
The Chinese Equity Risk Factor Model Based on Heavy Tailed Distributions, and Its Application in Risk Management and Portfolio Optimization

A Dissertation Presented

by

Tianyu (Daniel) Lu

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Applied Mathematics and Statistics

(Quantitative Finance)

Stony Brook University

December 2013

Stony Brook University

The Graduate School

Tianyu (Daniel) Lu

We, the dissertation committee for the above candidate for the
Doctor of Philosophy degree, hereby recommend
acceptance of this dissertation.

Svetlozar (Zari) Rachev
Advisor
Department of Mathematics and Statistics

Wei Zhu
Chairman
Department of Mathematics and Statistics

Xiaolin Li
Member
Department of Mathematics and Statistics

Christopher Bishop
Outside Member
Department of Mathematics

This dissertation is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Abstract of the Dissertation

The Chinese Equity Risk Factor Model Based on Heavy Tailed Distributions, and Its Application in Risk Management and Portfolio Selection

by

Tianyu (Daniel) Lu

Doctor of Philosophy

in

Applied Mathematics and Statistics

(Quantitative Finance)

Stony Brook University

2013

In financial risk and portfolio management, the heavy tailed distributions of one-dimensional asset returns and complex dependence structure among multi-dimensional asset returns are two widely discussed problems. Various distributions, which share the desired properties, such as heavy tails, skewness and kurtosis, have been introduced and applied into practice. On the other hand, factor modeling as a technique of dimension reduction, has been used in multivariate statistical analysis and econometric model building, especially in high-dimensional world of financial risk management

In this dissertation, we study these two popular problems in risk and portfolio management. The findings are empirically examined through their applications to the Chinese stock markets. More specifically, we study the behavior of stock returns in the Chinese market from 2002 to 2012 and build the advanced risk factor model for risk and portfolio management. Firstly we give an empirical examination of the Chinese market with testing the Gaussian hypothesis and alternative non-Gaussian distribution hypotheses under the (1) unconditional homoscedastic distribution assumption and (2) conditional heteroscedastic distribution assumption. An ARMA-GARCH model with non-Gaussian distributed innovations is applied to the index and backtested during highly volatile market periods from 2006 to 2012. The model provides a strong capacity of forecasting and possible warning signals of coming market crash. Secondly, we build the equity multi-factor model covering the entire Chinese stock markets. Risk factor returns, including market factors and fundamental factors, are estimated though time series regression and cross-sectional regression. The forecasting methods are created with multivariate ARMA-GARCH models for different distributed innovations, and compared with industrial standard methods. Thirdly, we applied the risk factor models for portfolio and risk management, including risk monitoring, risk budgeting and portfolio optimization. Different risk measures and optimization strategies are tested to provide the most suitable tools.

To my parents

Table of Contents

List of Figures

List of Tables

Acknowledgement

I would like to express my profound gratitude to my advisor, Professor Svetlozar Rachev, for suggesting this important and exciting thesis topic and for his advice, support, encouragement and guidance toward my Ph. D. degree. He taught me not only the way to do scientific research, but also the way to become a professional scientist. He was always enthusiastic about my education. He is my advisor and a lifetime role model for me.

I would like to thank Professor Andrew Mullhaupt. Sharp from whom I have learned many important financial and mathematical skills. I would also like to thank Professor Wei Zhu, Professor Xiaolin Li and Professor Christopher Bishop for being on my dissertation committee.

I would like to thank my co-authors: Dr. Frank Fabozzi, Dr. Aaron Kim and Dr. Ivan Mitov for their kindness help in writing the papers.

I would like to thank all my friends during my years of study as a graduate student at Stony Brook for their friendship and encouragement. In particular, I would like to mention Drs. Yuan Shang, Xiaoping Zhou, Xu Dong, Tengjie Jia, Pengyuan Shao, Yu Mu and many others. They have shared with me many interesting and inspiring ideas.

Throughout my academic career, the constant support of my parents and my girlfriend has always motivated me to strive forward. Their unconditional love has never been affected by the physical distance between us. My dissertation is dedicated to them.

1.Probabilistic Models for Assets Return

1.1 Univariate Distributions

1.1.1 Introduction

In this study, all probability distributions discussed here are continuous probability distributions. Although the price or return of a financial asset is not a continuous random variable, there is no loss in describing the random variables as continuous for a substantial gain in mathematical tractability and convenience.

Numerous studies of return and price distributions of different assets and national financial markets reject the notion that the distributions are normal. These investigations led to the conclusion that the multivariate normal distribution is not appropriate to capture the complex dependence structure between assets, since it does not allow for modeling tail dependence between assets and leptokurtosis as well as heavy tails of marginal distributions. Since Mandelbrot (1963) introduced the Lévy stable (or α-stable) distribution to model the empirical distribution of asset prices, the α-stable distribution became the most popular alternative to normal distribution. Rachev and Mittnik (2000) and Rachev et al. (2008) have developed financial models with α-stable distributions and applied them to market and credit risk management, option price and portfolio selection as well as discussing the major attacks on the α-stable models.

While the α-stabel distribution has certain desirable properties that mentioned above and will be discussed in more detail in Section 1.1.2, it is not suitable modeling applications such as the modeling of option prices. In order to obtain a well-defined model for pricing options, the mean, variance and exponential moments of the return distribution must exit. In addition, a fair conclusion of the literature is that while the empirical evidence does not support the normal distribution, it is also not consistent with an α-stable distribution. The distribution of returns for assets has heavier tails relative to the normal distribution and thinner tails than the α-stable distribution. For these reasons, the so-called tempered stable distributions obtained from tempering the tail properties of the α-stable distributions have been proposed for financial modeling. In this study we focus on the classical tempered stable (CTS) distribution (Koponen, 1995; Boyarchenko and Levendorskiĭ, 2000; Carr et al., 2002). Other examples of the class of tempered stable distributions include and the modified tempered stable (MTS) distribution (Kim et al., 2009), the normal tempered stable (NTS) distribution (Barndorff-Nielsen and

Levendorskiĭ, 2001; Kim et al. 2008), the Kim-Rachev tempered stable (KRTS) distribution (Kim et al., 2007) and the rapidly decreasing tempered stable (RDTS) distribution (Bianchi et al., 2010; Kim et al., 2010). These distributions have not only heavier tails than the normal distribution and thinner than the α-stable distribution, but also have finite moments for all orders.

1.1.2 Gaussian Distributions

The class of normal distributions, or Gaussian distributions, is certainly one of the most important probability distributions in statistics, and due to some of its appealing properties, also the class that is used in applications in finance.

Definition 1. (Gaussian distribution)

The random variable $X \in \mathbb{R}$ is said to follow a Gaussian distribution $N(\mu, \sigma)$, if the density function of the random variable is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{\sigma^2}} \qquad 1.1$$

where μ is called the location parameter and σ is called the scale parameter.

A Gaussian distribution with $\mu = 0$ and $\sigma = 1$ is said to be a standard Gaussian distribution. Gaussian distribution has some interesting properties, which are useful in finance. Here we only discuss the properties that are shared with some particular classes of distributions, such as elliptical distributions, stable distributions, and infinitely divisible distributions.

First, it is *closed under linear combination*, i.e. a random variable Y, obtained as $Y = aX + b$, is Gaussian distributed if X is a Gaussian distributed random variable. Second, it follows *a stable law*, i.e. if X_1, X_2 are two independent copies of a Gaussian distributed random variable X, then for $a, b, c > 0$ and $d \in \mathbb{R}$, the linear combination $aX_1 + bX_2$ has the same distribution as $cX + d$, which is also a Gaussian distribution.

Third, it is *infinitely divisible*, i.e. for a Gaussian distribution with distribution function (d.f.) F and for every positive integer n, there exist n independent, distributed (i.i.d.) random variable $X_1, \dots, X_n$ whose sum $\sum_{i=1}^{n} X_i$ also has the d.f. F.

The last important property is the fact that the Gaussian distribution possesses a *domain of attraction*. A mathematical result called the *Central Limit Theory* states that under certain technical conditions the distribution of a large sum of random variables behaves necessarily like a normal distribution. It is often misunderstand that the Gaussian distribution is the unique class of probability distributions that has this property. In fact, it is the class of stable distribution (containing the Gaussian distribution), which is unique in the sense that a large sum of random variables can only converge to a stable distribution. All these properties will be discussed in details in the next sub-sections.

1.1.3 α-Stable Distributions

As mentioned in Section 1.1.1, there are several reasons for popularity of the α-stable distribution for modeling asset returns

(i) α-stable distributions are leptokurtic; i.e., compared to the Gaussian distributions, they are typically heavy-tailed and more peaked around the center.

(ii) α-stable distributions have domains of attraction. The CLT for normalized sums of i.i.d random variables determine the domain of attraction of each stable law.

(iii) α-stable distributions belong to their own domain of attraction, i.e., the set of α-stable distributions is stable with respect to the n-fold convolution and scaling.

(iv) The family of α-stable distributions is fairly flexible, given that it is characterized with four parameters.

Here we given the definition of α-stable distributions. (see, Samorodnitsky and Taqqu, 1944**).**

Definition 2 (stable law)

Suppose $X_1, \dots, X_n$ are i.i.d random variables with d.f. H. The d.f. H is said to be stable if there exist constants $a_n > 0$ and $b_n \in \mathbb{R}$, such that for any n:

$$a_n(X_1 + \cdots + X_n) + b_n \overset{d}{=} X_1 \tag{1.2}$$

D.f. H is said to be strictly stable if **(1.2)** holds with $b_n = 0$. A stable d.f. is called symmetric if $H(x) = 1 - H(-x)$. A symmetric stable d.f. is obviously strictly stable.

Definition 3 (α-stable distribution)

The definition can also be represented explicitly with characteristic function :

$$\phi_{stable}(u) = \phi_{stable}(u; \alpha, \beta, c, m) \tag{1.3}$$
$$= \begin{cases} \exp\left(-c^{\alpha}|u|^{\alpha}\left(1 - i\beta sign(u)\tan\left(\frac{\pi\alpha}{2}\right)\right) + imu\right), \text{if } a \neq 1, \\ \exp\left(-c|u|\left(1 + i\beta\frac{2}{\pi}sign(u)\ln(\mathrm{u})\right) + imu\right), \quad \text{if } a = 1, \end{cases}$$

where $\alpha \in (0,2], \beta \in [-1,1], c > 0$ and $m \in \mathbb{R}$, $sign(u)$ is 1 if $u > 0$, 0 if $u = 0$, and -1 if $u < 0$.

The parameter α is called the index of stability and can also be interpreted as a shape parameter. The parameter β is called the skewness of the distribution. If $\beta = 0$, the distribution is symmetric, if $\beta > 0$, it is skewed to the right, if $\beta < 0$, it is skewed to the left. The distribution is said to be totally skewed to the right if $\beta = 1$, and totally skewed to the left if $\beta = -1$. The parameter c is called the scale parameter. Finally the parameter m is called the location parameter. We shall often assume for simplicity that $\mu = 0$, because it affects only location. Since the characteristic function of an α-stable distribution is determined by these four parameters, we denote stable distribution by $S_\alpha(c, \beta, m)$. Here denote the scale parameter as c instead of σ, as in general, α-stable distributions don't have second moment, except some special cases.

Property 1. (Domain of attraction)

D.f. H is said to be in the domain of attraction of α-stable distributions S_α, if for any sequence of $X_1, X_2, \ldots$ of i.i.d random variables with common d.f. H, there are sequence of constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$Z_n := a_n(X_1 + \cdots + X_n) - b_n \overset{d}{\rightarrow} Y_\alpha \qquad 1.4$$

where Y_α is an S_α-stable distributed random variables.

It follows from the definition of α-stable that S_α belongs to its own domain of attraction, i.e., if X_k is assumed to be α-stable distributed then, under certain normalozation, the sums $X_1 + \cdots + X_n$ has the same distribution. On the orther hand if the d.f of X_k belongs to the domain of attraction of S_α the normalized sum Z_n is asymptotically S_α-distributed.

There are three special cases of α-stable distribution, which has closed-form solution for the densities. They are (1) the Gaussian case ($\alpha = 2$), (2) the Cauchy case ($\alpha = 1$), and (3) the Lévy case ($\alpha = 1/2, \beta = \pm 1$) with the following respective densities:

Gaussian:
$$f(x) = \frac{1}{\sqrt{2\pi c^2}} e^{-\frac{(x-m)^2}{c^2}}, -\infty < x < \infty \qquad 1.5$$

Cauchy:
$$f(x) = c/(\pi((x-m)^2 + c^2)), -\infty < x < \infty \qquad 1.6$$

Lévy:
$$f(x) = \sqrt{c}/\left(\sqrt{2\pi}(x-\mu)^{\frac{3}{2}}\right) e^{-\frac{c}{2(x-m)}}, m < x < \infty \qquad 1.7$$

Here we briefly describe four basic properties of α-stable distributions.

(1) The tail of the distributions decays like a power function (slower than the exponential decay), which can be described as

$$P(|X| > x) \propto C \cdot x^{-\alpha}, x \to \infty \qquad 1.8$$

for some constant C.

(2) The raw moments of the distributions satisfy the property:

$$E|X|^p < \infty \text{ for any } 0 < p < \alpha \qquad 1.9$$

$$E|X|^p = \infty \text{ for any } p \geq \alpha$$

(3) The mean of the distribution is finite only for $\alpha > 1$

$$E(X) = m, \text{for } \alpha > 1 \qquad 1.10$$

$$E(x) = \infty, \text{for } 0 < \alpha \leq 1$$

1.1.4 Classical Tempered Stable Distributions

In this section, the CTS distribution and the generalized CTS distribution are discussed with their definitions and properties.

Definition 4 (Classical tempered stable distribution)

A random variable X is said to follow the classical tempered stable (CTS) distribution if the characteristic function of X is given by

$$\begin{aligned} \phi_X(u) &= \phi_{CTS}(u; \alpha, C, \lambda_+, \lambda_-, m) \qquad 1.11 \\ &= \exp(ium - iuC\Gamma(1-\alpha)(\lambda_+^{\alpha-1} - \lambda_-^{\alpha-1}) \\ &+ C\Gamma(-\alpha)((\lambda_+ - iu)^\alpha - \lambda_+^\alpha + (\lambda_-^\alpha - iu)^\alpha - \lambda_-^\alpha)) \end{aligned}$$

where $\alpha \in (0,1) \cup (1,2), C, \lambda_+, \lambda_- > 0$, and $m \in \mathbb{R}$, and we denote $X \sim CTS(\alpha, C, \lambda_+, \lambda_-, m)$.

Using the *n*-th derivative of $\psi(u) = \log \phi(u)$ evaluated around zero, the cumulants of X are obtained by

$$\begin{aligned} &c_1(X) = m \qquad 1.12 \\ &c_n(X) = C\Gamma(n-\alpha)(\lambda_+^{\alpha-n} + (-1)^n \lambda_-^{\alpha-n}), \qquad \text{for } n = 2, 3, \ldots \end{aligned}$$

The parameter α is the shape parameter. Along with parameter α, the parameter λ_+ and λ_- control the rate of decay on the positive and negative tails, respectively. If $\lambda_+ > \lambda_-$ $(\lambda_+ < \lambda_-)$, then the distribution is skewed to the left (right), and if $\lambda_+ = \lambda_-$, then it is symmetric. The parameter C is the scale parameter and parameter m is the location parameter.

If we take a special parameter C defined by

$$C = \left(\Gamma(2-\alpha)(\lambda_+^{\alpha-2} + \lambda_-^{\alpha-2})\right)^{-1} \qquad 1.13$$

Then $X \sim \text{CTS}(\alpha, C, \lambda_+, \lambda_-, 1)$ has zero mean and unit variance. In this case, X is called the standard CTS distribution with parameter $(\alpha, \lambda_+, \lambda_-)$ and denoted by

$X \sim std\text{CTS}(\alpha, \lambda_+, \lambda_-)$.

A more general for of the characteristic function of the CTS distribution is given by

$$\phi_{GCTS}(u) = \exp(imu - iu\Gamma(1-\alpha)(C_+\lambda_+^{\alpha_+-1} - C_-\lambda_-^{\alpha_+-1}) \quad 1.14$$
$$+C_+\Gamma(-\alpha_+)\big((\lambda_+ - iu)^{\alpha_+} - \lambda_+^{\alpha_+}\big)$$
$$+C_-\Gamma(-\alpha_-)((\lambda_- - iu)^{\alpha_-} - \lambda_-^{\alpha_-})$$

where $\alpha_+, \alpha_- \in (0,1) \cup (1,2)$, $C_+, C_-, \lambda_+, \lambda_- > 0$, and $m \in \mathbb{R}$. This distribution is referred to as the generalized classical tempered stable (GTS) distribution and we denote it by $X \sim \text{GTS}(\alpha_+, \alpha_-, C_+, C_-, \lambda_+, \lambda_-, m)$.

The cumulants of X are

$$c_1(X) = m, \quad 1.15$$
$$c_n(X) = C_+\Gamma(n-\alpha_+)\lambda_+^{\alpha_+-n} + (-1)^n C_-\Gamma(n-\alpha_-)\lambda_-^{\alpha_--n}$$

for $n = 2, 3, \ldots$.

Other members of tempered stable distributions are introduced in literature (see, for example, Rachev, et al, 2011). In next section, we talk about the class of infinitely divisible distribution and the relationship among all the distributions we have already discussed**.**

1.1.5 Infinitely Divisible Distributions

Definition 5. (Infinitely divisible)

A probability distribution F on $\mathbb{R}^d$ is said to be infinitely divisible if for any integer $n \geq 2$, there exists n i.i.d. random variables $Y_1, \ldots, Y_n$ such that $Y_1 + \cdots + Y_n$ has the distribution F.

The most common examples of infinitely divisible distributions are: the Gaussian distribution, the gamma distribution, the Poisson distribution, α-stable distribution and the tempered stable distribution. A random variable have any of these distributions can be decomposed into a sum of n i.i.d. parts having the same distribution with modified parameters. For our study, if $X \sim N(\mu, \sigma^2)$, then X can be rewritten as $X = \sum_{i=1}^{n} Y_i$, where Y_i are i.i.d random variable following $N\left(\frac{\mu}{n}, \frac{\sigma^2}{n}\right)$. The relations of X and Y_i for α-stable and CTS distributions are: $X \sim S_\alpha(C, \beta, m)$, $Y_i \sim S_\alpha\left(\frac{C}{n}, \beta, \frac{m}{n}\right)$; and $X \sim \text{CTS}(\alpha, C, \lambda_+, \lambda_-, m)$, $Y_i \sim \text{CTS}\left(\alpha, \frac{C}{n}, \lambda_+, \lambda_-, \frac{m}{n}\right)$ (for other distributions, see, for example, Rachev, et al. 2011).

In the literature, the characteristic function of one-dimensional infinitely divisible distribution is generated by the Lévy-Khinchin formula:

$$\exp\left(i\gamma u - \frac{1}{2}\sigma^2 u^2 + \int_{-\infty}^{\infty}\left(e^{iux} - 1 - iux1_{|x|\le 1}\right)\nu(dx)\right) \qquad 1.16$$

In the formula, the measure ν is referred to as Lévy measure. The measure is a Borel measure satisfying the condition that $\nu(0) = 0$ and $\int_{\mathbb{R}}(1 \wedge |x^2|)\nu(dx) < \infty$. The parameter γ and σ are real numbers and γ is referred to as the center or drift and determines the location. This triplet (σ^2, ν, γ) is uniquely defined for each infinitely divisible distribution, and called a Lévy triplet.

If $\nu(dx) = 0$, the infinitely divisible distribution becomes the Gaussian distribution with mean γ and variance σ^2, which is easy to check from the characteristic function of the Gaussian distribution.

For α-stable distribution, the Lévy measure is given by

$$\nu_{\mathrm{stable}}(dx) = \left(\frac{C_+}{x^{1+\alpha}}1_{x>0} + \frac{C_-}{|x|^{1+\alpha}}1_{x<0}\right)dx \qquad 1.17$$

Using the Lévy-Khinchine formula we can obtain the characteristic function in **(1.3)**. The Lévy measure of the tempered stable distributions can be obtained by multiplying tempering function to the Lévy measure of α-stable distribution. For instance, if we take $q(x) = e^{-\lambda_+ x}\mathbf{1}_{x>0} + e^{-\lambda_-|x|}\mathbf{1}_{x<0}$ as the tempering function, then we obtain the Lévy measure of the CTS distribution as

$$\nu_{\mathrm{CTS}}(dx) = \left(\frac{C_+ e^{-\lambda_+ x}}{x^{1+\alpha}}1_{x>0} + \frac{C_- e^{-\lambda_-|x|}}{|x|^{1+\alpha}}1_{x<0}\right)dx \qquad 1.18$$

For this reason, they are referred to as the tempered stable distribution. For more references of the multivariate cases of α-stable distributions and TS distributions, see Rachev and Mittnik (2000), and Kim, et al. (2012).

1.2 Multivariate Distributions

1.2.1 Introduction

Financial models are inherently multivariate. The value change of a portfolio of assets over a fixed time horizon depends on a random vector of risk-factor changes or returns. In this section, we consider some models for random vectors that are particular useful for financial data. We do this from a static, distributional point of view without considering about time series aspects, which will be discussed in Section 1.3.

A stochastic model for a random vector can be thought of as simultaneously providing probabilistic descriptions of the behavior of the components in the random vector and of their dependence structure. In Section 1.2.2 we review the widely used multivariate normal distribution and then discuss a generalization of it known as a multivariate normal mixture distribution, which shares much of the structure of the multivariate normal distribution and retains many of its properties. We treat both variance mixture, which belongs to the wider class of elliptical distributions, and mean-variance mixture, which allows asymmetry. Concrete examples include student-t distributions and generalized hyperbolic distributions. In Section 1.2.3, we look more closely at the issue of dependence structure among a random vector of financial risk factors using the concept of copula. Then we provide an approach to create the so-called meta distribution, which is a multivariate distribution defined by a particular copula with some marginal distributions.

1.2.2 Multivariate Mixture Distributions

Definition 6. (Multivariate normal distribution)

A random vector $\boldsymbol{X} = (X_1, \dots, X_d)'$ is said to follow a multivariate normal or Gaussian distribution if

$$\boldsymbol{X} = \boldsymbol{\mu} + A\boldsymbol{Z} \qquad 1.19$$

where $Z = (Z_1, \dots, Z_k)'$ is a vector of i.i.d. univariate standard normal random variables, and $A \in \mathbb{R}^{d \times k}$ and $\boldsymbol{\mu} \in \mathbb{R}^d$ are a matrix and a vector of constants, respectively.

It is easy to verify that $\mathrm{E}[\boldsymbol{X}] = \boldsymbol{\mu}$ and $\mathrm{cov}[\boldsymbol{X}] = \Sigma = AA'$. We only focus on the non-singular cases where $\mathrm{rank}(A) = \mathrm{d} \leq \mathrm{k}$. In these cases, the covariance matrix $\boldsymbol{\Sigma}$ has full rank and is positive-definite. We denote the multivariate normal distribution by $\boldsymbol{X} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and the standard multivariate normal distribution by $\boldsymbol{X} \sim N_d(\boldsymbol{0}, \boldsymbol{I}_d)$ if the components of $\boldsymbol{X}$ are mutually independent and follow the standard normal distributions.

The characteristics function of $\boldsymbol{X}$ can be derived from that of the univariate standard normal r.v.s $\boldsymbol{Z}$. Here we provide the characteristics function and also the continuous distribution function of $\boldsymbol{X}$ as following:

$$\phi_X(\boldsymbol{u}) = \exp\left(i\boldsymbol{u}'\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{\mu}'\boldsymbol{\Sigma}\boldsymbol{\mu}\right), \boldsymbol{u} \in \mathbb{R}^d \qquad 1.20$$

$$f(\boldsymbol{X}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}(\boldsymbol{x} - \boldsymbol{\mu})\right) \qquad 1.21$$

where $|\boldsymbol{\Sigma}|$denotes the determinant of Σ.

It can be seen from the density that the points lie on ellipsoids determined by the equation of the form $(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}(\boldsymbol{x}-\boldsymbol{\mu})=c$, for constant $c>0$. Moreover, a multivariate density $f(\boldsymbol{x})$ depended on $\boldsymbol{x}$ only through the quadratic form $(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}(\boldsymbol{x}-\boldsymbol{\mu})$ is the density of a so-called elliptical distribution.

An elegant character of multivariate normality is that $\boldsymbol{X}$ is multivariate normal if and only if $\boldsymbol{a}'\boldsymbol{X}$ is univariate normal for all $\boldsymbol{a}\in\mathbb{R}^d\backslash\{\boldsymbol{0}\}$.

A generalization of multivariate normal distribution is normal mixture distributions, which introduce randomness into the mean vector and covariance matrix of a multivariate normal distribution by introducing a positive random variable W.

First we introduce the norm variance mixtures which only adopt randomness into the covariance matrix.

- Normal variance mixtures

Definition 7. (Normal variance mixtures)

The random vector $\boldsymbol{X}$ is said to follow a multivariate normal variance mixture distribution if

$$\boldsymbol{X}=\boldsymbol{\mu}+\sqrt{W}A\boldsymbol{Z} \quad 1.22$$

where $\boldsymbol{Z}\sim N_d(\boldsymbol{0},\boldsymbol{I}_k)$; $W\geq 0$ is a non-negative, scalar-valued random variable which is independent of $\boldsymbol{Z}$; and $A\in\mathbb{R}^{d\times k}$ and $\boldsymbol{\mu}\in\mathbb{R}^d$ are a matrix and a vector of constants, respectively.

Conditioning on the random variable W, we observe that $\boldsymbol{X}|W=w\sim N_d(\boldsymbol{\mu},w\boldsymbol{\Sigma})$, where $\Sigma=AA'$. Provided that W has a finite expectation, we may easily calculate that

$$\begin{aligned}\mathrm{E}[W]&=\mathrm{E}\left[\boldsymbol{\mu}+\sqrt{W}A\boldsymbol{Z}\right]\\&=\boldsymbol{\mu}+\mathrm{E}\left[\sqrt{W}\right]A\mathrm{E}[\boldsymbol{Z}]=\boldsymbol{\mu}\end{aligned} \quad 1.23$$

$$\begin{aligned}\mathrm{cov}[\boldsymbol{X}]&=\mathrm{E}\left[\left(\sqrt{W}A\boldsymbol{Z}\right)\left(\sqrt{W}A\boldsymbol{Z}\right)'\right]\\&=\mathrm{E}[W]A\mathrm{E}[\boldsymbol{Z}\boldsymbol{Z}']A'=\mathrm{E}[W]\boldsymbol{\Sigma}\end{aligned} \quad 1.24$$

The characteristic function of $\boldsymbol{X}$ is given by

$$\begin{aligned}\phi_{\boldsymbol{X}}(u)&=\mathrm{E}\left[\exp\left(i\boldsymbol{u}'\boldsymbol{\mu}+\frac{1}{2}W\boldsymbol{\mu}'\boldsymbol{\Sigma}\boldsymbol{\mu}\right)\right]\\&=\exp(i\boldsymbol{u}'\boldsymbol{\mu})\widehat{H}\left(\frac{1}{2}\boldsymbol{\mu}'\boldsymbol{\Sigma}\boldsymbol{\mu}\right)\end{aligned} \quad 1.25$$

where $\widehat{H}(\theta)=\int_0^\infty e^{-\theta\nu}dH(\nu)$ is the Laplace – Stieltjes transform of the df H of W. Based on (1.25) we use the notation $\boldsymbol{X}\sim M_d(\boldsymbol{\mu},\boldsymbol{\Sigma},\widehat{H})$ for normal variance mistues.

If we take W in (1.22) to be a random variable with an inverse gamma distribution

$W \sim \text{Ig}\left(\frac{1}{2}\nu, \frac{1}{2}\nu\right)$, then $\boldsymbol{X}$ has a multivariate t distribution with ν degrees of freedom. Another example is the symmetric generalized hyperbolic distributions, which are obtained by taking W in (1.22) to have a generalized inverse Gaussian (GIG) distribution, $W \sim N^{-}(\lambda, \chi, \psi)$.

All of the multivariate distributions we have considered so far have elliptical symmetry, which implies that all one-dimensional marginal distributions are symmetric and thus contradict to the different tail behaviors of positive and negative returns from financial risk factor observations. The asymmetry of the normal mixture distributions can be obtained by mixing normal distributions with different means as well as different variances. We give the definition of the class of normal mean-variance mixture distributions as following:

- Normal mean-variance mixtures

Definition 8. (Normal mean-variance mixtures)

A random vector $\boldsymbol{X} = (X_1, \dots, X_d)'$ is said to follow a multivariate normal mean-variance mixture distribution if

$$\boldsymbol{X} = \boldsymbol{m}(W) + \sqrt{W} A \boldsymbol{Z} \qquad 1.26$$

where $\boldsymbol{Z} \sim N_d(\boldsymbol{0}, \boldsymbol{I}_k)$; $W \geq 0$ is a non-negative, scalar-valued random variable which is independent of $\boldsymbol{Z}$; and $A \in \mathbb{R}^{d \times k}$ and $\boldsymbol{\mu} \in \mathbb{R}^d$ are a matrix and a vector of constants, respectively; $\boldsymbol{m}: [0, \infty) \rightarrow \mathbb{R}^d$ is a measurable function.

One possible concrete specification for the function $\boldsymbol{m}(W)$ in (**1.26**) is:

$$\boldsymbol{m}(W) = \boldsymbol{\mu} + W \boldsymbol{\gamma} \qquad 1.27$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}$ are parameter vector in $\mathbb{R}^d$.

Since $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}'$, we have that:

$$\mathrm{E}[\boldsymbol{X}] = \mathrm{E}\left[\mathrm{E}[\boldsymbol{X}|W]\right] = \boldsymbol{\mu} + \mathrm{E}[W]\boldsymbol{\gamma} \qquad 1.28$$

$$\begin{aligned} \mathrm{cov}[\boldsymbol{X}] &= \mathrm{E}\left[\mathrm{cov}[\boldsymbol{X}|W]\right] + \mathrm{cov}\left[\mathrm{E}[B|W]\right] \qquad 1.29 \\ &= \mathrm{E}[W]\boldsymbol{\Sigma} + \mathrm{var}[W]\boldsymbol{\gamma}\boldsymbol{\gamma}' \end{aligned}$$

When the mixing variable W has finite variance, we observe from **(1.28)** and (1.29) that the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are not, in general, the mean vector and covariance matrix. The is only the case when $\boldsymbol{\gamma} = \boldsymbol{0}$, so that **(1.22)** is a simpler formula of (**1.27**).

A very flexible family of normal mean-variance mixtures is the so called generalized hyperbolic (GH) distributions, where $W \sim N^{-}(\lambda, \chi, \psi)$, a GIG distribution. An important property of GH distributions is their link to Lévy processes, i.e. processes with independent cases in continuous time. In fact, similar to the α-stable distributions and TS distributions discussed in Section 1.1.3 and 1.1.4, GH distributions are also infinitely divisible distribution, a property that it inherits from the GIG mixing distribution of W.

The characteristic function of GH distributions may be calculated as following:

$$\phi_{\boldsymbol{X}}(\boldsymbol{u}) = e^{i\boldsymbol{u}'\boldsymbol{\mu}}\widehat{H}(\frac{1}{2}\boldsymbol{u}'\boldsymbol{\Sigma}\boldsymbol{u} - i\boldsymbol{u}'\boldsymbol{\gamma}) \quad 1.30$$

where $\widehat{H}$ is the Laplace – Stieltjes transform of the GIG distribution.

We adopt the notation $\boldsymbol{X} \sim \mathrm{GH}_d(\lambda, \chi, \psi, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\gamma})$. Note that this notation is not unique. For instance, $\boldsymbol{X} \sim \mathrm{GH}_d(\lambda, \chi, \psi, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\gamma})$ and $\boldsymbol{X} \sim \mathrm{GH}_d(\lambda, \chi/k, k\psi, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\gamma})$ are identical.

GH distribution family is extremely flexible and contains many special cases known by other names. Here we give some examples of these special cases.

- If $\lambda = 1$, the multivariate distribution has univariate margins of one-dimensional hyperbolic distributions. The one-dimensional hyperbolic distributions have been widely studied in univariate analysis of financial returns;
- If $\lambda = \frac{1}{2}(d+1)$, the multivariate distribution is referred to as a d-dimensional hyperbolic distribution whose margins are not one-dimensional hyperbolic distributions. For more references of hyperbolic distributions, see Eberlein and Keller (1995), and Eberlein, et al. (1997).
- If $= -\frac{1}{2}$, the distribution is known as an normal inverse Gaussian (NIG) distribution. Its functional form is similar to hyperbolic with a slightly heavier tails. For more references about applying NIG model with financial return data, see Barndorff-Nielsen (1997).
- If $\lambda > 0$ and $\chi = 0$, we get a limiting cases of the distribution known variously as a generalized Laplace, Bessel function or variance-gamma distribution. For more references, see Madan, et al. (1998), and Kotz, et al. (2001).
- If $\lambda = -\frac{1}{2}\nu,\ \chi = \nu,$ and $\psi = 0$ we get another limiting cases suggested by McNeil, et al. (2005), known as asymmetric or skewed t distribution. For estimation and application of skewed t distribution, see Aas and Haff (2005), Hu and Kercgeval (2006), and McNei and Demarta (2007). For alternative skewed extensions of the multivariate t, see Kotz and Nadarjah (2004), and Genton (2004).

Univairate and multivariate *t* distributions have been widely used in modeling the financial return data, since they can model the heavy tails of the return distributions and dependence structures among them, for instance, through a *t* copula. Due to the simplicity of those symmetric distributions, they cannot capture the observed asymmetry of the multivariate financial return data. For instance, large losses occur more frequently and simultaneously than large returns for equities. Thus it is natural and necessary to consider a skewed *t* distribution model in some circumstances.

Following the notation in McNeil, et al. (2005), the joint probability density function of multivariate skewed *t* distribution is given as following:

$$f(\boldsymbol{X}) = c\frac{K_{\frac{(\nu+d)}{2}}\left(\sqrt{(\nu+Q(\boldsymbol{x}))(\boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma})}\right)e^{(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}}}{\left(\sqrt{(\nu+Q(\boldsymbol{x}))(\boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma})}\right)^{-\frac{(\nu+d)}{2}}\left(1+\left(\frac{Q(\boldsymbol{x})}{\nu}\right)\right)^{\frac{(\nu+d)}{2}}} \quad 1.31$$

where $Q(\boldsymbol{x}) = (\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ and the normalizing constant is

$$c = \frac{2^{1-(\nu+d)/2}}{\Gamma\left(\frac{1}{2}\boldsymbol{\mu}\right)(\pi\nu)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \quad 1.32$$

This density reduces to the standard multivariate *t* density as $\boldsymbol{\gamma} \to \boldsymbol{0}$. We denote the multivariate skewed *t* distribution as $\boldsymbol{X} \sim skewedt(\nu, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\gamma})$.

In this case, we obtain a special case of $W \sim \text{InverseGamma}(\nu/2, \nu/2)$ as an inverse gamma random variable. Thus it is easy to calculate the mean vector and covariance matrix of $\boldsymbol{X}$ as following:

$$\mathrm{E}[\boldsymbol{X}] = \boldsymbol{\mu} + \boldsymbol{\gamma}\frac{\nu}{\nu-2} \quad 1.33$$

$$\mathrm{cov}[\boldsymbol{X}] = \frac{\nu}{\nu-2}\boldsymbol{\Sigma} + \boldsymbol{\gamma}\boldsymbol{\gamma}'\frac{2\nu^2}{(\nu-2)^2(\nu-4)} \quad 1.34$$

The GH distributions can be fitted to data by the algorithms of the EM (expectation-maximization) type (see, Protassov, 2004; Barnforff-Nielsen and Shephard, 2005). Fitting the multivariate skewed *t* distributions with the EM-type algorithms with unknown degrees of freedom has been studied in many literatures, such as Meng and Rubin (1993), Liu and Rubin (1994), Liu (1997) and Meng and van Dyk (1997). Here we adopt the algorithms suggested by McNeil et al. (2005) and details given by Hu and Kercheval (2006).

1.2.3 Copula and Meta Distributions

A bottom-up approach to multivariate model building is particularly useful in such scenario that we have more information about the behavior of individual risk factors than their dependence structure. The copula approach, in general, provides a way of isolating the description of dependence structure and thus allows us to combine our more developed marginal model with a variety of possible dependence models and to investigate the sensitivity of risk to the dependence specification.

Copula helps in the understanding of dependence at a deeper level than the linear correlation, though the later one plays a central role in financial history. A common misunderstanding about linear correlation is that the marginal distributions and pairwise correlation determines its joint distribution. This statement is only true when our attentions are restricted to the elliptical distributions. A contradictory example is the normal mean-variance mixture given in (1.23) and (1.24) when $\boldsymbol{\gamma} \neq \mathbf{0}$. However, in later of this section, we will show that this statement is true in general if we replace correlation with copula.

A d-dimensional copula is defined as a distribution function on $[0,1]^d$ with standard uniform marginal distributions. A fundamental theory of copula is given by the Sklar (1995) theorem as following:

Theorem 1. (Sklar 1959)

Let F be a joint distribution function with margins $F_1, \dots, F_d$. Then there exists a copula : $[0,1]^d \to [0,1]$ such that, for all $x_1, \dots, x_d$ in $\overline{\mathbb{R}} = [-\infty, \infty]$,

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_1))) \qquad 1.35$$

Proof. For the proof, see McNeil et al. (2005).

