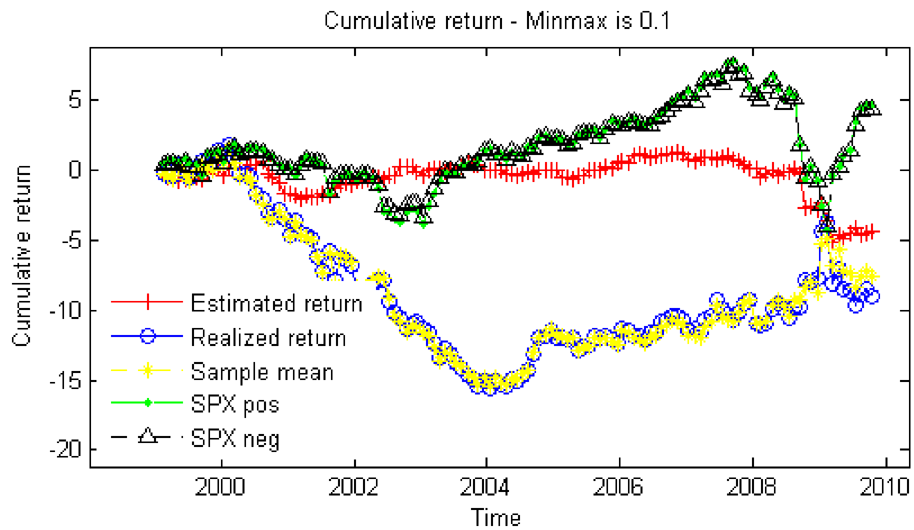


Figure 3-2: Cumulative returns of the five portfolios ($\gamma = 0.05$)



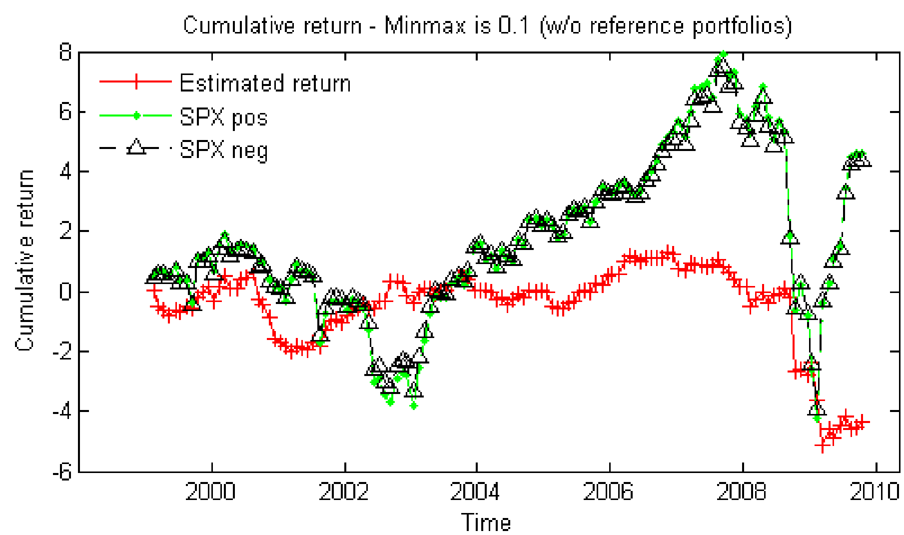


Figure 3-3: Cumulative returns of the five portfolios ($\gamma = 0.10$)

