Extensions of the VaR Approach to Portfolio Selection with Non-normal Returns

by

Wai Tse

B.Sc. (Mathematics), Hong Kong University of Science and Technology, 1995

M.Sc. (Economics), Hong Kong University of Science and Technology, 1997

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

THE FACULTY OF GRADUATE STUDIES

(Department of Statistics)

we accept this thesis as conforming
to the required standard

The University of British Columbia

May 1999

Department of Statistics

The University of British Columbia
Vancouver, Canada

Date May 21, 1999

Abstract

The well-known mean-variance approach to portfolio selection problem proposed by Markowitz (1952) is often citicized for its use of variance as a measure of risk exposure. Recently, Value at Risk (VaR) has become a popular alternative for measuring risk in many firms. Using the idea of VaR, we formulated a chance constrained programming problem for portfolio selection. Untill recently, most real life applications rely on the normality distributional assumption of the asset returns which seems to be inconsistent with the empirical distributions. To relax this assumption, our study focused on the extensions of the VaR approach to portfolio selection to the class of Elliptically Contoured Distributed returns, and time-varying distributed returns. For the later case, we proposed a new solution via empirical distributions. Moreover, a profile map of returns versus risks was proposed so that the optimal portfolio could be identified for various time-window sizes. The performance of various portfolios over different time periods were evaluated by means of off-sample cumulative returns and a new return-to-risk measure.

Contents

List of Tables

List of Figures

Acknowledgements

I would like to acknowledge the valuable supports and comments of my supervisor, Professor Jian Liu. I am also indebted to Professor Harry Joe for many of his helpful advice.

I wish to thank Fred and Kelly for always helping me and making my stay in Vancouver more enjoyable. I also have to say that I am so lucky to have many friends in Vancouver, especially Don, Maggie, Alice, Phoebe, Ida, Katy, Tiffany, Leonard, Wayne, Dieter, Marc, Jochen, Alex, Matías and Kenneth. Furthermore, I would like to thank for all the support and help from all the graduate students, faculty members and staff in the department.

Finally, I would like to thank my family for all the support and love they always give me.

RYAN W. TSE

The University of British Columbia

April 1999

To my family

Chapter 1

Introduction

In the development of portfolio theory, a well-known approach is the mean-variance approach proposed by Markowitz (1952). His idea was to choose a portfolio such that it could maximize the expected returns under certain risk control. In his paper, variance was used for controlling the risk. However, variance takes into account the fluctuations of returns in both directions, high and low. It may not be a good measure of risk since the former direction is not something the investors worry about. In the following, we will define a portfolio selection problem consisting of maximization of expected returns subject to another risk control.

Having this concern in mind, researchers have tried to find other measurements of risk exposure. These measures include the mean absolute deviation and the frequency of "down sides". Recently, a popular measurement called *Value at Risk* (VaR) is being widely used in risk management by the industry. Since VaR is an appropriate measure of risk, it may be applied to the portfolio selection problem as well. In fact, a similar idea has been suggested by Roy (1952), where he called it safety-first principle. Different forms of such a principle were suggested for various objectives. In addition to the development of *chance constrained programming*, first proposed by Charnes and Cooper (1959), portfolio selection with VaR as the risk measure has become an alternative to the mean-variance portfolio selection model.

In the process of actual portfolio selection, the issue of the distributional assumption arises. So far, most existing practices used the normality assumption. Certainly, the violation of the assumption will not give us the best choice of portfolios. In chapter 2, we will extend the VaR approach (or the chance constrained approach) to our portfolio selection problem, with specific attention to a more general class of distributions, the *Elliptically Contoured Distributions.* An introduction of this distribution class will be presented, and we will discuss how it can be used in practice. Furthermore, we will allow the distribution to be dependent on time and will make use of the empirical distributions.

In chapter 3, we will illustrate the results from an empirical study on a data set of international stock exchange indices. We will attempt to find the best portfolio selection scheme among the traditional VaR approach and our extended VaR approaches by using some new comparison criteria. In addition, we will also examine the effect of using different window sizes of data in forming portfolios. By window size we mean the number of daily values we used in constructing the portfolio. We will also test the risk management ability of the models by the performances of various portfolio selection schemes for the period including the Asian financial crisis beginning late October of 1997.

Although we have extended the VaR approach of portfolio selecting problem to a much more general situation, there are still many areas where further research can be done. We will discuss them in chapter 4, and suggest some possible directions for future research.

Finally, chapter 5 concludes with a summary of our main ideas and findings.

Chapter 2

Portfolio Selection Theory

2.1 Introduction

Consider an investor who has certain amount of capital to invest in k different risky assets and one riskless asset. Suppose the continuously compounded rates of return of these k risky assets are the random variables $r_1, r_2, ..., r_k$, respectively, and the riskless asset has a fixed rate of return r_0. The investor would like to know how to optimally allocate his capital into these assets. Let $a_1, a_2, ..., a_k$ be the portions of the capital allocated to the k risky assets respectively, and a_0 be the portion invested in the riskless asset. The well-known mean-variance approach of portfolio selection due to Markowitz (1952) suggests the investor to solve the following problem:

$$\mathrm{Max}_{\{\mathbf{a_{(0)}}\}} \mathrm{E}(\sum_{i=0}^{k} a_i r_i)$$

$$\text{s.t. } \mathrm{Var}(\sum_{i=0}^{k} a_i r_i) = \mathrm{V}_0,$$

$$\sum_{i=0}^{k} a_i = 1,$$

$$\text{and } a_i \geq 0, \; i = 0, \ldots, k.$$

where $\mathbf{a}_{(0)} = (a_0, a_1, \ldots, a_k)'$, and V_0 is the risk (or variance here) the investor is exposed to. The resulting expected portfolio return and V_0 will be on the efficient portfolio frontier. Markowitz (1959, p.22) defined a portfolio to be efficient if "it is impossible to obtain a greater average [expected] return without incurring greater standard deviation; it is impossible to obtain smaller standard deviation without giving up return on the average."