If the margins are continuous, then C is unique; otherwise C is uniquely determined on Ran $F_1 \times$ Ran $F_2 \times \dots \times$ Ran F_d, when Ran $F_i = F_i(\overline{\mathbb{R}})$ denotes the range of F_i. Conversely, if C is a copula and $F_1, \dots, F_d$ are univariate distribution of functions, then the function F is a joint distribution function with margins $F_1, \dots, F_d$.

Moreover, if the random vector F has joint df F continuous marginal distributions $F_1, \dots, F_d$, then the copula of F (or X) is the df C of $F_1(X_1), \dots, F_d(X_d)$.

If we evaluate (1.35) at $x_u = F_i^{-1}(u_i), 0 \leq u_i \leq 1, i = 1, \dots, d$ with satisfying conditions of the generalized inverse, we obtain

$$C(u_1, \dots, u_d) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)) \qquad 1.36$$

which gives an explicit representation of C in terms of F and its margins.

For the converse statement assume that C is a copula and that $F_1, \dots, F_d$ are univariate dfs. By taking $\boldsymbol{U}$ to be a random vector with df C and setting $\boldsymbol{X} \coloneqq (F_1^{-1}(U_1), \dots, F_d^{-1}(U_d))$, we verify through **(1.31)** that

$$P(X_1 \le x_1, \dots, X_d \le x_d) \quad 1.37$$
$$= P(F_1^{-1}(U_1) \le x_1, \dots, F_d^{-1}(U_d) \le x_d)$$
$$= P\big(U_1 \le F_1(x_1), \dots, U_d \le F_d(x_d)\big)$$
$$= C(F_1(x_1), \dots, F_d(x_d))$$

The converse statement of Sklar's Theorem provides a very powerful technique for constructing multivariate distributions with arbitrary margins and copulas. Starting with a copula C and margins $F_1, \dots, F_d$, then $F(\boldsymbol{X}) \coloneqq C(F_1(x_1), \dots, F_d(x_d))$ defines multivariate df with margins $F_1, \dots, F_d$. Such multivariate distributions are known as meta distributions. For instance, if the distribution is constructed with a Gauss copula and arbitrary margins, then it is called meta-Gauss distribution. In Chapter 3, we use this technique to model multivariate factor returns with the distribution constructed with classical tempered stable distributions as margins and a skewed *t* copula.

A unique copula is contained in every multivariate distribution with continuous marginal distributions, and a useful class of parametric copulas are those contained in the normal mixture distributions in Section 1.2.2. Here we give an example of the skewed *t* copula.

Example 1. (Skewed *t* copula)

Consider $\boldsymbol{Y} \sim skewedt(\nu, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\gamma})$ is a skewed *t* random vector, then its copula is a so-called skewed *t* copula with the form as following:

$$C_{\nu,\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\gamma}}^{skewedt}(\boldsymbol{u}) \quad 1.38$$
$$= skewedt_{(\nu,\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\gamma})}(skewedt_{\nu,\mu_1,\sigma_1^2,\gamma_1}^{-1}(u_1), \dots, skewedt_{\nu,\mu_d,\sigma_d^2,\gamma_d}^{-1}(u_d))$$

where $skewedt_{\nu,\mu_i,\sigma_i^2,\gamma_i}^{-1}$ is the inverse cumulative probability function of one-dimensional skewed *t* distribution with parameters $(\nu, \mu_i, \sigma_i^2, \gamma_i)$. In general skewed *t* copula does not have a close form.

The estimation method may be the EM-type approach discussed in Section 1.2.2. Here we provide an algorithm for simulation of skewed *t* copulas. Consider generated N independent d-dimensional vectors from the multivariate skewed *t* distribution with obtained parameter estimates $(\hat{\nu}, \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}, \widehat{\boldsymbol{\gamma}})$. The generated scenarios form a $N \times n$ matrix $\boldsymbol{C} = \{C_{ij}\}$.

(1) Generate N independent d-dimensional vectors $\boldsymbol{Z}$ from multivariate normal distribution $N(\boldsymbol{0}, \widehat{\boldsymbol{\Sigma}})$.

(2) Generate N independent random variables W from the inverse gamma distribution $InverseGamma(\hat{\nu}/2, \hat{\nu}/2)$.

(3) Obtain N independent d-dimensional vectors $\boldsymbol{S} = \widehat{\boldsymbol{\mu}} + \widehat{\boldsymbol{\gamma}}W + \sqrt{W}\boldsymbol{Z}$.

(4) Transfer simulated $\boldsymbol{S}$ into uniform scenarios $\boldsymbol{C}$ either through the sample distribution function of the margins obtained from $\boldsymbol{S}$,

$$\hat{F}_k(x) = \frac{1}{N}\sum_{j=1}^{N} I\{X_{jk} \leq x\} \tag{1.39}$$

where $I\{\cdot\}$ stands for the indicator function; or, ideally, from the cdf of margins

$$F_k(x) = \int_{-\infty}^{x} f_k(t)dt \tag{1.40}$$

Where $f_k(\cdot)$ is the density of one-dimensional skewed *t* distribution with parameters $(\nu, \hat{\mu}_k, \hat{\sigma}_k^2, \hat{\gamma}_k)$. Then

$$C_{jk} = \hat{F}(S_{jk}) \tag{1.41}$$

or

$$C_{jk} = F(S_{jk})$$

While using the cdf of margins is more accurate, it is not analytically tractable.

1.3 Time Series Models

1.3.1 Introduction

This section provides a short summary of classical time series analysis with a focus on that is relevant to modeling risk-factor return series. More specifically, we introduce some examples of models used in this study, such as ARMA (autoregressive moving-average) processes, and GARCH (generalized autoregressive conditionally heteroscedastic) processes. There are several variations of the form of these time series models, here we follow the definitions given in McNeil et al. (2005).

In general, the processes we consider will be stationary in one or both of the following two senses:

Definition 8. (strict stationary)

The time series $(X_t)_{t\in\mathbb{Z}}$ is strictly stationary if

$$(X_{t_1}, \ldots, X_{t_n}) \overset{d}{=} (X_{t_1+k}, \ldots, X_{t_n+k}) \tag{1.42}$$

for all $t_1, \ldots, t_n$, $k \in \mathbb{Z}$ and for all $n \in \mathbb{Z}$.

Definition 9. (covariance stationary)

The time series $(X_t)_{t\in\mathbb{Z}}$ is covariance stationary (or weakly or second-order stationary) if the first two moments exist and satisfy

$$\mu(t) = \mu, \qquad t \in \mathbb{Z} \qquad 1.43$$
$$\gamma(t, s) = \gamma(t + k, s + k), \qquad t, s, k \in \mathbb{Z}$$

Definition 10. (white noise)

$(X_t)_{t\in\mathbb{Z}}$ is a white noise process if it is covariance stationary with autocorrelation function

$$\rho(h) = \begin{cases} 1, h = 0 \\ 0, h \neq 0 \end{cases} \qquad 1.44$$

A white noise process centered to have zero mean and variance $\sigma^2 = \text{var}(X_t)$ will be denoted as $\text{WN}(0, \sigma^2)$.

Definition 11. (strict white noise)

$(X_t)_{t\in\mathbb{Z}}$ is a strict white noise process if it is a series of iid, finite-variance random variables.

1.3.2 ARMA Processes

The family of classical ARMA process are widely used in financial applications of time series analysis. They are covariance-stationary processes that are constructed using white noise as a basic building block. A general notation in this section and reminder of this chapter we donate while noise by $(\varepsilon_t)_{t\in\mathbb{Z}}$ and strict white noise by $(Z_t)_{t\in\mathbb{Z}}$.

Definition 12. (ARMA process)

Let $(\varepsilon_t)_{t\in\mathbb{Z}}$ be $\text{WN}(0, \sigma^2)$. The process $(X_t)_{t\in\mathbb{Z}}$ is a zero-mean $\text{ARMA}(p, q)$ process if it is a covariance stationary process satisfying difference equations of the form

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \qquad 1.45$$

X_t is an ARMA process with mean μ if the centered series $(X_t - \mu)_{t\in\mathbb{Z}}$ is a zero-mean $\text{ARMA}(p, q)$ process.

Here we give an example of a non-zero mean $\text{ARMA}(1,1)$ process. The form of the process is given by:

$$X_t - \mu - \phi(X_{t-1} - \mu) = \varepsilon_t + \theta \varepsilon_{t-1} \qquad 1.46$$

Suppose an ARMA(1,1) model of the form (1.42) has been fitted with the data and its parameters have been determined. The one-step ahead prediction for X_{t+1} is

$$\text{E}[X_{t+1} | \mathcal{F}_t] = \mu_{t+1} = \mu + \phi(X_t - \mu) + \theta \varepsilon_t \qquad 1.47$$

Since $\text{E}[\varepsilon_{t+1} | \mathcal{F}_t] = 0$. And in general we have a h-step prediction

$$\text{E}[X_{t+h} | \mathcal{F}_t] = \mu + \phi^h (X_t - \mu) + \phi^{h-1} \theta \varepsilon_t \qquad 1.48$$

1.3.3 GARCH Processes

ARCH and its modification GARCH processes are important for capturing the phenomena of volatility clustering observed in financial risk factor data.

Definition 13. (GARCH)

Let $(Z_t)_{t\in\mathbb{Z}}$ be SWN(0,1). The process $(X_t)_{t\in\mathbb{Z}}$ is a GARCH(p,q) process if it is strictly stationary and if it satisfies, for all $t\in\mathbb{Z}$ and some strictly positive-valued process $(\sigma_t)_{t\in\mathbb{Z}}$, the equations

$$X_t = \sigma_t Z_t \tag{1.49}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$$

where $\alpha_0 > 0, \alpha_i \geq 0, i = 1, \dots, p$, and $\beta_j \geq 0, j = 1, \dots, q$.

In practice, low-order GARCH models are most widely used and we concentrate on the GARCH(1,1) model. A GARCH(1,1) process is said to be a covariance stationary process if and only if $\alpha_1 + \beta_1 < 1$.

Since we have seen that the ARMA processes are driven by a white noise $(\varepsilon_t)_{t\in\mathbb{Z}}$ and a covariance stationary process is a white noise. It is nature to extend the basic GARCH model related to an ARMA model by setting the ARMA error ε_t equal to $\sigma_t Z_t$. Hence, for instance, a ARMA(1,1)-GARCH(1,1) model has the form as following:

$$X_t = \mu + \phi_1(X_{t-1} - \mu) + \theta_1 \sigma_{t-1} Z_{t-1} + \sigma_t Z_t \tag{1.50}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1} Z_{t-1} + \beta_1 \sigma_{t-1}^2$$

where $\alpha_0 > 0, \alpha_1 \geq 0, \beta_1 \geq 0$ and $\alpha_1 + \beta_1 < 1$.

And predictions for this model can be easily calculated as

$$\mathrm{E}[X_{t+h} | \mathcal{F}_t] = \mu + \phi^h (X_t - \mu) + \phi^{h-1} \theta \varepsilon_t \tag{1.51}$$

$$\mathrm{var}[X_{t+h} | \mathcal{F}_t] = \alpha_0 \sum_{i=0}^{h-1} (\alpha_1 + \beta_1)^i + (\alpha_1 + \beta_1)^{h-1} (\alpha_1 \varepsilon_t^2 + \beta \sigma_t^2) \tag{1.52}$$

1.4 Some Statistical Tests

1.4.1 Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov (K-S) test is a goodness-of-fit test used to decide if a sample comes from a hypothesized continuous distribution. Assume that a random sample $x_1, \dots, x_N$ comes from some continuous distribution with cumulative distribution function (cdf) $F(x)$. The empirical cdf is denoted by $\hat{F}(x)$.

- K-S test

H_0: The data follow the specified distribution.

H_1: The data do not follow the specified distribution.

The Kolmogorov-Smirnov statistic for a given cdf $F(x)$ is

$$D = \sup_x \left| \hat{F}(x) - F(x) \right| \quad 1.53$$

where $\sup_x|\cdot|$ is the supremum of the set of distances.

By the Glivenko-Cantelli theorem, if the sample comes from distribution $F(x)$, then D converges to 0 almost surely. The hypothesis regarding the distribution form is rejected if the test statistic, D, is greater than the critical value. There are several variations of the critical value tables in literature. We do not present the critical value table since many software programs that perform a K-S test will provide the relevant critical values.

1.4.2 Anderson-Darling Test

The Anderson-Darling (A-D) test is a goodness-of-fit test used if a sample of data came from a population with a specific distribution (see, Anderson, T.W. and Darling, D.A., 1952). It is a modification of the K-S test and gives more weight to the tails than does the K-S test. The K-S test is distribution free in the sense that the critical values do not dependent on the specific distribution tested. The A-D test males use of the specific distribution in calculating critical values. This has the advantage of allowing a more sensitive test and the disadvantage that critical values must be calculated each time.

- A-D test

H_0: The data follow the specified distribution.

H_1: The data do not follow the specified distribution.

The A-D test statistic for a given c.d.f.$F(x)$ is

$$A^2 = -N - S \quad 1.54$$

$$S = \sum_{i=1}^{n} \frac{2i-1}{N} [\ln F(Y_i) + \ln(1 - F(Y_{N+1-i}))]$$

where Y_i are the ordered data.

1.4.3 Christoffersen Likelihood Ratio Test

The Christoffersen Likelihood Ratio (CLR) test is a backtesting approach to test the accuracy of a putative VaR model. Specifically, consider a time series of daily ex post portfolio returns, R_t, and a corresponding time series of ex ante VaR forecast, $\mathrm{VaR}_\alpha(t)$ with promised coverage rate α, such that ideally $P_{t-1}\big(R_t < -\mathrm{VaR}_\alpha(t)\big) = \alpha$.

Define the hit sequence of $\mathrm{VaR}_\alpha(R_t)$ violations as following:

$$I_t(\alpha) = \begin{cases} 1 \text{ if } R_t \leq \mathrm{VaR}_\alpha(t) \\ 0 \text{ if } R_t > \mathrm{VaR}_\alpha(t) \end{cases} \quad 1.55$$

Christfoffersen (1998) points out that the problem of determining the accuracy of a VaR model can be reduced to the problem of determining whether the hit sequence, $[I_t(\alpha)]_{t=1}^{t=T}$ satisfy two following properties:

1. Unconditional coverage property. The probability of number of violations for a fixed time period should equal the coverage rate. A higher or lower rate of violations would suggest the reported VaR measure systematically understates or overstates the portfolio's actual level of risk, respectively.

2. Independence property. This property places a strong restriction on the time dependence of violations. Specifically, any two elements of the hit sequence should be independent from each other.

This two property lead to the null hypothesis in CLR test that the elements in the hit sequence should be iid random variable following a Bernoulli distribution with parameter α, i.e. $I_t \sim iid\ Bernoulli(\alpha)$

The test of correct unconditional coverage (uc) is suggested by Kupiec (1995). Denoting $n_1 = \sum_{t=1}^{T} I_t$ as the number of violations in the time interval $[1, T]$, the hypotheses are given as following:

- Unconditional coverage test

H_0: $\alpha = \hat{\alpha} = \frac{n_1}{T}$.

H_1: $\alpha \neq \hat{\alpha}$.

And $\hat{\alpha} = \frac{\sum_{t=1}^{T} I_t}{T} = \frac{n_1}{T}$

According to Kupiec (1995) this test is best conducted as a likelihood ratio test. The test statistic takes the form:

$$\mathrm{LR}_{\mathrm{uc}} = -2\ln\left(\frac{(1-\alpha)^{T-n_1}p^{n_1}}{\left(1-\frac{n_1}{T}\right)^{T-n_1}\left(\frac{n_1}{T}\right)^{n_1}}\right) \quad 1.56$$

Under the null hypothesis $\mathrm{LR}_{\mathrm{uc}}$ is asymptotically χ^2 distributed with one degree of freedom. If the value of the $\mathrm{LR}_{\mathrm{uc}}$ exceeds the critical value of the χ^2 distribution, the null hypothesis will be rejected and the model is deemed as inaccurate.

The test of independence is defined where the hit sequence follows a first-order Markov sequence with switching probability matrix

$$\Pi = \begin{bmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{bmatrix} \quad 1.57$$

Where π_{ij} is the probability of an state i on time $t-1$ being followed by a state j on time t, and state 0 stands for no violation occurred and state 1 stands for a violation occured. The test of independence (ind) is then:

- Independence test

H_0: $\pi_{01} = \pi_{11}$.

H_1: $\pi_{01} \neq \pi_{11}$.

In implementation, denote n_{ij} as the number of observation with a state j following a state i and the calculate the ML estimates as

$$\hat{\pi}_{01} = \frac{n_{01}}{n_0} \quad 1.58$$

$$\hat{\pi}_{11} = \frac{n_{11}}{n_1}$$

And

$$\mathrm{LR_{ind}} = -2\ln\left(\frac{\left(1-\frac{n_1}{T}\right)^{n_0}\left(\frac{n_1}{T}\right)^{n_1}}{(1-\pi_{01})^{n_{00}}\pi_{01}^{n_{01}}(1-\pi_{11})^{n_{10}}\pi_{11}^{n_{11}}}\right) \quad 1.59$$

Under the null hypothesis, $\mathrm{LR_{ind}}$ is asymptotically χ^2 distributed with one degree of freedom By combing the independence test with the unconditional coverage test we obtain a joint test of conditional coverage (cc) with

$$\mathrm{LR_{cc}} = \mathrm{LR_{uc}} + \mathrm{LR_{ind}} \quad 1.60$$

And the test statistic $\mathrm{LR_{cc}}$ is asymptotically χ^2 distributed with two degree of freedom.

1.4.4 Berkowitz Likelihood Ratio Test

Unlike CLR test focusing on violations of VaR model, Berkowitz (2000) suggest density evaluation methods that make use of the full distribution of outcomes and thus extract a greater amount of information from the available data.

As suggested in Rosenblatt (1952), it is possible to transfor all realizations into a series of iid random variable via the transformation as following

$$x_t = \int_{-\infty}^{y_t} \hat{f}(u)du = \hat{F}(y_t) \quad 1.61$$

where y_t is the ex post portfolio profit or loss realization and $\hat{f}(\cdot)$ is the ex ante forecasted loss density.

A wide variety of tests would be then available for independence and for uniformity. One approach is to transform the uniformity into normality and then calculate the Gaussian likelihood and construct LR tests. Such a transformation is given by

$$z_t = \Phi^{-1}(x_t) = \Phi^{-1}\left(\int_{-\infty}^{y_t} \hat{f}(u)du\right) \quad 1.62$$

And thus the null hypothesis is that z_t are iid $N(0,1)$ random variable.

Loosely speaking, when backtesting a VaR model, we focus of the shape of the left tail of the loss distribution rather than the shape of the entire distribution. In particular, a LR tests is constructed on a censored observed tail and any observation which do not fall in the tail will be truncated.

Consider a VaR_α value calculated from a standard normal distribution, for instance, $\mathrm{VaR}_{95\%}$ is 1.64. Then the new variable z_t^* is defined as following:

$$\mathrm{z}_\mathrm{t}^* = \begin{cases} -\mathrm{VaR}_\alpha \text{ if } z_t \geq -\mathrm{VaR}_\alpha \\ z_t \text{ if } z_t < -\mathrm{VaR}_\alpha \end{cases} \quad 1.63$$

The log likelihood function for joint estimation of μ and σ of the normal distribution is

$$\mathrm{L}(\mu, \sigma | z^*) = \sum_{z_t^* < -\mathrm{VaR}_\alpha} \ln \frac{1}{\sigma} \Phi(\frac{z_t^* - \mu}{\sigma}) + \sum_{z_t^* = -\mathrm{VaR}_\alpha} \ln(1 - \Phi\left(\frac{-\mathrm{VaR}_\alpha - \mu}{\sigma}\right)) \quad 1.64$$

And the LR test statistics is given as

$$\mathrm{LR}_\mathrm{tail} = -2(\mathrm{L}(0,1) - \mathrm{L}(\hat{\mu}, \hat{\sigma}^2)) \quad 1.65$$

Under the null hypothesis, the test statistics is χ^2 distributed with two degree of freedom.

The test of independence can be construted with a AR(1) processes with untruncated z_t:

$$z_t - \mu = \rho(z_{t-1} - \mu) + \varepsilon_t \quad 1.66$$

And the independent test statistics is calculated as following:

$$\mathrm{LR}_\mathrm{ind} = -2(\mathrm{L}(\hat{\mu}, \hat{\sigma}^2, 0) - \mathrm{L}(\hat{\mu}, \hat{\sigma}^2, \hat{\rho})) \quad 1.67$$

where $\mathrm{L}(\mu, \sigma^2, \rho)$ the well-known likelihood function associated with (1.66).

To construct a test of AVaR or ETL, $\mathrm{E}[z_t | z_t < -\mathrm{VaR}_\alpha]$, one might instead devise a one-sided test of the null. For instance, consider H_0: $z_t \sim N(\mu, \sigma^2), \mu < 0, \sigma^2 > 1$. If either $\mu < 0$ or $\sigma^2 > 1$, the density places greater probability mass in the tail region that does an iid $N(0,1)$.

2.An Empirical Examination of Return Distributions for Chinese Stocks

2.1 Introduction

Since the establishment of the Shanghai Stock Exchange in December 1990 and the Shenzhen Stock Exchange in April 1991 in China, some foreign investors have viewed these exchanges as being inefficient and conservative in the sense that there was a restriction on foreign investors in the Chinese A-shares prior to 2003.[1] Mookerjee and Yu (1999) argue that this view is attributable to three unfavorable characteristics of these exchanges during the decade when they were introduced: a restricted supply of stocks, abrupt regulatory policy changes resulting in excessive volatility, and limited regulatory enforcement resulting in a scarcity of information needed by investors to evaluate companies. However, more recent empirical evidence supports improvements in market efficiency (e.g., Li, 2003a, 2003b; Fifield and Jetty, 2008).

In this chapter, we document the behavior of stock returns and risk in the Chinese stock market. We start by examining whether the identically and independent distributed hypothesis of stock returns is valid for the Chinese stock market. For the U.S. stock market, Rachev et al. (2005b) provide empirical evidence that fails to support this hypothesis. Rachev et al. (2007) found the same for returns for the Taiwan stock market. For the Chinese stock market, Charles and Darn é(2009) and Hou (2013) report similar results. The development of the Chinese stock market since the turn of the century, especially the efforts by financial regulators in China to increase market efficiency due to China's accession to the World Trade Organization in 2001, suggest that a re-examination of stock return behavior is warranted.

The autocorrelation of return and clustering of volatility behavior that have been observed in developed equity markets are also present in the Chinese stock market. In order to improve forecasting for the Chinese stock return process, an ARMA-GARCH model is used in this

---

[1] Foreign investor were only allowed to trade in B-share market, where shares are issued by mainland Chinese companies but traded in foreign currencies with some different regulations from the A-share (main) market. Although the restriction of Chinese domestic investors investing in B shares was lifted on February 19, 2001 in order to increase the mobility and liquidity, the B-share market is still a marginal market compared to the A-share market. As of this writing, the government is experimenting with transferring B shares to H shares (shares traded in the Hong Kong stock market). Since December 1, 2012 a more convenient approach for foreign investors investing in the entire Chinese stock market, namely, Qualified Foreign Institutional Investors (QFII), was approved by the China Securities Regulation Commission (CSRC). See www.csrc.gov.

chapter. As we discussed in Chapter 1, Section 1.3, the autocorrelation of returns is explained by the autoregressive-moving average (ARMA) part and the volatility clustering is explained by the generalized autoregressive conditional heteroscedaticity (GARCH) part.

The Gaussian distribution assumption for the innovations of GARCH models is not supported by empirical evidence and therefore this assumption is inappropriate for describing the stock return process, especially in forecasting market crashes. Kim et al. (2011) have shown that alternative distribution-based models, such as α-stable distribution family and tempered stable distribution family, perform better than the Gaussian model. This is because the Gaussian model fails to capture the skewness and leptokurtosis observed in stock returns. Since the Gaussian distribution is a special case in the α-stable distribution family, and the latter distribution is also embodied in the more general tempered stable distribution family, in this chapter we use the enhanced ARMA-GARCH models with tempered stable distributed innovations to study the behavior of the Chinese stock returns.

As for risk measures, the standard deviation has long been criticized as a measure of risk, despite its key role in mean-variance analysis for optimal portfolio construction. An alternative measure that has been proposed is the value-at-risk (VaR) measure. However, although VaR has been adopted as the risk measure in the financial industry because of its acceptance by bank regulators, it has a number of well-known limitations as a risk measure (e.g., Bookstaber, 2009). For example, VaR is not a coherent risk measure because of its non-subadditivity property (Artzner et al., 1999), and it ignores the extreme losses exceeding VaR. An alternative risk measure that offers a more reliable assessment of risk and the one we use in this chapter is average value-at-risk (AVaR). Also called conditional value-at-risk (CVaR) (Rockafellar and Uryasev, 2000, 2002), AVaR is the average of losses larger than the VaR for a given tail probability. In contrast to VaR, AVaR satisfies all axioms of coherent risk measure (Rachev et al., 2008).[2] The closed-form solution for AVaR for the α-stable and the infinitely divisible distributions containing tempered stable distributions have been derived by Stoyanov et al. (2006) and Kim et al. (2011).

In this chapter, we examine the return and volatility observed in the A-share sector of the Chinese stock market. We do not explore the B-share market or possible interactions between the A- and B-share sectors. As of June 29, 2012, the number and total capitalization of B-share stocks are, respectively, 5.1% and 0.72 % of those of A-share stocks. We find that the A-share

[2] If the distribution is continuous, AVaR is equivalent to expected tail loss (ETL).

returns are better characterized as a non-Gaussian distribution, such as α-stable and tempered stable distribution families. In addition, we look at the forecasting performance of ARMA-GARCH models with different distributed innovations during periods of high volatility. We find that the tail-risk spread, which we define as the daily difference between the values of AVaR for the Gaussian and non-Gaussian ARMA-GARCH models, may offer an early warning signal for a forthcoming market crash. Moreover, we find that for the time period studied, January 2010 to June 2012, the Chinese stock market became more sensitive to tailrisk than it was in 2005 and 2006. Although other studies have investigated the use of tempered stable distribution family for several markets, we believe that this is the first to apply this distribution family to the Chinese stock market.

The reminder of Chapter 2 is organized as following. In Section 2.2, we present the basic pattern of returns and volatility in the Chinese stock market. In Section 2.3, we provide tests for the distribution of stock returns under the white noise model and the ARMA-GARCH model, along with an explanation of the methodology and a description of our data. The ARMA-GARCH model with different distributed innovations is described in Section 2.4, along with a comparison of their forecasting performance. The backtesting of the values for VaR and AVaR generated by different ARMA-GARCH models are also presented and compared in that section. We summarize our principal findings regarding the behavior of Chinese stock returns in Section 2.5.

2.2. Return and Volatility for the Chinese Stock Market

A-share stocks are traded on the Shanghai and Shengzhen exchanges. In this chapter, we investigate the daily returns for the CSI 300 Index from January 8, 2002 to June 29, 2012 (2,535 observations).[3] We analyze the returns and volatility of this index as well as those of three market indices from more developed stock markets — a world index representing 24

[3] The CSI 300 Index, a capitalization-weighted stock market index that is compiled by the China Securities Index Company, Ltd., is designed for use as a performance benchmark and serves as a reference index for derivatives and a benchmark for equity indexing. As of year-end 2012, there were 12 exchange-traded funds, 24 index funds, and one index futures contract tied to the CSI 300 Index in the Chinese financial market. The components of the CSI 300 Index belong to the top 300 A-share stocks by market capitalization and the most liquid 50% of all A-share stocks in both exchanges. The index has been calculated from January 8, 2002 with a base level of 1000 on November 31, 2004.

.

developed markets (the Morgan Stanley Capital International World Index), the U.S. stock market (the S&P Index), and the Hong Kong stock market (the Hang Seng Index). All data were obtained from Bloomberg L.P. Before using this data source, we checked the index provider's website and other Chinese data provider's open data sources. We found that no other source provides a longer time period for the index data.

Some sample distribution statistics for the four stock market indices are presented in Table 2.1. Statistics include daily return sample mean and standard deviation, coefficients of skewness and kurtosis, and Ljung-Box Q test statistic. The Ljung-Box Q (10) statistic is used to test the significance of serial autocorrelation at lag 10.

There are five findings based on these summary statistics. First, the mean and standard deviation of returns for the CSI 300 Index is higher during the study period than those of the other three indices. Since the Chinese stock market is viewed as riskier than the other three markets, this finding is consistent with the notion of greater potential return in exchange for accepting greater risk. Second, minimum return for the Chinese index is less than that for the MSCI world index and the U.S. index, while its maximum return is less than all of the others. Large market crashes can still be observed in the Chinese stock market, but not as significant as it was in the 1990s (Wang et al., 2004). Third, the skewness coefficient for the Chinese stock market is negative and statistically significant, indicating asymmetry of the return distribution. Fourth, although the kurtosis coefficient is lower for the Chinese stock market than those for the other stock markets, the return distribution can still be characterized as exhibiting "fat tails". Finally, the Ljung-Box Q(10) statistics are highly significant for all four stock indices, indicating that daily returns are positively autocorrelated.

Figure 2.1 presents a more intuitive view of the return distributions for each stock market by comparing the normalized sample distributions against the standard Gaussian distribution. The sample return distributions deviate from the Gaussian distribution, especially the tails part. Moreover, the sample return distribution for the Chinese market implies a relatively thinner tail than those for the other stock markets, which, of course, is consistent with the sample kurtoses reported in Table 1. One possible reason for this finding is suggested by the study by Friedman and Sanddorf-Köhle (2002) who in comparing volatility before and after the imposition by Chinese regulators of daily price change limits in 1996, found a reduction in volatility.

2.3. Examination of the Distribution Hypotheses for Chinese Stock Returns

2.3.1 Methodology

In this section, we employ two models to test whether the return distribution of Chinese stocks follow a Gaussian distribution. In addition, we select and test three non-Gaussian distributions as alternatives. The Chinese stocks in our study are the 300 component stocks in the CSI 300 Index.[4]

In the first model, we assume that the daily return observations are independent and identically distributed (iid), thus they are assumed to follow a white noise (WN) process. In the second model, the daily stock returns are assumed to follow an ARMA(1,1)-GARCH(1,1) process. The WN model concerns itself with the unconditional, homoscedastic distribution model while the ARMA-GRACH model belongs to the class of conditional heteroscedastic distribution models (Rachev et al., 2005a). For both models, we consider four distribution assumptions: the Gaussian distribution, the α-stable distribution,[5] the classical tempered stable (CTS) distribution,[6] and the normal tempered stable (NTS) distribution.[7] The last three distributions are referred to as non-Gaussian distributions in this chapter.

Let $(S_t)_{t\geq 0}$ be the asset price process and $(y_t)_{t\geq 0}$ be the return process of $(S_t)_{t\geq 0}$ defined by $y_t = \log\frac{S_t}{S_{t-1}}$. We propose the following ARMA(1,1)-GARCH(1,1) model:

$$\begin{cases} y_t = ay_{t-1} + b\sigma_{t-1}\varepsilon_{t-1} + \sigma_t\varepsilon_t + c, \\ \sigma_t^2 = \alpha_0 + \alpha_1\sigma_{t-1}^2\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2, \end{cases}$$

where $\varepsilon_0 = 0$, and a sequence $(\varepsilon_t)_{t\in N}$ of iid real random variables. If we set a, b, α_1 and β_1 equal to zero, then the model becomes a WN process model, which assumes that stock returns are iid random variables. Without the preset of the parameters, we will have the ARMA(1,1)-GARCH(1,1) time-series model. The innovations ε_t are assumed to follow different distributions in this chapter. For convenience, we refer to these models with respect to

[4] We select all the 300 companies that are introduced in the CSI 300 Index at June 29, 2012, and select the full time series from January 8, 2002 to June 29, 2012, with 2,535 observations. The missing data are backfilled with the CSI300 index and MSCI Chinese sector indices, as the CSI300 sector indices do not have long enough time series. The backfill process is a linear regression model.

[5] Extensive analysis of α-stable distributions and their properties can be found in Samorodnitsky and Taqqu (1994), Rachev and Mitnik (2000).

[6] The CTS distribution has been introduced under different names including: truncated Lévy flight by Koponen (1995), the KoBoL distribution by Boyarchenko and Levendorskiĭ (2000), and the CGMY distribution by Carr et al. (2002).

[7] The NTS distribution was originally obtained using a time-changed Brownian motion with a tempered stable subordinator by Barndorff-Nielsen and Levendorskiĭ (2001). Later, Kim (2008) defined the NTS distribution by the exponential tilting for symmetric MTS distribution.

their underlying distributions as following:

WN models: normal-WN, stable-WN, CTS-WN, and NTS-WN models.

ARMA(1,1)-GARCH(1.1) models: normal-ARMA-GARCH, stable-ARMA-GARCH model, CTS-ARMA-GARCH and NTS-ARMA-GARCH models.

The parameters of the normal-ARMA-GARCH model are estimated using the maximum likelihood estimation (MLE) method. For the non-Gaussian ARMA-GARCH models, the parameters of the ARMA-GARCH part are estimated using a methodology described in Bianchi et al (2010). More specifically, the parameters of the ARMA-GARCH part are estimated using the quasi-maximum likelihood (QML) method assuming the innovations follow a Student's *t* distribution. Then we extract residuals using the estimated parameters, and fit the parameters of the innovation distribution to the extracted residuals using MLE.

For the assessment of the goodness-of-fit, we use the Kolmogorov-Smirnov (KS) test and the Anderson-Darling (AD) test. We use the later statistic in order to obtain a better test to evaluate the tail fit.

2.3.2 Test Results

Table 2.2 provides the test results using the standard KS test with different significance levels. The Gaussian hypothesis is rejected in almost all of the 300 stocks for both the WN model and ARMA-GARCH model at a significance level of 5%. Even at an extremely low significance level of 0.1%, the Gaussian hypothesis is rejected for 94% of the stocks for the WN model and 71% for the ARMA-GARCH model. For the α-stable distribution hypothesis, only around 25% of stock cases is rejected at the 0.1% significance level, and about half of all the stock cases are rejected at the significance level of 5% for both models tested. Moreover, there are much fewer cases of stocks being rejected under the two tempered stable distribution hypotheses for both models. The NTS distribution hypothesis is rejected for less than 10% of the stocks for the ARMA-GARCH model; the CTS distribution performs best among all the four tested distributions, with only 1% of the stocks being rejected for the ARMA-GARCH model.

We believe that there are three fair conclusions that can be reached based on these results. First, returns in the Chinese stock market do not follow the Gaussian distribution. This holds regardless of whether the WN model or the ARMA-GARCH model is used. Second, the ARMA-GARCH model can better capture the characteristics of the stock return process than the WN model, which fails to account for the observed autocorrelations and volatility clustering phenomenon. Finally, the three non-Gaussian distributions perform better than the Gaussian distribution in modeling Chinese stock returns, and the CTS-ARMA-GARCH model is believed to be the best among all tested models.

Figure 2.2 provides more evidence against the Gaussian distribution hypothesis. In the figure, we plot KS statistic and AD statistic for all stocks under the Gaussian distribution hypothesis and the CTS distribution hypothesis for the WN models and ARMA-GARCH models. The four panels in Figure 1 show that for most stocks the KS statistics for the CTS distribution models are less than that for the Gaussian distribution models. This phenomenon is even more obvious in comparing the AD statistic of the two distribution models. The KS statistic implies that for our sample stocks there is a better fit of the CTS model around the center of the distribution while the AD statistic implies a better fit in the tails. The large difference between the AD statistics computed for the CTS model relative to the Gaussian model strongly suggests a much better ability to forecast extreme events using the CTS model.

Summary statistics of the various statistical tests for the entire sample are provided in Table 3.3. Again, the results follow our previous analysis that the Gaussian-based model is the worst among the four models in describing the stock return process, and the other three models capture the asymmetry and heavy-tail characteristics. Generally, the KS statistic and AD statistic of the Gaussian models are about five times larger than those of the two tempered stable models. The CTS distribution performs better than the α-stable distribution and NTS distribution for both the WN and ARMA-GARCH models.

2.4. Backtesting of ARMA-GARCH Model

In this section, we first compare the ARMA-GARCH models with the Gaussian, α-stable, CTS, and NTS distributed innovations by analyzing their ability to identify a market crash. Then we perform a backtesting analysis of the four ARMA-GARCH models using two risk measures, VaR and AVaR. Here we use the daily time series of the CSI 300 Index from January 8, 2002 to June 29, 2012. In the analysis, we adopt the same parameter estimation methods and the goodness-of-fit tests as in Section 3. We use a moving time window of three years of daily data for the parameter estimation.

2.4.1 Forecasting Capability of the ARMA-GRACH Models

We exam the forecasting capability of the four ARMA-GARCH models for the largest 10 daily declines in the CSI 300 Index during the period 2002 to 2012. These bear market periods are listed in Table 2.4. For estimating parameters, we use three years of historical data until the closest trading day of each selected date.

In Table 2.5, we report the goodness-of-fit test statistics for the four ARMA–GARCH models, along with the forecasting probability of occurrence for these bear markets. The average time of occurrence of market crash is calculated by

$$\frac{1}{250 \cdot P(\varepsilon_t \leq \varepsilon_t^*)}$$

where ε_t^* is the observed residual at time *t*. We provide the observed residual on the first day of the bear market, and the probability and average time that the price decline would happen under each different model.