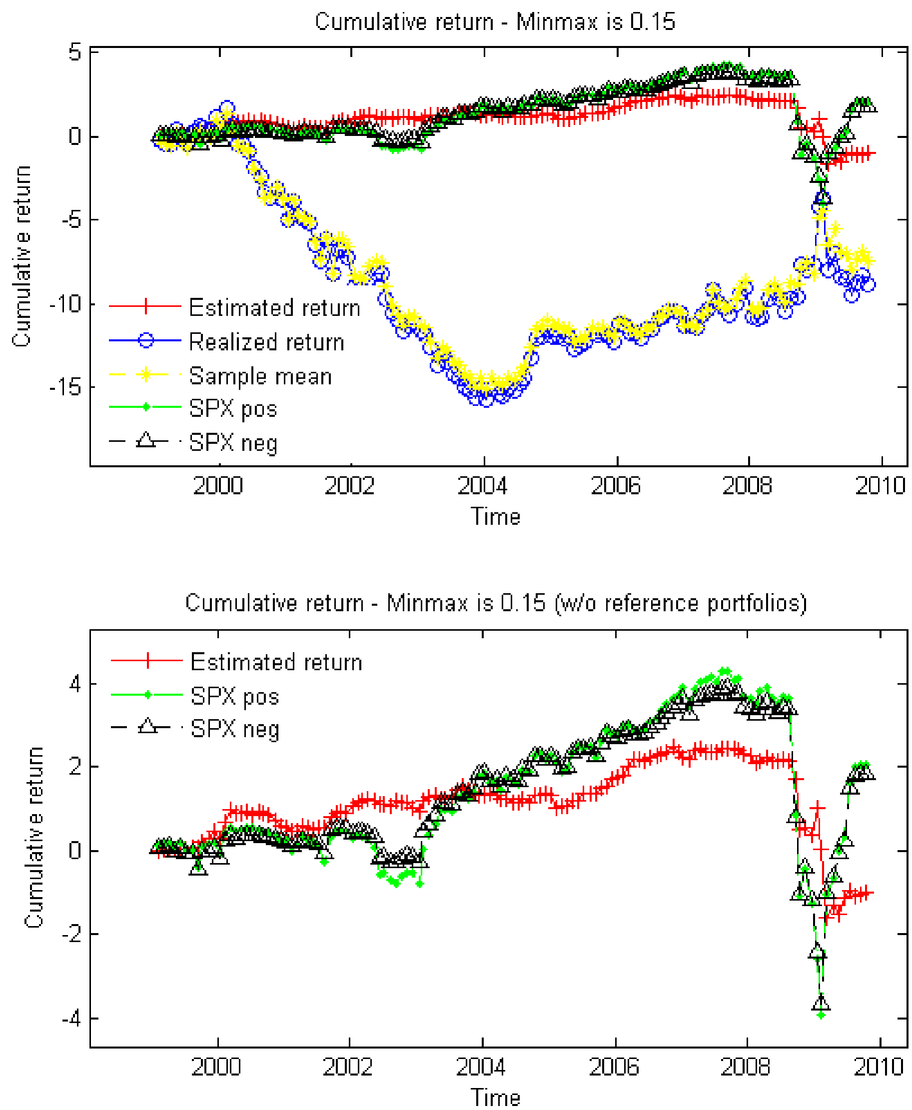


Figure 3-4: Cumulative returns of the five portfolios ($\gamma = 0.15$)