As mentioned in the previous chapter, variance or standard deviation does not seem to be a good measure of risk because it treats both up-and-down sides equally as risk. Thus, we will form our portfolio selection problem on the following basis:

Maximize the portfolio's expected return

subject to a certain risk control.

In our study, we will assume the measure of risk to be Value at Risk (VaR). Intuitively, VaR is the amount of loss from the original capital to a disaster level at certain probability. Disasters are events like market crashes and bankruptcies.

In fact, some early use of VaR idea in portfolio selection can be traced back to Roy (1952), who proposed the so-called *safety-first principle.* Among the early work, there are several forms of the safety-first principle:

- Min $\Pr\{\sum_{i=0}^{k} a_i r_i \leq d\}$;
- Max d s.t. $\Pr\{\sum_{i=0}^{k} a_i r_i \leq d\} \leq \alpha$;
- Max $E(\sum_{i=0}^{k} a_i r_i)$ s.t. $\Pr\{\sum_{i=0}^{k} a_i r_i \leq d\} \leq \alpha$;

where d is the disaster level of returns and α is the probability that a disaster comes true. The first form was used by Roy (1952). The second one was suggested by Kataoka (1963) and the last form was proposed by Telser (1955-56).

Pyle and Turnovsky (1970) had illustrated these three forms graphically in the context of finding their solutions in the (mean, standard deviation), or (μ, σ), plane with the use of the efficient portfolio frontier. They also described the relationships of the solutions of the 3 forms of safety-first approaches and the mean-variance approach. Baumol (1963) used a slightly different approach than Telser's. Under the assumption of normally distributed returns, he considered the plane (μ, d) instead of the (μ, σ) plane. The conclusion is that not all efficient portfolios in the (μ, σ) plane are reasonable. He then defined efficient portfolios in his (μ, d) plane in the way that a portfolio is efficient if it is impossible to obtain a greater expected return without incurring a greater disaster level (smaller d) with the same given probability, and it is impossible to obtain a lesser disaster level (larger d) with the same given probability without giving up some expected return. Moreover, the efficient portfolios in the (μ, d) plane will be an appropriate subset of the efficient portfolios in the (μ, σ) plane which contains reasonably efficient portfolios. This gives another reason for us to use safety-first approach instead of the mean-variance approach.

In this study, our portfolio selection model via VaR coincides with the form suggested by Telser (1955-56), which in turn is a special case of the chance constrained programming model proposed by Charnes and Cooper (1959). It can be seen that our model is very comparable with the mean-variance approach, and is commonly used in practice.[1] Due to the situation that we have a constraint which is a probabilistic statement, we need to transform it into an equivalent deterministic one in order to solve the optimization problem. To do so, many authors (e.g. Pyle and Turnovsky (1970)) have noticed that we can handle the cases where the portfolio return follows some distributions containing only 2 parameters - mean and variance. However, most papers assumed normally distributed returns when real data were used.

It has been noted that many return series tend to have heavier tails in their distributions and hence multivariate normal distribution model for the returns is not appropriate. In section 2.2, we will consider the class of Elliptically Contoured (EC) Distributions, which includes both the multivariate normal and multivariate-t families. With the assumption that the returns have a joint distribution in this class, we can simplify the portfolio selection problem, which will be stated formally in section 2.3, into a solvable one. In addition to the generalization to the class of EC Distributions, we consider

[1]The paper by Agnew, Agnew, Rasmussen and Smith (1969) is a good reference of the application of Telser's model in portfolio selection in a casualty insurance firm. Similar to Baumol's paper, they also assumed normality.

to relax the assumption of time-invariant distributional form for the returns. The motivation is that the distribution may be changed when the outside environments change. An example is the distribution of stock returns before and after market crashes. To have such a generalization, in section 2.4, we extend the distributional model-based VaR criterion to that based on the empirical distribution of returns and consequently solve the portfolio optimization problem via the extended VaR approach.

In section 2.5, we will implement the Cutting Plane method, suggested by Kelley (1960) in the context of chance-constrained problem, to find the numerical solution to our portfolio selection problem. The actual computer algorithm used corresponds to the Supporting Hyperplane algorithm of Veinott (1967).

2.2 Elliptically Contoured (EC) Distributions

In classical multivariate analysis, a commonly used basic distributional assumption is the multivariate normal distribution. For more general cases, statisticians studied a class of distributions, namely *Elliptically Contoured (EC) Distributions*, which can be considered as an extension of the multivariate normal distribution. Two excellent references of this class of distributions are the books by Fang, Kotz and Ng (1989), and Fang and Zhang (1989). In this section, we will only highlight some properties of this class of distributions which are useful in our portfolio selection problem. The proofs of the theorems and the detail descriptions of EC Distributions can be found in the above references.

It is well known that a k-dimensional random vector $\mathbf{x}$, which follows $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, has the same distribution function as $\boldsymbol{\mu}+\mathbf{A}'\mathbf{y}$, where $\mathbf{y}$ follows the k-dimensional multivariate standard normal distribution $N(\mathbf{0}, \mathbf{I})$ and $\boldsymbol{\Sigma} = \mathbf{A}'\mathbf{A}$. Therefore, $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a generalization of $N(\mathbf{0}, \mathbf{I})$ and many properties of $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ are parallel to those of $N(\mathbf{0}, \mathbf{I})$. As we mentioned before, the class of EC distributions is the extension of $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then the extension of $N(\mathbf{0}, \mathbf{I})$ is the class of *Spherical Distributions*, a sub-class of EC distributions. Similar to $N(\mathbf{0}, \mathbf{I})$ and $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, these two classes of distributions also have many similar properties. Thus, we will first focus on the discussion of spherical distributions.

Definition 2.1: (Fang, Kotz and Ng (1989, p.27))
A $k \times 1$ random vector $\mathbf{x}$ has a spherical distribution if for every $\boldsymbol{\Gamma} \in O(k)$, $\boldsymbol{\Gamma}\mathbf{x}$ has the same distribution as $\mathbf{x}$, where $O(k)$ denotes the set of $k \times k$ orthogonal matrices.

The above definition may not be helpful in checking whether a distribution is in the class of spherical distributions. However, we can make use of the following theorem.