From Table 2.5, we see that the probabilities that the market would drop on the 10 days studied are smaller under the normal-ARMA-GARCH model than under the other three models. For example, on February 27, 2007, the CSI 300 Index dropped 9.24%, its worst daily performance during our study period, 2002 to 2012. Under the normal-ARMA-GARCH model, such a major market decline is expected to occur every 860 years, which provides no realistic warning. In contrast to the Gaussian model, the stable-ARMA-GARCH, CTS-ARMA-GARCH, and NTS-ARMA-GARCH models predict much shorter average times for a market decline of this type, which are 3.7 years, 17.6 years, and 15 years, respectively. Moreover, for four of the other nine days studied in this chapter, the average times for the corresponding market drops under the normal-ARMA-GARCH model are more than 10 times longer than those under the three non-Gaussian model. In real world markets, such major market declines should be expected to occur in much shorter time periods. Similar results are observed for the other five days studied, but the differences of the times for the Gaussian model and the non-Gaussian models are relatively smaller.

Among the three non-Gaussian models, the stable-ARMA-GARCH model gives the shortest average time of occurrence for the 10 days studied. However, the stable-ARMA-GARCH model is rejected based on the KS test at the 1% significance level for seven of the10 days studied, and the statistics from the AD test are much larger than those for the two tempered stable models for the seven days, suggesting that the α-stable model does not explain the tail property better than the tempered stable models. The CTS-ARMA-GARCH model and the NTS-ARMA-GARCH model are not rejected by the KS test at the 1% significance level on any of the 10 days studied. The average times of occurrence for the two tempered stable models are typically close to each other except for May 30, 2012. Thus, we conclude that the two tempered stable models provide better forecasting ability than the Gaussian and α-stable models, even though the normal-ARMA-GARCH model does consider the volatility clustering effect. Later we will see similar results in the backtesting of the four models during the highly volatile market period in Section 2.4.2.

2.4.2 VaR, AVaR and Backtesting

We define the VaR and AVaR with significance level η, given the information until time *t*, as following:

$$\mathrm{VaR}_{t,\eta}(y_{t+1}) = -\inf\{x \in \mathbb{R} | P_t(y_{t+1} \le x) > \eta\}$$

where $P_t(A)$ is the conditional probability of a given event A for the information until time t.

$$\mathrm{AVaR}_{t,\eta}(y_{t+1}) = \frac{1}{\eta}\int_0^{\eta} \mathrm{VaR}_{t,\epsilon}(y_{t+1}) d\epsilon \,,$$

For AVaR, if the distribution of y is continuous, then we have

$$\mathrm{AVaR}_{t,\eta}(y_{t+1}) = -E[y_{t+1} | y_{t+1} < -\mathrm{VaR}_{t,\eta}(y_{t+1})].$$

For evaluating the accuracy of forecasting VaR and AVaR for the four ARMA-GARCH models, we perform backtesting using the Christoffersen Likelihood Ratio (CLR) test (Christoffersen, 1998) and the Berkowitz Likelihood Ratio (BLR) test (Berkowitz, 2001) discussed Chapter 1., Section 1.4.. The time periods considered in backtesting are one-year observations from 2006 to 2011. The CLR test accounts for the conditional coverage of the VaR measure, assuming the violations of VaR for a given period follow an iid Bernoulli distributed random variables. For the CLR test, we first perform the unconditional coverage test and the independence test, which account for the cumulative probability and independence of violations, respectively, and then perform a joint CLR test for the conditional coverage hypothesis.

Table 2.6 provides the CLR test statistics and the corresponding *p*-values for the four models in the unconditional coverage tests and the joint tests. Only the normal-ARMA-GARCH model is rejected at the 1% significance level for 2007 and 2008; for the other years, none of the four models is rejected. For the independence test, none of the four models is rejected at the 1% significance level for all the years.

Instead of relying on the violations of VaR in the CLR test, the BLR test makes use of the full distributions of the outcomes and thus uses more information from the available data. As suggested by Berkowitz, we use the likelihood ratio test for the independence property and the accuracy of forecast of the tail distribution, by comparing the forecast AVaR to the realized AVaR.

From the BLR test results in Table 2.7, we find that none of the four models is rejected at the 1% significance level in the test of independence. In the test of tail distribution, the normal-ARMA-GARCH model is rejected at the 1% significance level for 2007 and 2008, while the other three models are not rejected at the same significance level.

Because the three non-Gaussian ARMA-GARCH models typically provide larger values for VaR than that for the normal-ARMA-GARCH model in the test period, we compute the average of the relative difference (ARD) between the normal-ARMA-GARCH model and the other three models as following:

$$\mathrm{ARD} = \mathrm{E}\left[\frac{\mathrm{VaR}_{t,0.01}^{\mathrm{non-normal}}(y_{t+1}) - \mathrm{VaR}_{t,0.01}^{\mathrm{normal}}(y_{t+1})}{\mathrm{VaR}_{t,0.01}^{\mathrm{normal}}(y_{t+1})}\right]$$

where $\mathrm{VaR}_{t,0.01}^{\mathrm{normal}}(y_{t+1})$ is the value of VaR for the normal-ARMA-GARCH model and the $\mathrm{VaR}_{t,0.01}^{\mathrm{non-normal}}(y_{t+1})$ is the value of for the other three models (Kim et al. 2011).

As VaR is related to the eligible capital requirement in risk control, the ARD metric is regarded as the difference in the capital requirement based on a non-Gaussian model calculated VaR and that based on a Gaussian model calculated VaR. Hence, for the non-Gaussian models, a smaller ARD is more economically efficient.

Table 2.8 presents the ARD values computed for the six time periods. From the results we conclude that for year 2006 and 2007, the CTS-ARMA-GARCH model has larger ARD values than those for the stable-ARMA-GRACH model, but they are not markedly different. For years after 2007, the CTS-ARMA-GARCH model has ARD values that are roughly 19% while the stable-ARMA-GARCH model has ARD values exceeding 25%, except for 2010 when the ARD value is 22%. The ARD values for the NTS-ARMA-GARCH model are always just slightly larger than those for the CTS-ARMA-GARCH model. Taken together, the results of the CLR test and BLR test, in which the normal model is rejected in 2007 and 2008 when the market was highly volatile, we conclude that the two tempered stable ARMA-GARCH models are better than the other two models not only because they are not rejected in backtesting, but also because they are relatively economically efficient. This conclusion follows from the test results that we observed in Section 3.

In Figure 2.3, we show the daily return series for the CSI 300 Index model from January 4, 2005 to November 11, 2013 and the corresponding daily values of AVaR forecasted using the normal-ARMA-GARCH model and CTS-ARMA-GARCH. For comparison, we also plot the time series of SPX in the same time period. We calculate the spread $\mathrm{d}_{\mathrm{CTS}\&Normal}$ as following:

$$\mathrm{d}_{\mathrm{CTS}\&Normal} = \mathrm{AVaR}_{t,0.01}^{\mathrm{CTS}} - \mathrm{AVaR}_{t,0.01}^{\mathrm{normal}}$$

where $\mathrm{AVaR}_{t,0.01}^{\mathrm{CTS}}$ is the value of 1% AVaR for the CTS-ARMA-GARCH model and $\mathrm{VaR}_{t,0.01}^{\mathrm{normal}}$ is the value for 1% AVaR for the normal-ARMA-GARCH model.

In Figure 2.4, we present the time series of $\mathrm{d}_{\mathrm{CTS}\&Normal}$ for the CSI 300 Index and the SPX index, from January 4, 2005 to November 11, 2013.

The daily spreads for the entire year of 2008, a year in which the CSI 300 Index lost 65% of its value, are typically larger than those in the years before and after 2008, which matches the highly volatile market period observed in Figure 2.3. Moreover, by regressing the daily spreads from 2005 to 2011 with respect to an arbitrary constant, we divide the sample results into three regimes which corresponding to the daily spreads from 2005 to 2006, 2007 to 2009, and 2010 to 2011. Each regime is tested against others using the Chow test under both homoscedasticity and heteroscedasticity assumptions. Also, we applied the Chow test within each regime by dividing the regime into shorter time periods according to the years. Panel a of Table 2.9 shows the results for the tests between each regime. The null hypotheses of no existing structure breaks are rejected based on a 1% level of significance, indicating that the daily spreads in 2007 significantly increased to a higher level than that in the years 2005 and 2006. Thus, not only the increasing value of AVaR, but also the increasing value of the spreads indicates the increasing risk of the Chinese stock market. Additionally, the daily spreads after 2010 decreased to a lower level than that from 2007 to 2009, but remained higher than that from 2005 to 2006, which indicates that the market are more sensitive to tail risk (extreme events) after the recent financial crisis. The results for the within each regime test shown in panel b of Table 2.9 indicate that the null hypotheses are generally not rejected at the 1% level of significance. These results once again indicate the shift of spread level in 2007 and 2010, and the Chinese stock market remained highly sensitive to tail risk in 2010 and 2011. The similar results can be found for the SPX index.

2.5 Conclusions

In this chapter, we use the CSI 300 Index and its component stocks to investigate the return process for the Chinese stock market. We reject the Gaussian distribution hypothesis under either the unconditional homoscedastic distribution assumption or the conditional heteroscedastic distribution assumption. Instead, we first introduce an empirical

ARMA-GARCH model with α-stable and tempered stable distributed innovations. We then provide an assessment of the model's forecasting power and compared it to the Gaussian distribution model. We find that the ARMA-GARCH model with CTS distributed innovations provides the best approach for modeling Chinese stock returns.

In addition, we analyze the accuracy of two risk measures for the different ARMA-GARCH models. Applying both the Christoffersen likelihood ratio and Berkowitz Likelihood ratio tests, we reject the Gaussian model during highly volatile market periods. In contrast, the three non-Gaussian models investigated are not rejected by these two tests. Our investigation of the relative difference between VaR values of the three non-Gaussian models and the Gaussian model indicates that from a practical risk control perspective, the capital requirement as suggested by the VaR calculated from the CTS and NTS models are lower than that from the α-stable model but higher than that from the Gaussian model.

Based on our statistical tests coupled with our backtesting results, we conclude of the four models studied in this chapter, the CTS-ARMA-GARCH and NTS-ARMA-GARCH models provide superior modeling capability and forecasting for Chinese stock returns. Finally, we find two structural breaks of the daily spreads between AVaR values for the normal-ARMA-GARCH and the CTS-ARMA-GARCH models. The daily spreads increased one year before the market crash of 2008, suggesting increasing market risk.

Table 2-1 Distributional characteristics of returns in the Chinese stock market and other world stock markets: Daily returns from 2002 to 2012.

	CSI 300	MXWO	S&P 500	HSI
Mean return	0.0254%	0.0074%	0.0061%	0.0198%
Std.dev	1.80%	1.14%	1.37%	1.60%
Sharpe ratio	1.41%	0.65%	0.44%	1.22%
Minimum	- 9.70%	- 7.33%	- 9.47%	- 13.58%
Maximum	8.97%	9.10%	10.96%	13.41%
Skewness	- 0.2386*	- 0.3403*	- 0.1857*	0.0589
Kurtosis	6.1667*	10.4593*	11.2674*	11.8310*
Ljung-Box Q(10)	20.1525*	51.3604*	54.4398*	25.0047*

[a] * Statistically significant at the 5% level.

CSI 300: Equity index that consists of 300 A-share stocks listed on the Shanghai or Shenzhen Stock Exchanges in China.
MXWO: The MSCI Word Index equity index including developed word markets, and not including emerging markets.
S&P 500: Standard and Poor's 500 Index, equity index that consists of 500 stocks listed on the NYSE or NASDAQ in U.S.
HSI: Hang Seng Index, equity index that consists of 50 stocks listed on the Stock Exchange of Hong Kong.

Table 2-2 Percentages of stocks for which the Gaussian, α-stable, CTS and NTS distribution hypotheses are rejected at different significance levels using Kolmogorov-Smirnov test.

Significance Level	5%	1%	0.5%	0.1%
WN model				
Normal	100.00%	98.33%	97.67%	94.00%
Stable	57.00%	39.33%	32.67%	23.67%
CTS	8.67%	4.67%	2.67%	2.67%
NTS	14.00%	6.67%	5.67%	2.67%
ARMA-GARCH model				
Normal	98.33%	93.00%	90.00%	71.00%
Stable	64.00%	43.00%	34.33%	25.67%
CTS	1.33%	0.33%	0.00%	0.00%
NTS	8.00%	5.00%	4.00%	3.33%

Table 2-3 Summary of the statistics of the various statistical tests of the WN models and ARMA-GARCH models for the entire sample

	KS Statistic				AD Statistic			
	Normal (p-Value)	Stable (p-Value)	CTS (p-Value)	NTS (p-Value)	AD Norm	AD Stable	AD CTS	AD NTS
WN								
Mean	0.0609 (0.0007)	0.0322 (0.1394)	0.0236 (0.4017)	0.0196 (0.4013)	0.0908	0.0384	0.0265	0.0234
Median	0.0555 (0.0000)	0.0289 (0.0328)	0.0185 (0.3665)	0.0181 (0.3693)	0.0819	0.0320	0.0203	0.0207
25% quantile	0.0484 (0.0000)	0.0214 (0.0013)	0.0154 (0.1809)	0.0145 (0.1229)	0.0710	0.0240	0.0164	0.0161
75% quantile	0.0645 (0.0000)	0.0382 (0.2074)	0.0221 (0.5962)	0.0231 (0.6392)	0.0958	0.0443	0.0252	0.0272
ARMA-GARCH								
Mean	0.0506 (0.0055)	0.0326 (0.0878)	0.0133 (0.7469)	0.0146 (0.6873)	0.0727	0.0365	0.0129	0.0159
Median	0.0442 (0.0001)	0.0298 (0.0220)	0.0124 (0.8279)	0.0118 (0.8370)	0.0621	0.0315	0.0117	0.0113
25% quantile	0.0373 (0.0000)	0.0242 (0.0009)	0.0107 (0.6115)	0.0100 (0.4722)	0.0531	0.0264	0.0101	0.0093
75% quantile	0.0545 (0.0017)	0.0390 (0.1016)	0.0150 (0.9319)	0.0160 (0.9579)	0.0776	0.0428	0.0144	0.0159

Table 2-4 Dates of major market declines for the CSI 300 Index from January 8, 2002 to June 29, 2012

	Date	Return(%)
1	February 27, 2007	-9.2
2	June 10, 2008	-8.1
3	June 4, 2007	-8.0
4	January 22, 2008	-7.9
5	November 18, 2008	-7.4
6	June 19, 2008	-7.2
7	October 27, 2008	-7.1
8	August 31, 2009	-7.1
9	January 28, 2008	-6.8
10	March 30, 2007	-6.7

Table 2-5 Daily analysis of major market declines of the CIS 300 Index under four different ARMA-GARCH models.

	KS (p-value)	AD	Residual	Probability	Average Time
1. February 27, 2007. The CSI 300 Index dropped by 9.2%					
	KS (p-value)	AD	Residual	Probability	Average Time
Gaussian	0.0509 (0.0401)	0.0502	-4.43	4.65E-06	859.71
Stable	0.0454 (0.0889)	0.0378	-4.55	1.11E-03	3.60
CTS	0.0206 (0.9019)	0.0224	-4.55	2.27E-04	17.59
NTS	0.0249 (0.7317)	0.0243	-4.55	2.66E-04	15.06
2. June 10, 2008. The CSI 300 Index dropped by 8.1%					
	KS (p-value)	AD	Residual	Probability	Average Time
Gaussian	0.0603 (0.0083)	0.0785	-3.55	1.80E-04	22.21
Stable	0.0888 (0.0000)	0.1209	-3.61	9.72E-03	0.41
CTS	0.0126 (0.9997)	0.0120	-3.61	4.02E-03	1.00
NTS	0.0289 (0.5494)	0.0321	-3.61	4.23E-03	0.94
3. June 4, 2007. The CSI 300 Index dropped by 8%					
	KS (p-value)	AD	Residual	Probability	Average Time
Gaussian	0.0521 (0.0329)	0.0583	-3.13	7.91E-04	5.06
Stable	0.0329 (0.3829)	0.0278	-3.40	5.44E-03	0.74
CTS	0.0290 (0.5452)	0.0250	-3.40	4.74E-03	0.84
NTS	0.0378 (0.2286)	0.0281	-3.40	4.56E-03	0.88
4. January 22, 2008. The CSI 300 Index dropped by 7.9%					
	KS (p-value)	AD	Residual	Probability	Average Time
Gaussian	0.0509 (0.0397)	0.0671	-3.72	8.48E-05	47.17
Stable	0.0700 (0.0012)	0.0912	-3.78	8.26E-03	0.48
CTS	0.0154 (0.9932)	0.0151	-3.78	4.66E-03	0.86
NTS	0.0272 (0.6275)	0.0293	-3.78	3.22E-03	1.24
5. November 18, 2008. The CSI 300 Index dropped by 7.4%					
	KS (p-value)	AD	Residual	Probability	Average Time
Gaussian	0.0606 (0.0078)	0.0790	-2.42	7.96E-03	0.50
Stable	0.0990 (0.0000)	0.1271	-2.49	2.57E-02	0.16
CTS	0.0192 (0.9407)	0.0146	-2.49	2.02E-02	0.20
NTS	0.0464 (0.0770)	0.0562	-2.49	1.77E-02	0.23
6. June 19, 2008. The CSI 300 Index dropped by 7.2%					
	KS (p-value)	AD	Residual	Probability	Average Time
Gaussian	0.0570 (0.0148)	0.0779	-2.46	6.85E-03	0.58
Stable	0.0931 (0.0000)	0.1246	-2.21	3.24E-02	0.12
CTS	0.0162 (0.9878)	0.0129	-2.21	2.81E-02	0.14
NTS	0.0428 (0.1255)	0.0573	-2.21	2.57E-02	0.16
7. October 27, 2008. The CSI 300 Index dropped by 7.1%					
	KS (p-value)	AD	Residual	Probability	Average Time
Gaussian	0.0579 (0.0126)	0.0769	-2.21	1.41E-02	0.28
Stable	0.0926 (0.0000)	0.1213	-2.25	3.24E-02	0.12
CTS	0.0152 (0.9943)	0.0127	-2.25	2.75E-02	0.15
NTS	0.0420 (0.1390)	0.0502	-2.25	2.45E-02	0.16
8. August 31, 2009. The CSI 300 Index dropped by 7.1%					
	KS (p-value)	AD	Residual	Probability	Average Time
Gaussian	0.0644 (0.0038)	0.0811	-2.51	6.22E-03	0.64
Stable	0.1052 (0.0000)	0.1424	-2.51	2.58E-02	0.16

CTS	0.0159 (0.9899)	0.0160	-2.51	1.96E-02	0.20
NTS	0.0554 (0.0193)	0.0725	-2.51	1.77E-02	0.63
9. January 28, 2008. The CSI 300 Index dropped by 6.8%					
	KS (p-value)	AD	Residual	Probability	Average Time
Gaussian	0.0521 (0.0333)	0.0719	-2.60	4.26E-03	0.94
Stable	0.0755 (0.0004)	0.0975	-2.59	1.83E-02	0.22
CTS	0.0143 (0.9975)	0.0123	-2.59	1.46E-02	0.27
NTS	0.0187 (0.9520)	0.0180	-2.59	1.51E-02	0.26
10. March 30, 2007. The CSI 300 Index dropped by 6.7%					
	KS (p-value)	AD	Residual	Probability	Average Time
Gaussian	0.0472 (0.0690)	0.0548	-3.88	4.57E-05	87.45
Stable	0.0376 (0.2330)	0.0301	-4.58	2.26E-03	1.77
CTS	0.0300 (0.4993)	0.0268	-4.58	9.58E-04	4.17
NTS	0.0232 (0.8073)	0.0240	-4.58	4.04E-04	9.90

Table 2-6 Summary of Christoffersen Likelihood Ratio tests of the CSI 300 Index for 6 years.

Model	2006(January 4, 2006 – December 29, 2006)				2007(January 4, 2007 – December 28, 2007)			
	N	CLR_{uc} (p-value)	CLR_{ind} (p-value)	CLR_{cc} (p-value)	N	CLR_{uc} (p-value)	CLR_{ind} (p-value)	CLR_{ind} (p-value)
normal-ARMA-GARCH	5	0.1354 (0.7129)	0.0760 (0.7829)	0.2113 (0.8997)	9	10.6646 (0.0011)	0.6985 (0.4033)	11.3631 (0.0034)
Stable-ARMA-GARCH	2	0.0748 (0.7845)	0.0336 (0.8545)	0.1084 (0.94720	6	3.7987 (0.0516)	0.3064 (0.5799)	4.0961 (0.1290)
CTS-ARMA-GARCH	2	0.0748 (0.7845)	0.0336 (0.8545)	0.1084 (0.94720	6	3.7987 (0.0516)	0.3064 (0.5799)	4.0961 (0.1290)
NTS-ARMA-GARCH	2	0.0748 (0.7845)	0.0336 (0.8545)	0.1084 (0.94720	6	3.7987 (0.0516)	0.3064 (0.5799)	4.0961 (0.1290)
	2008 (January 2, 2008 – December 31, 2008)				2009 (January 5, 2009 – December 31, 2009)			
	N	CLR_{uc} (p-value)	CLR_{ind} (p-value)	CLR_{cc} (p-value)	N	CLR_{uc} (p-value)	CLR_{ind} (p-value)	CLR_{cc} (p-value)
normal-ARMA-GARCH	8	7.9155 (0.0049)	1.3553 (0.2443)	9.2708 (0.0097)	6	3.7003 (0.0544)	0.3026 (0.5823)	4.0028 (0.1351)
Stable-ARMA-GARCH	3	0.8188 (0.3655)	4.0758 (0.0435)	4.89469 (0.0865)	2	0.0891 (0.7654)	0.0331 (0.8557)	0.1221 (0.9408)
CTS-ARMA-GARCH	3	0.8188 (0.3655)	4.0758 (0.0435)	4.89469 (0.0865)	2	0.0891 (0.7654)	0.0331 (0.8557)	0.1221 (0.9408)
NTS-ARMA-GARCH	3	0.8188 (0.3655)	4.0758 (0.0435)	4.89469 (0.0865)	2	0.0891 (0.7654)	0.0331 (0.8557)	0.1221 (0.9408)
	2010(January 5, 2010 – December 31, 2010)				2011(January 4, 2011 – December 30, 2011)			
	N	CLR_{uc} (p-value)	CLR_{ind} (p-value)	CLR_{cc} (p-value)	N	CLR_{uc} (p-value)	CLR_{ind} (p-value)	CLR_{cc} (p-value)
normal-ARMA-GARCH	5	2.1463 (0.1429)	0.2128 (0.6446)	2.3591 (0.3074)	5	2.0916 (0.1491)	0.2101 (0.6467)	2.2917 (0.3179)
Stable-ARMA-GARCH	5	2.1463 (0.1429)	0.2128 (0.6446)	2.3591 (0.3074)	3	0.1210 (0.7280)	0.0750 (0.7842)	0.1960 (0.9067)
CTS-ARMA- GARCH	5	2.1463 (0.1429)	0.2128 (0.6446)	2.3591 (0.3074)	4	0.08445 (0.3581)	0.1339 (0.7144)	0.9784 (0.6131)
NTS-ARMA- GARCH	5	2.1463 (0.1429)	0.2128 (0.6446)	2.3591 (0.3074)	4	0.08445 (0.3581)	0.1339 (0.7144)	0.9784 (0.6131)

[b] CLR_{uc}: Christoffersen Likelihood Ratio test of unconditional coverage.
CLR_{ind}: Christoffersen Likelihood Ratio test of independence.
CLR_{cc}: Christoffersen Likelihood Ratio joint test of coverage and independence.

Table 2-7 Summary of Berkowitz Likelihood Ratio tests of the CSI 300 index for 6 years.

Model	2006		2007		2008	
	BLR_{ind} (p-value)	BLR_{tail} (p-value)	BLR_{ind} (p-value)	BLR_{tail} (p-value)	BLR_{ind} (p-value)	BLR_tail (p-value)
normal-ARMA-GARCH	0.0760 (0.7829)	0.2113 (0.8997)	0.6985 (0.4033)	11.3631 (0.0034)	1.3553 (0.2443)	9.2708 (0.0097)
Stable-ARMA- GARCH	0.0336 (0.8545)	0.1084 (0.9472)	0.3064 (0.5799)	4.0961 (0.1290)	4.0758 (0.0435)	4.89469 (0.0865)
CTS-ARMA- GARCH	0.0336 (0.8545)	0.1084 (0.9472)	0.3064 (0.5799)	4.0961 (0.1290)	4.0758 (0.0435)	4.89469 (0.0865)
NTS-ARMA- GARCH	0.0336 (0.8545)	0.1084 (0.9472)	0.3064 (0.5799)	4.0961 (0.1290)	4.0758 (0.0435)	4.89469 (0.0865)
	2009		2010		2011	
	BLR_{ind} (p-value)	BLR_{tail} (p-value)	BLR_{ind} (p-value)	BLR_{tail} (p-value)	BLR_{ind} (p-value)	BLR_{tail} (p-value)
normal-ARMA- GARCH	0.3026 (0.5823)	4.0028 (0.1351)	0.2128 (0.6446)	2.3591 (0.3074)	0.2101 (0.6467)	2.2917 (0.3179)
Stable-ARMA- GARCH	0.0331 (0.8557)	0.1221 (0.9408)	0.2128 (0.6446)	2.3591 (0.3074)	0.0750 (0.7842)	0.1960 (0.9067)
CTS-ARMA-GARCH	0.0331 (0.8557)	0.1221 (0.9408)	0.2128 (0.6446)	2.3591 (0.3074)	0.1339 (0.7144)	0.9784 (0.6131)
NTS-ARMA-GARCH	0.0331 (0.8557)	0.1221 (0.9408)	0.2128 (0.6446)	2.3591 (0.3074)	0.1339 (0.7144)	0.9784 (0.6131)

[c] BLR_{ind}: Berkowitz Likelihood Ratio test of independence.
BLR_{tail}: Berkowitz Likelihood Ratio test of tail distribution.

Table 2-8 Summary of the average relative difference from the stable-ARMA-GARCH model, CTS-ARMA-GARCH model and NTS-ARMA-GARCH model to the normal-ARMA-GARCH model, respectively.

ARD		2006	2007	2008	2009	2010	2011
Normal-ARMA-GARCH	stable-ARMA-GARCH	0.1466	0.1256	0.2585	0.2582	0.2246	0.2913
	CTS-ARMA-GARCH	0.1515	0.1411	0.1956	0.1922	0.1942	0.1892
	NTS-ARMA-GARCH	0.1637	0.1523	0.2063	0.2013	0.2043	0.1977

Table 2-9 Summary of Chow test for structure breaks in daily spread of AVaR from 2005 to 2011.Test statistic of Chow test between each two of the three divided regimes which are 2005-2006, 2007-2009 and 2010-2011. (a) Test statistic of Chow test between each two of the three divided regimes which are 2005-2006, 2007-2009 and 2010-2011. (b) Test statistic of Chow test within each one of the three divided regimes which are 2005-2006, 2007-2009 and 2010-2011

	F statistics	P-value	Wald statistics	P-value
$\lambda_{1,2}$	48.3070	3.1970e-12	89.2257	0
$\lambda_{2,3}$	42.9570	2.0228e-11	25.0453	5.6000e-07
$\lambda_{1,3}$	116.1464	0	94.9391	0

[d] $\lambda_{1,2}$: Structure breaks between 2005-2006 and 2007-2009.
$\lambda_{2,3}$: Structure breaks between 2007-2009 and 2010-2011.
$\lambda_{1,3}$: Structure breaks between 2005-2006 and 2010-2011.
(a) .

	F statistics	P-value	Wald statistics	P-value
$\lambda_{1,1}$	1.1458	0.3265	7.8104	0.0052
$\lambda_{2,2}$	2.0810	0.1359	0.0372	0.8471
$\lambda_{3,3}$	31.8083	0	4.7849	0.0287

[e] $\lambda_{1,1}$: Structure breaks between 2005 and 2006.
$\lambda_{2,2}$: Structure breaks between 2007 and 2008-2009.
$\lambda_{1,3}$: Structure breaks between 2010 and 2011.
The size of the sample in each test is 50 and the observations are uniformly sampling from the daily spreads

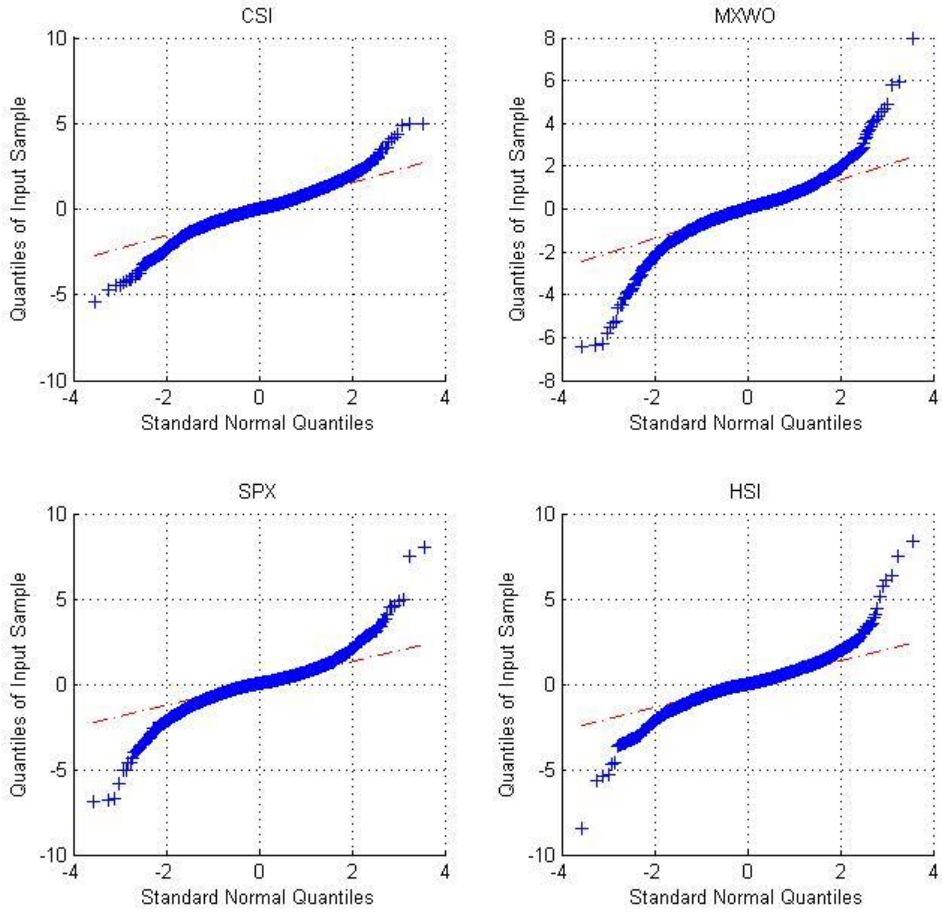


Figure 2-1 Q-Q plot of sample return distributions against standard Gaussian distribution for the Chinese stock market and other world stock markets

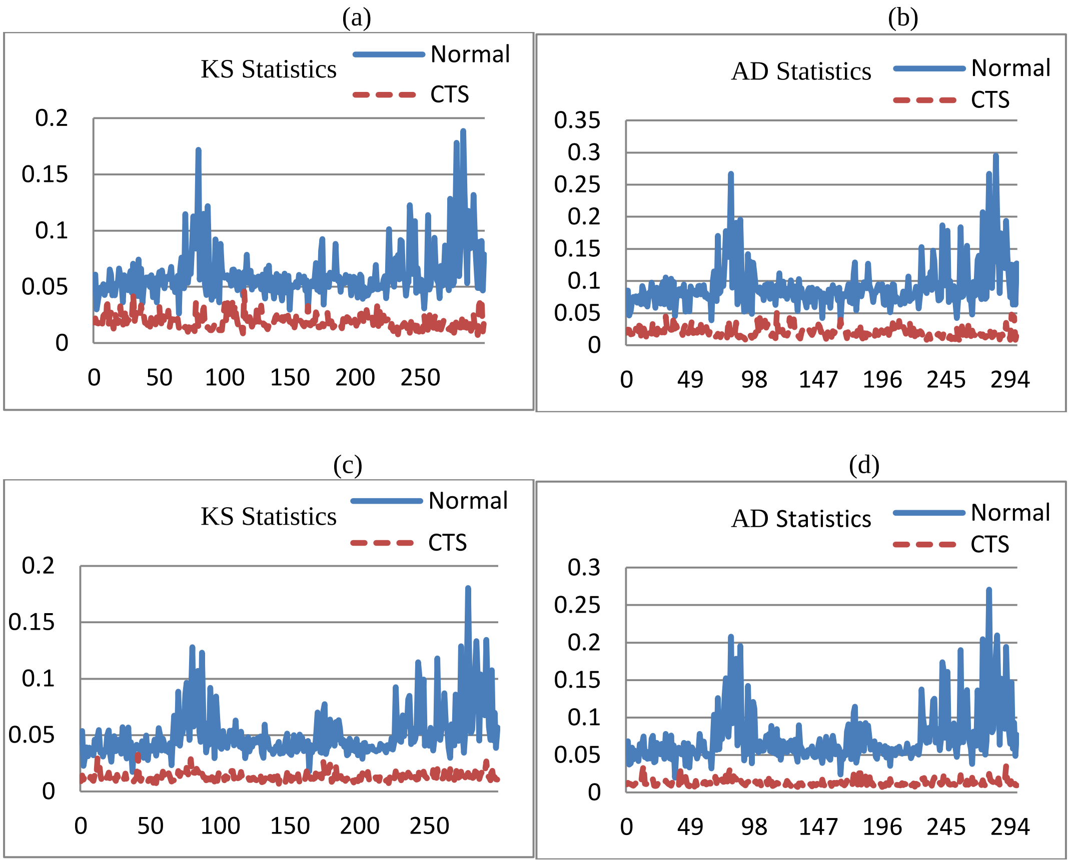
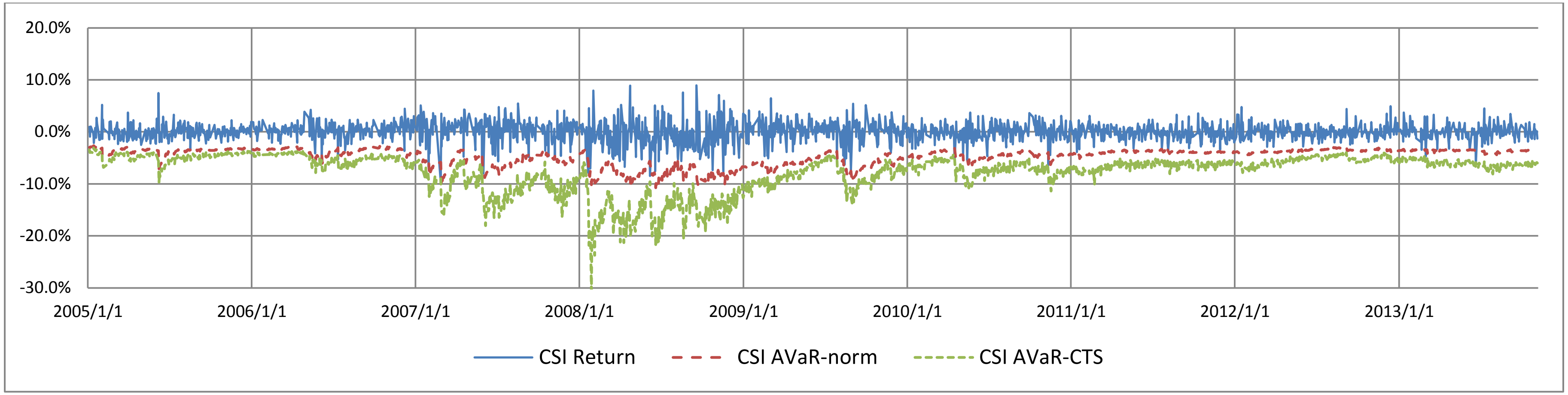


Figure 2-2 The Kolmogorov-Smirnov test statistic and Anderson-Darling test statistic of the Gaussian distribution hypothesis and the CTS distribution hypothesis, for WN model (a) (b), and for ARMA-GARCH model (c) (d).