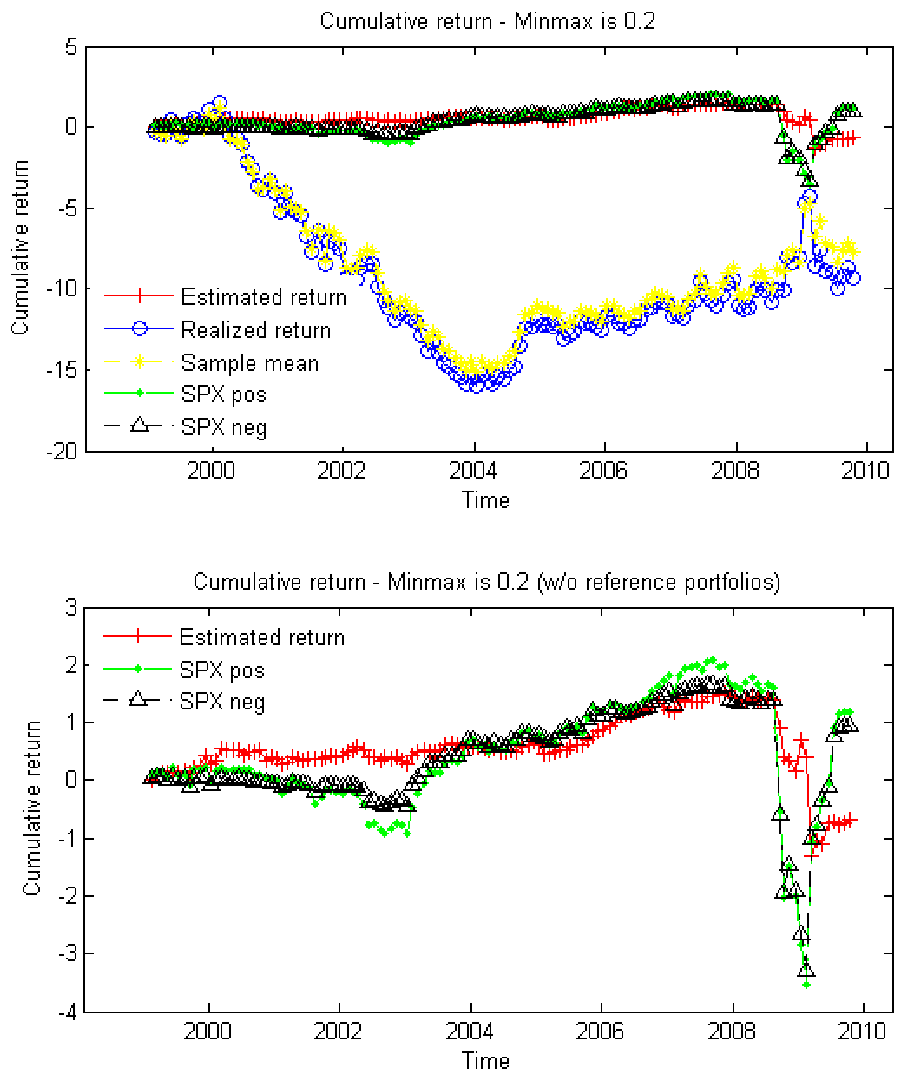


Figure 3-5: Cumulative returns of the five portfolios ($\gamma = 0.20$)