Theorem 2.1: (Fang, Kotz and Ng (1989, p.27))

A $k \times 1$ random vector $\mathbf{x}$ has a spherical distribution if and only if its characteristic function (c.f.) $\psi(\mathbf{t})$ satisfies one of the following equivalent conditions:

1. $\psi(\mathbf{\Gamma}'\mathbf{t}) = \psi(\mathbf{t})$, $\forall \mathbf{\Gamma} \in O(k)$;

2. $\exists$ a scalar function $\phi(\cdot)$ such that $\psi(\mathbf{t}) = \phi(\mathbf{t}'\mathbf{t})$.

The function $\phi(\cdot)$ is called the *characteristic generator* of the spherical distribution. Using similar notations as in Fang, Kotz and Ng (1989), the family of all possible characteristic generators for all $k \times 1$ random vectors is

$$\mathbf{\Phi_k} = \{\phi(\cdot) : \ \phi(t_1^2 + \ldots + t_k^2) \text{ is a } k\text{−dimensional c.f.}\}.$$

The second condition is very helpful when we want to know whether a $k \times 1$ random vector $\mathbf{x}$ has a spherical distribution. If the condition is satisfied, we write $\mathbf{x} \sim S_k(\phi)$. Let us illustrate the idea with the following examples.

Example 2.1: (Fang, Kotz and Ng (1989, p.28))

Suppose $\mathbf{x} \sim N_k(\mathbf{0}, \mathbf{I})$, i.e., multivariate normal, the c.f. of $\mathbf{x}$ is

$$\psi(\mathbf{t}) = \exp\{-\frac{1}{2}(t_1^2 + \cdots + t_k^2)\} = \exp\{-\frac{1}{2}\mathbf{t}'\mathbf{t}\}.$$

We can see that $\psi(\mathbf{t}) = \phi^*(\mathbf{t}'\mathbf{t})$, where $\phi^*(u) = \exp(-\frac{u}{2})$. Therefore, $\mathbf{x} \sim S_k(\phi^*)$ with the characteristic generator $\phi^*(\cdot)$.

Example 2.2: (Fang, Kotz and Ng (1989, p.86-87))

Suppose $\mathbf{x} \sim Mt_k(m, \ \mathbf{0}, \ \mathbf{I})$, i.e., multivariate t-distribution with degree of freedom m,[2] the c.f. is

1. *m is odd:*

$$\psi_1(\mathbf{t};m)=\frac{\sqrt{\pi}\Gamma(\frac{m+1}{2})\exp(-\sqrt{\mathbf{t'}\mathbf{t}})}{2^{m-1}\Gamma(\frac{m}{2})}\times\sum_{p=1}^{q}[\frac{(2q-p-1)!}{(q-p)!(q-1)!}\frac{(2\sqrt{\mathbf{mt'}\mathbf{t}})^{p-1}}{(p-1)!}],$$

where $q=\frac{m+1}{2}$. We can see that $\psi_1(\mathbf{t};m)=\phi_1^*(\mathbf{t'}\mathbf{t})$. Therefore, $\mathbf{x}\sim S_k(\phi_1^*)$ with the characteristic generator $\phi_1^*(\cdot)$.

2. *m is even:*

$$\psi_2(\mathbf{t};m)=\frac{(-1)^{q+1}\Gamma(\frac{m+1}{2})}{\sqrt{\pi}\prod_{p=1}^{q}(q+\frac{1}{2}-p)\Gamma(\frac{m}{2})}\times\sum_{n=0}^{\infty}\{(\frac{m\mathbf{t'}\mathbf{t}}{4})^n\frac{1}{(n!)^2}(\sum_{p=0}^{q-1}\prod_{o=0,o\neq p}^{q-1}(n-o))$$

$$+\prod_{p=0}^{q-1}(n-p)[\log\frac{m\mathbf{t'}\mathbf{t}}{4}-h(n+1)]\},$$

where $q=\frac{m}{2}$ and $h(n+1)=\frac{\Gamma'(n+1)}{\Gamma(n+1)}$. Also, $\psi_2(\mathbf{t};m)=\phi_2^*(\mathbf{t'}\mathbf{t})$. Therefore, $\mathbf{x}\sim S_k(\phi_2^*)$ with the characteristic generator $\phi_2^*(\cdot)$.

3. *m is a fraction:*

$$\psi_3(\mathbf{t};m)=\frac{(-1)^q\Gamma(\frac{m+1}{2})(\frac{\pi}{\sin\epsilon\pi})}{2^\epsilon\Gamma(\frac{m}{2})\Gamma(\epsilon+\frac{1}{2})\prod_{p=1}^{q}(\frac{m+1}{2-p})}m^{m/2-q}\times$$

$$\sum_{n=0}^{\infty}[(\frac{m\mathbf{t'}\mathbf{t}}{4})^n\frac{1}{(n!)}(\frac{2^\epsilon\prod_{p=0}^{q-1}(n-\epsilon-p)}{m^\epsilon\Gamma(n+1-\epsilon)}-\frac{(\mathbf{t'}\mathbf{t})^\epsilon\prod_{p=0}^{q-1}(n-p)}{2^\epsilon\Gamma(n+1+\epsilon)})],$$

where $q=\frac{m+1}{2}$ is the integer part of $\frac{m+1}{2}$, $\epsilon=\frac{m}{2}-s$, and $0<|\epsilon|<\frac{1}{2}$. Thus, we can also write $\psi_3(\mathbf{t};m)=\phi_3^*(\mathbf{t'}\mathbf{t})$, and $x\sim S_k(\phi_3^*)$ with the characteristic generator $\phi_3^*(\cdot)$.

[2]Stochastic representation of $Mt_k(m,\ \mathbf{0},\ \mathbf{I})$:
If $\mathbf{z}\sim N_k(\mathbf{0},\mathbf{I})$, $s\sim\chi_m^2$, and $\mathbf{z}$ is independent with s, then $\mathbf{y}=\frac{m^{1/2}\mathbf{z}}{\sqrt{s}}$ has a multivariate t-distribution with degree of freedom m.