(a)



(b)

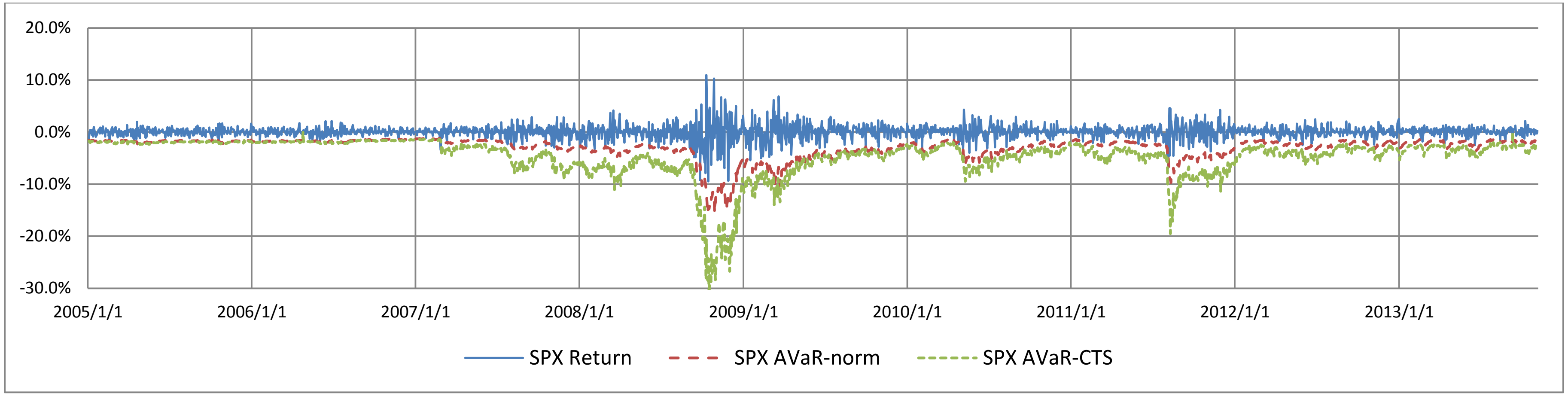
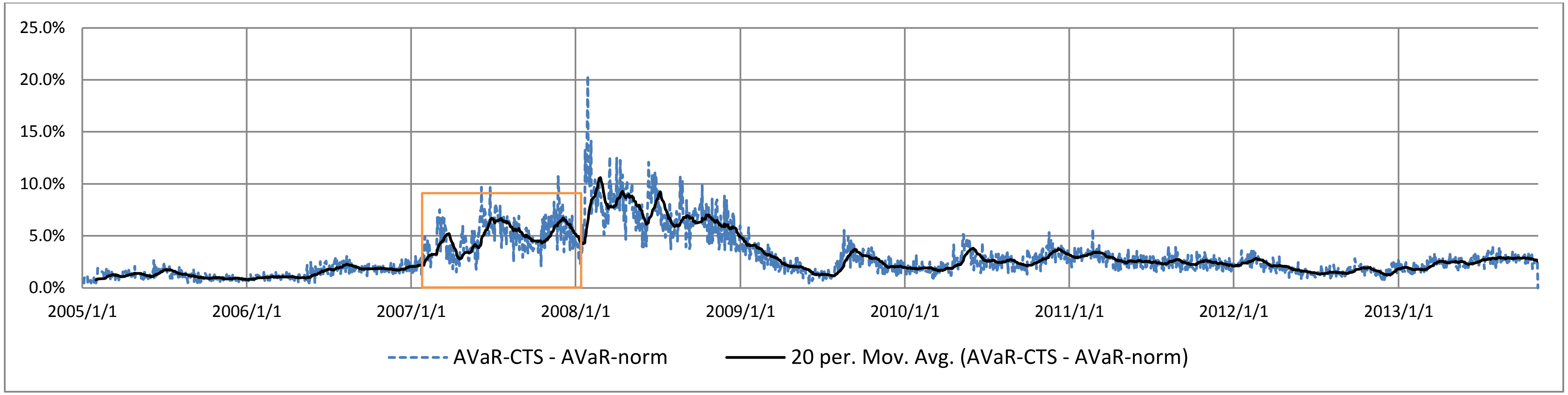


Figure 2-3 The log return and 1% AVaR for the normal-ARMA-GARCH and the CTS-ARMA-GARCH model. (a) The CSI 300 Index. (b) The S&P 500 Index.

(a)



(b)

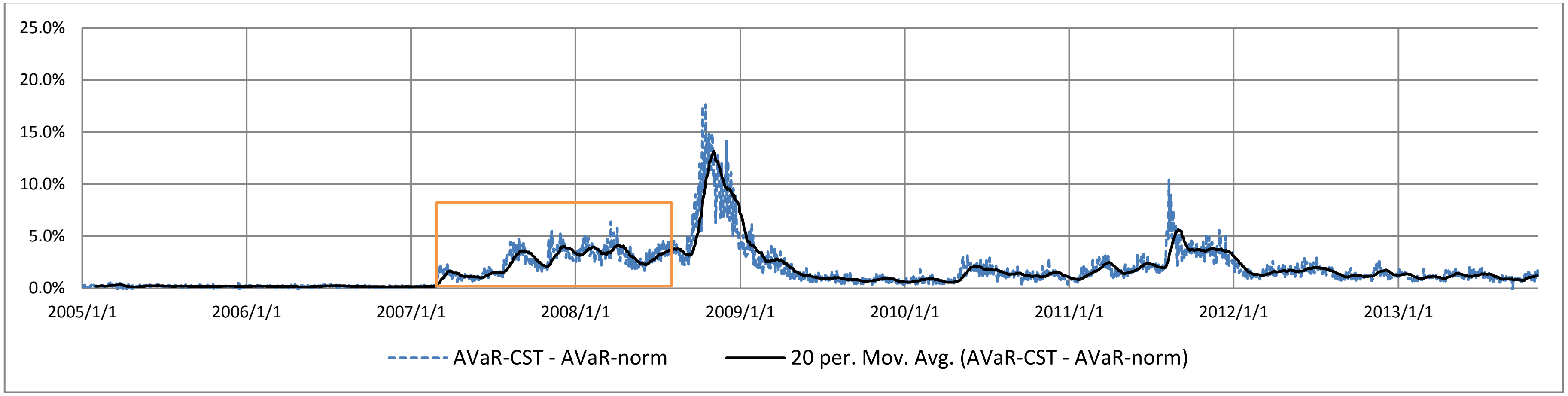


Figure 2-4 Daily spread between the 1% AVaR for the normal-ARMA-GARCH and CTS-ARMA-GARCH model. (a) The CSI 300 Index. (b) The S&P 500 Index.

3.Chinese Equity Factor Model

3.1 Introduction

Building on the pioneering work of Markowitz (1952) of the modern portfolio theory, Sharpe (1964) proposed the Capital Asset Pricing Model for pricing risky assets. Under certain assumptions, Sharpe shown that the single super-efficient portfolio in Markowitz mean-variance framework was the market portfolio itself. Thus, the expected return of an asset only depends on the expected return of the market portfolio and the beta of the asset relative to the market. Based on the CAPM theory, the return of any asset can be decomposed into a systematic component that is perfectly correlated to the market, and a residual component that is uncorrelated with the market. The risk raised from the residual components of assets in a portfolio, called specific or idiosyncratic risk, can be reduced through diversification, while the risk raised from the systematic components, called systematic risk, can not be diversified (within one market) but can be hedged.

The assumption of uncorrelated residuals of the CAPM can be challenged by empirical examination, even if there is only one source of expected returns. One extension is multi-factor risk model first development by Rosenberg (1974) and Rosenberg and Marathe (1975) to estimate the co-movement of stock returns, while in the CAPM theory the market portfolio return is treated as the only one factor. Returns that cannot be explained by the factors are deemed specific returns and are assumed to be uncorrelated.

The applications of multi-factor models are various and are based on the analysis and prognosis of portfolio risk. Besides modeling the returns and covariance of the assets and constructing portfolios, multi-factor models can give valuable insights especially in performance and risk attribution. In this study we provide a Chinese equity factor (CEF) model for a universe of 2012 Chinese A-share stocks listed in both Shanghai and Shenzhen stock exchanges. In Section 3.2, we give the general methodology of the multi-factor risk model. Factor exposures and factor returns are computed in Section 3.3 and Section 3.4. In Section 3.5 we discuss the time series forecasting model of factor returns and specific returns and extensive back-tests for both portfolio and assets level. Finally we give the conclusion in Section 3.5.

3.2 Multi-factor Risk Models

3.2.1 General Structure of Multi-factor Models

Under the frame work of Markowitz portfolio theory, the expected return and covariance are estimated and applied in portfolio construction. A key challenge in estimating the asset covariance matrix lies in the sheer dimensionality of the problem. For instance, an active portfolio containing 2000 stocks[8] requires more than two million independent elements. If the asset covariance matrix is computed naively, then the matrix is likely to be extremely ill-conditioned. The larger the number of assets the more severe estimation error and hence portfolio instability arise. A typical phenomena for a large scale portfolio is that the number of time observations is less than the number of stocks, then the covariance matrix is said to be "rank deficient", meaning that it is possible to construct apparently riskless portfolios

Multi-factor models were development to provide a more robust solution to this problem. For instance, computing the covariance matrix of a 2000 stocks portfolio with a 20-factor model requires estimation of roughly 2200 elements. An intuitive interpretation of multi-factor model follows the basic idea of the CAPM theory, that the returns of different stocks are "explained" by common factors and partly estimated through their sensitivity or exposures to the factors in a linear model. Within a factor model, the returns and risk of stocks are divided into two distinct sources. The first source, due to the factors, represents the systematic components. The second source represents the diversifiable components that cannot be explained by the factors, and is therefore deemed idiosyncratic or asset specific. More specifically, the stock return is explained as

$$r_i = \sum_k \gamma_{i,k} f_k + e_i \qquad 3.1$$

where r_i is the return of stock i, $\gamma_{i,k}$ is the exposure of stock i to factor k, f_k is the return to the factor, and e_i is the stock specific return.

Consider a portfolio containing n stocks with weight w_i, and return given by

$$R_p = \sum_i w_i r_i \qquad 3.2$$

with the return of stocks given by the factor model, we write portfolio return as

$$R_p = \sum_i \sum_k w_i \gamma_{i,k} f_k + \sum_i w_i e_i \qquad 3.3$$

Two important assumptions need to be fulfilled in factor risk modeling are (a) the factor returns are uncorrelated with specific returns and (b) the specific returns are uncorrelated among themselves, i.e. the correlations of stock returns are only explained by the common

[8] Up to January 2013, there are 2473 A-share companies listed in Shanghai stock exchange and Shenzheng stock exchanges.

factors. We do not in general require independence, only uncorrelatedness. This allows the variance of portfolio to be expressed as

$$var(R_p) = \sum_{i,j} w_{i,j}^2 \gamma_{i,j,k}^2 \sigma_{k,l}^f + \sum_i w_i^2 \sigma_{i,i}^e \tag{3.4}$$

where $\sigma_{k,l}^f$ is the predicted covariance between factor k and l, and $\sigma_{i,i}$ is the predicted variance of specific return e_i.

We can rewrite the equation (3.3) and (3.4) in matrix format as

$$\boldsymbol{r} = \boldsymbol{\Gamma f} + \boldsymbol{e} \tag{3.5}$$

where $\boldsymbol{r} = (r_1, \dots, r_n)'$ is a n-demensional random vector representing the observations of asset returns;

$\boldsymbol{f} = (f_1, \dots, f_k)'$ is a k-demensional random vector of common factors with $k < n$;

$\boldsymbol{e} = (e_1, \dots, e_n)'$ is a n-demensional random vector of specific, or idiosyncratic terms, which are uncorrelated;

$\boldsymbol{\Gamma} \in \mathbb{R}^{n\times k}$ is a matrix of constant factor exposures; and

$\mathrm{Cov}(\boldsymbol{f}, \boldsymbol{e}) = 0$.

In the same way, we can write the return of portfolio as:

$$R_p = \boldsymbol{w^T \Gamma f} + \boldsymbol{w^T e} \tag{3.6}$$

where $\boldsymbol{w} = (\mathrm{w}_1, \dots \mathrm{w}_\mathrm{d})'$ is the n-by-1 portfolio weight vector.

And the variance of portfolio return is computed as

$$Var(R_p) = Var(\boldsymbol{w^T \Gamma f} + \boldsymbol{w^T e}) \tag{3.7}$$
$$= Var(\boldsymbol{w^T \Gamma f}) + Var(\boldsymbol{w^T e})$$
$$= \boldsymbol{w}(\boldsymbol{\Gamma \Theta \Gamma^T})\boldsymbol{w^T} + \boldsymbol{w \Omega w^T}$$

where $\boldsymbol{\Theta} = cov(\boldsymbol{f})$ is the k-by-k covariance matrix and $\boldsymbol{\Omega} = var(\boldsymbol{e})$ is the n-by-n diagonal matrix with variance of specific returns.

From the matrix formation, it is easy to see the tremendous dimension reduction effect by the factor model. For instance, the covariance matrix of stocks returns Σ is a n-by-n matrix with $\frac{1}{2}n(n+1)$ different elements. By the factor model, we can compute $\boldsymbol{\Sigma} = \boldsymbol{\Gamma \Theta \Gamma} + \boldsymbol{\Omega}$ which only need $(n + \frac{1}{2}k)(k+1)$ elements to be estimated.

3.2.2 Statistical Calibration Strategies

Now we have the data $\boldsymbol{X}_1, \dots, \boldsymbol{X}_k \in \mathbb{R}^n$ representing risk stock returns. Each vector observation $\boldsymbol{X}_t$ recorded at a time t is assumed to be generated by the factor model (3.5) for some common factor return vector $\boldsymbol{f}_t$ and some specific return vector $\boldsymbol{e}_t$.

There are mainly three different methodologies to estimate factor models depending on whether or not the factor is observable or unobservable:

1. time series analysis,
2. cross-sectional analysis,
3. statistical factor analysis.

In time-series analysis, we assume that appropriate factors for the stock returns have been identified and collected in advance and thus they are observable. A simple instance would be the CAPM case, in which the market index representing the efficient market portfolio is the only factor. Given the time series of individual stock return observations $\boldsymbol{X}_i$ and index return observations $\boldsymbol{f}^M$, we can estimate the factor exposure γ^{M} which is usually named β using the standard ordinary least-squares (OLS) method in linear regression. The regression model of X_i take the form as

$$\begin{pmatrix} x_{i,t_0} \\ \vdots \\ x_{i,t} \end{pmatrix} = \begin{bmatrix} f_{1,t_0} & \cdots & f_{k,t} \\ \vdots & \ddots & \vdots \\ f_{1,t} & \cdots & f_{k,t} \end{bmatrix} \begin{pmatrix} \gamma_{i,1} \\ \vdots \\ \gamma_{i,k} \end{pmatrix} + \begin{pmatrix} e_{i,t_0} \\ \vdots \\ e_{i,t} \end{pmatrix} \tag{3.8}$$

The advantage of this approach is the control of the factors which can be interpreted easily. Typical factors are considered relevant in many studies. For instance, the factors in the studies of Berry et al. (1988) are the excess return of long term bonds, exchange rates, price changes of raw materials and inflation. Another famous example is the study of Fama and French (1993) and later Fama and French (2012), in which factors are Rm-Rf, SMB, HML and Mom.

Usually a linear regression is performed with the additional assumption that the factor exposures, known as effects or regression coefficients in linear regression, are constant over time. In the original formulation the factor model, this constancy was not required. It is reasonable to assume that the factor exposures change over time. Even if no constancy of the parameters is assumed, this methodology always need some time to identify and adapt to changes in these parameters due to some abrupt changes, such as mergers or even unexpected financial reports of companies. A shorter estimation period, for instance, daily update instead of monthly, can mitigate the weakness.

In cross-sectional analysis, the factors are not unobservable while the factor exposures are taken as given. Intuitively all stock return observations $\boldsymbol{X}_{n,t}$ at a time t are regressed against some pre-selected factor exposures $\boldsymbol{\Gamma}_t$, which are regarded as regressors or independent variables in a linear regression language. The regression model in period t thus take the form as:

$$\begin{pmatrix} x_{1,t} \\ \vdots \\ x_{n,t} \end{pmatrix} = \begin{bmatrix} \gamma_{1,1,t} & \cdots & \gamma_{1,k,t} \\ \vdots & \ddots & \vdots \\ \gamma_{n,1,t} & \cdots & \gamma_{n,k,t} \end{bmatrix} \begin{pmatrix} f_{1,t} \\ \vdots \\ f_{k,t} \end{pmatrix} + \begin{pmatrix} e_{1,t} \\ \vdots \\ e_{n,t} \end{pmatrix} \qquad 3.9$$

This regression is performed for several periods and the estimated regression coefficients are collected as the time series of factor returns F_t. Starting from these time series we can then estimate future factor returns $\hat{F}_{t+s}$ and corresponding covariance Θ_{t+s}, where $s \geq 1$. For instance, the multi-factor model developed by BARRA[9] use fundamental descriptors to calculate factor exposures, which representing the sensitivity of stock returns to the economic profile of the corresponding companies. Compare to time series regression, cross-sectional regression has little problem with companies change their financial profile overnight. However, it suffer from with econometric problems that not all observations carry the same information. Observations with large errors should be weighted less, but there is little available theory guiding model builders. Also multivariate outliers are difficult to spot and might have a nontrivial effect on factor returns (Scherer, et al. 2010).

In statistical factor analysis, appropriate factors alone with factor exposures are themselves estimated simultaneously from the data $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$. In the case of stock risk factor model, it is not clear a priori what the best set of factors would be. There are two general strategies for finding factors. The first strategy, which is quite common in finance, is to use the method of principle components to construct factors (See McNiel, et al, 2005). We note that the factors we obtain can on the one hand, “optimally” explain the past, but on the other hand may not have obvious economical interpretations. In the second approach, classical statistical factor analysis, it is assumed that the data are identically distributed with a distribution whose covariance matrix have the factor structure. Various techniques are used to find the estimates and then use them to construct factors. The statistical models have a good in sample it, while they are more challenged with bad out of sample performance. Empirical test by Miller (2006) have shown that statistical models have a poor record in identifying changing risk structure. Using statistical models in isolation might not be a very good idea, Instead they are increasingly used to safeguard against hidden factors in residuals. This idea gives rise to so called hybrid models try to combine different models. Using fundamental factors either by time

[9] BARRA, founded by Rosenberg in 1975, later acquired and combined with MSCI in 2004, developed several versions of multi-factor equity model for markets in several countries and regions, as well as a global market version.

series analysis or cross-sectional analysis provides context for risk and portfolio management as well as aligning the covariance matrix of asset returns. On the other hand, applying a statistical factor model on the residuals imposes coverage of un-modeled common factors.

3.3 Model Scope and Factors

This section describes how to construct a factor risk model in a time series framework and the factors used in our CHE factor model. Technical details about descriptors and factor exposures are discussed at the end.

In the context of risk management, the goal of all approached to factor models is to obtain factor returns $\boldsymbol{f}_t$ and factor exposures $\boldsymbol{\Gamma}_t$ (and the constant part where relevant). After this approach, we can then concentrate on modeling the distribution or dynamics of $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_k$, which is a lower-dimensional problem then modeling $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$. The unobservable residuals $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ are of secondary importance. In situation we have many risk factors the risk embodied in the residuals is partly mitigated by a diversification effect, whereas the risk embodied in the common factors remains.

In our work we focus on the factor model with pre-defined factors, which means either factors or factor exposures are observable and collected first. More specifically, except for one market factor collected from the market, all other factors are unobservable, but the exposure matrix Γ is assumed to be known.

In a time series framework, we define of the factor model as following:

$$\boldsymbol{R}_t = \boldsymbol{\Gamma}_{t-1}\boldsymbol{f}_t + \boldsymbol{e}_t \qquad 3.10$$

where $\mathbf{X}_\mathrm{t}$ = n-dimensional vector of one-period stock returns;
$\boldsymbol{\Gamma}_t$ = $n \times k$ matrix of stock exposures to factors as of time $\mathrm{t}-1$;
$\boldsymbol{f}_t$ = k-dimensional vector of one-period factor returns; and
$\boldsymbol{e}_t$ = n-dimensional vector of one-period specific returns;

Some considerations must be emphasized in building a high-quality factor risk model. A key assumption is that factors capture all systematic drivers of stock returns, thus implying that specific returns are mutually uncorrelated. Thus, on one side, a high-quality factor structure should explain as fully as possible the co-movement among stock returns. Missing important factors could impact the portfolio through diversification as hidden risk resources spuriously appeared to be diversified. On the other side, the factor structure should be parsimony, meaning systematic component in stock returns is explained with the fewest possible number of factors. Combining weak or spurious factors makes the model more susceptible to noise and less robust in capturing the underlying relations of stock returns. One way to avoid weak or spurious factors is to ensure the statistical significance of factor returns. In particular, the statistical significance should be persistent across time, and not due to isolated events that are unlikely to occur in the future.

Collinearity is another important consideration in building factor risk model. If the factor structure is excessively collinear, which means one or more factors can be approximately replaced by linear combinations of other factors, estimation errors in regressions can be very large and the factor returns difficult to interpret. A frequently used measure of collinearity is Variance Inflation Factor (VIF). A detailed discussion can be found in Section 3.4.

Last, but not least, risk factors should be intuitive. In other words, they must be transparent, easily interpretable, and consistent with investors' views about what these factors represent.

The CEF model provides risk forecasts for a broad coverage universe of total over 2000 A-share stocks listed in Shanghai Stock Exchange (SHSE) and Shengzhen Stock Exchange (SZSE), excluding small cap stocks listed in ChinNext market[10]. We collect the data from January 4, 2002 to December 29, 2012 and provides forecast since 2005, as only a limited number of stocks available before then. Figure 3.1 presents the estimation universe, which shows a blooming growth of stocks in later half of first decade in 21th century. The returns of stocks are calculated on a daily base as

$$r_t = \log\left(\frac{r_{t+1}}{r_t}\right) \qquad 3.11$$

where r_t is the log-return at time t, p_{t+1} and p_t are the adjusted closing price at time $t+1$ and t, respectively.

The Chinese equity model uses total 18 equity factors:

- A market factor
- 7 style factors

[10] As an independent market of China's multi-tier capital market system, ChiNext offers a new capital platform tailor-made for the needs of enterprises engaged in independent innovation and other growing venture enterprises. The difference between ChiNext and the main board lies in their mechanisms of financing, investment and risk management for issuers at various stages of development, rather than simply the sizes. The ChiNext market was launched in SZSE on October 23, 2009. Due to the short period of available data, we exclude stocks listed in ChiNext market in this study.

- 10 industry factors

The return of CSI 300 index[11] is used as the market factor, which representing the efficient market portfolio same as in CAPM. The market factor exposure is estimated ahead of other factors with time series regression, providing intuitively a beta-like parameter for portfolio management.

Style factors and industry factors are referred as fundamental factors used to capture fundamental and market-driven systematic sources of risk and return. Referring to equity models developed by Barra (see Briner, et al., 2009 and Menchero, et al., 2011) and other multi-factor models (see Connor, 1995; Campbell et al., 1997; Alexander, 2001, Zivot and Wang, 2003), the 7 factors are size, value, growth, momentum, liquidity, volatility and leverage. Style factor exposures are built by combining multiple descriptors with similar characteristics. The formula of combination and definition of all descriptors used can be found in Table 3.1. Here we summarize the qualitative characteristics of the styles.

The *Size* factor captures systematic return and risk difference between large cap and small cap stocks. The exposures of the stocks to this factor are calculated based on their market capitalization. Most stocks in the CEF universe have negative size exposure. This is a consequence of centering the exposures such that the cap-weighted portfolio has zero exposure. Size factor returns generally exhibit positive correlations with the market factor.

The *Value* factor uses the price-to-book descriptor. It indicates whether or not a company is undervalued based on its trading price and book value. The value factor does not consider a company's ability of generating positive future income. Some companies have positive exposure to value factor could indicate nice investment possibility as the market undervalues their potential earning ability, while some companies with positive exposure because the market grades them negatively due to a failed business model.

The *Growth* factor combines three descriptors widely used in fundamental analysis of stocks, which are dividend per share, earning per share and return on equity.

The *Momentum* factor is related to the traditional technique analysis of stocks, which uses the past performance of stocks over a pre-defined time period to measure the relative strength.

The *Liquidity* factor represents the liquidity in the market. The factor exposures are

---

[11] Description of CSI 300 can be found in Chapter 3. More details are provided on http://www.csindex.com.cn.

defined as the natural logarithm of 20-day average volume over the 20-day average market capitalization.

The *Volatility* factor measures the stock's relative volatility over time according to its historical behavior.

The *Leverage* factor provides a measure of the stock's exposure to debt level. The exposure of a given stock is calculated as its total debt over its market capitalization.

Industries represent another major class of risk model factors. Here we adopt industry factors based on the Global Industry Classification Standard(GICS®), a widely used hierarchical industry scheme. The factor exposure of a given stock to a particular industry factors is either 1 if the stock belongs to that industry or 0 otherwise.

Some technical details about handling descriptors and factor exposures are discussed next.

- Truncation of outliers

There are three types of outliers.

The first group represents values so extreme that they are treated as potential data errors and removed from the estimation process. For instance, a stock's market capitalization and book to price ration must be positive.

The second group represents values that are regarded as legitimate, but nonetheless so large that their impact on the model must be limited. We trim these observations to three standard deviations from the mean or to their certain quintiles.

The third group of observations, forming the bulk of the distribution, consists of values that are less than three standard deviations from the mean; these observations are left unadjusted.

- Standardize factor exposures

Style factor exposures γ^s are centered and standardized over the core market part of the estimation universe. $\gamma_{i,l}^s$ is denoted as standardized factor exposure of stock i to factor l, and calculated as following:

$$\gamma^s{}_{il} = \frac{\gamma_{il}^{s(raw)} - \mu_l}{\sigma_l} \tag{3.12}$$

$$\mu_l = \sum_{i=1}^{N} w_i \gamma_{il}^{s(raw)}$$

$$\sigma_l = \sqrt{var(\gamma_l^{s(raw)})}$$

where $\gamma_{il}^{s(raw)}$ denotes the raw exposure of stock i o factor l, $w_i = MC_i / \sum_j MC_j$ is the market capital weight of stock i, μ_l is the cap-weighted average raw exposure of factor l, and σ_l is the equal-weighted standard deviation of all raw exposures of factor l.

- Missing data

In some cases that not all descriptor exposures are available for a stock, exposures are replaced by industry capitalization weighted average exposures of stocks with valid exposures in the same industry. This method to replace missing exposure is based on the notion that many industry exhibit a distinctive style signature. For example, technology industry contains many small companies with high fundamental growth and small value. Size exposures are not replaces as every stock in the universe of estimation must have size exposures, which means its market capitalization should be available.

3.4 Factor Returns

In this section, we first describe the methodology of estimate market factor exposures and fundamental factor returns. An R squared statistic and in-sample goodness-of-fit tests are provide after then. With the time series of factor returns, we compare different forecast models widely used in both academy and industry. Out-of-sample tests are provided in both individual stock level and portfolio level.

3.4.1 Factor returns estimation and in-sample tests

On a daily base, the exposure of market factor is obtained through a robust linear regression of stock returns against one index returns representing market factor returns. To be consistent, we still use the CSI 300 index as the market factor. The regression model for stock i from time t_0 to t has the form as following:

$$\begin{pmatrix} r_{i,t_0} \\ \vdots \\ r_{i,t} \end{pmatrix} = \widehat{\gamma^M} \begin{pmatrix} f_{t_0}^M \\ \vdots \\ f_{t_1}^M \end{pmatrix} + \begin{pmatrix} \hat{\varepsilon}_{i,t_0} \\ \vdots \\ \hat{\varepsilon}_{i,t} \end{pmatrix} \quad 3.13$$

where $(r_{i,t_0}, \dots, r_{i,t})'$ is the time series of daily return of stock i from time t_0 to t, $(f_{t_0}^M, \dots, f_t^M)'$ is the time series of daily return of the CSI 300 Index, $\widehat{\gamma^M}$ is the estimator of exposure of market factor, and $(\hat{\varepsilon}_{i,t_0}, \dots, \hat{\varepsilon}_{i,t})'$ is the residuals left.

Also we can rewrite **(3.13)** in a matrix form and define the daily excess return of stocks $\tilde{r}$ as following:

$$\tilde{\boldsymbol{r}} = \boldsymbol{r} - \gamma^M \boldsymbol{f}^M \quad 3.14$$

Fundamental factor returns are estimated by regressing daily excess returns against the style and industry factor exposures. The cross-sectional regression can be described as a geometric projection; it projects the excess returns from the N-deminsional asset space to the much smaller K-dimensianl factor space. The cross-sectional regression model at a time t can be described as following:

$$\tilde{\boldsymbol{r}}_t = \tilde{\boldsymbol{\Gamma}}_{t-1}\hat{\boldsymbol{f}}_t + \hat{\boldsymbol{e}}_t \quad 3.15$$

where $\tilde{\boldsymbol{r}}_t = (\tilde{r}_{1,t}, \dots, \tilde{r}_{n,t})'$ is the vector representing stock returns at a time t; $\tilde{\boldsymbol{\Gamma}}_{t-1} = \begin{pmatrix} \gamma_{1,1,t-1} & \cdots & \gamma_{1,k-1,t-1} \\ \vdots & \ddots & \vdots \\ \gamma_{n,1,t-1} & \cdots & \gamma_{n,k-1,t-1} \end{pmatrix}$, is the matrix representing style and industry factor exposures calculated at time $t-1$;

$\hat{\boldsymbol{f}}_t = (\hat{f}_{1,t}, \dots, \hat{f}_{k-1,t})'$ is the vector containing factor return estimators; and $\hat{\boldsymbol{e}}_t = (\hat{e}_{1,t}, \dots, \hat{e}_{n,t})'$ is the vector containing specific return estimators.

For cross-sectional regression, we use *ordinary least squares* (OLS) method, where the residuals are not assigned with any distribution assumption but only assumed to be uncorrelated and have finite variance. Thus we assume that $E(\boldsymbol{e}) = 0, E(\boldsymbol{ee}') = \sigma^2\boldsymbol{I}$, where σ^2 is the variance of the residuals and I is the identity matrix; and $\tilde{\boldsymbol{R}}$ is distributed independently of the residual $\boldsymbol{e}$. We estimate $\hat{\boldsymbol{f}}$ by minimize the sum of the squared residuals (SSR) and have $\hat{\boldsymbol{f}} = (\tilde{\boldsymbol{\Gamma}}'\tilde{\boldsymbol{\Gamma}})^{-\mathbf{1}}\tilde{\boldsymbol{\Gamma}}'\tilde{\boldsymbol{r}}$.

The assumption that σ, the standard deviation of the residual, is constant over all values of the explanatory variables. In other words, each data point in the regression provides equally information. This assumption, however, clearly does not hold, even approximately, in factor modeling. For example, large cap returns exhibit more common behavior than small cap returns and it makes sense to favor them in regression. *Weighted least squares* (WLS) can be used to maximize the efficiency of parametric estimation. This is done by attempting to give each data point its proper amount of influence over the parameter estimates. A procedure that treats all of the data equally would give less precisely measured points more influence than they should have and would give highly precise points too little influence.

WLS method reflects the behavior of the random errors in the model. It works by incorporating extra nonnegative constants, or weights, associated with each data point, into the fitting criterion. The size of the weight indicates the precision of the information contained in the associated observation. Optimizing the weighted fitting criterion to find the parameter estimates allows the weights to determine the contribution of each observation to the final parameter estimates. It is important to note that the weight for each observation is given relative to the weights of the other observations; so different sets of absolute weights can have identical effects.

In the CHE model, regression weights are set to $w_i = \frac{MC_i}{\sum_j MC_j}$, where MC is the market capitalization. The WLS criterion minimized to obtain the parameter estimators is:

$$Q = \sum_{i=1}^{n} w_i \left(\tilde{r}_i - \tilde{\boldsymbol{\gamma}}_{\boldsymbol{i}} \hat{\boldsymbol{f}}\right)^2 \quad 3.16$$

This problem can be transformed and solved in an OLS form as:

$$\sqrt{\boldsymbol{w}} \otimes \tilde{\boldsymbol{r}}_t = (\sqrt{\boldsymbol{w}} \otimes \tilde{\boldsymbol{\Gamma}}_{t-1}) \hat{\boldsymbol{f}}_t + \hat{\boldsymbol{e}}_t \quad 3.17$$

where $\boldsymbol{w} = (w_1, \ldots, w_n)'$ is the vector representing the weights in WLS, and $\otimes$ represents out products.

Next, we present and discuss some of the quantitative characteristics of the CHE factors. In particular, we investigate the degree of collinearity among factor exposures, and the statistical significance, performance and volatility of the factor returns

A feature of the multi-factor model is that it disentangles the effects of many variables acting simultaneously. One measure of collinearity is pair-wise cross-sectional correlations between factor exposures. In Table 3.2, we report regression-weighted correlations among style factors and industries, averaged over the period January 2004 to December 2012. In general, the correlations are intuitive in signs. For instance, *value* is positive correlated with *energy*, *financial*, *utilities* and negatively correlated with *basic materials* and *non-cyclical consumer*. *Volatility* is positively correlated with *liquidity* and negatively correlated with *size*. Although none of the correlations are particularly large, the correlations between pairs of style factors are typically larger, on average, than those between styles and industries.

Another frequently used measure of cross-sectional multi-collinearity is *Variance Inflation Factor* (VIF), which is defined as following:

$$VIF_i = \frac{1}{1 - R_j^2} \qquad 3.18$$

where R_j is the R squared by regressing factor exposure i to the rest of other factor exposures, and VIF_i is the VIF of factor exposure i.

Typically VIF greater than 5 is an indication for multi-collinearity. In Figure 3.2, we report the VIF of all 7 style factors from September 2005 to December 2012, as style factor exposures exhibit larger correlations than those of industry factor exposures.

The goodness of fit parameter, R^2, of the factor regression indicates the percentage of the variance of stock returns that can be explained by the model.We adjust the formula from normal R^2 in OLS to follow the estimation method in WLS.

Figure 3.3 presents the trailing one-year average R^2 of CHE. Results are obtained on the full CHE estimation university and use regression weights. The solid lines indicates that the full model has an averaged R^2 of 54% over the 8-year backtesting period. R^2 peaks in the period of high market volatility such as the recent financial crisis starting in 2008. The market recovering from the bottom in 2009 is characterized by lower volatility. In this period the model has a lower average R^2, indicating that a smller percentage of the stock volatility can be attributed to common factors. In addition, the second peak in 2010 corresponds to the end of the short market recovery and a continuous bear market from middle 2010 to 2012.

The additional R^2 time series in Figure 3.3 (a) gives valuable insight into the in-sample relevance of different factor categories. The dotted line indicates the explanatory power of the CHE market factor alone, with an average R^2 of 17%. While this indicated that the market factor has significant power in explaining, it also shows that a single market model misses a lot of systematic factor risk compared to the full CHE model. The dashed line shows the explanatory power given industry factors and the market factor. It is clear that the industry factors have significant in-sample explanatory power with an increment of 24% of the average R^2. Finanlly the style factors add 7% to the total average R^2.

Isolating the explanatory power of style factors and industry factors helps to understand the changing investment characteristics of the market. The solid line in Figure 3.3 (b) is the time series of isolated R^2 given by style factors and the dashed line is the isolated R^2 given by industry factors. The R^2 by style factors peaks in 2007 and declines to a relatively constant level after the financial crisis in 2008, while the R^2 by industry factors drops in 2007 and grows to a highest level in 2009. These together indicating that in the period of market bubble investigating the financial profile of the companies is more important than choosing the industries to the investors, and in the period of post financial crisis, choosing the right industries becomes more important and attractive as different industries having different time for recovering. This change of investment characteristics of the market becomes more obvious in 2012.

3.5 Measuring Risk with Factor Model and Backtests

3.5.1 Risk Measures and Approaches to Measure Risk

Before presenting the forecasting methodology in CHE model, it is helpful to discuss some challenges in quantitative risk management (QRM) (see McNail, et al. 2005). A very important challenge in QRM is the need to address unexpected, abnormal or extreme outcomes, rather than expected normal or average outcomes that are the focus of many classical applications. A further important challenge is presented by multivariate nature of risk. A particular concern in the multivariate modeling is the phenomenon of dependence between extreme outcomes, when many risk factors move simultaneously. Another challenge is the typical scale of the portfolio under consideration, in the most general case a portfolio may represent the entire position in risky assets of a financial institution. Calibration of high-dimensional multivariate model is a well-nigh impossible task and hence dimension reduction techniques are required. All these challenges must be considered carefully when modeling risk in QRM.

Central issues in QRM include the measurement of risk, approaches for measuring risk and backtesting the performance of risk-measurement system.

In the CHE model, we model the return of asset value in a short time horizon with a linear function with the risk factor returns by **(3.1).** The next question arises as modeling factor returns in conditional or unconditional return distributions. The differences between conditional or unconditional return distributions are strongly related to time series properties of the series of risk factor returns $(\boldsymbol{f_t})_{t\in N}$. Suppose risk factor returns form a stationary time series with some stationary distribution, and for a fixed point in time t, denote by $\mathcal{F}_{\mathrm{t}} = \sigma(\{\boldsymbol{f}_s: s < t\})$, the sigma field generated by past and present factor returns. In most stationary time series models relevant to QRM, the distribution of $\boldsymbol{f_{t+1}}|\mathcal{F}_t$ is not equal to the distribution of $\boldsymbol{f}$, unless $(\boldsymbol{f_t})_{t\in N}$ are iid series. In CHE model, we observe that the $(\boldsymbol{f_t})_{t\in N}$ are not iid series, hence we model the conditional distribution of factor returns and further the conditional distribution of stock returns.

Historically the variance of asset return distribution has been the dominating risk measure in finance, to a large extent, due to the huge impact of the portfolio theory of Makowitz (see, for example, Markowitz 1952). Variance is a well-understood concept which is easy to use analytically. However, as a risk measure it has at least two drawbacks. On one side, the assumption that the second moment of return distribution exists may not be guaranteed in practice. For instance, modeling a heavy-tailed r.v.s with student *t* distribution requires the degree of freedom larger than 2. On the other side, variance makes not distinction between positive and negative deviations from the mean, which makes variance as a good risk measure only in the case of symmetric distribution assumptions of asset returns.

Value-at-Risk (VaR) is probably the most widely used risk measures in financial institutions nowadays and has made its way into the Basel II capital-adequacy frame work. The idea is straight stated as “maximum loss which is not exceed with a given high probability, the so-called confidence level”. In probability terms, VaR is simply a quantile of the return distributions. Here we give an example of VaR for Gaussian distributions

Suppose that the return distribution F_X is Gaussian with mean μ and variance σ^2, then

$$\mathrm{VaR}_\alpha = -\mu + \sigma\Phi^{-1}(\alpha) \qquad 3.19$$

where α is the confidence level and $\Phi^{-1}(x)$ is the inverse distribution function of standard Gaussian.