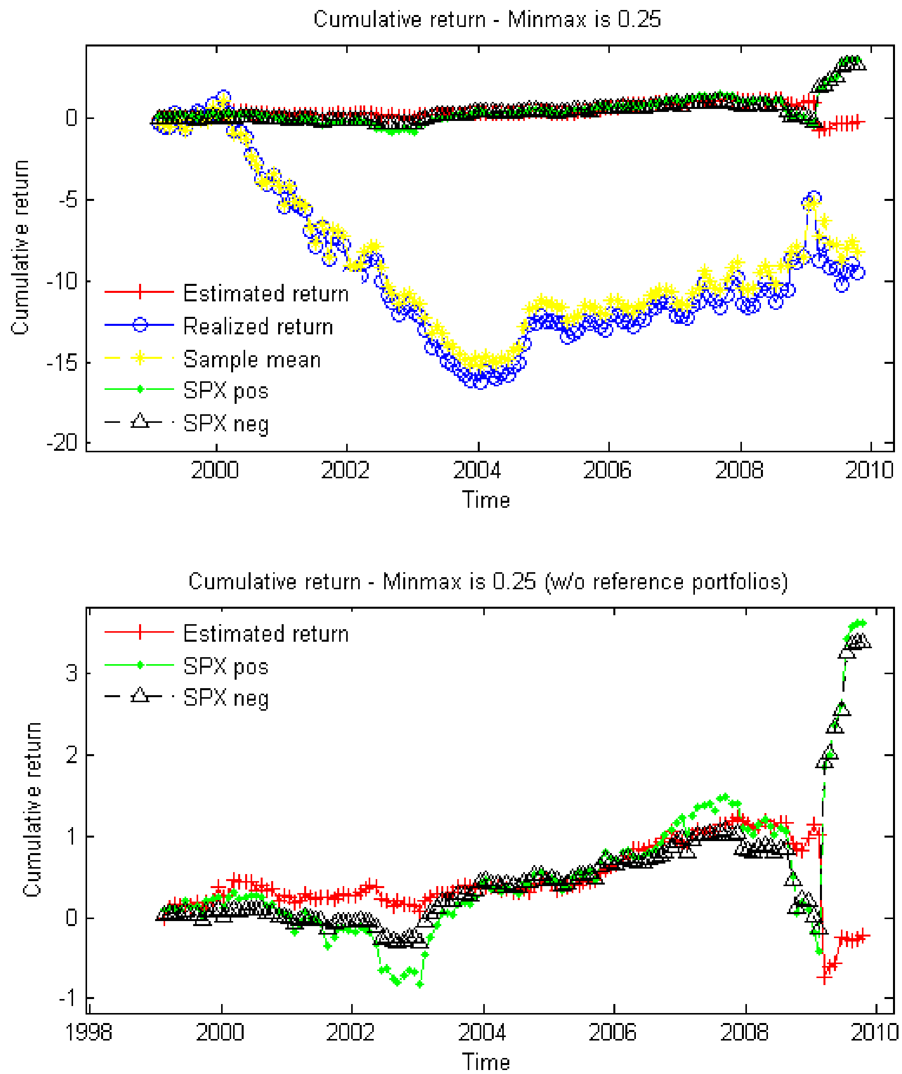


Figure 3-6: Cumulative returns of the five portfolios ($\gamma = 0.25$)