As a result, we have just shown that the multivariate t-distribution belongs to the class of spherical distributions for any degree of freedom m.

Notice that a $k \times 1$ random vector $\mathbf{x} \sim S_k(\phi)$, in general, does not necessarily have a density. If it does, the density must be of the form $g(\mathbf{x}'\mathbf{x})$ for some nonnegative function $g(\cdot)$ of a scalar variable. Moreover, a nonnegative function $g(\cdot)$ can be used to define a density of a spherical distribution if and only if

$$\int_0^\infty y^{\frac{n}{2}-1} g(y)dy < \infty.$$

We call $g(\cdot)$ a **density generator** or **p.d.f. generator** of the spherical distribution.

Now, we will discuss the family of EC distribution. We are going to start with the following definition.

Definition 2.2: (Fang, Kotz and Ng (1989, p.31))
A $k \times 1$ random vector $\mathbf{x}$ has an EC distribution with parameters $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)'$ and $\boldsymbol{\Sigma}$ if $\mathbf{x}$ has the same distribution as $\boldsymbol{\mu} + \mathbf{A}'\mathbf{y}$, where $\mathbf{y} \sim S_l(\phi)$ and $\mathbf{A}$ is an $l \times k$ matrix such that $\mathbf{A}'\mathbf{A} = \boldsymbol{\Sigma}$ with rank($\boldsymbol{\Sigma}$)=l. We write $\mathbf{x} \sim EC_k(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$.

From the above definition, it can be easily verified that if $\mathbf{x} \sim EC_k(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ with rank($\boldsymbol{\Sigma}$)=l, we can find some scalar function $\phi(\cdot)$ such that the c.f. of $\mathbf{x}$,

$\psi(\mathbf{t}) = E(e^{i\mathbf{t}'\mathbf{x}})$ is of the following form

$$\psi(\mathbf{t}) = e^{i\mathbf{t}'\boldsymbol{\mu}}\phi(\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}).$$

Among the theorems and properties of the family of EC distributions, the following one will be very useful for our investor's problem.

Theorem 2.2: (Fang, Kotz and Ng (1989, p.43))

Suppose $\mathbf{x} \sim EC_k(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ with rank($\boldsymbol{\Sigma}$)=l, $\mathbf{B}$ is a $k \times m$ matrix and $\mathbf{v}$ is an $m \times 1$ vector, then

$$\mathbf{v} + \mathbf{B}'\mathbf{x} \sim EC_m(\mathbf{v} + \mathbf{B}'\boldsymbol{\mu}, \mathbf{B}'\boldsymbol{\Sigma}\mathbf{B}, \phi).$$

A special case of this Theorem is that any linear combination of the components of $\mathbf{x}$, $\mathbf{a}'\mathbf{x}$ where $\mathbf{a} = (a_1, \ldots, a_k)'$, follows $EC_1(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}, \phi)$.

In addition, we will also find that the following theorem is useful in the later sections. This theorem is a generalized version of the one presented in Weintraub and Vera (1991).

Theorem 2.3:

Let $\mathbf{x} = (x_1, \ldots, x_k)' \sim EC_k(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$. If a vector $(a_0^0, \ldots, a_k^0)$, where $a_i^0 \geq 0$ for $i = 0, 1, \ldots, k$, is such that

$$a_0^0 x_0 + \sum_{i=1}^{n} a_i^0 x_i^0 \geq d, \tag{2.1}$$

where x_0 is a constant, and x_i^0 satisfies $\Pr(x_i \geq x_i^0) \geq 1-\alpha$, for $i = 1, \ldots, k$, then the vector $(a_0^0, \ldots, a_k^0)$ will also satisfy the following inequality:

$$a_0^0 x_0 + \mathbf{a^0}'\boldsymbol{\mu} + F^{-1}(\alpha)\sqrt{\mathbf{a^0}'\boldsymbol{\Sigma}\mathbf{a^0}} \geq d, \tag{2.2}$$

where $\mathbf{a^0} = (a_1^0, \ldots, a_k^0)'$, $F(\cdot)$ is the cumulative distribution function of the distribution $EC_1(0, 1, \phi)$ or $S_1(\phi)$, and α is chosen such that $F^{-1}(\alpha) \leq 0$.

Proof. Let us first denote the (i, l) element of the variance-covariance matrix $\boldsymbol{\Sigma}$ as s_{il}, for $i = 1, \ldots, k$ and $l = 1, \ldots, k$. Then we have $s_{il} = \rho_{il} s_i s_l$, where ρ_{il} is the correlation between x_i and x_l, and s_i and s_l are the standard deviations of x_i and x_l respectively.

Applying Theorem 2.2 with $\mathbf{v} = \mathbf{0}$ and $\mathbf{B}$ being a $k \times 1$ vector with value 1 in the i^{th} element and values 0 elsewhere, we have

$$x_i \sim EC_1(\mu_i, s_i^2, \phi) \quad \text{for} \quad i = 1, \ldots, k.$$

For each x_i^0 satisfying $\Pr(x_i \geq x_i^0) = 1 - \alpha$, we have

$$\Pr\left(\frac{x_i - \mu_i}{s_i} \geq \frac{x_i^0 - \mu_i}{s_i}\right) \geq 1 - \alpha,$$

where $\frac{x_i - \mu_i}{s_i} \sim EC_1(0, 1, \phi) = S_1(\phi)$. So, we can re-write the above probability statement as the following:

$$\frac{x_i^0 - \mu_i}{s_i} \leq F^{-1}(\alpha),$$

$$\Rightarrow x_i^0 \leq \mu_i + F^{-1}(\alpha) s_i.$$

Therefore, if $a_0^0 x_0 + \sum_{i=1}^k a_i^0 x_i^0 \geq d$, then

$$a_0^0 x_0 + \sum_{i=1}^k a_i^0 \mu_i + F^{-1}(\alpha) \sum_{i=1}^k a_i^0 s_i \geq d \tag{2.3}$$