Average-Value-at-Risk (AVaR), also named expected shortfall (see Artzner et al. 1997, 1999) or conditional VaR (see Rockafellar and Uryasev, 2000), is closely related to VaR. Instead of fixing a particular confidence level α we average VaR over all levels $u \geq \alpha$ and thus “look further into the tail” of the return distributions. Formal definition is discussed in Section 2.4. Here, again, we only give an example of AVaR for Gaussian distributions.

$$\mathrm{AVaR}_\alpha = -\mu + \sigma\frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha} \qquad 3.20$$

where ϕ is the density of standard Gaussian distribution.

Next we discuss several approaches to measure risk in CHE model.

Firstly, we model the factor returns with conditional multivariate distributions as the conditional distributions are not equal to their stationary distributions especially for a small time horizon, such as one day. Thus the stock returns are also modeled with conditional distributions, which may be the same as the distributions of factor returns, depending on whether these distributions are closed under linear operations.

There three methods in the CHE model to measure factor returns, which are:

- VC-EWMA: A conditional version of the variance-covariance method in which the multivariate exponentially weighted moving average (EWMA) method is used to model the conditional covariance matrix of the next day's risk factor returns.
- VC-MGARCH: A conditional version of the variance-covariance method in which the multivariate GARCH with constant conditional correlation method is used to model the conditional covariance matrix of the next day's risk factor returns.
- MC-MGARCH: A dynamic version of the Monte Carlo method in which the multivariate ARMA-GARCH with constant conditional correlation method is used to model the conditional distribution of the next day's risk factor returns.

In VC-EWMA and VC-MGARCH method, we assume that given the information until today, tomorrow's distribution of factor returns follows a multivariate Gaussian distribution, denoted by $\boldsymbol{f}_{t+1}|\mathcal{F}_t \sim N_k(\boldsymbol{\mu}_{t+1}, \boldsymbol{\Sigma}_{t+1})$, where $\boldsymbol{\mu}_{t+1}$ and $\boldsymbol{\Sigma}_{t+1}$ are the conditional mean and covariance matrix. Thus the conditional distribution of stock returns is also Gaussian with additional Gaussian assumption for the specific returns. Then, given a portfolio with weights $\boldsymbol{W} = (w_1, \dots, w_n)'$, the tomorrow's variance, VaR and AVaR of the portfolio have the form as following:

$$\mathrm{Var}\left(R_{p,t+1}|\mathcal{F}_t\right) = \boldsymbol{W_t}(\boldsymbol{\Gamma_t}\boldsymbol{\Theta_{t+1}}\boldsymbol{\Gamma_t}^{\boldsymbol{T}})\boldsymbol{W_t}^{\boldsymbol{T}} + \boldsymbol{W_t}\boldsymbol{\Omega_{t+1}}\boldsymbol{W_t^T} \qquad 3.21$$

where $\boldsymbol{\Theta_{t+1}}$ is a *k*-by-*k* matrix representing the conditional covariance matrix of factor returns and $\boldsymbol{\Omega_{t+1}}$ is a *n*-by-*n* diagonal matrix with elements representing the conditional variance of specific returns, of which the estimation method is presented in the end of this section.

It is easy to calculate the VaR and AVaR at a confidence level α for the portfolio of which the return follows a Gaussian distribution.

$$\mathrm{VaR}_\alpha(R_{p,t+1}|\mathcal{F}_t) = -\mu_{t+1} + \sigma_{t+1}\Phi^{-1}(\alpha) \qquad 3.22$$

and

$$\mathrm{AVaR}_\alpha(R_{p,t+1}|\mathcal{F}_t) = -\mu_{t+1} + \sigma_{t+1}\frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha} \qquad 3.23$$

where μ_{t+1} is the portfolio's forecasted return, which can be calculated separately with other return forecast model, such as an ARMA(*p*,*q*) model, or simply assumed to be zero for in the risk management world. σ_{t+1} is the portfolio's forecasted standard deviation.

In VC-EWMA method, we forecast $\boldsymbol{\Theta_{t+1}}$ with a EWMA method, which is also adapted by BARRA equity factor with some modification of the general form.

$$\theta_{k,l}^{t+1} = cov(f_k, f_l)_{t+1} \tag{3.24}$$

$$= \frac{m(1-\lambda)}{(m+1)(1-\lambda^m)} \sum_{s=t-m}^{t} \lambda^{t-s}(f_{k,s} - \overline{f_k})(f_{l,s} - \overline{f_l})$$

where m denotes the sample size and λ denotes the exponential weight, which reduces the influence of past observations on the present forecast. Here we set $\lambda = 0.99$.

In VC-MGARCH method, we estimate the time series of factor returns with ARMA-MGARCH model and use the corresponding forecasting method. An ARMA(1,1)-MGARCH(1,1) model for the factor returns has the following form:

$$\boldsymbol{f}_{t+1} = \boldsymbol{c} + \boldsymbol{\phi}\boldsymbol{f}_t + \boldsymbol{\theta}\boldsymbol{\varepsilon}_t + \boldsymbol{\varepsilon}_{t+1} \tag{3.25}$$

$$\boldsymbol{\varepsilon}_{t+1} = \boldsymbol{\Sigma}_{t+1}^{\frac{1}{2}}\boldsymbol{Z}_{t+1}$$

$$\boldsymbol{\Sigma}_{\mathrm{t+1}} = \boldsymbol{\Delta}_{t+1}\boldsymbol{P}_c\boldsymbol{\Delta}_{t+1}, \quad \boldsymbol{\Delta}_{t+1} = \mathrm{diag}(\boldsymbol{\sigma}_{\mathrm{t+1}}^2)$$

$$\boldsymbol{\sigma}_{\mathrm{t+1}}^2 = \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_{1,i}\boldsymbol{\varepsilon}_t + \boldsymbol{\beta}\boldsymbol{\sigma}_{\mathrm{t}}^2$$

where

$\boldsymbol{f}_{t+1}$	=	k-dimensional vector of factor returns at time $t+1$;
$\boldsymbol{c}, \boldsymbol{\phi}, \boldsymbol{\theta}$	=	k-dimensional vector of ARMA(1,1) coefficients;
$\boldsymbol{\alpha_0} > \mathbf{0}$ $\boldsymbol{\alpha_1}, \boldsymbol{\beta} \geq \mathbf{0}$	=	k-dimensional vector of GARCH(1,1) coefficients;
$\boldsymbol{P}_c$	=	$k \times k$ constant, positive definite correlation matrix;
$\boldsymbol{\Delta}_{t+1}$	=	$k \times k$ diagonal matrix with elements $\boldsymbol{\sigma}_{\mathrm{t+1}}^2$ follow the univariate GARCH model;
$\boldsymbol{\Sigma}_{t+1}^{\frac{1}{2}}$	=	$k \times k$ Cholesky factor of a positive-definite matrix $\boldsymbol{\Sigma}_{t+1}$, which is measurable with respect to $\mathcal{F}_t$; and
$\boldsymbol{Z}_{t+1}$	=	$(\mathbf{Z}_s)_{s\in\mathbb{Z}}$ is the SWN$(\mathbf{0}, \boldsymbol{I}_k)$ process.

It is easy to calculate the conditional mean and covariance matrix of $(\boldsymbol{f}_{t+1}|\mathcal{F}_t)$:

$$\mathrm{E}(\boldsymbol{f}_{t+1}|\mathcal{F}_t) = c + \boldsymbol{c} + \boldsymbol{\phi}\boldsymbol{f}_t + \boldsymbol{\theta}\boldsymbol{\varepsilon}_t \tag{3.26}$$

$$\mathrm{cov}(\boldsymbol{f}_{t+1}|\mathcal{F}_t) = \boldsymbol{\Sigma}_{\mathrm{t+1}} = \boldsymbol{\Delta}_{t+1}\boldsymbol{P}_c\boldsymbol{\Delta}_{t+1} \tag{3.27}$$

In CHE model, the ARMA(1,1)-MGARCH(1,1) model is fitted with a two-stage method, in which the individual factor return process is estimated by fitting univariate ARMA(1,1)-GARCH(1,1) model using MLE. In the second stage we construct an estimate the innovation process denoted by $\widehat{\boldsymbol{Y}}_t = \widehat{\boldsymbol{\Delta}}_t^{-1}\boldsymbol{\varepsilon}_t$, where $\widehat{\boldsymbol{\Delta}}_t$ is the estimate of $\boldsymbol{\Delta}_t$. Assuming the adequacy of the model, the $\widehat{\boldsymbol{Y}}_t$ data should behave like a realization from a standard $\mathrm{WN}(\boldsymbol{0}, P_c)$ process, and the conditional correlation matrix $\boldsymbol{P}_c$ can be estimated from $\widehat{\boldsymbol{Y}}_t$ with sample correlation estimate $\widehat{\boldsymbol{P}}_c$.

The conditional variance of specific returns, denoted as $\boldsymbol{\Omega}_{\mathbf{t+1}}$ are forecasted with univariate GARCH(1,1) model, which is simpler compared to forecast of factor returns. Then, together with the forecasted conditional covariance matrix of factor returns, the VaR and AVaR of a portfolio are calculated with (3.23) and (3.24).

To our knowledge of finance and previous empirical examination of the Chinese stock market (see Lu et al. 2013), we know that the Gaussian distribution hypothesis of stock returns are generally rejected in neither (1) unconditional homoscedastic distribution assumption or (2) conditional heteroscedastic distribution assumption. Thus model risk would arise if the factor returns are assumed to follow the multivariate Gaussian distribution.

To answer the challenges in QRM discussed at the beginning of Section 3.5.1, we suggest a state-of-art methodology for modeling the behavior of stocks in the Chinese and possible extension to other stock markets.

To better capture the non-Gaussian characteristics of stock return distributions, such as heavy-tail, skewness and kurtosis, the ARMA-GARCH model with non-Gaussian distributed innovations, especially Classical Tempered Stable (CTS) distributed innovations, is suggested in Chapter 2, with both in-sample and out-of-sample tests**.** Hence, the CHE model adopts the same method in modeling factor returns, and finally the stock returns through linearized operator.

In MC-MGARCH method, the factor return process are fitted with CTS-ARMA-MGARCH model and the risk forecasts are calculated by Monte Carlo simulation.

ARMA-MGARCH-CTS is defined as following:

$$
\begin{aligned}
&\boldsymbol{f}_{t+1} = \boldsymbol{c} + \boldsymbol{\phi}\boldsymbol{f}_t + \boldsymbol{\theta}\boldsymbol{\varepsilon}_t + \boldsymbol{\varepsilon}_{t+1} \\
&\boldsymbol{\varepsilon}_{t+1} = \boldsymbol{\Delta}_{t+1}\boldsymbol{u}_{t+1}, \\
&\boldsymbol{\Delta}_{t+1} = \mathrm{diag}(\boldsymbol{\sigma}_{t+1}^2) \\
&\boldsymbol{\sigma}_{t+1}^2 = \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_{1,i}\boldsymbol{\varepsilon}_t + \boldsymbol{\beta}\boldsymbol{\sigma}_t^2
\end{aligned}
\tag{3.28}
$$

where innovations, denoted by $\boldsymbol{u}_{t+1} \in (\boldsymbol{u}_s)_{s\in\mathbb{Z}}$**,** is the $\mathrm{SWN}(\boldsymbol{0}, \boldsymbol{\Sigma})$ process, and other notations remains the same as in (3.26).

Consider a fixed point at time t, $\boldsymbol{u}_t$ follows a *k*-dimensional multivariate distributions, denoted as CTS-skewedt distribution, with margins of the CTS distributions and dependence structure described by the skewed *t* copula. Thus, the extreme outcomes and asymmetry of individual factor returns are estimated through marginal distributions and the different tail dependences are estimated by the copula instead of a simple linear correlation.

The parameters of the CTS-ARMA-MGARCH model are estimated through a two-stage method known as the quasi-maximum likelihood (QML) method; first we fit the ARMA-GARCH part with MLE, assuming univariate student *t* distributed innovations. In the second stage, we fit individual series of the extracted $\boldsymbol{u} = \boldsymbol{\varepsilon}/\boldsymbol{\sigma}$ with standard CTS distribution and the entire sample series of $\boldsymbol{u}$ with skewed *t* copula.

Next we generate scenarios, typically 10,000 for daily forecast, from CTS-ARMA-MGARCH model by Monte Carlo simulation. This process can be viewed as an inverse process of model estimation.

Algorithm (simulation of MC-MGARCH)

(1) Generate $\boldsymbol{U}^{(s)} = \left(\boldsymbol{U}_1^{(s)}, \ldots, \boldsymbol{U}_k^{(s)}\right)'$, and $1 \leq s \leq 10{,}000$ from *k*-dimensional skewed *t* copula;

(2) Generate the $\boldsymbol{u}^{(s)} = \left(F_{CTS}^{-1}\left(\boldsymbol{U}_1^{(s)}\right), \ldots, F_{CTS}^{-1}(\boldsymbol{U}_k^{(s)})\right)'$, where F_{CTS}^{-1} is the inverse distribution function of CTS distribution. The random vector $\boldsymbol{u}^{(s)}$ has the margins of CTS distribution and skewed *t* copula.

(3) Return $\boldsymbol{f}^{(s)} = \left(\Psi\left(\boldsymbol{u}_1^{(s)}\right), \ldots, \Psi\left(\boldsymbol{u}_k^{(s)}\right)\right)'$, where Ψ is the function of ARMA-MGARCH with estimated parameter.

With generated scenarios, it is easy to calculate variance, VaR and AVaR for either individual stock or portfolios. A practice is given in Chapter 4.

For the assessment of the goodness-of-fit, we use the Kolmogorov-Smirnov (KS) test and the Anderson-Darling (AD) test for the hypotheses of marginal distribution of innovations. We use the later statistic in order to obtain a better test to evaluate the tail fit. Similar work has been done for returns of stock in the Chinese stock market in Chapter 2.

Table 3.3 provides the test results using the standard KS test with different significance levels. The Gaussian hypothesis is rejected by KS test at a significance level of 5% for fifteen of eighteen factors, and twelve of eighteen factors at a smaller significance level of 1%. The α-stable hypothesis is rejected for twelve factors at a significance level of 5%, and ten factors at a significance level of 1%. While the CTS hypothesis is not rejected for any factor by KS test at a significance level of 1%.

Figure 3.4 uses *growth* and *basic material* factors as examples, and presents a more intuitive view of the distributions of innovations by comparing the sample distribution of the innovations against the standard Gaussian standard α-stable and standard CTS distributions. The Q-Q plots for *growth* and *basic material* factors indicate that Gaussian hypothesis and α-stable hypothesis are rejected by goodness-of-fit test for different reasons. For instance, the tail parts indicate the Gaussian distribution underestimates the extreme outcomes from the sample while α-stable distribution overestimate the extreme outcomes by generating excessive extreme scenarios. Comparing with Gaussian andα-stable distributions, the samples studied here are better fitted with CTS distribution.

As we mentioned in Section 3.3, modeling the specific returns $\boldsymbol{e}$ is of second importance as the risk embodied in $\boldsymbol{e}$ can be mitigated through diversification in the Markowitz portfolio framework. In addition, each series of specific returns $\boldsymbol{e}$ can be estimated by univariate time series model, due to its idiosyncratic character. In VC-EWMA and VC-MGARCH method, specific returns are fitted by standard ARMA(1,1)-GARCH(1,1) model; in MC-MARCH method, they are fitted by CTS-ARMA(1,1)-GARCH(1,1) model.

The challenge with using time series estimation is that not all stocks in a broad universe lend themselves to this modeling approach. For instance, recent IPOs have only a short history of specific returns, which implies that their sample volatility has a high estimation error. Another example would be the stocks of recent resumption.

In the CHE model, specific returns are divided into three categories with respect to the length of their time series: long-history stocks with more than 212 (85% of the 250) last daily specific returns present; short-history stocks with less than 212 but more than 60 last daily specific returns present; shortest-history stocks with less than 60 last daily specific returns present.

Long-history specific returns are fitted directly from ARMA-GARCH models. Short-history and shortest-history specific returns are treated as iid r.v.s from $N(0,\sigma^2)$, where σ in the former case is estimated by sample standard deviation for the former case while in the latter case it is estimated by the factor model. Standard deviation estimates for long-history specific returns in one industry are regressed against their *size* factor exposures, and the regression coefficient is used in calculating the σ estimate for shortest-history specific returns in the same industry.

3.5.2 Backtests

This section discusses the characteristics of the CHE risk factor forecasts. It provides insight into differences among three forecasting methods and overall forecasting capability of the CHE model.

Bias tests are a standard method of assessing the out-of-sample performance of risk models. Essentially, bias tests determine how often the risk forecast of a test item (asset, factor, or portfolio) falls within a 95% confidence interval (CI) over a back-test period.

The z-scores z_t and bias statistics $b_{t,T}$ over a time horizon T, is defined as following:

$$\mathrm{z_t} = \frac{r_t}{\hat{\sigma}_t} \quad 3.29$$

$$b_{t,T} = \left(\frac{1}{T-1}\sum_{s=t-T+1}^{t}(z_s - \bar{z})\right)^{1/2} \quad 3.30$$

where r_t is the realization of return at time t, $\hat{\sigma}_t$ is the forecasted standard deviation at $t-1$; then z_t is standardized return with forecasted risk.

If forecasts are perfect, $(z_t)_{t\in\mathbb{Z}}$ have a standard deviation of one. A standard deviation below one is indicative of over-forecasting and a standard deviation above one indicates under-forecasting. Under simplifying norm assumption of return r_t, a 95% confidence interval can be given for the basic statistic:

$$C_T = \left[1 - \sqrt{\frac{2}{T}}, 1 + \sqrt{\frac{2}{T}}\right] \quad 3.31$$

This means that, with perfect forecasts and normal returns, 95% of a broad group of bias statistic value fall within the confidence interval C_T.

From (3.32) it can be seen that the confidence interval is determined by the time horizon T. In this backtesting, T is set to be 22 days, which corresponds to a confidence interval of $[0.6984, 1.3015]$. Though increasing T leads to narrowing C_T, calculating bias statistic over a long time horizon does not necessarily increase accuracy. Instead, a phase where the model consistently over-forecast can be followed by another phase where the model under-forecasts, producing an average deceptively close to one. In practice T should be specified according to the investor's interest.

A main disadvantage of a short time horizon is a high sensitive to outliers. For instance, either a large positive return or large losses r_t in (3.31) results in an outlier z_t with a large absolute value, and all bias statistics in (3.32) containing this outlier will be out of confidence irrespective of the value of the other z-scores used. This shortcoming is related to the assumption of normally distributed returns in (3.33). If it is possible to precisely model the size and frequency of outliers, the confidence interval could be stretched accordingly to take into account the outlier distribution. An alternative way suggested by Barra EUE3 to limit influence of outliers is by truncating the z-scores:

$$\tilde{z}_t = \max(-3, \min(+3, z_t)) \qquad 3.32$$

Then robust statistics are calculated with truncated z-scores $\tilde{z}_t$. This truncation approach probes the ability of the model to predict the core of the return distribution and mitigates the influence of outliers to the testing results. In spite of their shortcomings, raw bias statistics do reflect the outliers that should not be ignored by investors. In addition, the difference between raw bias statistics and robust statistics can provide insight about the severity and frequency of outliers. Notice that a covariance model cannot predict the timing and size of future outliers and tend to under-forecast when an outlier is observed.

Bias tests are applied for stock returns on both individual stock level and portfolio level. The test universe is stocks with valid *price* and *size* factor exposures on the first trading day in 2007, such that the universe is not too small to represent the market and the testing period could cover the highly volatile market periods as well. A universe portfolio is constructed with equally weighted stocks of the entire test universe.

Table 3.4 provides the detailed bias tests results over the testing period from 2007 to 2012. In Table 3.4(a) the 22-days rolling raw bias statistics and robust bias statistics based on three different forecasting methods are compared by the percentage of those in the same CI. First looking at testing results of robust bias statistics, the VC-MGARCH method has the largest percentage in CI, followed by MC-MGARCH and then VC-EWMA. Although the difference of percentage in CI are not significant among three methods, it indicates that the VC-MGARCH method, which is the combination of EWMA estimate for correlation matrix and GARCH estimate for individual variance, performs best in forecasting portfolio's variance among three studied methods. Moreover, the model tends to over-forecasting more frequently than under-forecasting regardless of the forecasting methods. The differences between pair values of robust and raw bias statistics indicate influence of outliers which cannot be predicted by a variance model.

A close look of bias statistics on the individual stock level presents more obvious differences of percentage of raw bias statistics in certain intervals among different forecasting methods. Simply calculating the percentages in the intervals $(-\infty, 0.9]$ and $[1.1, \infty)$, the CHE factor model tends to more frequently over-forecasting than under-forecasting for the variance of individual stocks. More specifically, MC-MGARCH method gives a larger percent of over-forecasting than the VC-MGARCH method, due to the different estimation methods of conditional correlation of factor returns.

Figure 3.5 present a more intuitive view of the bias statistics calculated from VC-EWMA and MC-GARCH method. An interesting phenomenon shows that the bias statistics from the two methods diverges increasingly after 2010. Together with Figure 3.6 which plots the forecasted volatility by VC-EWMA and MC-MGARCH methods and realized volatility, we may conclude that the two methods performs differently in predicting variance in a relatively stable market time period while they are more consistent in a relatively volatile market time period. This may indicate a regime switch of dependence structure in the market of different time periods.

A major disadvantage of variance or standard deviation as a risk measure is that it penalize symmetrically both the negative and positive deviations from the mean. Instead of the classical bias test based on a variance model, we consider VaR and AVaR backtests as alternatives. Here we use the joint Christoffersen Likelihood Ratio (CLR) test (Christoffersen, 1998) and the Berkowitz Likelihood Ratio (BLR) test (Berkowitz, 2001) as the same as in Chapter 2. The CLR test accounts for the conditional coverage of the VaR measure, assuming the violations of VaR for a given period follow an iid Bernoulli distributed random variables. And, loosely speaking, the BLR test compares the shape of forecasted tail of density of risk factors to the observed tail. The time periods considered in backtesting are one-year observations from 2007 to 2011.

Table 3.5 and Table 3.6 provide CLR test statistics and corresponding *p*-values for the three forecasting methods in the unconditional coverage tests, independence tests and the joint tests, and VaR are calculated at 95% and 99% confidence level, respectively. In Table 3.5, the MC-MGARCH method is rejected at a 1% significance level in the unconditional tests for 2009 and 2010; none of the three methods is rejected at a 1% significance level in the independence tests or the joints for all the years. In Table 3.6, none of the three methods is rejected at a 1% significance level in all tests. Although the number of violations given by the MC-MGARCH method is smaller than that given by the VC-EWMA method, the differences are not significant in general.

Table 3.7 and Table 3.8 provide BLR test statistics and corresponding *p*-values with AVaR calculated at 95% and 99% confidence level, respectively. Table 3.7 shows that in the BLR tail test VC-MGARCH method and MC-MGARCH are rejected at a 1% significance level for most years except 2008 and 2011, respectively; while VC-EWMA method is only rejected for 2007 and 2009. Furthermore, none of the three methods is rejected at a 1% significance level in BLR independence test for all years. Compared to Table 3.7, Table 3.8 shows fewer rejections at a 1% significance level. In the BLR tail tests, MC-MGARCH method is rejected for 2007 and 2010; VC-MGARCH method is rejected for 2009 and 2010; and VC-EWMA method is rejected for 2007 and 2009. And none of the three methods are rejected in the BLR independence tests.

From the VaR backtest results we may conclude that the CHE factor model has a strong capability in describing the time dependence structure as well as the tail behaviors of the stock returns.

3.6 Conclusions

In this chapter, we provided the CHE factor model for risk and portfolio management. Developed for modeling the entire Chinese stock markets, this factor model includes 17 specified factors from the market and companies financial profiles. Different time series models have been developed for describing risk factor returns and residuals. More specifically, a multi-dimensional AMRM-GARCH process with innovations following a heavy-tailded and skewed multivariate distribution has been applied for the factor returns and the one-dimensional ARMA-GARCH process has been applied for the individual series of residuals from the factor models.

Sophisticated in-sample and out-of-sample statistical tests have been applied for model validation. And the test results suggested the CHE model with strong capacity in describing and forecasting the risk of the Chinese stock markets.

Table 3-1 The definition of all descriptors and formula of calculating style factor exposures.

Factor	Formulation	Description
Size	$= \ln MC$	Market Capitalization – last month average
Value	$= 1/PB$	Book-to-Price ratio – last month
Growth	$= (1 - \frac{dps}{eps})roe$	Dividends per share – last year trailing Earnings per share – last year trailing Income before extraordinary income – last year trailing Common equity value – last year
Momentum	$= \frac{Mom(\epsilon_{eq})}{Mom(r_m)}$ $Mom(r) = \prod_{i=1}^{n}(1 + r_i)$	Stock price time-series for one year Market time-series for one year
Liquidity	$= \frac{ln\,TV}{lnMC}$, $TV = \frac{1}{n}\sum_{i=0}^{n-1} PX_{t-i}V_{(t-i)}$	Traded Volume time-series – last 20 days Market Capitalization – last 20 days
Volatility	$= \frac{high - low}{high + low}$	Stock price time-series for one month
Leverage	$= \frac{debt}{MC}$	Long-term Debt – most recent report data Short-term debt – most recent report data Market Capitalization – last month average

Table 3-2 Regression-weighted cross-sectional correlation of style and industry factor exposures. Results are averages over the period January 2004 to December 2012. Correlations above 0.10 in absolute value are shaded in gray.

Factor	Size	Value	Growth	Momentum	Liquidity	Volatility	Leverage
Size	1.000	-0.255	-0.002	0.071	-0.898	-0.354	0.114
Value	-0.255	1.000	0.226	-0.365	0.240	0.265	-0.365
Growth	-0.002	0.226	1.000	-0.204	-0.016	0.047	-0.107
Momentum	0.071	-0.365	-0.204	1.000	-0.099	-0.163	0.110
Liquidity	-0.898	0.240	-0.016	-0.099	1.000	0.436	-0.098
Volatility	-0.354	0.265	0.047	-0.163	0.436	1.000	-0.119
Leverage	0.114	-0.365	-0.107	0.110	-0.098	-0.119	1.000
Basic Materials	0.497	-0.130	-0.007	0.019	-0.455	-0.130	-0.055
Communications	-0.022	0.030	0.015	-0.010	-0.007	0.000	-0.043
Consumer, Cyclical	0.051	-0.034	0.045	-0.017	-0.086	-0.020	0.144
Consumer, Non-Cyclical	0.296	-0.133	0.039	0.017	-0.314	-0.153	0.133
Diversified	-0.007	0.003	0.016	0.002	0.016	0.007	0.091
Energy	-0.016	0.065	-0.001	-0.027	0.000	0.001	-0.035
Financial	0.095	0.119	0.089	-0.063	-0.091	0.000	-0.031
Industrial	0.203	-0.069	0.003	0.027	-0.240	-0.078	0.062
Technology	-0.014	-0.046	0.018	-0.001	0.003	-0.011	-0.010
Utilities	0.049	0.151	0.135	-0.102	-0.066	0.020	-0.091

Table 3-3 Summary of the statistics of the various goodness-of-fit tests of univariate ARMA-GARCH models with different distributed innovations for factor returns from January 2005 to December 2012. P-values less than 0.05 are shaded and those less than 0.01 are bolded.

	Gaussian			Stable			CTS		
	p-value	ks	ad	p-value	ks	ad	p-value	ks	ad
Market	**0.000**	0.059	0.069	**0.000**	0.072	0.088	0.468	0.022	0.019
Size	0.308	0.030	0.048	0.339	0.030	0.025	0.418	0.028	0.026
Value	0.503	0.026	0.033	0.105	0.038	0.050	0.586	0.024	0.024
Growth	**0.002**	0.058	0.086	**0.006**	0.054	0.053	0.868	0.019	0.022
Momentum	0.012	0.050	0.055	**0.004**	0.056	0.064	0.804	0.020	0.024
Turnover	0.280	0.031	0.037	0.183	0.034	0.047	0.930	0.017	0.019
Volatility	0.049	0.043	0.046	0.143	0.036	0.037	0.982	0.015	0.020
Leverage	**0.006**	0.054	0.071	0.154	0.036	0.042	0.447	0.027	0.026
Basic Materials	**0.001**	0.060	0.074	0.010	0.052	0.064	0.325	0.030	0.027
Communications	**0.006**	0.054	0.063	0.031	0.045	0.055	0.065	0.041	0.029
Consumer, Cyclical	**0.004**	0.056	0.069	**0.000**	0.066	0.073	0.444	0.027	0.026
Consumer, Non-Cyclical	**0.000**	0.070	0.080	**0.000**	0.079	0.093	0.534	0.025	0.026
Diversified	0.031	0.045	0.057	0.259	0.032	0.036	0.950	0.016	0.016
Energy	**0.001**	0.064	0.080	**0.000**	0.076	0.099	0.200	0.034	0.039
Financial	**0.001**	0.064	0.075	**0.003**	0.057	0.054	0.832	0.020	0.022
Industrial	**0.003**	0.058	0.069	**0.000**	0.064	0.069	0.451	0.027	0.028
Technology	**0.001**	0.064	0.077	**0.002**	0.058	0.067	0.640	0.023	0.026
Utilities	**0.000**	0.066	0.079	**0.001**	0.061	0.071	0.525	0.026	0.028

Table 3-4 Bias tests results of the universe portfolios and single stocks from January 2007 to January 2012. (a) 22-days rolling bias statistics for universe portfolios based on different forecasting methods. Table shows percentage of robust bias statistics and raw bias statistics (in brackets) in confidence. (b) Entire raw bias statistics for single stocks in the universe portfolios based on different forecasting methods. Table shows the percentage of bias statistics in different interval.

(a)

Universe Portfolio	VC-EWMA		VC-MGARCH		MC-MGARCH	
over-forecast	5.63%	(5.27%)	4.72%	(4.38%)	4.97%	(4.52%)
in-forecast	91.25%	(88.30%)	92.00%	(89.02%)	91.32%	(88.70%)
under-forecast	3.12%	(6.43%)	3.28%	(6.60%)	3.71%	(6.68%)

(b)

Single Stocks	VC-EWMA	VC-MGARCH	MC-MGARCH
$(-\infty, 0.7)$	20.49%	19.46%	18.28%
[0.7,0.9)	32.96%	33.15%	56.33%
[0.9,1.1)	28.24%	28.24%	17.33%
[1.1,1.3)	10.36%	11.00%	5.64%
$[1.3, +\infty)$	7.75%	7.95%	2.22%

Table 3-5 Summary of Christoffersen Likelihood Ratio tests of the universe portfolio for 5 years. VaR is calculated at the confidence level of 95%.

Model	2007(January 4, 2007 – December 28, 2007)				2008 (January 2, 2008 – December 31, 2008)			
	N	CLR_{uc} (p-value)	CLR_{ind} (p-value)	CLR_{cc} (p-value)	N	CLR_{uc} (p-value)	CLR_{ind} (p-value)	CLR_{ind} (p-value)
VC-EWMA	12	0.0000 (1.0000)	0.2471 (0.6191)	0.2471 (0.8838)	16	1.2764 (0.2586)	0.7622 (0.3826)	2.0386 (0.3608)
VC-MGACH	11	0.0901 (0.7640)	0.4222 (0.5158)	0.5124 (0.7740)	7	2.5629 (0.1094)	0.5538 (0.4568)	3.1167 (0.2105)
MC-MGACH	10	0.3711 (0.5424)	0.6530 (0.4190)	1.0241 (0.5993)	15	0.7340 (0.3916)	1.0658 (0.3019)	1.7997 (0.4066)
	2009 (January 5, 2009 – December 31, 2009)				2010(January 5, 2010 – December 31, 2010)			
	N	CLR_{uc} (p-value)	CLR_{ind} (p-value)	CLR_{cc} (p-value)	N	CLR_{uc} (p-value)	CLR_{ind} (p-value)	CLR_{cc} (p-value)
VC-EWMA	9	0.8610 (0.3535)	0.9474 (0.3304)	1.8084 (0.4049)	8	1.5823 (0.2084)	0.5542 (0.4566)	2.1366 (0.3436)
VC-MGACH	10	0.3711 (0.5424)	0.6530 (0.4190)	1.0241 (0.5993)	9	0.8610 (0.3535)	0.7045 (0.4013)	1.5656 (0.4571)
MC-MGACH	4	7.4886 (0.0062)	0.1362 (0.7121)	7.6248 (0.0221)	4	7.4886 (0.0062)	0.1362 (0.7121)	7.6248 (0.0221)
	2011(January 4, 2011 – December 30, 2011)							
	N	CLR_{uc} (p-value)	CLR_{ind} (p-value)	CLR_{cc} (p-value)				
VC-EWMA	12	0.0000 (1.0000)	1.2693 (0.2599)	1.2693 (0.5301)				
VC-	12	0.0000	1.2693	1.2693				

MGACH		(1.0000)	(0.2599)	(0.5301)
MC-MGACH	8	1.5823 (0.2084)	0.5542 (0.4566)	2.1366 (0.3436)

Table 3-6 Summary of Christoffersen Likelihood Ratio tests of the universe portfolio for 5 years. VaR is calculated at the confidence level of 99%.

Model	2007(January 4, 2007 – December 28, 2007)				2008 (January 2, 2008 – December 31, 2008)			
	N	CLR_{uc} (p-value)	CLR_{ind} (p-value)	CLR_{cc} (p-value)	N	CLR_{uc} (p-value)	CLR_{ind} (p-value)	CLR_{ind} (p-value)
VC-EWMA	4	0.8974 (0.3435)	0.1362 (0.7121)	1.0336 (0.5964)	2	0.0714 (0.7893)	0.0338 (0.8542)	0.1051 (0.9488)
VC-MGACH	1	1.0573 (0.3038)	0.0084 (0.9270)	1.0657 (0.5869)	1	0.1404 (0.7079)	0.2043 (0.6513)	0.3447 (0.8417)
MC-MGACH	2	0.0714 (0.7893)	0.0338 (0.8542)	0.1051 (0.9488)	1	1.0573 (0.3038)	0.0084 (0.9270)	1.0657 (0.5869)
	2009 (January 5, 2009 – December 31, 2009)				2010(January 5, 2010 – December 31, 2010)			
	N	CLR_{uc} (p-value)	CLR_{ind} (p-value)	CLR_{cc} (p-value)	N	CLR_{uc} (p-value)	CLR_{ind} (p-value)	CLR_{cc} (p-value)
VC-EWMA	3	0.1404 (0.7079)	0.0763 (0.7824)	0.2167 (0.8973)	3	0.1404 (0.7079)	0.0763 (0.7824)	0.2167 (0.8973)
VC-MGACH	4	0.8974 (0.3435)	0.1362 (0.7121)	1.0336 (0.5964)	4	0.8974 (0.3435)	0.1362 (0.7121)	1.0336 (0.5964)
VC-MGACH	0	4.8242 (0.0281)	0.0000 (1.0000)	4.8242 (0.0896)	4	0.8974 (0.3435)	0.1362 (0.7121)	1.0336 (0.5964)
	2011(January 4, 2011 – December 30, 2011)							
	N	CLR_{uc} (p-value)	CLR_{ind} (p-value)	CLR_{cc} (p-value)				
VC-EWMA	3	0.1404 (0.7079)	0.0763 (0.7824)	0.2167 (0.8973)				
VC-MGACH	3	0.1404 (0.7079)	0.0763 (0.7824)	0.2167 (0.8973)				
MC-MGACH	2	0.0714 (0.7893)	0.0338 (0.8542)	0.1051 (0.9488)				

Table 3-7 Summary of Berkowitz Likelihood Ratio tests of the universe portfolio for 5 years. VaR is calculated at the confidence level of 95%.

Model	2007		2008		2009	
	BLR_{ind} (p-value)	BLR_{tail} (p-value)	BLR_{ind} (p-value)	BLR_{tail} (p-value)	BLR_{ind} (p-value)	BLR_tail (p-value)
VC-EWMA	2.4415 (0.1182)	20.4803 (0.0000)	0.1825 (0.6692)	1.1319 (0.5678)	0.9674 (0.3253)	16.2260 (0.0003)
VC-MGACH	2.2489 (0.1337)	20.5868 (0.0000)	0.4931 (0.4825)	0.8099 (0.6670)	1.0081 (0.3154)	17.8236 (0.0001)
MC-MGACH	1.2714 (0.2595)	13.4807 (0.0012)	1.9881 (0.1585)	17.4405 (0.0002)	3.1226 (0.0772)	22.3617 (0.0000)
	2010		2011			
	BLR_{ind} (p-value)	BLR_{tail} (p-value)	BLR_{ind} (p-value)	BLR_{tail} (p-value)		
VC-EWMA	1.8014 (0.1795)	8.3540 (0.0153)	0.2491 (0.6177)	7.1538 (0.0280)		
VC-MGACH	1.9186 (0.1660)	9.3207 (0.0095)	0.0870 (0.7681)	7.3252 (0.0257)		
MC-MGACH	2.1913 (0.1388)	12.3798 (0.0021)	3.1120 (0.0777)	8.3059 (0.0157)		

Table 3-8 Summary of Berkowitz Likelihood Ratio tests tests of the universe portfolio for 5 years. VaR is calculated at the confidence level of 99%.