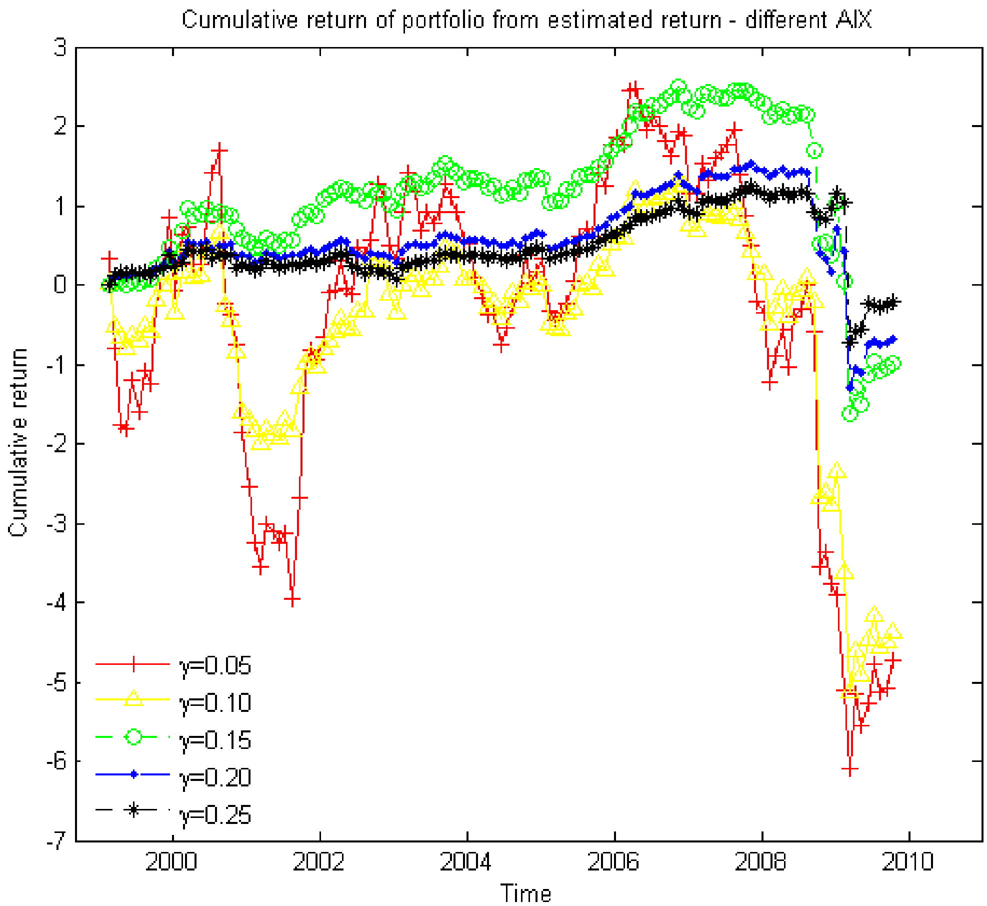


Figure 3-7: Estimated-return portfolio at different risk level γ (AIX=MINMAXVAR)

	ERP	*RRP*	*MRP*	*SPX+*	*SPX-*
$\gamma = \mathbf{0.05}$	-0.0176	-0.031	-0.0319	0.0156	0.015
$\gamma = \mathbf{0.10}$	-0.0175	-0.0873	-0.0824	0.0156	0.0174
$\gamma = \mathbf{0.15}$	0.0078	-0.2722	-0.2272	0.0156	0.0226
$\gamma = \mathbf{0.20}$	0.0264	-0.4262	-0.3671	0.0156	0.0193
$\gamma = \mathbf{0.25}$	0.0394	-0.3704	-0.3829	0.0156	0.0252

Table 3.1: Mean of portfolio return at different risk level (January 1999 - October 2009)

	ERP	*RRP*	*MRP*	*SPX+*	*SPX-*
$\gamma = \mathbf{0.05}$	0.2895	0.4373	0.4159	0.5964	0.5666
$\gamma = \mathbf{0.10}$	0.2740	0.9312	0.9438	0.5964	0.5406
$\gamma = \mathbf{0.15}$	0.2797	2.4156	2.5823	0.5964	0.4539
$\gamma = \mathbf{0.20}$	0.3155	3.4108	3.5483	0.5964	0.3989
$\gamma = \mathbf{0.25}$	0.3599	3.9118	3.9843	0.5964	0.3429

Table 3.2: Std. of portfolio return at different risk level (January 1999 - October 2009)

3.4 Conclusion

A new method is proposed to the classic portfolio selection problem. Several new approaches in this method are employed: Portfolios are constructed in a non-Gaussian environment; the FGC technique is employed to construct the complicate dependence structure; a new estimator for expected return is used, which is expected to provide a better and precise estimation; finally, new criteria, the distorted expectation and the acceptability index, are employed to mathematically formulate the optimization and evaluate the portfolio performance.

Three kinds of portfolios are built with the same setting in the portfolio selection procedure, except the estimator for the expected return input. These portfolios are compared to two leveraged SPX. Comparison is conducted at each of the five risk levels for the five portfolios. We observed that the estimated-return portfolio outperforms the other two reference portfolios, which have consistent loss This indicates the estimated return using the numeraire-portfolio method is an effective estimator for the asset's expected return, compared to the estimators from the historical data. Furthermore, this estimator may also serve as a good input in the portfolio optimization at higher acceptable level (γ is near or above 0.20) because, the portfolio has the similar performance as the market index at these levels.

Future work following this study can be the portfolio performance testing under various scenarios, including different optimization criteria, e.g., utility function as the objective function, more variety of components in portfolios, such as options and bonds. The purpose is to find out whether, in more general scenario, the numeraire-portfolio method provides an effective estimator for the expected return in portfolio selection.

Bibliography

[1] Artzner, P., F. Delbaen, J. Eber, and D. Heath (1999). *Coherent Measures of Risk*, Mathematical Finance, 9: 203–228.