On the other hand, from the facts that $\rho_{il} \leq 1$, $s_i \geq 0$, and $a_i^0 \geq 0$, we also have

$$(\sum_{i=1}^{k} a_i^0 s_i)^2 = \sum_{i=1}^{k}\sum_{l=1}^{k} a_i^0 a_l^0 s_i s_l,$$

$$\Rightarrow (\sum_{i=1}^{k} a_i^0 s_i)^2 \geq \sum_{i=1}^{k}\sum_{l=1}^{k} \rho_{il} a_i^0 a_l^0 s_i s_l,$$

$$\Rightarrow \sum_{i=1}^{k} a_i^0 s_i \geq \sqrt{\mathbf{a^{0\prime}\Sigma a^0}}.$$

Hence, with $F^{-1}(\alpha) \leq 0$ we have

$$F^{-1}(\alpha)\sum_{i=1}^{k} a_i^0 s_i \leq F^{-1}(\alpha)\sqrt{\mathbf{a^{0\prime}\Sigma a^0}} \tag{2.4}$$

From (2.3) and (2.4), we get

$$a_0^0 x_0 + \sum_{i=1}^{k} a_i^0 \mu_i + F^{-1}(\alpha)\sqrt{\mathbf{a^{0\prime}\Sigma a^0}} \geq d.$$

Q.E.D.

Theorem 2.3 states how the inequality (2.2) can be approximated by a linear one (2.1) when the random variables have an EC distribution. We will see how it is applied when we describe the cutting plane method.

In the next section, we will illustrate how to handle the portfolio selection problem when the returns follow an EC distribution.

2.3 Portfolio Selection Model with Deterministic Constraints

Recall that our portfolio selection problem is:

$$\text{Max}_{\{\mathbf{a}_{(0)}\}}\text{E}(\sum_{i=0}^{k} a_i r_i)$$

$$\text{s.t.} \quad \Pr\{\sum_{i=0}^{k} a_i r_i \geq d_j\} \geq 1-\alpha_j, \;\; j = 1,\ldots,m,$$

$$\sum_{i=0}^{k} a_i = 1,$$

$$\text{and} \;\; a_i \geq 0, \;\; i = 0,\ldots,k.$$

where $\mathbf{a}_{(0)} = (a_0, a_1, \ldots, a_k)'$, The non-negative constraints for a_i's assume that no short-selling is allowed. Moreover, we can see that there are m probabilistic constraints in the model. Constraint j specifies a certain disaster level of return d_j an investor can tolerate given a certain probability of occurrence α_j. This can capture the situation of multiple levels of risk.

In order to solve the above maximization problem, we first rewrite the probabilistic constraints into their deterministic forms by assuming the returns r_i's for $i = 1,\ldots,k$ following some multivariate distributions. Here, we will assume that they have an EC distribution, i.e.,

$$\mathbf{r} \sim EC_k(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi),$$

where $\mathbf{r} = (r_1, r_2, \ldots, r_k)'$. Using theorem 2.2, we have

$$\mathbf{a}'\mathbf{r} \sim EC_1(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}, \phi),$$

where $\mathbf{a} = (a_1, a_2, \ldots, a_k)'$. Hence, for each probabilistic constraint j, we can write it as

$$\Pr\{y \geq \frac{d_j - a_0 r_0 - \mathbf{a}'\boldsymbol{\mu}}{\sqrt{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}}}\} \geq 1 - \alpha_j,$$

where $y = \frac{\mathbf{a}'\mathbf{r} - \mathbf{a}'\boldsymbol{\mu}}{\sqrt{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}}} \sim EC_1(0, 1, \phi) = S_1(\phi)$. Assuming $F(\cdot)$ is the cumulative distribution function of y, we have the following equivalent deterministic form:

$$\frac{d_j - a_0 r_0 - \mathbf{a}'\boldsymbol{\mu}}{\sqrt{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}}} \leq F^{-1}(\alpha_j),$$

$$\Leftrightarrow a_0 r_0 + \mathbf{a}'\boldsymbol{\mu} + F^{-1}(\alpha_j)\sqrt{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}} \geq d_j.$$

Therefore, the deterministic version of the portfolio selection problem will be:

$$\begin{aligned} &\mathrm{Max}_{\{\mathbf{a}_{(0)}\}} a_0 r_0 + \sum_{i=1}^{k} a_i \mu_i \\ \text{s.t. } \; & a_0 r_0 + \mathbf{a}'\boldsymbol{\mu} + F^{-1}(\alpha_j)\sqrt{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}} \geq d_j, \; j = 1, \ldots, m, \\ & \sum_{i=0}^{k} a_i = 1, \\ & \text{and } \; a_i \geq 0, \; i = 0, \ldots, k. \end{aligned}$$

With the Cutting Plane Algorithm, which will be presented in section 2.5, we will be able to solve the above problem.

However, before we get into the algorithm, we will first discuss in details how the empirical distribution is used in the next section.

2.4 Empirical Distributions

In section 2.3, although we allow for the more general class of distributions to model the returns which can capture to a certain degree the heavier tails, it should be noted that choosing an appropriate distribution for a data set is non-trivial. Also, the distributional form may change as a result of some influential market movements such as market crashes and rallies. This section mainly focuses on using the historical data in constructing the empirical distributions, so that the portfolio selection problem can be solved.

First of all, suppose we have a total of k series of returns data under consideration. We will have k different marginal empirical distributions. The empirical distribution function of the i^{th} series is defined as

$$\hat{H}_i(r) = \frac{1}{t_2 - t_1 + 1} \sum_{j=t_1}^{t_2} \mathbf{I}(r_{ij} \leq r),$$

where r_{ij} is the return of the i^{th} index at date j starting from t_1 to t_2, and r is some values of returns ranging from $-\infty$ to ∞. The function $\mathbf{I}(r_{ij} \leq r)$ is an indicator function which equals 1 when $r_{ij} \leq r$, and 0 otherwise.