Model	2007		2008		2009	
	BLR_{ind} (p-value)	BLR_{tail} (p-value)	BLR_{ind} (p-value)	BLR_{tail} (p-value)	BLR_{ind} (p-value)	BLR_tail (p-value)
VC-EWMA	2.4415 (0.1182)	12.6328 (0.0018)	0.1825 (0.6692)	0.0521 (0.9743)	0.9674 (0.3253)	13.5565 (0.0011)
VC-MGACH	2.2489 (0.1337)	5.2388 (0.0728)	0.4931 (0.4825)	-0.2338 (0.8897)	1.0081 (0.3154)	13.0136 (0.0015)
MC-MGACH	1.2714 (0.2595)	10.9016 (0.0043)	1.9881 (0.1585)	4.0757 (0.1303)	3.1226 (0.0772)	6.0372 (0.0489)
	2010		2011			
	BLR_{ind} (p-value)	BLR_{tail} (p-value)	BLR_{ind} (p-value)	BLR_{tail} (p-value)		
VC-EWMA	1.8014 (0.1795)	8.5490 (0.0139)	0.2491 (0.6177)	4.6937 (0.0957)		
VC-MGACH	1.9186 (0.1660)	10.4664 (0.0053)	0.0870 (0.7681)	4.3856 (0.1116)		
MC-MGACH	2.1913 (0.1388)	12.1373 (0.0023)	3.1120 (0.0777)	5.5948 (0.0610)		

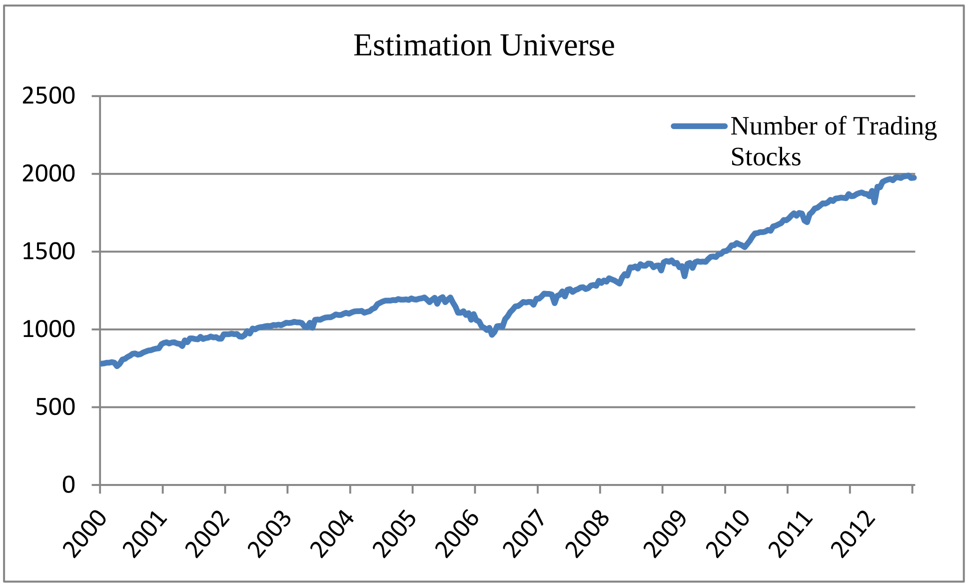


Figure 3-1 Number of trading stocks listed in Shanghai Stock Exchange and Shengzhen Stock Exchange, from 2000 to 2012.

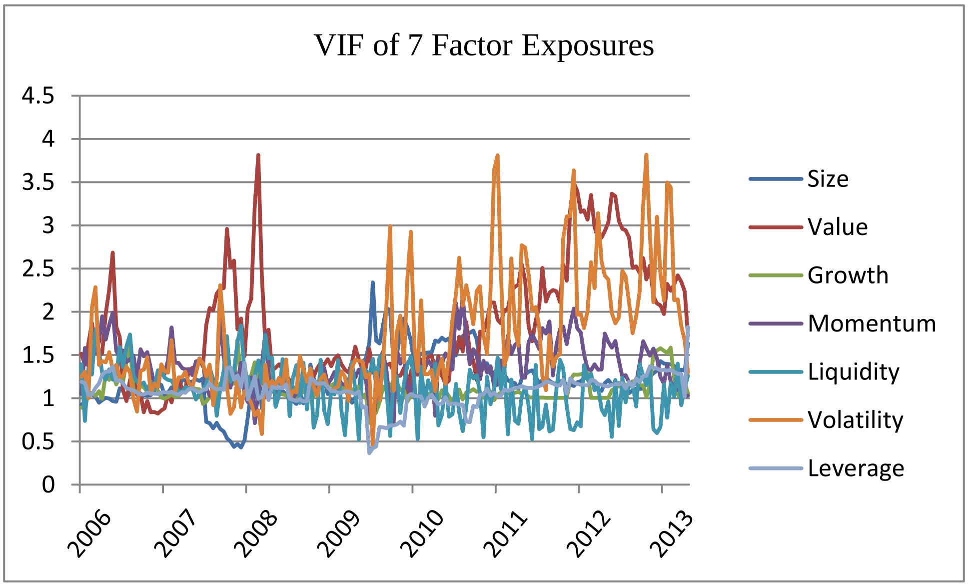
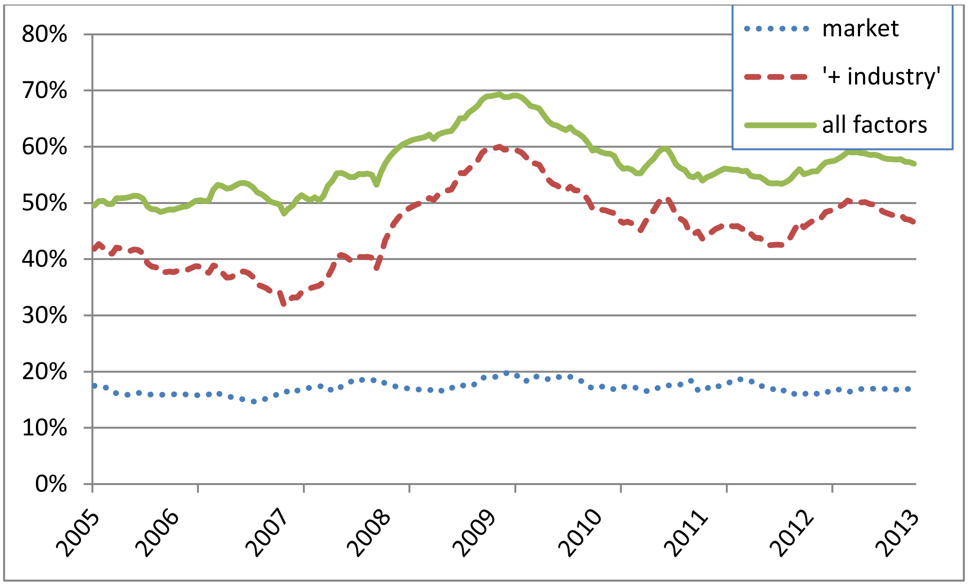


Figure 3-2 Cross-sectional multi-collinearity among factor exposures is measured by Variance Inflation Factor (VIF). VIFs of 7 style factor exposures are plotted from September 2005 to December 2012.

(a)



(b)

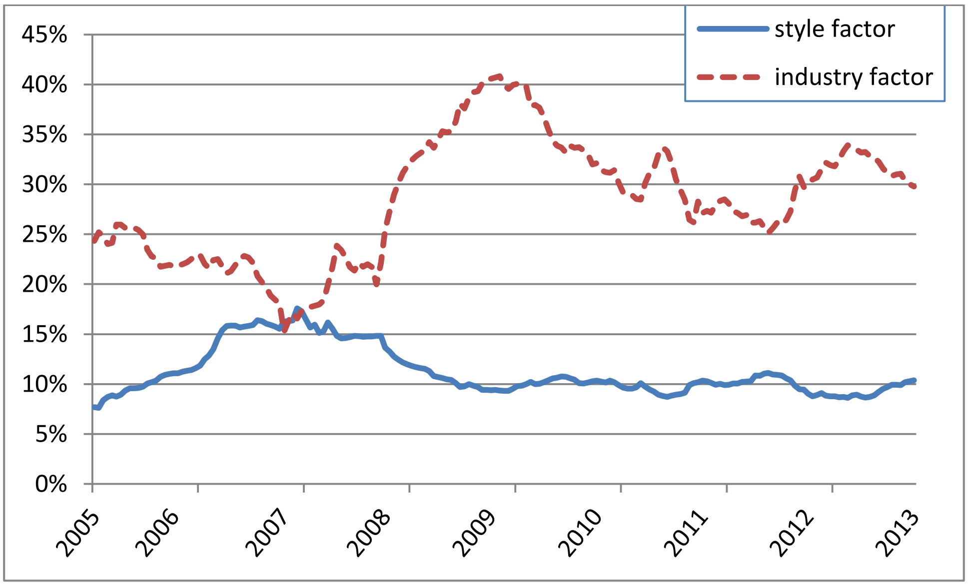
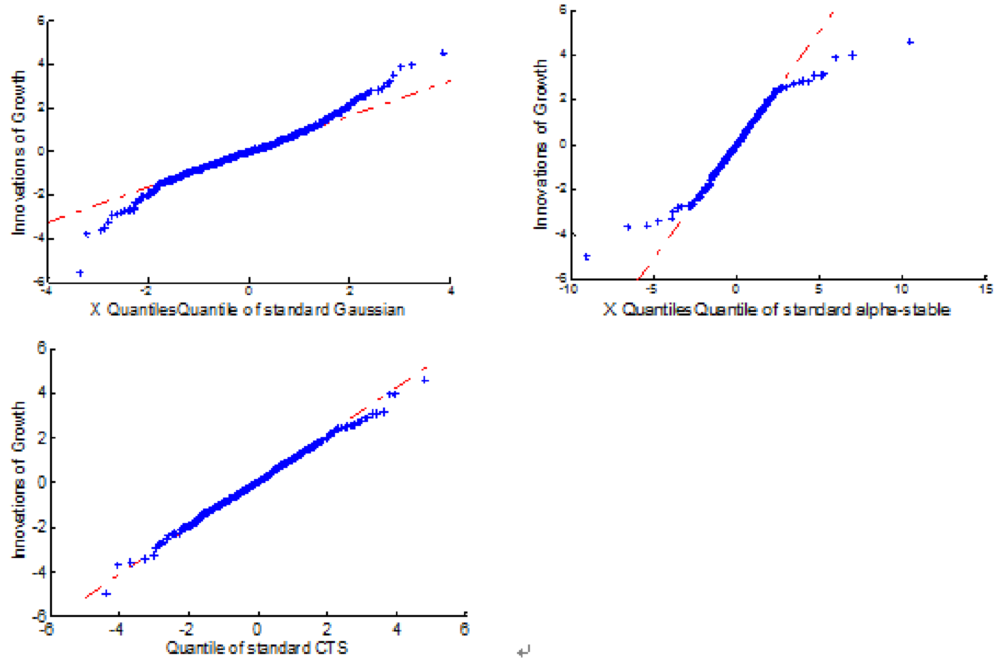


Figure 3-3 CHE one-year trailing average of R^2 from January 2005 to December 2012.

(a) The decomposition of R^2 into factor categories highlights the incremental explanatory power of the CHE market, industry and style factors.

(b) The isolated explanatory power given by style factor or industry factor along helps to understand the investment characteristics of the market.

(a)



(b)

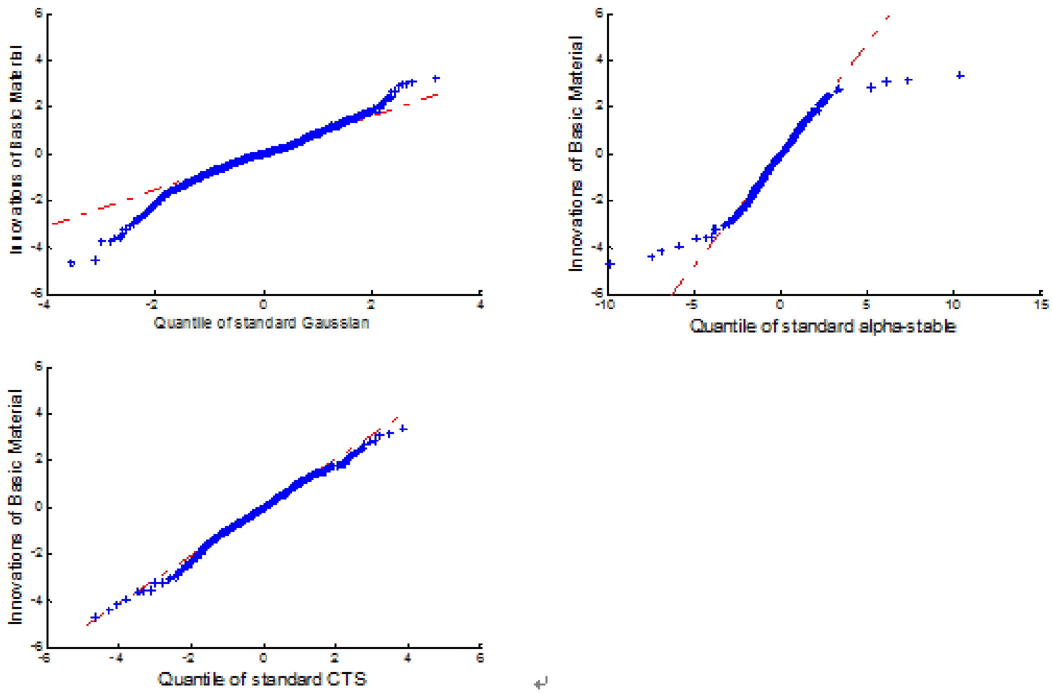


Figure 3-4 Q-Q plot of sample distributions of innovations against standard Gaussian, α-stable and CTS distributions for the Chinese stock market and other world stock markets.

(a) Q-Q plot for *growth* factor. (b) Q-Q plot for *basic material* factor

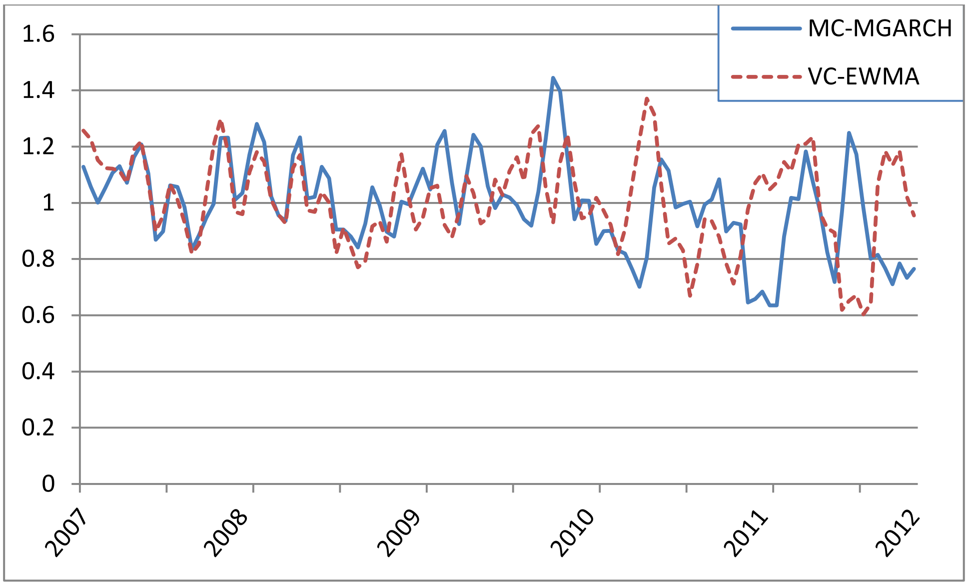


Figure 3-5 12-month rolling window bias statistics of the universe portfolio returns for different forecasting methods from January, 2007 to December 2012.

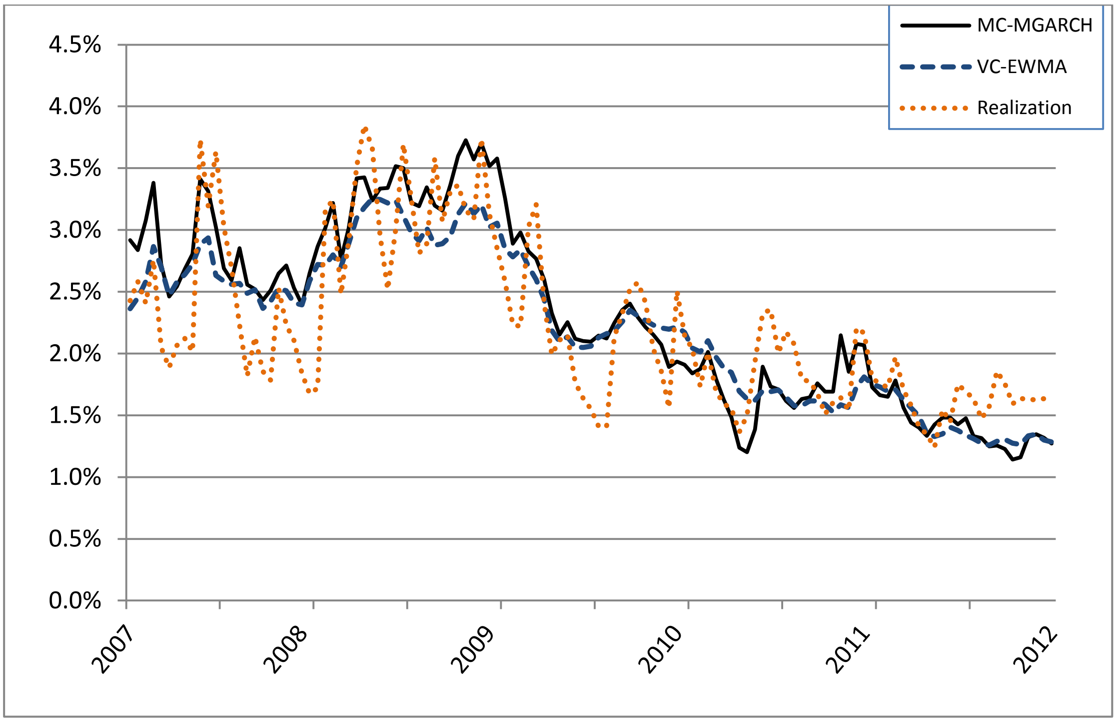


Figure 3-6 Forecasted volatility by VC-EWMA and MC-MGARCH methods and realized volatility of the universe portfolio from January, 2007 to December 2012.

4. Portfolio Optimization and Risk Budgeting

4.1 Introduction

Optimal portfolio selection concerns prudent decision making about the portfolio composition, in order to obtain the investor's objectives. An objective is a feature of a given portfolio for a given investment horizon. For instance, an objective is represented by final wealth at the horizon, or net gains, or wealth relative to some benchmark. The objective is a random variable dependent on the portfolio. One method of evaluating the portfolios, or more precisely, the distribution of objective relative to a portfolio is through the definition of stochastic dominance, a criterion by which we are allowed to evaluate the distribution of the objective as a whole. While the theory of stochastic dominance is solid, the practical application is limited as the resulting optimization problems are very difficult to solve. Thus, the investor might not be able to compare the results of different portfolios and make the right investment decision.

As a consequence, in Section 4.2, an alternative approach is achieved by defining an index of satisfaction which summarizes all the properties of a distribution in a single number. If the index of satisfaction is properly defined the investor can in all circumstances choose the portfolio that best suits him. The potential features of a proper satisfaction index are displayed by Fruttelli and Rosazza Gianin (2002) and Meucci (2007).

A major category of such satisfactory index is based on the intuitive concept of expected utility. In Section 4.3.1, we discuss some properties and examples of the utility functions. From this point, the famous mean-variance analysis introduced by Harry Markowitz in the 1950 can be viewed as an approximation, according to which the investor's satisfaction is determined by the first two moments of the distribution of his objective. The mean-variance optimization has significant advantages as well as some well-known pitfalls. In Section 4.3.2. In general, the mean-variance analysis can be extended to a broader framework, namely, the mean-risk analysis. In section 4.3.4, Instead of taking variance as a proper measure of risk, quantile, such as Value-at-Risk (VaR) and coherent indices, such as Average Value-at-Risk (AVaR) are selected as a risk measure.

4.2 Stochastic Dominance

In this section, we present the stochastic dominance approach to portfolio optimization. For further references, see Ingersoll(1987), Lévy (1998), Yamai and Yoshiba (2002), and

Meucci (2007).

An investor's objectives are described as the quantities that the investor perceives as beneficial and therefore he desires in the largest possible amounts from investing in a portfolio. Given a portfolio with allocation α, the objectives denoted by $\Psi_B = \boldsymbol{\alpha}'\boldsymbol{M}$ can be the final wealth over the investment horizon and $\boldsymbol{M}$ denotes the market vector, for instance. In such ways, two portfolios with different allocation α and β can be selected by comparing the corresponding objectives Ψ_α and Ψ_β. If the objective Ψ_α is always larger than the objective Ψ_β in all scenarios, then the former one is said to strongly dominate the later one:

$$\text{strong dom}: \Psi_\alpha \geq \Psi_\beta \text{ in all scenarios} \qquad 4$$

In other words, the difference between two objectives is a positive random variable. Therefore an equivalent definition of strong dominance is given by the form of cumulative probability function as following:

$$\text{strony dom}: F_{\Psi_\alpha - \Psi_\beta}(0) \equiv \boldsymbol{P}\left(\Psi_\alpha - \Psi_\beta \leq 0\right) = 0 \qquad 4.2$$

The strong dominance is also called order zero dominance, for the reasons become clear below.

Nevertheless, the strong dominance cannot be a proper criterion because if strong dominance takes place, there will be an arbitrage opportunity. Instead, in general an objective Ψ_α from an allocation α is lager in some scenarios and smaller in others that the objective Ψ_β from another allocation β. Thus we would be prone to choose an allocation α over another allocation β if the probability density function of the ensuing objective were concentrated around larger value than for the other allocation. This condition is expressed more easily in terms of cumulative probability function. The objective Ψ_α, or the allocation α, is said to weakly dominate the objective Ψ_β, or the allocation β, if the following condition holds:

$$\text{weak dom}: F_{\Psi_\alpha}(\psi) \leq F_{\Psi_\beta}(\psi) \text{ for all } \psi \in \mathbb{R} \qquad 4.3$$

Weak dominance is also called first-order stochastic dominance (FSD), and other form of the definition may display.

Although less restrictive than the strong dominance, the weak, or first order dominance hardly ever happens. An even weaker type of dominance, such as second-order stochastic dominance (SSD) is introduced.

The objective Ψ_α, or the allocation α, is said to weakly dominate the objective Ψ_β, or the allocation β, if the following condition holds:

$$\text{SSD: } \mathcal{I}^2\left[f_{\Psi_\alpha}\right](\psi) \leq \mathcal{I}^2\left[f_{\Psi_\beta}\right](\psi) \text{ for all } \psi \in \mathbb{R} \quad 4.4$$

Where $\mathcal{I}^2$ is the iterated integral of the pdf:

$$\mathcal{I}^2[f_\Psi](\psi) \equiv \mathcal{I}[F_\Psi](\psi) \equiv \int_{-\infty}^{\psi} F_\Psi(s)ds \quad 4.5$$

We can extend the second-order dominance if weaker types of dominance are need. In general, we say the objective Ψ_α, or the allocation α, order-q dominates the objective Ψ_β, or the allocation β, if the following condition holds:

$$q\text{-dom: } \mathcal{I}^q\left[f_{\Psi_\alpha}\right](\psi) \leq \mathcal{I}^q\left[f_{\Psi_\beta}\right](\psi) \text{ for all } \psi \in \mathbb{R} \quad 4.6$$

Applying the integration operator to both sides of (4.6) we see that order-q dominance implies order $(q+1)$ dominance, but the opposite is not necessarily to be true.

Taking advantage of the criteria of stochastic dominance, we can characterize the efficient sets of the corresponding categories of investors. The efficient set of a given class of portfolios is defined as the set of portfolios not dominated with respect to the corresponding stochastic dominance relation. Any portfolio which is not in the efficient set will be discarded by all investors in that class. Once efficient set is obtained, the investor can rank all possible portfolios in the efficient set with respect to some criteria and maximize his objectives.

Nevertheless, in practice the explicit efficient set is difficult to obtain. In addition, there is no guarantee that there exists an order q such that any two portfolio can be ranked, i.e. the two portfolio cannot be compared in the sense that neither of the respective objectives dominates the other.

It should cause attention whether adopting the stochastic dominance approach for the payoff or the return. For instance, the SSD on the set of payoff distributions implies the same order dominance of the set of log-return distributions but not vice versa. For more references, see Rachev, et at, 2008.

4.3 Satisfaction and Mean-Risk Analysis

Alternative to the stochastic dominance approach for portfolio selection, we summarize all the features of a given portfolio α into one single number S that indicates the respective degree of satisfaction:

$$\boldsymbol{\alpha} \rightarrow S(\boldsymbol{\alpha}) \quad 4.7$$

The investor will then chooses the portfolio that corresponds to the highest degree of satisfaction. In Meucci 2007, the satisfaction index can be mainly divided into three categories, namely certainty-equivalent (expected utility), quantile (Value-at-Risk), and coherent indices (expected shortfall). Here we adopt the concept of satisfaction index but combine the latter two categories into a general mean-risk approach, which is derived from the famous mean-variance approach by Markowitz.

The properties and methods of construction of a satisfaction index can be found in Meucci, 2007.

4.3.1 Expected Utility

Consider a generic portfolio α that gives rise to the investor's objective Ψ_{α}, a utility function $u(\psi)$ describes the extent to which the investor enjoys the outcome $\Psi_{\alpha} = \psi$ of the objective, in case that realization takes place. By applying the expectation of all possible outcomes of the utility function, we obtain the satisfaction index as the expected utility from the given portfolio:

$$\boldsymbol{\alpha} \to \mathrm{E}[u(\Psi_{\alpha})] \equiv \int_{\mathbb{R}} u(\psi) f_{\Psi_{\alpha}}(\psi) d\psi \qquad 4.8$$

where f_{Ψ} is the probability density function of the objective. Indeed, this is the Von Newumann-Morgenstern specification of expected utility as an index of satisfaction, that the higher value is preferred by investors.

Here we pick some properties of the utility function that are derived from common arguments valid for investors belonging to a certain category. For non-satiation investors who prefer more to less have the utility functions with the following properties:

- Sensibility, i.e., the utility function must be an increasing function of the objective. Assuming that the utility function is smooth then the first derivative of the utility must be positive:

$$Du \geq 0 \qquad 4.9$$

where D is the derivative operator.

- Consistence with stochastic dominance, i.e., the expected utility is consistent with first-order stochastic dominance. In other words, if u is increasing, then:

$$F_{\Psi_{\alpha}}(\psi) \leq F_{\Psi_{\beta}}(\psi) \text{ for all } \psi \in \mathbb{R} \Rightarrow \mathrm{E}[u(\Psi_{\alpha})] \geq \mathrm{E}\left[u(\Psi_{\beta})\right] \qquad 4.10$$

However, consistence with second-order stochastic dominance is not guaranteed unless the utility function is increasing and concave. For a smoothing utility function, this condition is displayed as following:

$$Du \geq 0, D^2 u \leq 0 \qquad 4.11$$

- Risk aversion.

  Consider a fair game, i.e., a portfolio $\mathbf{f}$ such that its objective $\Psi_{\mathbf{f}}$ has zero expected value. Then the expected utility is risk averse if the risk-free portfolio $\mathbf{b}$ is preferred to the risky portfolio $\mathbf{b}+\mathbf{f}$ for any given level of the risk-free outcome $\psi_{\boldsymbol{b}}$ and any fair game $\mathbf{f}$:

$$\Psi_{\boldsymbol{b}} \equiv \psi_b, E[\Psi_{\mathrm{f}}] \equiv 0 \Rightarrow E[u(\psi_b)] \geq E[u(\psi_b + \Psi_{\mathbf{f}})] \qquad 4.12$$

  and the expected utility is risk seeking if :

$$\Psi_{\boldsymbol{b}} \equiv \psi_b, E[\Psi_{\mathrm{f}}] \equiv 0 \Rightarrow E[u(\psi_b)] \leq E[u(\psi_b + \Psi_{\mathbf{f}})] \qquad 4.13$$

  and the expected utility is risk neutral if:

$$\Psi_{\boldsymbol{b}} \equiv \psi_b, E[\Psi_{\mathrm{f}}] \equiv 0 \Rightarrow E[u(\psi_b)] \equiv E[u(\psi_b + \Psi_{\mathbf{f}})] \qquad 4.14$$

  To better describe the investor's attitude towards risk we define a more loca measure, namely, the Arrow-Pratt absolute risk aversion, as following:

$$A(\psi) \equiv -\frac{D^2 u(\psi)}{Du(\psi)} \qquad 4.15$$

  The Arrow-Pratt absolute risk aversion is positive if and only if the second derivative of the utility function is negative, which means that the utility function is concave.

4.3.2 Mean-variance Approach

A different approach toward portfolio optimization was introduced by Harry Markowitz in the 1950, mean-variance analysis (M-V analysis) and popularly referred to as modern portfolio theory (MPT), Makowitz (1952) suggested that the portfolio choice be made with respect to two criteria: the expected portfolio return and the variance of the portfolio, the latter used as a proxy for risk. Not only is M-V analysis intuitive, but also it is easy to apply in practice. There are convenient computational recipes and geometric interpretations of the trade-off between the two criteria.

Like the expected utility approach, we look for the consistence of M-V analysis with stochastic dominance. M-V analysis is consistent with first-order stochastic dominance, i.e., higher return is always appreciated by all non-satiation investors. However, generally M-V analysis is not consistent with second-order stochastic dominance, unless the distribution of the investor's market follows multivariate normal distribution, which is a very restrictive assumption. Alternatively, M-V analysis describes the correct choices by investors with quadratic utility function. Again the assumption of quadratic utility function is very restrictive even though we can extend it and consider all utility function that can be sufficiently well approximated by quadratic utilities.

Another well-known drawback is that in M-V analysis variance is used as a measure of risk. The deficiency was recognized by Markowitz (1959) and he suggested the downside semistandard deviation as a measure of risk. In contrast to variance, the downside semistandard deviation is consistent with SSD. With carefully choosing alternative risk measure, we can extend the mean-variance analysis to a more general approach to portfolio optimization, namely mean-risk analysis, which could be consistent with SSD. Two widely used measure of risk, namely, Value-at-Risk and average Value-at-risk, will be discussed under the mean-risk analysis in Section 4.3.3

The main principle behind M-V analysis can be summarized in two ways.

1. Given a certain level of expected return, find the optimal portfolio that has the minimal variance. The allocation $\boldsymbol{\alpha}$ is the solution to the optimization problem defined as following:

$$\begin{aligned} \min_{\alpha} \quad & \mathrm{Var}(\Psi_\alpha) \\ \text{subject to} \quad & \boldsymbol{\alpha} \in \mathrm{C} \\ & \mathrm{E}[\Psi_\alpha] = \psi_* \end{aligned} \qquad 4.16$$

where C denotes the set of constraints. The constraints are determined by the investors for strategy specific or liquidation considerations. For instance, a maximum allocation to a given industry, or transaction cost.

2. Alternatively, given a certain level of variance, find the optimal portfolio with maximal expected return. The allocation $\boldsymbol{\alpha}$ is the solution to the optimization problem defined as following:

$$\begin{aligned} \max_{\alpha} \quad & \mathrm{E}[\Psi_\alpha] \\ \text{subject to} \quad & \boldsymbol{\alpha} \in \mathrm{C} \\ & \mathrm{Var}(\Psi_\alpha) = \nu^* \end{aligned} \qquad 4.17$$

where $\nu^* \geq 0$. Making use of the Lagrangian formulation in (5.16), we can express the optimization problem as following:

$$\begin{aligned} \max_{\alpha} \quad & \mathrm{E}[\Psi_\alpha] - \lambda(\mathrm{Var}[\Psi_\alpha]) \\ \text{subject to} \quad & \boldsymbol{\alpha} \in \mathrm{C} \end{aligned} \qquad 4.18$$

The Lagrange coefficient λ that solves (4.18) can be interpreted as a coefficient of risk aversion. If λ is null the investor is risk neutral: indeed, the argument in the objective function of (4.18) becomes the expected value. On the other hand, if λ is positive the investor is risk averse: indeed, portfolios with the same expected value but with larger variance are penalized. Similarly, if λ is negative the investor is risk seeking. In fact, the objective function in (4.18) with fixed λ is the expected utility of an investor with a quadratic function,

$$\begin{aligned} \mathrm{E}[\Psi_\alpha] - \lambda(\mathrm{Var}[\Psi_\alpha]) &= \mathrm{E}[\Psi_\alpha] - \lambda(\mathrm{E}[\Psi_\alpha - \mathrm{E}(\Psi_\alpha)]^2) \\ &= \mathrm{E}[\Psi_\alpha] - \lambda(\mathrm{E}[\Psi_\alpha]^2 - (\mathrm{E}[\Psi_\alpha])^2) \\ &= \mathrm{E}(-\lambda\mathrm{E}[\Psi_\alpha]^2 + \Psi_\alpha + \lambda(\mathrm{E}[\Psi_\alpha])^2) \\ &= \mathrm{E}[u(\Psi_\alpha)] \end{aligned} \quad 4.19$$

Where the utility $u(x) = -\lambda x^2 + x + \lambda b$ with b equals to the squared expected value of the portfolio.

The M-V analysis can be also interpreted in the way of the index of satisfaction. Since the satisfaction index is a function of the distribution of the investors objective, which is in general univocally determined by its moments:

$$S(\boldsymbol{\alpha}) \equiv \mathcal{H}(\mathrm{E}[\Psi_\alpha], \mathrm{CM}_2[\Psi_\alpha], \mathrm{CM}_3[\Psi_\alpha], \dots) \quad 4.20$$

where $\mathrm{CM_k}$ denotes the central moment of order k of a univariate distribution.

Suppose that we can focus on the two first moments only and neglect all the higher moments. In other words, assume that (**4.20**) can be approximated as following:

$$S(\boldsymbol{\alpha}) \approx \widetilde{\mathcal{H}}(\mathrm{E}[\Psi_\alpha], \mathrm{Var}[\Psi_\alpha]) \quad 4.21$$

for a suitable bivariate function $\widetilde{\mathcal{H}}$. In this approximation, the mean-variance analysis arises as maximize the index of satisfaction. In fact, the only index of satisfaction S such that the approximation (4.21) is exact no matter the market is the case of quadratic utility. Nevertheless the quadratic utility is not flexible enough to model the whole spectrum of the investor’s preferences. For more references, see Meucci 2008.

As mentioned in the beginning of Section 4.3.2, there are several pitfalls of the M-V analysis. In general M-V analysis is not consistent with SSD, which concerns the non-satiation, risk-averse investors. It is only under specific conditions concerning the multivariate distribution of the market variables, such as multivariate normal distribution and the broader class of elliptical distributions. Alternatively, M-V analysis is consistent with the stochastic order arising from quadratic utilities, which also limits the practical application of the problem (4.16), (4.17) and (4.18).

4.3.3 General Mean-risk Approach

In Section 4.3.2, we discussed about the mean-variance (M-V) analysis to portfolio optimization and some pitfalls in applying it into practice. The principle reason is that it leads to correct decisions only when the market vector $\boldsymbol{M}$ follows the multivariate normal distribution, and, as we noted, there is strong empirical evidence against that assumption. The extensions involve including different risk measures in the optimization problem.

The M-V analysis can be extended for a general risk measure $\rho(\cdot)$, and the corresponding optimization problems can be re-stated in a similar way to M-V analysis:

1. Given a certain level of expected return, find the optimal portfolio that has the minimal risk. The allocation $\boldsymbol{\alpha}$ is the solution to the optimization problem defined as following:

$$\begin{aligned} \min_{\alpha} \quad & \rho(\Psi_{\alpha}) \\ \text{subject to} \quad & \boldsymbol{\alpha} \in \mathrm{C} \\ & \mathrm{E}[\Psi_{\alpha}] = \psi_{*} \end{aligned} \qquad 4.22$$

2. Alternatively, given a certain level of risk, find the optimal portfolio with maximal expected return. The allocation $\boldsymbol{\alpha}$ is the solution to the optimization problem defined as following:

$$\begin{aligned} \max_{\alpha} \quad & \mathrm{E}[\Psi_{\alpha}] \\ \text{subject to} \quad & \boldsymbol{\alpha} \in \mathrm{C} \\ & \rho(\Psi_{\alpha}) = \nu^{*} \end{aligned} \qquad 4.23$$

where $\nu^{*} \geq 0$. Again, making use of the Lagrangian formulation in (**5.22**)**,** we can express the optimization problem in a way of utility function as following:

$$\begin{aligned} \max_{\alpha} \quad & \mathrm{E}[\Psi_{\alpha}] - \lambda(\rho[\Psi_{\alpha}]) \\ \text{subject to} \quad & \boldsymbol{\alpha} \in \mathrm{C} \end{aligned} \qquad 4.24$$

The Lagrange coefficient λ that solves (4.24) can be interpreted as a coefficient of risk aversion. If λ is null the investor is risk neutral; if λ is positive the investor is risk averse; if λ is negative the investor is risk seeking.

M-R optimization problems are different from their counterparts in M-V analysis. In order to calculate the risk of the portfolio's objective $\rho[\Psi_{\alpha}]$, we need to know the multivariate distribution of the market vector $\boldsymbol{M}$. Otherwise, it will not be possible to calculate the distribution of the portfolio's objective, and, as a result, portfolio risk will be unknown. This requirement is not so obvious in the M-V optimization problems where we only need the covariance matrix as input.