[2] Bachelier, L. (1900), *Théorie de la spéculation*, Annales de l'Ecole Normale Supérieure, 17: 21–86. (English translation: Cootner, P., ed. 1964, *The random character of stock market Prices*, MIT Press: Cambridge, MA).

[3] Barndorff-Nielsen, O. E. (1997). *Normal inverse Gaussian distributions and stochastic volatility modelling*, Scandinavian Journal of statistics, 2: 41–68.

[4] Bartholdy, J. and P. Peare (2005). *Estimation of expected return: CAPM vs. Fama and French*, International Review of Financial Analysis, 14: 407–427 .

[5] Biglova, A., S. Ortobelli, S. Rachev and S. Stoyanov (2004). *Different approaches to risk estimation in portfolio theory*, Journal of Portfolio Management, 31: 103–112.

[6] Black, F. (1995). *Estimating expected return*, Financial Analysts Journal, 51: 168–171.

[7] Black, F. and M. Scholes (1973), *The pricing of options and corporate liabilities*, Journal of Political Economy, 81: 637–654.

[8] Breeden, D. (1979). *An intertemporal asset pricing model with stochastic consumption and investment opportunities*, Journal of Financial Economics, 7: 265–296.

[9] Bühlmann, H. and E. Platen (2002). *A discrete time benchmark approach for finance and insurance*, Working paper.

[10] Carr, P., H. Geman, and D. B. Madan (2001). *Pricing and hedging in incomplete markets*. Journal of Financial Economics, 62:131–67.

[11] Carr, P., H. Geman, D. B. Madan, and M. Yor (2002). *The fine structure of asset returns: An empirical investigation*, Journal of business, 75: 305–332.

[12] Carr, P., H. Geman, D. B. Madan, and M. Yor (2007). *Self-decomposability and option pricing*, Mathematical Finance, 17: 31–57.

[13] Carr, P. and D. B. Madan (1998). *Option valuation using the fast Fourier transform*, Journal of Computational Finance, 2: 61–73.

[14] Cherny, A. and D.B. Madan (2009), *New measures for performance evaluation*, Review of Financial Studies, 22: 2571–2606.

[15] Claus, J. and J. Thomas (2001). *Equity premia as low as three percent? Evidence from analysts' earnings forecasts for domestic and international stock markets*, Journal of Finance, 56: 1629–1666.

[16] Cochrane, J. H. (2001). *Asset Pricing*, Princeton University Press.

[17] Cont, R. (2001). *Empirical properties of asset returns: stylized facts and statistical issues*, Quantitative Finance, 1: 223–236.

[18] Cont, R., J.-P. Bouchaud, and M. Potter (1997). *Scaling in stockmarket data: stable laws and beyond*, in Scale Invariance and Beyond, Dubrulle, B., F. Graner, and D. Sornette, eds., Springer: Berlin.

[19] Cont, R. and P. Tankov (2004). *Financial modelling with jump processes*, Chapman & Hall/CRC.

[20] Cox, J. C. and S. A. Ross (1976). *The valuation of options for alternative stochastic processes*, Journal of Financial Economics, 3: 145–166.

[21] Delbaen, F. (2002). *Coherent risk measures on general orobability spaces*, in K. Sandmann and P. Schönbucher (eds.), Advances in Finance and Stochastics: Essays in Honor of Dieter Sondermann, Berlin: Springer 1–37.

[22] Eberlein, E. and D. B. Madan (2010). *The distribution of returns at longer horizons*, Working paper.

[23] Eberlein, E. and F. Ökzan (2003). *Time consistency of Lévy models*, Quantitative Finance, 3: 40–50.

[24] Eberlein, E., and K. Prause, (2000). *The generalized hyperbolic model: financial derivatives and risk measures*, Mathematical Finance-Bachelier Congress.

[25] Edwards, E. O. and P. W. Bell (1961). *The theory and measurement of business income*, New York, John Wiley and Sons.