However, without any assumption on the distribution of the data, we cannot proceed further as we need to deal with the distribution of the portfolio, i.e., some linear combinations of the returns. This is because before we optimize the portfolio holding, we need to know the distribution of the optimal portfolio. Recall that in the previous section, when we transform the probabilistic constraint into the deterministic form, we need to know the distribution of the linear combination of returns. However, we do not know what the optimal portfolio will be before optimization. Therefore, we need the following assumption:

Assumption: The returns follow a joint distribution (time-varying) which has the following properties:

$$y = \frac{\mathbf{a}'\mathbf{r} - \mathbf{a}'\boldsymbol{\mu}}{\sqrt{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}}}$$

follows the same distribution for any vector **a** at any specified time point.[3]

With the above assumption, it is easy to verify that the result of Theorem 2.3 still holds, and the deterministic version of the portfolio selection problem is the same as that presented in the previous section. The allowance of time-varying distributional form is actually a generalization of the portfolio selection model. The main advantage of using empirical distribution is appealing since we can have different approximations for the true distribution by using different portions of the historical data at different time points we are interested in. Thus the changes in the distributional form of the returns can be captured.

When we solve the portfolio selection problem, we may notice that the only place the distribution of returns being used is in the deterministic equivalence of each of the probabilistic constraints (e.g. constraint j):

[3]It is noted that all the distribution families in the class of EC distribution satisfy this assumption.

$$a_0 r_0 + \mathbf{a}'\boldsymbol{\mu} + F^{-1}(\alpha_j)\sqrt{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}} \geq d_j.$$

Further, we actually only need to know $F^{-1}(\alpha_j)$, the critical value of the standardized distribution (with zero mean and unit variance) of the linear combinations of the returns. When we are using the well-known distributions such as normal or t, the critical value can be easily found from a statistical table or software. However, when we use the empirical distribution for approximation, the problem becomes how to obtain an approximation for the critical value. To solve such problem we need to use the assumption that all linear combinations of the returns have the same distribution up to a scale change. We thus randomly generate 100 linear combinations. For each of them, we have an empirical distribution. Then we standardize them so that the means and variances of the empirical distributions are 0 and 1 respectively. Finally, we obtain the $100\alpha_j - th$ percentile point for each linear combination. Under the assumption, these percentile points should be the same for all the linear combinations. Certainly, when we use real data set, this will never happen. Therefore, we use the average of the $100\alpha_j - th$ percentile points of all linear combinations as the estimate of $F^{-1}(\alpha_j)$.

2.5 Cutting Plane Method

In this section, we will describe the cutting plane method for solving the portfolio selection problem. To be more specific, we mainly concentrate on the supporting hyperplane algorithm of Veinott (1967), Zangwill (1969), and

Weintraub and Vera (1991). Moreover, we will discuss its convergence properties.

In general, Veinott's algorithm can be applied to the following non-linear problem:

$$\text{Max}_{\{\mathbf{a}_{(0)}\}} a_0 r_0 + \sum_{i=1}^{k} a_i \mu_i$$

$$\text{s.t. } \mathbf{a}_{(0)} \in \mathbf{G} \equiv \{\mathbf{a}_{(0)} \mid g_j(\mathbf{a}_{(0)}) \geq d_j, \;\; j = 1, \ldots, m\},$$

where $g_j(\cdot)$, for $j = 1, \ldots, m$, are quasi-concave[4] and continuously differentiable. This means that the feasible set $\mathbf{G}$ is a convex set. We can see that the above non-linear problem has exactly the same objective as that in the portfolio selection problem. In fact, the portfolio selection problem is just a special case. We will discuss it in more detail after we have presented Veinott's algorithm.

We will first start with the assumptions of the algorithm:

1. There exists a compact set $\mathbf{U}$ such that the feasible set $\mathbf{G}$ is contained in it.

2. There exists an interior point $\mathbf{b}_{(0)}$ such that $g_j(\mathbf{b}_{(0)}) > d_j, \;\; j = 1, \ldots, m$.

3. For any $\mathbf{a}_{(0)}$ such that $g_j(\mathbf{a}_{(0)}) = d_j$, we have $\nabla g_j(\mathbf{a}_{(0)}) \neq \mathbf{0}$.

[4]A function $g(\cdot)$ is quasi-concave if and only if the set

$$G_\gamma = \{\mathbf{a} \mid g(\mathbf{a}) \geq \gamma\}$$

is convex for any scalar γ.

When these assumptions are satisfied, we can perform the following algorithm:

1. $\mathbf{Z^1} = \mathbf{U}$, $l = 1$

2. Solve the following linear programming problem for a solution ${\mathbf{u}_{(0)}}^l$:

$$\text{Max}_{\{\mathbf{a}_{(0)}\}} a_0 r_0 + \sum_{i=1}^{k} a_i \mu_i$$

$$\text{s.t. } \mathbf{a}_{(0)} \in \mathbf{Z}^l$$

3. Solution test: if ${\mathbf{u}_{(0)}}^l \in \mathbf{G}$, we can stop and ${\mathbf{u}_{(0)}}^l$ will be our solution. Otherwise, we will go to the next step.

4. Let $I_l \equiv \{1, \ldots, m_l\}$ be the index set where $g_j({\mathbf{u}_{(0)}}^l) < d_j$ for $j \in I_l$, we can find a scalar $\theta^l \in [0, 1]$ such that ${\mathbf{v}_{(0)}}^l = \mathbf{b}_{(0)} + \theta^l({\mathbf{u}_{(0)}}^l - \mathbf{b}_{(0)})$, where $g_j({\mathbf{v}_{(0)}}^l) \geq d_j$ for $j \in I_l$, and $g_{j_0}({\mathbf{v}_{(0)}}^l) = d_{j_0}$ for some $j_0 \in I_l$. Here, ${\mathbf{v}_{(0)}}^l$ is a boundary point of $\mathbf{G}$.

5. Create a set $\mathbf{H}^l \equiv \{\mathbf{a}_{(0)} \mid \nabla g_j({\mathbf{v}_{(0)}}^l)'(\mathbf{a}_{(0)} - {\mathbf{v}_{(0)}}^l) \geq 0, \text{ for } j \in I_l \}$.

6. Let $\mathbf{Z}^{l+1} = \mathbf{Z}^l \cap \mathbf{H}^l$. This will be the new constraint set.

7. $l \leftarrow l + 1$, and go back to step 2.