In Chapter 2, we discussed several types of risk measure, of which the quantile measures and coherent measures, such as Value-at-Risk (VaR) and average Value-at-Risk (AVaR) respectively, are most frequently discussed ones in both academy and industry. The question is that: are they natural candidates as risk measures adopted in the mean-risk (M-R) analysis of portfolio optimization? Basically, the answers to the question are different with respect to VaR and AVaR. For M-R analysis, a key concept inherited from M-V analysis is diversification, which is described by the sub-additive property of the risk measure. For all non-satiation, risk-averse investors with two different portfolios with allocation $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ the following inequity should be satisfied:

$$\rho(\boldsymbol{\alpha}+\boldsymbol{\beta}) \leq \rho(\boldsymbol{\alpha})+\rho(\boldsymbol{\beta}) \qquad 4.25$$

behind which is the effect of diversification. However, in general, VaR does not follow (4.25). This is the main reason why alternative measure of risk, such as AVaR were developed. Moreover, although VaR is consistent with first-order stochastic dominance, it is not consistent with second-order stochastic dominance (see, for example, Guthoff et al, 1997).

By definition, AVaR at tail probability ϵ, $\mathrm{AVaR}_{\epsilon}(\Psi_{\alpha})$, is the average of the VaR numbers larger than the VaR at tail probability ϵ. The formal definition is given in **in** Chapter 2**.** Substituting $\mathrm{AVaR}_{\epsilon}(\Psi_{\alpha})$ for $\rho[\Psi_{\alpha}]$ in **(4.22)**, we obtain the corresponding AVaR optimization problem. AVaR belongs to the class of coherent risk measure which is defined with the sub-additive property, thus AVaR as a risk measure implies the effect of diversification. Although in general, a coherent risk measure is not necessarily consistent with SSD, there exist some particular representatives where the consistency condition is true. For instance, AVaR and, more generally, a spectral risk measure are consistent with SSD. DeGiorgi (2005) provides more information and a formal proof of this fact.

The difficulty of solving in practice problem (4.22), (4.23) and (4.24) depends on the particular choice of the risk measure $\rho(\cdot)$. Generally, if the risk measure is convex function of $\boldsymbol{\alpha}$, then the optimization problems belongs to the convex problems and be solved using the methods of convex optimization. If additionally $\rho(\cdot)$ is differentiable, the one can take advantage of a method for numerical optimization of smooth function, for example gradient methods, or sequential quadratic programming methods. For some specific risk measure, the optimization problem can be further simplified. For instance, in the case of variance as risk measure, the optimization problem reduces to a quadratic programming problem. And there are examples in which the risk measures can be linearized and in consequence the optimization problem reduces to a linear programming problem. For instance this is the case when $\rho(\cdot)$ is the AVaR or a spectral measure of risk (see, Rockafellar and Uryasec, 2002, and Acerbi and

Simonetti, 2002).

Next we give an example of mean-AVaR portfolio optimization problem and discuss the approach of reducing it to a linear programming problem. In this example we choose the returns $\boldsymbol{r}$ of assets in the portfolio as the market vector and the portfolio weights $\boldsymbol{w}$ of the assets as the allocation $\boldsymbol{\alpha}$, then the portfolio optimization problem is given as following:

$$\min_{\boldsymbol{w}} \quad \mathrm{AVaR}_{\epsilon}(\boldsymbol{w}'\boldsymbol{r}) \tag{4.26}$$

$$\text{sbject to } \mathrm{E}[\boldsymbol{w}'\boldsymbol{r}] = R_*$$

$$\boldsymbol{w}^T\boldsymbol{e} = 1$$

where $\boldsymbol{r} = (r_1, \dots, r_n)'$, $\boldsymbol{w} = (w_1, \dots, w_n)'$, $\boldsymbol{e} = (1, \dots, 1)'$, $\epsilon \in (0,1)$ and $R_* \in \mathbb{R}$. If there are available scenarios for asset returns, we can reduce (**4.26**) to a linear programming problem by adopting a linearization of AVaR (see, Rockafellar and Uryasev, 2000). The scenarios can be historical observations of asset returns or independent, identically distributed (i.i.d.) scenarios generated from a multivariate model, for instance, a multi-factor model. These scenarios can be arranged in a $s \times n$ matrix H, in which each column represents the time series of historical observations or generated i.i.d. scenarios of each asset returns, given as following:

$$H = \begin{pmatrix} r_1^1 & \cdots & r_n^1 \\ \vdots & \ddots & \vdots \\ r_1^s & \cdots & r_n^s \end{pmatrix} \tag{4.27}$$

where r_j^k denotes the observation of asset return j at a time k, or k-th scenarios of asset return j. Therefore we can write AVaR in the linearized formulation as following:

$$\mathrm{AVaR}_{\epsilon}(Hw) = \min_{\theta, \boldsymbol{d}} \theta + \frac{1}{s\epsilon}\boldsymbol{d}'\boldsymbol{e} \tag{4.28}$$

$$\text{subject to} - \boldsymbol{Hw} - \theta\boldsymbol{e} \leq \boldsymbol{d}$$

$$\boldsymbol{d} \geq \boldsymbol{0}, \theta \in \mathbb{R}$$

where $\boldsymbol{d} = (d_1, \dots, d_s)'$ is a vector of auxiliary variables and $\theta \in R$ is the additional parameter. The first inequality in (**5.26**) concerns vectors and is to be interpreted in a component-by-component manner,

$$-\boldsymbol{Hw} - \theta\boldsymbol{e} \leq \boldsymbol{d} \Leftrightarrow \left| \begin{array}{c} -r^1 w - \theta \leq d_1 \\ \vdots \\ -r^s w - \theta \leq d_s \end{array} \right. \tag{4.29}$$

The objective function is linear and all constraints are linear equalities and inequalities, which construct the linear programming problem. In this way, we can re-write (4.26) in a form of linear programming problem as following:

$$\min_{\boldsymbol{w},\theta,\boldsymbol{d}} \quad \theta + \frac{1}{s\epsilon}\boldsymbol{d}'\boldsymbol{e} \tag{4.30}$$

$$\begin{aligned} \text{subject to} \; & -\boldsymbol{Hw} - \theta\boldsymbol{e} \leq \boldsymbol{d} \\ & \boldsymbol{w}'\boldsymbol{e} = \boldsymbol{1} \\ & \mathbf{E}[\boldsymbol{w}'\boldsymbol{r}] = R_* \\ & \boldsymbol{w} \geq \boldsymbol{0} \\ & \boldsymbol{d} \geq \boldsymbol{0}, \theta \in \mathbb{R} \end{aligned}$$

where $\boldsymbol{w} \geq 0$ represents a long-only portfolio. As a result, problem (4.30) has a more simple structure than (4.26). However, it turns out that (4.30) is not always superior as far as the computational burden is concerned. The simplification of (4.26) results in the cost of increasing the problem dimension. For instance, problem (4.26) has n variables and $n+2$ linear constraints, where n denotes the number of assets in the portfolio. In contrast, the corresponding linear problem (4.30) has $n+s+1$ variables and $2s+n+2$ linear constraints, in which k denotes the number of scenarios of each asset. A typical number of scenarios for calculating AVaR is 10,000. Furthermore, adding more scenarios makes the matrix defining the linear constraints in the linear programming problem become more nonsparse, which makes the numerical methods for linear programming less efficient.

4.4 Risk Budgeting

4.4.1 The Risk Budgeting Problem

Generally, risk budgeting is a way of thinking about investment and portfolio management, which takes for granted reliance upon probabilistic or statistical measures of risk and the use of modern risk and portfolio management tools to manage risk. Narrowly defined, risk budgeting is a process of measuring and decomposing risk, using the measures in asset allocation decisions, assigning portfolio managers risk budgets defined in terms of these measures, and using these risk budgets in monitoring the asset allocations and portfolio managers. A prerequisite for risk budgeting is risk decomposition, which involves identifying the various sources of risk, or risk factors, such as equity returns, interest rates, and ex-change rates and measuring each factors contribution to the total risk. More specifically, risk decomposition is a methodology to decompose the total risk of a portfolio into smaller units, each of which corresponds to each of the individual securities, or each of the subsets of securities in the portfolio. The smaller decomposed units of the total risk can be interpreted as the risk contribution. Tasche (2000) showed that there is only one definition for risk contributions which is suitable for risk budgeting performance measurement, namely as derivative of the underlying risk measure in direction of the asset weight of the portfolio.

After the primary source of the risk is identified, active portfolio hedging strategies can be carried out to hedge the significant risk already taken.

It is worthwhile to mention that we assume that there exists a portfolio already before the risk attribution analysis is taken. This pre-existing portfolio could be a candidate of or even an optimized portfolio. As the optimal portfolios are quite rare in practice, and because of the rapid changes of market environment, successful portfolio management is indeed a process consisting of small steps, which requires detailed risk diagnoses, namely risk budgeting. The process of portfolio optimization and risk attribution can be repeatedly performed until the satisfactory result is reached.

Consider a portfolio with the random returns set $\boldsymbol{X} = (x_1, \ldots, x_n)'$ of n assets and corresponding weight set $\boldsymbol{W} = (w_1, \ldots, w_k)$' of each asset in the portfolio. A three-step procedure for risk budgeting is used in practice:

Step 1. Compute the overall risk $\rho(r_p)$ where $r_p = \boldsymbol{W'X}$ and $\rho(\cdot)$ is a particular risk measure such as standard deviation, VaR or AVaR; note that at this stage we are not stipulating that $\rho(\cdot)$ must be coherent.

Step 2. Decompose the overall risk $\rho(r_p)$ into the individual asset in the portfolio according to some mathematical risk decomposition principle such that, if CR_i denotes the contribution to risk of asset i with potential return X_i, the sum of the risk contribution amount corresponds to the overall risk $\rho(r_p)$, which means that:

$$\boldsymbol{\rho(r_p) = \sum_{i=1}^{n} \mathrm{CR}_i = \sum_{i=1}^{n} w_i \mathrm{MCR}_{w_i}} \qquad 4.31$$

where MCR_i denotes the marginal contribution to risk of asset i in the portfolio.

Step 3. Adjust the weights of assets in the portfolio with a small non-significant amount according to some risk budgeting principle, such that the adjusted overall risk $\rho\left(\widetilde{r_p}\right) \le \rho\left(r_p\right)$, where $\rho\left(\widetilde{r_p}\right) = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \mathrm{MCR}_{w_i} \widetilde{w}_i$.

4.4.2 The Euler Principle and Examples.

Here we restrict our attention to risk measures that are positive homogenous, such as coherent risk measure, but also include standard deviation and VaR. Recall Euler's well-known rule that states that of $\rho(r_p)$ is positive homogeneous and differentiable at $\boldsymbol{W}$, we have

$$\rho(r_p) = \sum_{i=1}^{d} w_i \frac{\partial \rho}{\partial w_i}(r_p) \quad 4.32$$

$$= \sum_{i=1}^{n} w_i \mathrm{MCR}_{w_i}$$

Now we look at some examples of Euler's rule corresponding to different choices of risk measure ρ.

- Standard deviation risk decomposition

Consider the risk measure $\rho(r_p) = \sqrt{\mathrm{var}(r_p)}$ and Σ with elements $\sigma_{i,j}$ denote the covariance matrix of asset returns $\boldsymbol{X}$. Then the corresponding $\mathrm{MCR}_{\boldsymbol{w_i}}^{\mathrm{st.d.}}$ of asset i is calculated as following:

$$\mathrm{MCR}_{\boldsymbol{w_i}}^{\mathrm{st.d.}} = \frac{\partial \rho}{\partial w_i}(r_p) \quad 4.33$$

$$= \frac{\sum_{\mathrm{j=1}}^{\mathrm{n}} \mathrm{cov}[X_i, X_j] w_j}{\sqrt{\mathrm{var}(r_p)}}$$

$$= \frac{\mathrm{cov}[X_i, \sum_{j=1}^{d} X_j]}{\sqrt{\mathrm{var}(r_p)}}$$

$$= \frac{\mathrm{cov}[X_i, r_p]}{\sqrt{r_p}}$$

- VaR risk decomposition

Consider the risk measure $\rho(r_p) = \mathrm{VaR}_\epsilon(r_p)$. In general the $\mathrm{VaR}_\epsilon(r_p)$ is a complex expression of the allocation. Therefore the explicit decomposition cannot be computed analytically. Here we follow the conclusion in Tasche (2000). Under the simplifying assumption that the return distribution of $\boldsymbol{X}$ has a joint density, subject to technical conditions:

$$\mathrm{MCR}_{\boldsymbol{w_i}}^{\mathrm{VaR}} = \frac{\partial \rho}{\partial w_i}(r_p) \quad 4.34$$

$$= -\mathrm{E}[\mathrm{X_i} | r_p = -\mathrm{VaR}_\epsilon(r_p)]$$

- AVaR risk decomposition

$$\mathrm{MCR}_{\boldsymbol{w_i}}^{\mathrm{AVaR}} = \frac{\partial \rho}{\partial w_i}(r_p) \qquad 4.35$$

$$= \frac{1}{\epsilon}\int_0^{\epsilon} \frac{\partial VaR_u}{\partial w_i}(r_p)du$$

$$= \frac{1}{\epsilon}\int_0^{\epsilon} \mathrm{E}\left[\mathrm{X_i}|r_p = -\mathrm{VaR}_u(r_p)\right]du$$

$$= -\mathrm{E}\left[\mathrm{X_i}|r_p \leq -\mathrm{VaR}_{\epsilon}(r_p)\right]$$

4.4.3 Risk Budgeting

Risk decomposition helps identify the contribution to risk of each asset in the portfolio. Based on this identification, the portfolio can be adjusted with respect to some market or self-imposed risk constraints or budgets on either positions of certain assets or the entire portfolio. Combined with information about expected returns, the risk decomposition can also help optimize reward-risk tradeoff. For instance, it will help the portfolio manager identify the expected return forecasts explicit in the choice of a particular portfolio, that is, the implied views. It can also help to decide whether the benefits of altering a position are large enough to cover transaction cost. Moreover this process can be used to justify the efficiency of the optimizer used in selecting portfolio. Examples can be found in Pearson 2002.

Here we provide a general illustration about risk budgeting.

Suppose we have a general mean-risk portfolio. After exploiting a link between percentage contributions to risk, we can arrive at a rule for identifying risk diversifiers and risk contributors in the portfolio. The portfolio optimization problem is given as following:

$$\max_{\boldsymbol{W}} \mathrm{E}[\boldsymbol{W'X}] - \lambda\rho(\boldsymbol{W'X}) \qquad 4.36$$

$$\text{subject to } \boldsymbol{W} \in C$$

Where λ is the risk-aversion coefficient and C is some set of constraints. Solving this optimization problem with the allocation $\boldsymbol{W}$, we can calculate marginal contribution to risk and the contribution to risk of each asset with **equation 5.30** as following:

$$\mathrm{MCR}_{w_i} = \frac{\partial \rho}{\partial w_i}(\boldsymbol{W'X}) \qquad 4.37$$

$$\mathrm{CR}_{w_i} = w_i\frac{\partial \rho}{\partial w_i}(\boldsymbol{W'X})$$

In addition we can calculate the contribution to risk of each asset with respect to the total risk:

$$\mathrm{CR}_{\boldsymbol{w_i}}^{\%} = \frac{\mathrm{CR}_{\boldsymbol{w_i}}}{\rho(\boldsymbol{W'X})} \times 100\% \quad (5.36)$$

Assets in the portfolio with positive CR_{w_i} or $\mathrm{CR}_{\boldsymbol{w_i}}^{\%}$ are defined as risk contributors and those with negative CR_{w_i} or $\mathrm{CR}_{\boldsymbol{w_i}}^{\%}$ are defined as risk diversifiers, meaning that increasing the weight of risk diversifiers or decreasing the weight of risk contributors can reduce the total risk of portfolio. Note that the sign of MCR_{w_i} is not necessarily to be the same as the sign of w_i. In particular increasing the weight of short position means establishing a larger short position, which can be seen from the fact that for short positions the sign of CR_{w_i} is opposite of the sign of MCR_{w_i}.

The risk decomposition clearly highlights the positions that should be the focus of the portfolio or risk manager's attention. In the terminology of **Litterman** (1996), these positions are the portfolio's hot spots. The hot spots have no direct relationship with the size of the positions, meaning that large position may not be the focus of concern and vice versa.

Consider a target of risk η^* about the entire portfolio, if the risk of the current portfolio $\rho(\boldsymbol{W'X}) > \eta^*$ then we need to adjust the weight $\boldsymbol{W}$ to a new one $\boldsymbol{W}^*$ such that $\rho(\boldsymbol{W}^{*\prime}\boldsymbol{X}) \leq \eta^*$, namely, risk budgeting. This target can be approached by adjusting contribution to risk of certain positions in several ways, such as:

- Adjust certain positions with respect to risk contributor and risk diversifier

Suppose asset j with largest positive CR_{w_j}, the effect of reducing the weight from w_j to w_j^* can be obtained from the following equation:

$$\Delta\rho(\boldsymbol{W'X}) \approx \frac{\partial\rho}{\partial w_j}(\boldsymbol{W'X})w_j \times \frac{w_j^* - w_j}{w_j} \quad 4.38$$

reducing the total risk of approximately $\Delta\rho(\boldsymbol{W'X})$. The actual change in risk resulting from the reducing is not exactly the same as $\Delta\rho(\boldsymbol{W'X})$, due to the fact that the risk decomposition is a marginal analysis and only exactly correct for infinitesimally small changes in the positions.

This analysis can be reversed to approximately determine the trade necessary to have a desired effect on the total risk. Still taking asset j as an example, the new weight w_j^* can be determined as following:

$$w_j^* = \frac{\eta^* - \rho(\boldsymbol{W'X})}{\frac{\partial\rho}{\partial w_j}(\boldsymbol{W'X})} + w_j \quad 4.39$$

A more general approach to reducing total risk can be achieved by altering positions of both risk contributors and risk diversifiers simultaneously.

Suppose we do not have a risk budget and we do not want to change the total net value of current portfolio. Sort all positions with respect to MCR_{w_i} in a deceasing order denoted by $(w^1, \dots, w^n)'$, and set an integer $C < \frac{1}{2}n$ and a vector $\boldsymbol{\Delta} = (\Delta_1, \dots, \Delta_{2C})'$ such that the sum of elements in $\boldsymbol{\Delta}$ equals zero. Increase the top C weights $(w^1, \dots, w^C)'$ according to $(\Delta_1, \dots, \Delta_C)$ and decrease the bottom C weights $(w^{n-C}, \dots, w^n)'$ according to $(\Delta_{C+1}, \dots, \Delta_{2C})'$. The adjusted weight $\boldsymbol{W}^*$ reduces the total risk such that $\rho(\boldsymbol{W}^*) \leq \rho(\boldsymbol{W})$. This adjust process can be done with optimization process repeatedly.

- Adjust certain positions with respect to best hedging.

Another approach of risk budgeting is used to compute risk-minimizing trades, or best hedge, for each assets in the current portfolio. We illustrate this using the position of asset j. Letting Δw_j denote the change to current position w_j, the risk-minimizing trade is the Δw_j that minimize the portfolio risk, or the solution of the following equation:

$$\Delta w_j = \underset{\Delta w_j}{\operatorname{argmin}} \{\rho((\boldsymbol{W} + \Delta w_j \boldsymbol{I}_j)' \boldsymbol{X})\} \qquad 4.40$$

where $\boldsymbol{I}_j$ is a vector with j-th element equals 1 and others equal zero.

Best hedge can be applied to each position separately. It is also possible to be applied involving two or more position. In practice, a particular set H is selected and best hedge can be described by solving the optimization problem with changes in the positions included in H.

4.5 An Example of Portfolio Optimization and Risk Budgeting

In this section, we provide two back-testing experiments of long-only[12] optimal portfolio strategies using the CSI 300 universe. In the first experiment, different risk-reward strategies are applied and compared; and in the second experiment, risk budgeting process is adopted in

[12] The reason of choosing long-only strategy instead of long-short strategy is stated with the fact of market as following: A pilot program of margin trading and securities trading, also known as credit trading, was launched on March 31th, 2010, with 90 stocks eligible for such behaviors. The scope of eligible stocks was expanded to 278 in December 2011, and later to 494 in January 2013. The most recent expansion was on September 16th, 2013, with 700 eligible stocks, which account for approximately 28% of all A-share stocks in the Chinese stock markets. Thus only in the last year of our 6 year backtesting time period, most stocks in the portfolios are available for short trading. To obtain a realistic testing result, we only consider long-only portfolios in this circumstance. For more information about stock trading rules in the Chinese market, see http://sse.com.cn/tradmembership/trading/overview/ and http://www.szse.cn/main/en/Products/Trading.

the portfolio strategies. The back-testing time period is 6 years starting from January 4th, 2007 to December 29th, 2012. In these strategies, the CSI 300 Index is used as the benchmark. The optimization algorithm would take the form of mean-variance and mean-AVaR optimization problems described in (5.23) with the risk measures ρ being the standard deviation and AVaR, respectively.

In Section 4.5.1, we present and compare two daily trading portfolios both with the CHE factor model discussed in Chapter 3 as the risk-forecasting model, but with different optimization processes. In Section 4.5.2, we adopt the process of risk budgeting discussed in Section 4.4. Again, we compare the two portfolios, of which one is M-V portfolio and the other is M-AVaR portfolio.

To ensure a consistent analysis of all the experimental results, we first define the constraints, which are applied in the optimization problems for all tested portfolios, and some useful portfolio performance measures in comparing different portfolios.

The general mean-risk portfolio optimization problems with weight constraints and turnover constraints has the form as following:

$$\max_{\boldsymbol{W}} \mathrm{E}[\boldsymbol{W}'\boldsymbol{X}] - \lambda^{\rho}\rho(\boldsymbol{W}'\boldsymbol{X}) - d\|\boldsymbol{W} - \boldsymbol{W}_0\| \quad 4.41$$

$$\text{subject to } \boldsymbol{0} \leq \boldsymbol{W} \leq c\boldsymbol{I}$$

where λ^{ρ} is a positive constant representing the risk-aversion coefficient consistent for the particular risk measure ρ. $\boldsymbol{W}_0$ is a non-negative vector representing the weights of portfolio from last trade or initial portfolio which is a equal-weighted one. $\|\cdot\|$ is the Euclidean norm and d is a positive constant related to the turnover or transaction cost constraints. c is a positive constant representing the upper-bound of the weight of each asset in the portfolio, we drop the total investment constrain, which requires $\boldsymbol{I}'\boldsymbol{W} = 1$, assuming holding cash in the portfolios in some time periods.

The portfolio performance measures consider here are the Sharpe ratio and the STARR ratio. The Sharpe ratio, which was introduced by Sharpe (1966), is defined as following:

$$SR(\boldsymbol{W}) = \frac{\mathrm{E}[\boldsymbol{W}'\boldsymbol{X}] - r_b}{\sigma_{\boldsymbol{W}'\boldsymbol{X}}} \quad 4.42$$

where $\sigma_{\boldsymbol{W}'\boldsymbol{X}}$ is the standard deviation of the portfolio and r_b is a constant representing the expected return of the benchmark. In our cases, r_b is set to be 2.5% for simplicity.

The STARR ratio, which stands for stable tail-adjusted return ratio, is defined as following:

$$STARR_{\epsilon}(\boldsymbol{W}) = \frac{\mathrm{E}[\boldsymbol{W}'\boldsymbol{X} - r_b]}{\mathrm{AVaR}_{\epsilon}(\boldsymbol{W}'\boldsymbol{X} - r_b)} \qquad 4.43$$

where $\mathrm{AVaR}_{\epsilon}(\boldsymbol{W}'\boldsymbol{X} - r_b) = \mathrm{AVaR}_{\epsilon}(\boldsymbol{W}'\boldsymbol{X}) + r_b$ for a constant r_b. Again, r_b is set to be 2.5% in our cases.

4.5.1 A Portfolio Optimization Example

In this section we provide a comparison between two mean-risk portfolios constructed by different strategies, namely, M-V (mean-variance) strategy and M-AVaR (mean-average VaR) strategy with AVaR given at a 99% confidence level. The objective function takes the form of a utility function by maximizing the return with respect to a certain level of risk. The set of constraints includes weight constraints and turnover/transaction cost constraints.

A generic scenarios generation module provides scenarios with specified preferences on both margins and dependence structure of a multivariate model. In our experiments we estimate portfolio expected return and risk separately: expected returns of each stock are estimated from the mean-AVaR optimization problems model of individual time series of historical daily log-returns; and risk are calculated from scenarios generated from the CHE factor model discussed in Chapter 3.

Figure 4.1 provides the net value of investing ￥1000 in the M-V and M-AVaR portfolios and the CIS 300 Index on January 4th, 2007 and rebalancing the two portfolio every day. It is obvious that both portfolios outperform the index, especially after the financial crisis in 2008. In Figure 4.2, the realized volatility, given by the 60-day moving average of standard deviation of portfolio returns, indicates that the M-V portfolio is as volatile as the index in the highly volatile market periods from 2007 to 2009, and stays in a higher level of volatile than the index after 2009. However, the M-AVaR portfolio is much less volatile than the index in most years.

Figure 4.3 and 4.4 provide the distributions and time series of monthly returns of the two portfolios from 2007 to 2012, respectively. It can be observed from Figure 4.3 that the M-V portfolio has heavier tails of the monthly return distribution in both tails than those of the M-AVaR portfolio. Figure 4.4 may indicate that the two portfolios have the similar performance in terms of the months of profits and losses but the M-V portfolio has larger monthly profits and losses.

Another important characteristic of portfolio performance is the drawdown defined as the peak-to-trough decline during a specific record period. In Figure 4.5, we plot the so-called underwater curve of the drawdown of the two portfolios from 2007 to 2012. The M-V portfolio has a maximum drawdown of 98% in October, 2008. In contrast, the maximum drawdown of the M-AVaR portfolio is 43%.

The performance statistics of the two portfolios and the index and the net returns in each year are provided in Table 4.1. We may conclude that the M-AVaR portfolio outperforms the M-V portfolio in terms of all the test performance statistics. For instance, the higher Sharpe Ration and STARR of the M-AVaR portfolio may indicate a better trade-off between risk and reward. And the smaller values of drawdown of the M-V portfolio may indicate a better control in the downside risk. The monthly returns in each year are provided in Table 4.2.

Table 4.3 provides more information in the quantitative point of view. We applied the CAPM to the daily returns of the two portfolios, assuming an annual risk-free rate of 3.5%. The M-AVaR portfolio has a smaller value of Beta than that of the M-V portfolio. And both portfolios have the non-significant values of Alpha in terms of daily returns. The statistics of sample moments of the daily return distributions of the two may indicated that they are skewed to the left and concentrate around zero, which may be fitted with α-stable or tempered stable distributions. In addition, the quantiles of the returns also indicate that the return distributions are heavy-tailed.

4.5.2 A Risk Budgeting Example

In this section, we present a risk budgeting example through comparison of two M-AVaR portfolios denoted by P1 and P2. P1 and P2 are constructed as the same as the M-AVaR portfolio in Section 4.5.3, except that it is rebalanced every two days. The difference between P1 and P2 is that P2 has a risk budgeting process on the second day after optimization, while P1 simply hold the positions on the second day.

The risk budgeting process includes 3 steps:

(1) Calculate the MCR (marginal contribution to risk) with (4.35) of all positions that has weight $w_i > 0.1\%$.

(2) Sort MCRs in a descending order and choose the top 5% and bottom 5% of the group and divided them into two subgroups, namely, contributors and diversifiers.

(3) Adjust the weights of positions in contributors as following:

$$w_i^{ad} = w_i - \Delta \frac{\mathrm{MCR}_{w_i}}{\sum_{j \in C} \mathrm{MCR}_{w_j}} \qquad 4.44$$

where $i, j \in C$ and C stands for the subgroup of contributors.

And adjust the weights of positions in diversifiers as following:

$$w_i^{ad} = w_i + \Delta \frac{\mathrm{MCR}_{w_i}}{\sum_{j \in D} \mathrm{MCR}_{w_j}} \qquad 4.45$$

where $i, j \in D$ and D stands for the subgroup of diversifiers.

Figure 4.6 presents the cumulative return of portfolio P1 and P2 starting from January 4, 2007 to December 29, 2012. It is obviously that P2 outperforms P1 in almost all time periods after 2008. Noting that the optimal portfolio of P1 and P2 given by the M-AVaR optimization approach may have different positions, as the pre-optimization weights are different after risk budgeting.

Figure 4.7 presents the scalar plot of the pair $\mathrm{R_{P1}} - \mathrm{R_{P2}}$ and $\mathrm{AVaR_{P1}} - \mathrm{AVaR_{P2}}$, where $\mathrm{R_{P1}}, \mathrm{R_{P2}}$ are the daily returns of P1 and P2, and $\mathrm{AVaR_{P1}}, \mathrm{AVaR_{P2}}$ are the daily forecasted AVaR of P1 and P2, respectively. It can be observed that, with most positive values of $\mathrm{AVaR_{P1}} - \mathrm{AVaR_{P2}}$ representing reductions in total risk, there are more negative values of $\mathrm{R_{P1}} - \mathrm{R_{P2}}$ representing increase in total returns. Thus it may indicate that in a long time period, decreasing risk can increase returns. A more clearly investigation is provided in Figure 4.8, which plot the distributions of values of $\mathrm{R_{P1}} - \mathrm{R_{P2}}$ and $\mathrm{AVaR_{P1}} - \mathrm{AVaR_{P2}}$

In Table 4.4, we provide an example of risk budgeting process of P2 in an arbitrary day. In this example, all contributors have positive MCRs and all diversifiers have negative MCRs. We decreased the weights of contributors and decreased the weights of diversifiers according to **(4.4)** and **(4.5)**. Finally, the reduction in total risk is 12 base points with a slightly reduction in total return of 0.4 base points.

Table 4-1 Performance statistics of the M-V, M-AVaR portfolios and the CSI 300 Index starting from January 4, 2007 to December 29, 2012

(a) Performance statistics over the entire investing periods.

	M-V	M-AVaR	Index
Total Return	141.83%	171.74%	25.57%
Annualized Return	15.86%	18.13%	3.87%
Annualized Std. Deviation	34.40%	28.19%	31.74%
Correlation	0.033	0.032	1
Sharpe Ratio (rf=3.5%)	0.71	1.04	0.18

STARR(99%)	0.17	0.42	0.03
Worst Month	-25.67%	-16.77%	-19.80%
Date of Worst Month	09/2008	09/2008	06/2008
Worst Drawdown	-95.80%	-43.50%	-117.40%
Date of Worst Drawdown	01/08 - 11/08	01/08 - 11/08	10/07 - 11/08

(b) Yearly return of the two portfolios and the index over the entire investing periods

Year or YTD	M-V	M-AVaR	Index
2012	-27.16	-16.59	-14.40
2011	-22.53	-10.52	-23.87
2010	17.57	22.01	-6.31
2009	108.18	92.37	60.99
2008	-46.14	-8.86	-82.72
2007	111.91	93.32	91.87

Table 4-2 Monthly returns of the M-V and M-AVaR portfolios and the CSI 300 Index starting from January 4, 2007 to December 29, 2012.

(a) The M-V portfolio

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2012	-20.2	10.1	10.0	-8.2	9.9	-6.9	-7.0	-6.2	-0.4	-0.9	0.2	-7.7
2011	5.4	-2.3	6.5	1.8	-7.6	-4.8	3.7	-3.4	-6.2	-11.2	2.2	-6.5
2010	1.1	-9.2	5.9	9.4	-15.6	-1.7	-8.1	15.8	5.1	10.7	14.4	-10.3
2009	13.5	22.3	7.8	16.4	13.6	0.8	15.1	-2.3	2.9	2.0	9.7	6.3
2008	10.3	4.5	0.5	-10.6	-11.7	13.9	-25.0	13.0	-25.7	-13.9	-8.1	6.6
2007	17.8	18.7	16.5	30.7	1.9	-9.2	15.6	15.5	16.2	-17.0	-5.0	10.3

(b) The M-AVaR portfolio

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2012	-13.9	8.4	11.0	-6.1	8.8	-4.8	-3.6	-6.9	-2.7	-0.6	0.4	-6.7
2011	1.7	-2.7	7.4	1.1	-4.6	-4.2	5.0	-2.3	-4.3	-8.6	5.6	-4.9
2010	1.6	-3.9	4.1	7.8	-14.8	0.6	-3.6	11.3	3.2	10.8	13.0	-8.2
2009	9.1	18.4	6.2	12.6	7.3	4.6	13.2	0.8	-1.3	4.1	9.6	7.8
2008	7.4	-2.3	-1.0	-11.9	-5.2	14.3	-16.8	16.5	-16.8	-4.0	1.4	9.6
2007	18.6	6.8	8.8	18.9	9.9	-0.8	11.8	10.8	7.6	-4.0	-5.4	10.3

(c) The CSI 300 Index

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2012	-9.9	11.0	4.2	-4.3	5.4	-4.3	-2.7	-4.8	-6.4	4.2	0.7	-7.4
2011	-4.1	3.2	-0.3	4.7	-6.9	-5.7	5.5	-10.6	-1.6	-7.1	6.9	-7.8
2010	-4.4	-6.0	2.0	3.7	-16.4	-3.5	-3.6	7.4	4.6	10.6	0.6	-1.3
2009	-0.9	13.4	6.9	5.3	6.8	11.6	15.4	-12.7	5.7	2.9	7.4	-0.9
2008	-2.9	-7.4	-16.8	-3.5	-0.8	-19.8	0.7	-18.8	-15.6	-14.1	9.0	7.2
2007	6.8	9.2	11.9	24.5	0.3	2.4	14.4	19.1	3.9	6.3	-19.7	12.8

Table 4-3 Sample statistics of daily returns of the M-V and M-AVaR portfolios and the CSI Index wit total 1460 observations, starting from January 4, 2007 to December 29, 2012. Alpha and Beta are calculated from the CAPM model assuming annual risk-free rate to be

	M-V	M-AVaR	Index
Alpha (p-value)	0.002 (0.000)	0.002 (0.000)	
Beta (p-value)	0.067 (0.011)	0.043 (0.041)	
Mean	1.1%	1.3%	0.3%
Standard Deviation	2.2%	1.8%	02.1%
Skewness	-0.404	-0.338	-0.243
Kurtosis	4.209	4.719	5.149
Quantiles $\begin{bmatrix} 1\% & 5\% \\ 95\% & 99\% \end{bmatrix}$	$\begin{bmatrix} -6.3\% & -3.7\% \\ 3.4\% & 5.5\% \end{bmatrix}$	$\begin{bmatrix} -5.2\% & -3.1\% \\ 2.8\% & 4.6\% \end{bmatrix}$	$\begin{bmatrix} -5.5\% & -3.6\% \\ 3.2\% & 5.1\% \end{bmatrix}$

Table 4-4 Risk and return decomposition of current portfolio P2. Risk and return contributions are calculated as contribution of total risk and return, respectively. All numbers is shown in percentage.

	Return	Return Contribution	MCTR	Risk Contribution	Weight	Reduction in Total Risk	Adjusted Weight
C1	1.12	0.87	23.38	1.61	0.42	0.14	0.38
C2	1.28	1.06	23.32	1.72	0.45	0.14	0.41
C3	1.33	1.13	23.12	1.75	0.46	0.14	0.42
C4	1.47	1.32	22.87	1.84	0.49	0.14	0.45
C5	1.41	1.24	22.87	1.79	0.47	0.14	0.44
C6	1.29	1.07	22.85	1.69	0.45	0.13	0.41
C7	1.33	1.12	22.75	1.72	0.46	0.13	0.42
C8	1.49	1.35	22.74	1.85	0.49	0.13	0.46
C9	1.49	1.36	22.74	1.85	0.49	0.13	0.46
C10	1.33	1.13	22.73	1.72	0.46	0.13	0.42
C11	1.45	1.29	22.73	1.81	0.48	0.13	0.45
C12	1.38	1.19	22.72	1.75	0.47	0.13	0.43
C13	1.51	1.39	22.68	1.86	0.50	0.13	0.46
C14	1.51	1.39	22.58	1.85	0.50	0.13	0.46
D1	0.40	0.24	-1.09	-0.06	0.32	0.01	0.35
D2	0.41	0.24	-1.09	-0.06	0.32	0.01	0.35
D3	0.39	0.23	-1.13	-0.06	0.32	0.01	0.35
D4	0.39	0.23	-1.15	-0.06	0.32	0.01	0.35
D5	0.83	0.57	-1.17	-0.07	0.37	0.01	0.40
D6	0.51	0.32	-1.18	-0.07	0.33	0.01	0.36
D7	0.54	0.34	-1.27	-0.07	0.34	0.01	0.37
D8	0.53	0.33	-1.28	-0.07	0.34	0.01	0.37
D9	0.47	0.29	-1.31	-0.07	0.33	0.01	0.36
D10	0.61	0.39	-1.37	-0.08	0.34	0.01	0.38
D11	0.52	0.32	-1.41	-0.08	0.33	0.01	0.37
D12	0.53	0.33	-1.43	-0.08	0.34	0.01	0.37
D13	0.44	0.26	-1.44	-0.08	0.33	0.01	0.36
D14	0.44	0.27	-1.52	-0.08	0.33	0.01	0.36
D15	0.72	0.48	-1.57	-0.09	0.36	0.01	0.40
Total Return Reduction		0.43 BP					
Total Risk Reduction		12.09 BP					

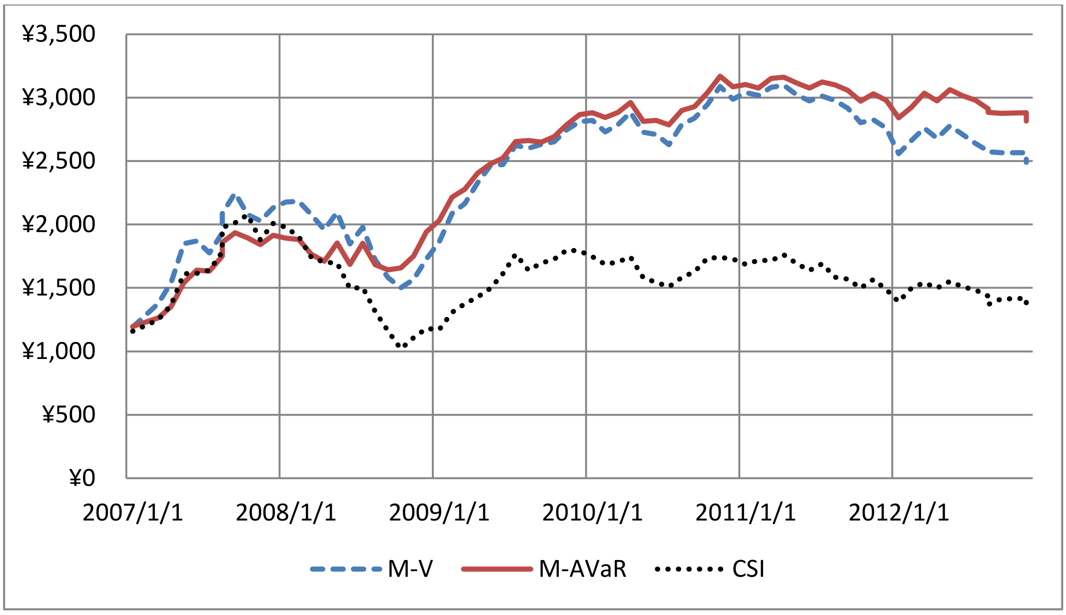


Figure 4-1 Performance comparisons. The M-V, M-AVaR portfolios and the CSI 300 Index starting with ￥1000 from January 4, 2007 to December 29, 2012.