[26] Elton, E. J. and M. J. Gruber (1997). *Modern portfolio theory, 1950 to date*, Journal of Banking & Finance, 21: 1743–1759.

[27] Embrechts, P., F. Lindskog, and A. McNeil (2003). *Modelling dependence with copulas and applications to risk management*, in Rachev, S. (Eds.) Handbook of Heavy Tailed Distributions in Finance, Elsevier, Chapter 8: 329–384.

[28] Fama, E. F. and K. R. French (1993). *Common risk factors in the returns on stocks and bonds*, Journal of Financial Economics 33: 3–56.

[29] Fama, E. F. and J. D. MacBeth (1974). *Long-term growth in a short-term market*, Journal of Finance, 29: 857–885.

[30] Gabaix, X., P. Gopikrishnan, V. Plerou, and H. E. Stanley (2003). *A theory of power-law distributions in financial market fluctuations*, Nature, 423: 267–270.

[31] Galloway, M. L., and C. A. Nolder (2007). *Option pricing with selfsimilar additive processes*, Working paper.

[32] Gopikrishnan, P., V. Plerou, L. A. Nunes Amaral, M. Meyer, and H. E. Stanley (1999). *Scaling of the distribution of fluctuations of financial market indices*, Physical Review E, 60: 5305–5316.

[33] Hansen, L. P. (1982). *Large sample properties of generalized method of moments estimators*, Econometrica, 50: 1029–1054.

[34] Harrison, J. M. and D. M. Kreps (1979). *Martingales and arbitrage in multiperiod securities markets*, Journal of Economic Theory, 20: 381–408.

[35] Ibbotson Associates (1995). *Stocks, bonds, bills and inflation*, Yearbook, Ibbotson Associates, Chicago, Ill.

[36] Jacod, J. and A. N. Shiryaev (1987). *Limit theorems for stochastic processes*, Berlin: Springer-Verlag.

[37] Keller, U. (1997). *Realistic modelling of financial derivatives*, Dissertation. Mathematische Fakultät der Albert-Ludwigs-Universität Freiburg im Breisgau.

[38] Kelly, J. R., Jr. (1956). *A new interpretation of information rate*, Bell System Technical Journal, 35: 917–926.

[39] Khanna, A. and D. B. Madan (2009). *Non Gaussian models of dependence in returns*, Working paper.

[40] Khintchine, A. Y. (1938). *Limit laws of sums of independent random variables*, ONTI, Moscow, Russia.

[41] Knight, F. B. (2001). *On the path of an inert object impinged on one side by a Brownian particle*, Probability Theory Related Fields 121: 577–598.

[42] Lévy, P. (1937): *Théorie de l'Addition des Variables Aléatoires*, Paris: Gauthier-Villars.

[43] Li, D. X. (2000). *On default correlation: a copula function approach*, Journal of Fixed Income, 9: 43–54.

[44] Lintner, J. (1965). *The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets*, Review of Economics and Statistics, 47: 13–37.

[45] Long, J. B. (1990). *The numeraire portfolio*, Journal of Financial Economics, 26: 29–69.

[46] Lucas, R. E., Jr. (1978). *Asset prices in an exchange economy*, Econometrica, 46: 1429–1445.

[47] Madan, D. B. (2010). *Pricing and hedging basket options to prespecified levels of acceptability*, Quantitative Finance, 10: 607–615.

[48] Madan, D., P. Carr, and E. Chang (1998). *The variance gamma process and option pricing*, European Economic Review, 2: 79–105.

[49] Madan, D. and E. Seneta, (1990). *The variance gamma (VG) model for share market returns*, Journal of Business, 63: 511–524.

[50] Malevergne, Y. and D. Sornette (2004). *Multivariate Weibull distributions for asset returns: I*, Finance Letters, 2: 16–32.

[51] Malevergne, Y. and D. Sornette (2005). *High-order moments and cumulants of multivariate Weibull asset return distributions: Analytical theory and empirical tests: II*, Finance Letters, 3: 54–63.