With all the conditions satisfied, the above algorithm was proven to be convergent, i.e.,

$${\mathbf{u}_{(0)}}^l \longrightarrow {\mathbf{u}_{(0)}}^\infty, \text{ as } l \longrightarrow \infty$$

where $\mathbf{u}_{(0)}{}^{\infty}$ passes the solution test. The details of the convergence property can be refered to Zangwill (1969).

In the portfolio selection problem, the feasible set is

$$\mathbf{G} \equiv \{\mathbf{a}_{(0)} \mid a_0 r_0 + \mathbf{a}'\boldsymbol{\mu} + F^{-1}(\alpha_j)\sqrt{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}} \geq d_j, \ \ j = 1, \ldots, m;$$
$$\sum_{i=0}^{k} a_i = 1, \ \text{and} \ \ a_i \geq 0, \ \ i = 0, \ldots, k\},$$

and we will let

$$g_j(\mathbf{a}_{(0)}) = a_0 r_0 + \mathbf{a}'\boldsymbol{\mu} + F^{-1}(\alpha_j)\sqrt{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}}, \ \ j = 1, \ldots, m,$$

which are quasi-concave (Appendix A) and continuously differentiable. Hence, the feasible set of our investor's problem is convex.[5]

Before proceeding with the supporting hyperplane algorithm, we need to find the set $\mathbf{U}$ and the interior point $\mathbf{b}_{(0)}$.

Existence of U:

To find a compact set $\mathbf{U}$ that contains $\mathbf{G}$, we consider the following compact set:

$$\mathbf{U} \equiv \{\mathbf{a}_{(0)} \mid a_0 r_0 + \mathbf{a}'\boldsymbol{\mu} \geq d_j, \ \ j = 1, \ldots, m;$$
$$\sum_{i=0}^{k} a_i = 1, \ \text{and} \ \ a_i \geq 0, \ \ i = 0, \ldots, k\}.$$

[5]Geometrically, as noted in Agnew, Agnew, Rasmussen and Smith (1969), the feasible set represents one nappe of a hyperboloid and its interior.

Since $F^{-1}(\alpha_j) < 0$ and $\sqrt{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}} \geq 0$, we have

$$a_0 r_0 + \mathbf{a}'\boldsymbol{\mu} \geq a_0 r_0 + \mathbf{a}'\boldsymbol{\mu} + F^{-1}(\alpha_j)\sqrt{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}}, \quad \forall \mathbf{a}_{(0)}.$$

Hence, $\mathbf{G} \subset \mathbf{U}$.

Existence of $\mathbf{b}_{(0)}$:

Applying Theorem 2.3, we can see that for any $\mathbf{a}_{(\mathbf{0})}$ satisfying

$$a_0 r_0 + \sum_{i=1}^{k} a_i r_{ij}^0 \geq d_j,$$

such that $\Pr(r_i \geq r_{ij}^0) \geq 1 - \alpha_j$ or $r_{ij}^0 \leq \mu_i + F^{-1}(\alpha_j)s_i$, for $i = 1, \ldots, k$, then

$$a_0 r_0 + \mathbf{a}'\boldsymbol{\mu} + F^{-1}(\alpha_j)\sqrt{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}} \geq d_j,$$

which is true for $j = 1, \ldots, m$. If we further consider $\tilde{r}_{ij}^0 = \mu_i + \tilde{F}_j^{-1} s_i < \mu_i + F^{-1}(\alpha_j)s_i = r_{ij}^0$, where $\tilde{F}_j^{-1}$ is chosen to be strictly smaller than $F^{-1}(\alpha_j)$, $j = 1, \ldots, m$, we can construct the following convex set:

$$\mathbf{L} \equiv \{\mathbf{a}_{(0)} \mid a_0 r_0 + \sum_{i=1}^{k} a_i r_{ij}^0 \geq d_j + \epsilon, \;\; j = 1, \ldots, m;$$

$$\sum_{i=0}^{k} a_i = 1, \text{ and } a_i \geq 0, \; i = 0, \ldots, k\},$$

where $\epsilon > 0$ is a very small value. Weintraub and Vera had shown that $\mathbf{L} \subset \mathbf{G}$ and the solution obtained from solving

$$\text{Max}_{\{\mathbf{a}_{(0)}\}} a_0 r_0 + \sum_{i=1}^{k} a_i \mu_i$$

$$\text{s.t. } \mathbf{a}_{(\mathbf{0})} \in \mathbf{L}$$

will be interior point of the feasible set $\mathbf{G}$.

One more technical issue in the algorithm we like to discuss here is how the boundary point ${\mathbf{v}_{(0)}}^l$ is determined after we find the solution ${\mathbf{u}_{(0)}}^l$ in the l^{th} iteration. Recall that we need to find ${\mathbf{v}_{(0)}}^l = \mathbf{b}_{(0)} + \theta^l({\mathbf{u}_{(0)}}^l - \mathbf{b}_{(0)})$ such that

$$g_j({\mathbf{v}_{(0)}}^l) \geq d_j \text{ for } j \in I_l,$$

$$\text{and } g_{j_0}({\mathbf{v}_{(0)}}^l) = d_{j_0} \text{ for some } j_0 \in I_l.$$

Now, for each $j \in I_l$, we can find at least one $\theta_j \in [0,1]$ such that

$$g_j[\mathbf{b}_{(0)} + \theta_j({\mathbf{u}_{(0)}}^l - \mathbf{b}_{(0)})] = d_j.$$