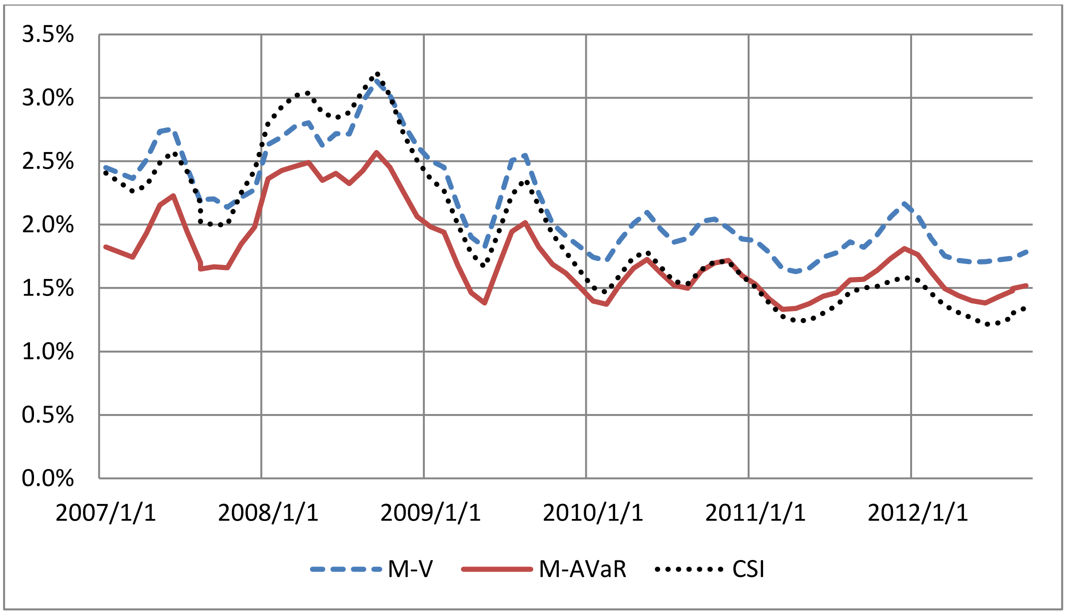
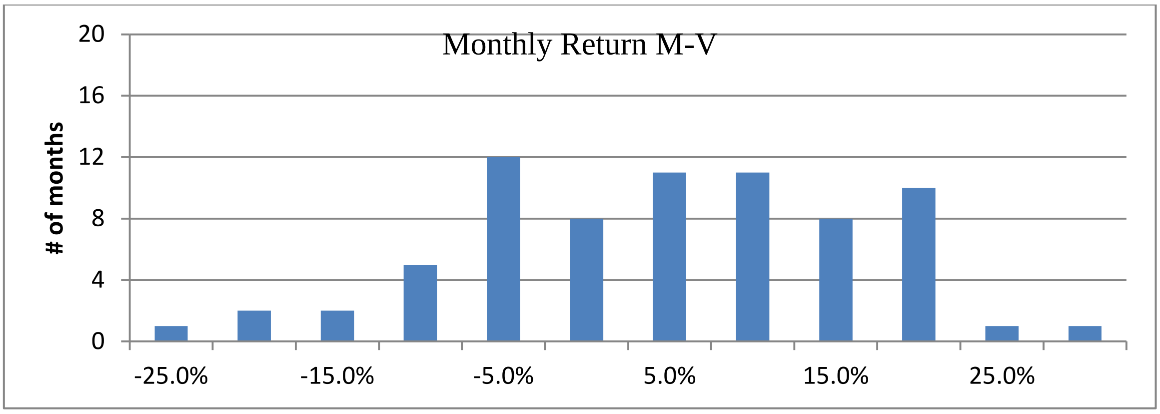


Figure 4-2 Volatility comparisons. The 60-day moving average of standard deviations of the M-V, M-AVaR and Index portfolios starting from January 4, 2007 to December 29, 2012

(a)



(b)

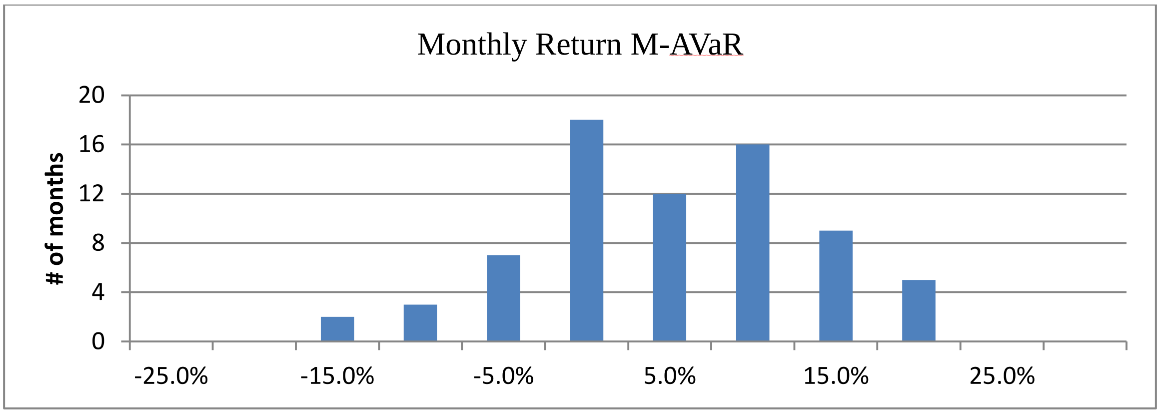
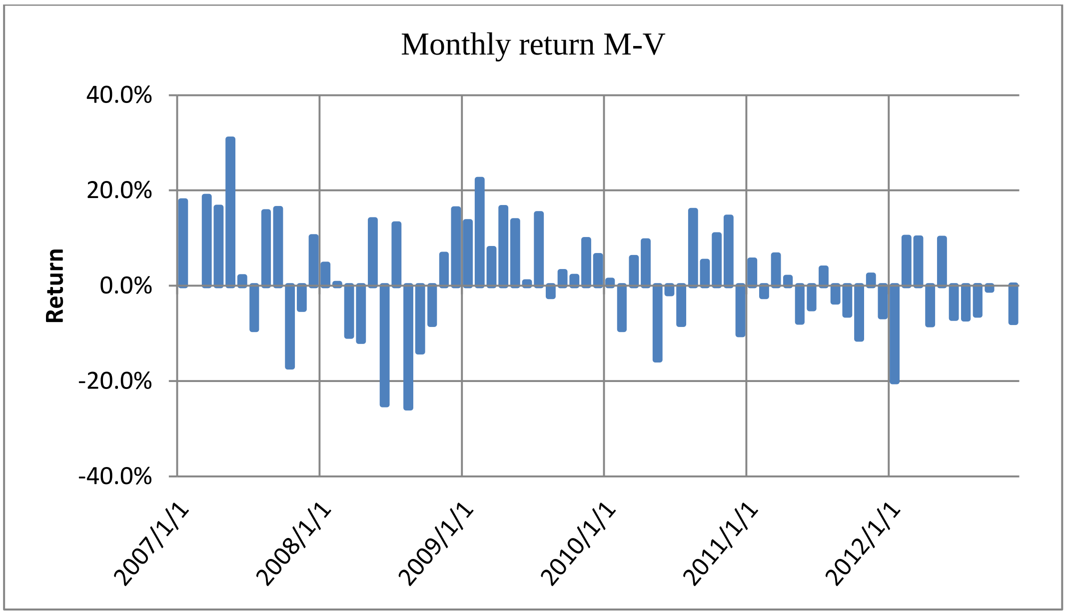


Figure 4-3 Distribution of monthly returns of the two portfolios starting from January 4, 2007 to December 29, 2012. (a) The M-V portfolio. (b) The M-AVaR portfolio

(a)



(b)

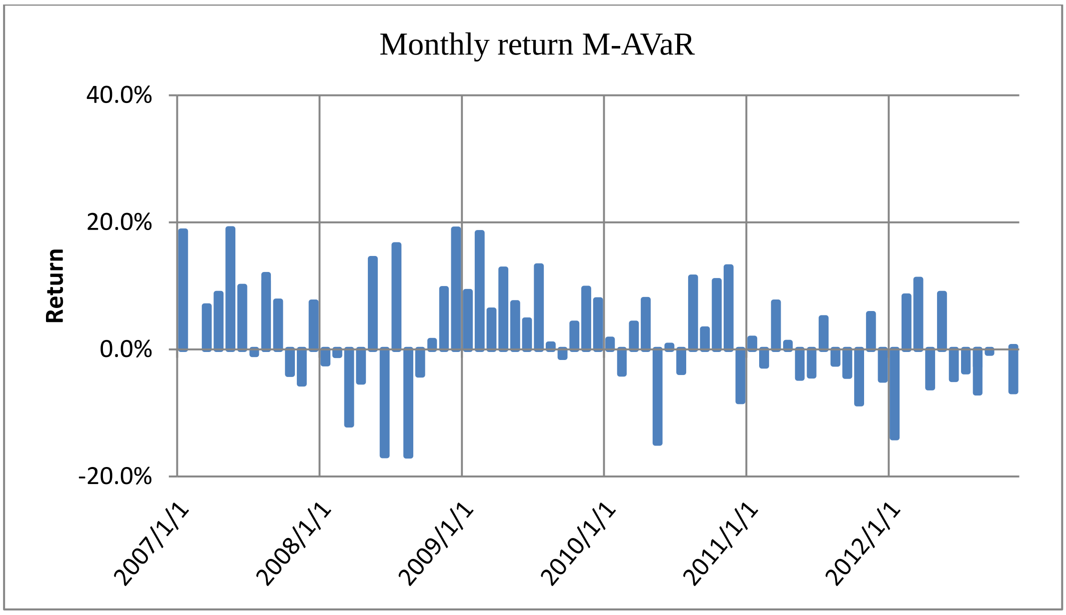
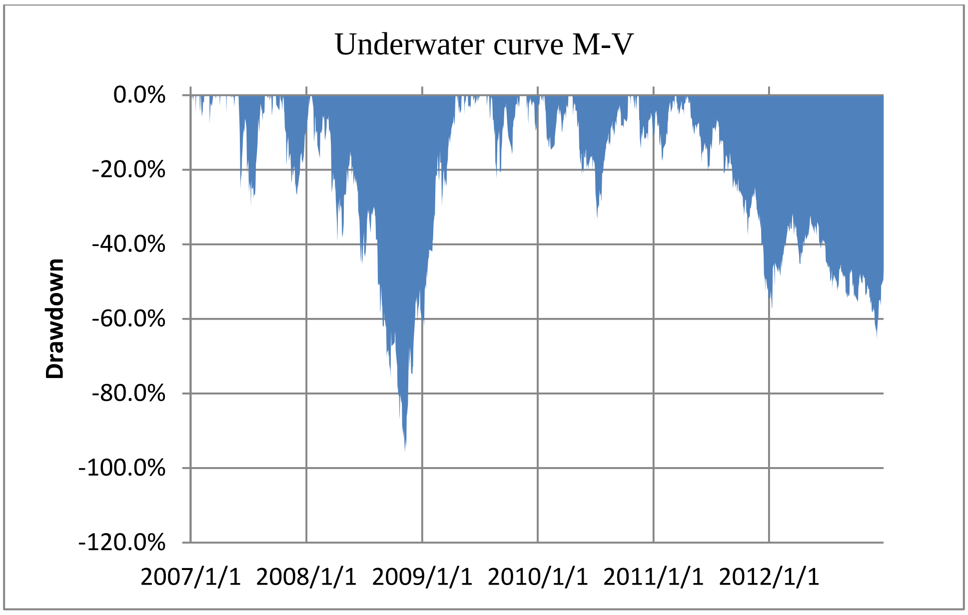


Figure 4-4 Time series of monthly returns of the two portfolios starting from January 4, 2007 to December 29, 2012. (a) The M-V portfolio. (b) The M-AVaR portfolio.

(a)



(b)

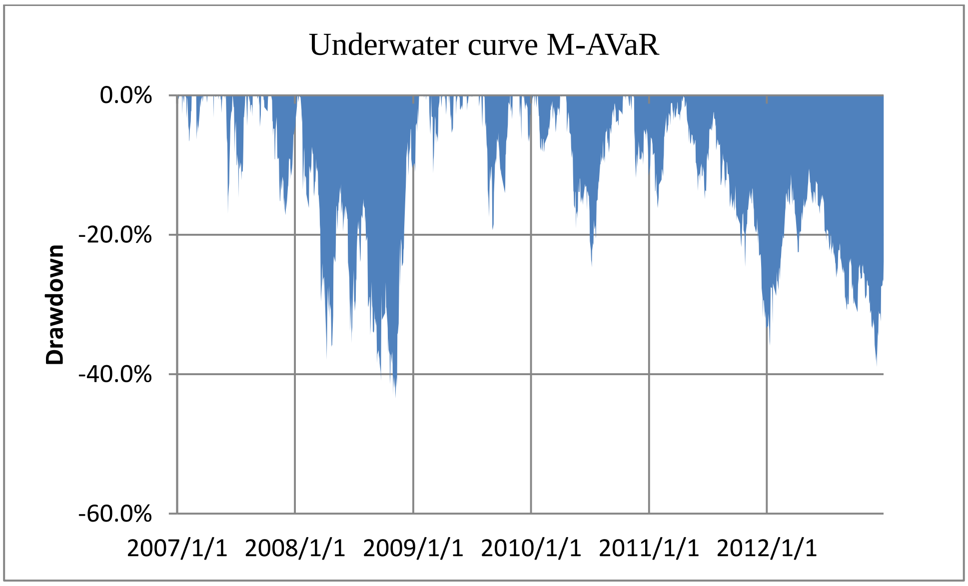


Figure 4-5 Underwater curves. The drawdowns of the two portfolios starting from January 4, 2007 to December 29, 2012. (a) The M-V portfolio. (b) The M-AVaR portfolio.

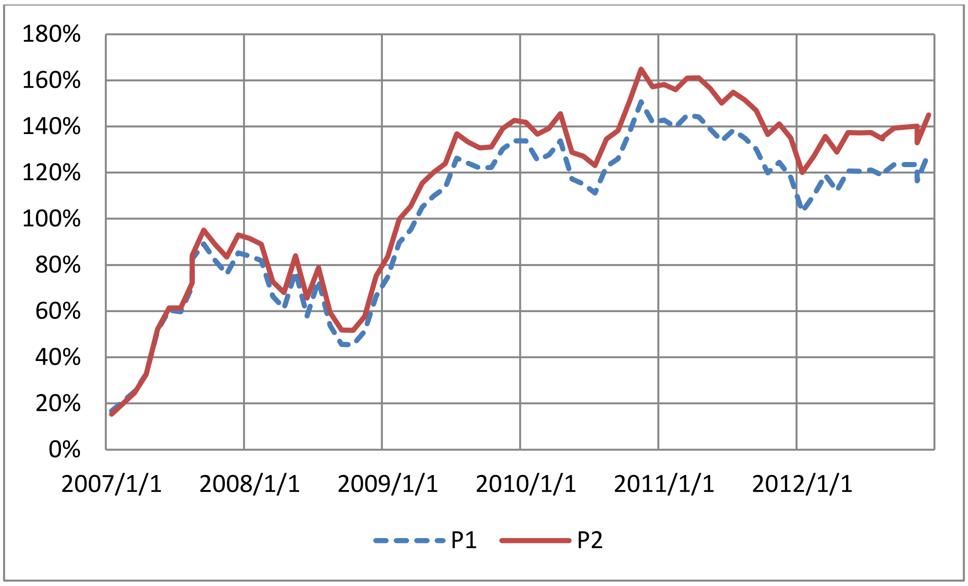


Figure 4-6 Cumulative return of portfolio P1 and P2 starting from January 4, 2007 to December 29, 2012.

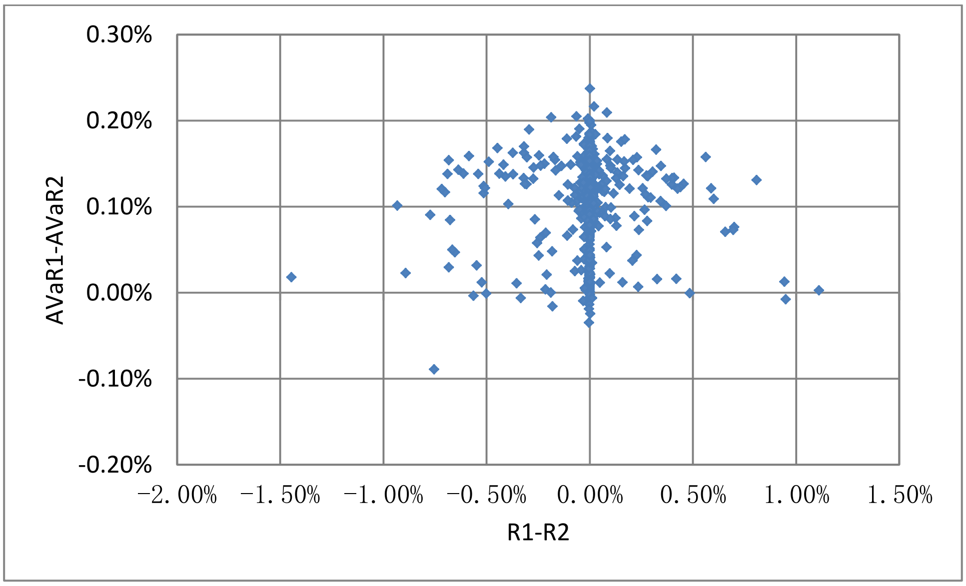
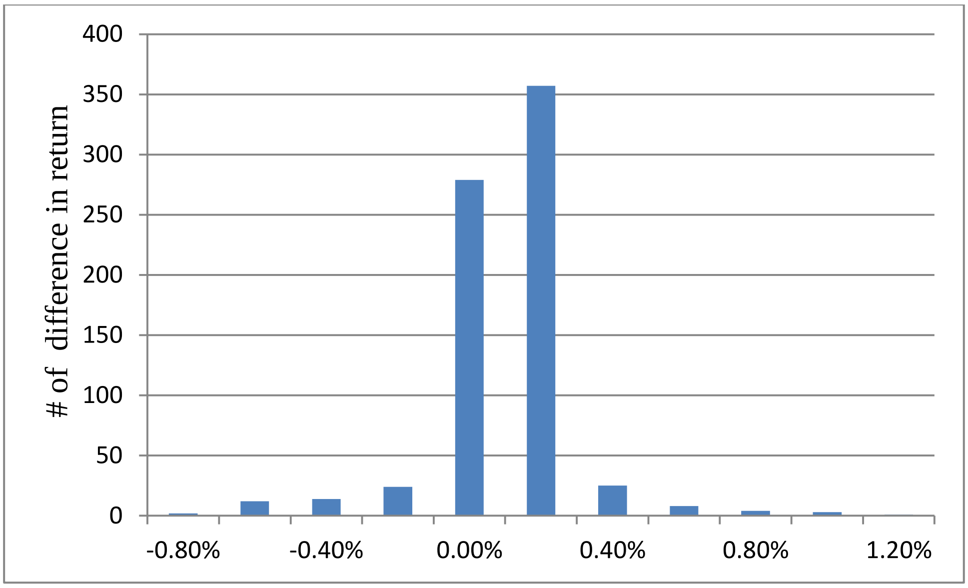


Figure 4-7 Scalar plot of $R_{P1} - R_{P2}$ and $AVaR_{P1} - AVaR_{P2}$



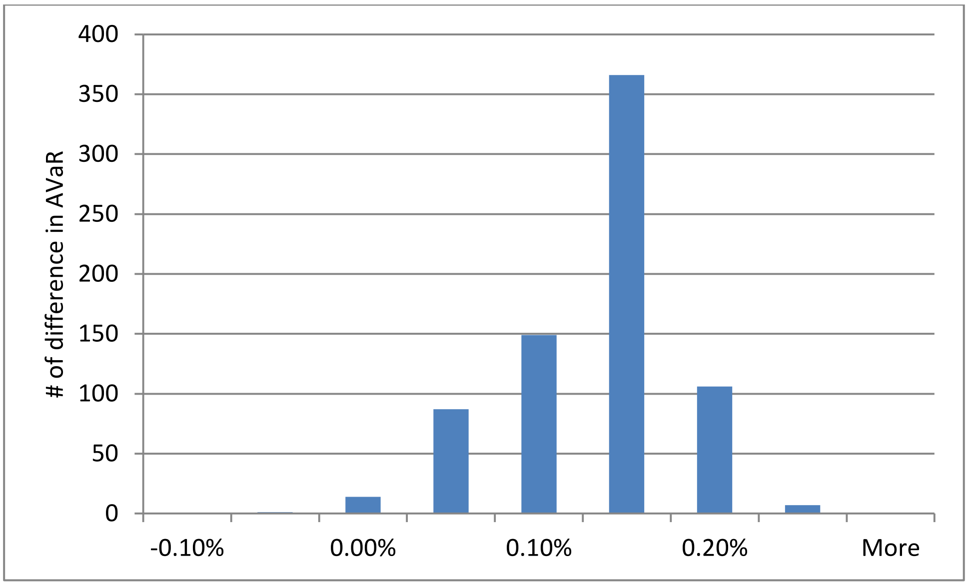


Figure 4-8 Distribution of $R_{P1} - R_{P2}$ and $AVaR_{P1} - AVaR_{P}2$. (a) $R_{P1} - R_{P2}$. (b) $AVaR_{P1} - AVaR_{P2}$. R_{P1} and R_{P2} stand for returns of P1 and P2, and $AVaR_{P1}$ and $AVaR_{P2}$ stand for the values of AVaR for P1 and P2.

5.Conclusions

In this thesis, we presented a factor model based framework for market risk estimation and portfolio optimization. The framework is based on multivariate factor model with advanced time series forecasting approach enhanced by heavy-tailed, skewed distributions and copula.

Started with an overview empirical examination of the return distributions for Chinese stock market, we found the appropriate time series model providing superior modeling capability and forecasting signals for market crashes.

A multi-factor model is then created for modeling returns and risk of stocks by introducing common factors to stocks returns, such as market factors, style factors and industry factors. As far as modeling stand-alone variables is concerned, we considered ARMA-GARCH process with innovations following the class of stable distributions and the class of classical tempered stable distributions. For capturing the dependence structure between variables, we considered a skewed *t* copula approach.

The described framework is a forward looking tool as it is based on the Monte Carlo method. All risk statistics are computed on the basis of generated scenarios from the fitted multivariate model. The downside risk measure we considered is the value-at-risk and average value-at-risk, and the latter one not only provides more information about the tail behavior beyond a certain point, but also is a coherent risk measure. As a consequence, we can incorporate the average value-at-risk with the heavy-tailed model and build risk budgeting tools and forward-looking portfolio optimization tools.

The recent turbulent evens in 2008 represent clear evidence that measuring and managing properly risk is a complicated task. This task has to be performed looking at the portfolio from different angles with different tools. For this purpose, practitioners need an integrated framework built upon realistic assumptions.

The empirical examples presented here and published in a number of papers support the concept.

Bibliography

[1] Aas, K., Haff, I.H., 2005. The generalized hyperbolic skew Student's t-distribution. *Journal of Financial Econometrics,* 4, 275-309.

[2] Acerbi, C., Tasche, D., 2002. On the coherence of expected shortfall. Journal of Banking and Finance, 26, 1487-1503.

[3] Alexander, C., 2001. *Market Models: A guide to Financial Data Analysis.* John Wiley & Sons, Inc, New Jersey.

[4] Anderson, T. W., and Darling, D. A., 1952, Asymptotic Theory of Certain "Goodness-of-Fit" Criteria Based on Stochastic Processes. *Annals of Mathematical Statistics,* 23, 193–212.

[5] Anderson, T.W., Darling, D.A., 1952. A test of goodness-of-fit. *Journal of the American Statistical Association,* 69, 765-769.

[6] Artzner, P., Delbaen, F., Eber, J-M., and Health, D., 1999, Coherent Measure of Risk. *Mathematical Finance,* 9, 203-228.

[7] Barndorff-Nielsen, O. E., and Levendorskiĭ, S. Z., 2001, Feller Processes of Normal Inverse Gaussian Type. *Quantitative Finance*, 1, 318-331.

[8] Barndorff-Nielsen, O.E., 1997. Normal inverse Gaussian and stochastic volatility modeling. *Scandinavian Journal of Statistics,* 24, 1-13.

[9] Barndorff-Nielsen, O.E., Shephard, N., 2005. *Continuous Time Approach to Financial Volatility.* Cambridge University Press. Cambridge, U.K.

[10] Berkowitz, J., 2001, Testing Density Forecasts, With Applications to Risk Management. *Journal of Business and Economic Statistics,* 19, 465-474.

[11] Bianchi, M. L., Rachev, S. T., Kim, Y. S., and Fabozzi, F. J., 2010, Tempered Stable Distributions and Processes In finance: Numerical Analysis. *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, 33-42.

[12] Bianchi, M. L., Rachev, S. T., Kim, Y. S., Fabozzi, F. J., 2010. Tempered infinitely divisible distributions and processes. *Theory of Probability and Its Applications (TVP)*, Society for Industrial and Applied Mathematics (SIAM). 55, 59-86.

[13] Bollerslev, T., 1986, Gneralized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, 307-327.

[14] Box, G. E. P., and Jenkins, G. M., 1970. *Time Series Analysis: Forecasting and Control* (John Wiley & Sons, Inc.).

[15] Boyarchenko, S. I., and Levendorskiĭ, S. Z., 2000, Option pricing for truncated Lévy processes. *International Journal of Theoretical and Applied Finance,* 3, 559-552.

[16] Briner, B. G., Smith, R. C., Ward, P., 2009. The Barra Europe equity model (EUE3), research notes. MSCI Barra.

[17] Campbell, J.Y., Lo, A.W., MacKinley, A.C., 1997. *The Econometrics of Financial Markets.* Princeton University Press, New Jersey.

[18] Carr, P., Geman, H., Madan, D., Yor, M., 2002. The fine structure of asset returns: an empirical investigation. *Journal of Business* 75, 305–332.

[19] Charles, A., and Darné, O., 2009, The random walk hypothesis for Chinese stick markets: Evidence from variance ratio tests. *Economic Systems*, 33, 117-126.

[20] Christoffersen, P. F., 1998, Evaluating interval forecasts. *International Economic Review*, 39, 841-862.

[21] Connor, G., 1995. The three types of factor models : a comparison of their explanatory power. *Financial Analysts Journal*, 84, 61-84.

[22] Davison, A. C., Hinkley, D. V., 1997. *Bootstrap Methods and Their Application*. Cambridge University Press. Cambridge, England.

[23] DeGiorgi, E., 2005. Reward-risk portfolio selection and stochastic dominance. Journal of banking and finance, 29(4), 895-926.

[24] Eberlein, E., Keller, U., 1995. Hyperbolic distributions in finance. *Bernoulli*, 1, 281-299.

[25] Eberlein, E., Keller, U., Prause, K., 1998. New insights into smile, mispricing, and value at risk: the hyperbolic model. *Journal of Business*, 38, 371-405.

[26] Embrechts, P., McNeil, A.J., Straumann, D., 2002. Correlation and dependence in risk management: properties and pitfalls. In *Risk Management: Value at Risk and Beyond*. 176-233. Cambridge Universe Press, Cambridge, UK.

[27] Engle, R. F,, 1982, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50, 987-1007.

[28] Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Ecomomics. 33, 3-56.

[29] Fama, E. F., French, K. R., 2012. Size, value and momentum in international stock returns. Journal of Financial Economics. 105, 457-472.

[30] Fifield, S. G. M., and Jetty, J., 2008, Further evidence on the efficiency of the Chinese stock markets: A note. *Research in International Business and Finance*, 22, 351-361.

[31] Friedmann, R. and Sanddorf-Köhle, W.G., 2002, Volatility Clustering and Nontrading Days in Chinese Stock Markets. *Journal of Economics and Business*, 54, 193-217.

[32] Frittelli, M., Rosazza Gianin, E., 2002. No contagion, only interdependence: measuring stock market co-movements. Journal of Finance. 67, 2223-2261.

[33] Genton, M.G., 2004. *Skew-Elliptical Distributions and Their Applications.* Chapman &

Hall/CRC. Boca Ration, Florida.

[34] Guthoff, A., Pfingsreb, A., Wolf, J., 1997. On the compatibility of value at risk, other risk concepts and expected utility maximization. Diskussionsbeitrag 97-01, Westfaelische wilhelms-Universitaet Muenster.

[35] Hou, A.J. 2013, Asymmetric effects of shocks in Chinese stock markets volatility: A generalized additive nonparametric approach. *Journal of International Financial Markets, Institutions & Money*, 23, 12-32.

[36] Kim, Y. S., Rachev, S. T., Bianchi, M. L., and Fabozzi, F. J., 2008, Financial market models with Lévy processes and time-varying volatility. *Journal of Banking & Finance*, 32, 1363-1378.

[37] Kim, Y. S., Rachev, S. T., Bianchi, M. L., Fabozzi, F. J., 2007. A new tempered stable distribution and its application to finance. In G. Bol, S. T. Rachev, and R. Wuerth (Eds.), *Risk Assessment: Decisions in Banking and Finance*, 77–110. Physika-Verlag, Springer.

[38] Kim, Y. S., Rachev, S. T., Bianchi, M. L., Mitov, I., and Fabozzi, F. J., 2011, Time series analysis for financial market meltdowns. *Journal of Banking & Finance*, 35, 1879-1891.

[39] Kim, Y., Rachev, S., Bianchi, M., Fabozzi, F. (2010). Tempered stable and tempered infinitely divisible GARCH models. *Journal of Banking and Finance*, 34, 2096-2109.

[40] Kim, Y.S., Giacametti, R., Rachev, S.T., Fabozzi, F.J., Mignacca, D., 2012. Measuring financial risk and portfolio optimization with a non-Gaussian multivariate model. *Annals of Operational Research*. 201, 325-343.

[41] Kim, Y.S., Rachev, S.T., Chung, D.M., Bianchi, M.L., 2009. The modified tempered stable distribution, GARCH-models and option pricing. *Probability and Mathematica Statistics*. 29, 91-117.

[42] Koponen, I., 1995. Analytic approach to the problem of convergence of truncated Levy flights towards the Gaussian stochastic process. *Physical Review*, 52, 1197-1199.

[43] Kotz, S., Kozubowski, T.J., Podgórski, K., 2001. *The Laplace Distribution and Generalizations*. Birkhäuser Boston, MA.

[44] Kotz, S., Nadarajah, S., 2004. *Multivariate t Distributions and Their Applications*. Cambridge University Press. Cambridge, U.K.

[45] Kupiec P., 1995. Techniques for verifying the accuracy of risk management models. *Journal of Derivatives*, 3,73-84.

[46] Li, X.M., 2003a, China: further evidence on the evolution of stock markets in transition economies. *Scottish Journal of Political Economy*, 50, 341-358.

[47] Li, X.M., 2003b, Time-varying informational efficiency in China's A-share and B-share market. *Journal of Chinese Economics and Business Studies*, 1, 33-56.

[48] Lu, T.Y., Mitov, I.K., Fabozzi, F.J., Kim, Y.S., 2013. An empirical examination of return

distributions for Chinese stocks. *Emerging Markets Review.* Under reviewing.

[49] MacNail, A. J., Frey, R., Embrechts, P., 2005. *Quantitative Risk Management, Concepts, Techniques, Tools.* Princeton University Press. New Jersey.

[50] Madan, D.B., Carr, P., Chang, E.C., 1998. The variance gamma process and option pricing. *European Finance Review,* 2, 74-105.

[51] Mandelbrot, B.B., 1963.The variation of certain speculative prices. *Journal of Buisiness*, 36, 392-417.

[52] Markowitz, H., 1952. Portfolio selection. Journal of Finance. 7, 77-91.

[53] McNeil, A.J., Demarta, S., 2007. The *t* copula and related copulas. *International Statistical Review.* 73, 119-129.

[54] Menchero, J., Orr, D .J., Wang, J., 2011. The Barra US equity model (USE4), methodology notes. MSCI Barra.

[55] Meng, X.L., Rubin, D.B., 1993. Maximum likelihood estimation of via the ECM algorithm: a general framework. *Biometrika*, 80, 267-278.

[56] Meucci, A., 2007. *Risk and Asset Allocation.* Springer Berlin Heidelberg, New York.

[57] Miller, G., 2006. Needles, haystacks, and hidden factors. Journal of Portfolio Management. 32, 25-32.

[58] Mookerjee, J. and Yu, X., 1999, An empirical analysis of the equity markets in China. *Review of Financial Economics*, 8, 41-60.

[59] Pearson, N.D., 2002. *Risk Budgeting: Portfolio problem with Value-at-Risk.* John Wiley & Sons, Inc. New Jersey.

[60] Protassov, R.S., 2004. EM-based maximum likelihood parameter estimation of multivariate generalized hyperbolic distributions with fixed λ. *Statistics and Computing,* 14, 67-77.

[61] Rachev, S. T., and Mittnik, S., 2000, *Stable Paretian Models in Finance* (John Wiley & Sons, Inc.)

[62] Rachev, S. T., Kim, Y. S., Bianchi, M. L., and Fabozzi, F. J., 2011, *Financial Models with Lévy Processes and Volatility Clustering* (John Wiley & Sons, Inc.).

[63] Rachev, S. T., Menn, C., and Fabozzi, F. J., 2005b, *Fat-Tailed and Skewed Asset Return Distributions: Implications for Risk Management, Portfolio Selection, and Option Pricing* (John Wiley & Sons, Inc.).

[64] Rachev, S. T., Stoyanov, S. V., and Fabozzi, F. J., 2008, *Advanced Stochastic Models, Risk Assessment, and Portfolio Optimization: The ideal Risk, Uncertainty, and Performance Measures* (John Wiley & Sons, Inc.).

[65] Rachev, S. T., Stoyanov, S., Biglova, A., and Fabozzi, F. J., 2005a, An Empirical

Examination of Daily Stock Return Distributions for U.S. Stock. *Data Analysis and Decision Support*, Springer Berlin Heidelberg, 269-281.

[66] Rachev, S. T., Stoyanov, S., Wu, C., and Fabozzi, F. J., 2007, Empirical Analyses of Industry Stock Index Return Distributions for Taiwan Stock Exchange. *Annals of Economics and Finance*, 8, 21-31.

[67] Rockafellar, R. T., and Uryasev, S., 2000, Optimization of Conditional Value-at-Risk. *Journal of Risk*, 2, 21-42.

[68] Rockafellar, R. T., and Uryasev, S., 2002, Conditional Value-at-Risk for general loss distribution. *Journal of Banking & Finance*, 26, 1443-1471.

[69] Rosenberg, B., 1974. Extra-market components of covariance among security prices. Journal of Financial and quantitative analysis. 9, 263-274.

[70] Rosenberg, B., Marathe, V., 1975. Tests of capital asset pricing hypotheses. Research Program in Finance. 32.

[71] Rosiński, J., 2007, Tempering stable processes. *Stochastic Processes and their Applications*, 117, 677-707.

[72] Samorodnitsky, G., Taqqu, M., 1994. Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance . CRC Press., Boca Raton, FL.

[73] Scherer, B., Balachander, B., Falk, R., Yen, B., 2010. Introducing Capital IQ's fundamental us equity risk model. S&P Capital IQ. July.

[74] Sharpe, W.F., 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. Journal of Finance. 19, 425-442.

[75] Stoyanov, S., Samorodnitsky, G., Rachev, S. T., and Ortobelli, S., 2006, Computing the portfolio conditional value-at-risk in the a-stable case. *Probability and Mathematical Statistics*, 26, 1-22.

[76] Tasche, D., 1999. Risk contribution and performance measurement. Preprint, department of Mathematics, TU-München.

[77] Tasche, D., 2000. Conditional expectation as a quantile derivative. Preprint, department of Mathematics, TU-München.

[78] Wang, P., Liu, A. and Wang, P.., 2004, Return and risk interactions in Chinese stock markets. *Journal of International Financial Markets, Institutions and Money*, 14, 367-384.

[79] Zivot, E., Wang, J., 2003. *Modeling Financial Time Series with S-PLUS*. Springer.