[52] Mandelbrot, B. (1963). *The variation of certain speculative prices*, Journal of business, 36: 394–419.

[53] Mandelbrot, B. (1997). *Fractals and scaling in finance: discontinuity, concentration, risk: selecta volume E.*

[54] Mantegna, R. N., and H. E. Stanley (1995). *Scaling behaviour in the dynamics of an economic index*, Nature, 376: 46–49.

[55] Markowitz, H. M. (1952). *Portfolio selection*, Journal of Finance, 7: 77–91.

[56] Markowitz, H. M. (1959). *Portfolio selection: efficient diversification of investments*, John Wiley & Sons., New York.

[57] Mehra, R. and E. C. Prescott (1985). *The equity premium: a puzzle*, Journal of Monetary Economics, 15: 145–161.

[58] Merton, R. C. (1973). *Theory of rational option pricing*, Bell Journal of Economics and Management Science, 4: 141–183.

[59] Merton, R. C. (1976). *Option pricing when underlying stock returns are discontinuous*, Journal of Financial Economics, 3: 125–144.

[60] Mossin, J. (1966). *Equilibrium in a Capital Asset Market*, Econometrica, 34: 768–783.

[61] Nelsen, R. B. (1999). *An introduction to copulas*, Springer-Verlag, New York.

[62] Ohlson, J. (1995). *Earnings, book values and dividends in security valuation*, Contemporary Accounting Research, 11: 661–687.

[63] Peters, E. E. (1999). *Complexity, risk and financial markets*, Wiley, New York.

[64] Philips, T. K. (2003). *Estimating expected returns*, Journal of Investing, 12: 49–57.

[65] Platen E. and D. Heath (2006). *A Benchmark Approach to Quantitative Finance*, Springer Finance, Springer.

[66] Platen, E. (2009). *A benchmark approach to investing and pricing*, working paper.

[67] Prause, K. (1999). *The generalized hyperbolic model: Estimation, financial derivatives, and risk measures*, Doctoral Thesis, University of Freiburg.

[68] Preinreich, G. (1938). *Annual survey of economic theory: The theory of depreciation*, Econometrica, 6: 219–241.

[69] Roll, R. (1973). *Evidence on the "Growth-Optimum" model*, Journal of Finance, 28: 551–66.

[70] Roll, R. (1977). *A critique of the asset pricing theory's tests Part I: On past and potential testability of the theory*, Journal of Financial Economics, 4: 129–176.

[71] Rubinstein, M. (1976). *The valuation of uncertain income streams and the pricing of options*, Bell Journal of Economics, 7: 407–425.

[72] Sato, K. (1999): *Lévy processes and infinitely divisible distributions*, Cambridge: Cambridge, University Press.

[73] Schoutens, W. (2001). *The Meixner processes in finance*, EURANDOM Report 2001–2001, EURANDOM, Eindhoven.

[74] Schoutens, W. (2003). *Lévy processes in finance: pricing financial derivatives*, John Wiley & Sons Inc.

[75] Sharpe, W. F. (1964). *Capital asset prices: A theory of market equilibrium under conditions of risk*, Journal of Finance, 19: 425–442.

[76] Sklar, A. (1959). *Fonctions de répartition à n dimensions et leurs marges*, Publications de I'Institut de Statistique de L'Université de Paris, 8: 229–231.

[77] Vasicek, O. (1977). *An equilibrium characterization of the term structure*, Journal of Financial Economics, 5: 177–188.

[78] Walker, J. S. (1996). *Fast Fourier Transforms*, CRC Press, Boca Raton, Florida.

[79] Welch, I. (2000). *Views of financial economists on the equity premium and on professional controversies*, The Journal of Business, 73: 501–537.

[80] http://en.wikipedia.org/wiki/Generalized_method_of_moments

[81] http://www.fftw.org

[82] http://www.mscibarra.com

[83] http://wrds.wharton.upenn.edu