Note that if we find more than one $\theta_j \in [0,1]$ satisfying the above equation, we will take the small value as our θ_j. To find θ_j which satisfies it, there are many methods available. Here, we will follow the one used by Weintraub and Vera. After some algebra, it is not hard to find that the above equation can be written as a quadratic equation in θ_j as the following:

$$A_j\theta_j^2 + B_j\theta_j + C_j = 0,$$

where

$$A_j = -(e_0^l r_0 + \sum_{i=1}^{k} e_i^l \mu_i)^2 + [F^{-1}(\alpha_j)]^2 \sum_{i=1}^{k}\sum_{t=1}^{k} s_{it} e_i^l e_t^l,$$

$$B_j = 2(d_j - b_0 r_0 - \sum_{i=1}^{k} b_i \mu_i)(e_0^l r_0 + \sum_{i=1}^{k} e_i^l \mu_i) + 2[F^{-1}(\alpha_j)]^2 \sum_{i=1}^{k}\sum_{t=1}^{k} s_{it} b_t e_i^l,$$

$$C_j = [F^{-1}(\alpha_j)]^2 \sum_{i=1}^{k}\sum_{t=1}^{k} s_{it} b_i b_t - (d_j - b_0 r_0 - \sum_{i=1}^{k} b_i \mu_i)^2,$$

$$e_i^l = u_i^l - b_i, \text{ for } i = 0, \ldots, k.$$

Then, we define

$$\theta^l = \min\{\theta_j \mid j \in I_l\}.$$

Hence we will obtain the boundary point ${\mathbf{v}_{(0)}}^l$. In addition to the fact that $\mathbf{G}$ is convex, ${\mathbf{v}_{(0)}}^l$ has the following property:

$$g_j({\mathbf{v}_{(0)}}^l) \geq d_j, \text{ for } j = 1, \ldots, m,$$

$$\text{and } g_j({\mathbf{v}_{(0)}}^l) = d_j, \text{ for some } j.$$

In the next chapter, we will start looking at some data and make use of the theory in this chapter. An empirical study will be presented and the results will be discussed in detail.

Chapter 3

Empirical Results

3.1 Introduction

In our study, we mainly consider an investor who is interested in investing certain amount of capital into the international markets. We have chosen 11 stock indices from different international markets. We use the daily price data (Oct 25, 1990 to Oct 27, 1998)[1] of different stock markets in the world. In section 3.2, some preliminary analysis of these data have been done in order to let us have some idea of their distributional behaviors. Next, we will use an example of the portfolio selection model described in chapter 2 for illustration and apply the cutting plane method to obtain the optimal allocation of capital. In this thesis, we consider the following simple model:

$$\mathrm{Max}_{\{\mathbf{a}_{(0)}\}} a_0 r_0 + \sum_{i=1}^{k} a_i \mu_i$$

[1]We obtained the data from **DataStream** provided in David Lam Library at University of British Columbia. Many thanks to Christina, the librarian in David Lam library, for her kind help when we were gathering the data.

$$\text{s.t. } a_0 r_0 + \mathbf{a}'\boldsymbol{\mu} + F^{-1}(\alpha)\sqrt{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}} \geq d,$$

$$\sum_{i=0}^{k} a_i = 1,$$

$$\text{and } a_i \geq 0, \ \ i = 0, \ldots, k.$$

Now, we have $k = 11$, and the values of α and r_0 are chosen to be 5% and 0 (assuming the riskless asset is just the cash held in hand) respectively. Moreover, we only have one constraint related to the risk exposure. The only disaster level of return, d, will be treated as an input variable chosen by the investor.

As stated before, many people assume the returns follow multivariate normal distribution, $MVN_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, when they use the chance constrained model, we will also use it but only treat it as a benthmark for comparison. Then, we assume that the returns have an EC distribution. In our study, we will consider the well known multivariate-t distribution, $Mt_k(m; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ where m is the degrees of freedom. In addition, we further consider the case that we do not know which distribution family they belong to and their joint distribution may be time-varying. Hence, we will use the empirical distributions described in section 2.4. Furthermore, we also consider the choice of d and window size in selecting portfolios. In section 3.3, we will clearly specify these factors and explain how we compare different portfolio schemes. Lastly, all empirical results, detailed analysis and interpretations will be presented in section 3.4, where we attempt to find the "best" portfolio scheme using a new approach. The steps of this approach can be summarized as the following:

1. For each portfolio scheme, we create a "portfolio line" which illustrates the relationship of portfolio returns and disaster levels of returns.

2. By connecting the envelop of these portfolio lines, we can obtain the "portfolio frontier".

3. Using a new return-to-risk ratio, we can find the "best" portfolio scheme, in terms of the rate of return and risk management ability, from the portfolio frontier.

3.2 Data

The returns data we are using here are calculated by using the log-difference of the index prices, i.e.,

$$r_{i,t} = \log P_t - \log P_{t-1} \ ,$$

for index i at time t. The followings are the summary and notations of the data set:

1. r_1: S&P Composite returns

2. r_2: Dow Jones Industrial Average returns

3. r_3: NASDAQ Composite returns

4. r_4: Hong Kong Heng Seng Index returns

5. r_5: Tokyo Nikkei 225 Composite returns

6. r_6: Seoul Composite returns

7. r_7: Australia All Ordinaries returns

8. r_8: London FTSE 100 Index returns

9. r_9: Frankfurt DAX Composite returns

10. r_{10}: Paris CAC 40 Composite returns

11. r_{11}: Toronto TSE 300 Composite returns

From the time series plots of these returns data, Figure B.1a - B.1k, we find the fluctuations of the returns vary both among different markets and different time periods. The most obvious observation is the exceptionally high volatilities in Hong Kong and Seoul markets after the 1997 Asian financial crisis. Some statistics for the whole data set are summarized in Tables 3.1 and 3.2.

In order to assess the validity of normality assumption for the returns, we simply look at the quantile-quantile plots (QQ plots) of our daily return data. Not surprisingly, we find serious violation of the normality assumption for each index return from Figure B.2a - B.2k. On the other hand, the QQ plots suggest that the data have a fatter tails compared to those of normal distribution. Based on the histograms in Figure B.3a - B.3k, it is not obvious that any series of the returns data is far from symmetry in terms of the distribution. Therefore, it is reasonable to assume that returns data follow some EC distributions.

Since multivariate-t is the most well-known distribution in the EC distribution family other than multivariate normal, and it is known for its fat-tail property, we will consider it as one possible candidate for the distribution of the returns. By looking at the QQ plots of the returns data and the t-distribution quantiles with different degrees of freedom, we found that t-distribution with degrees of freedom 5 is a pretty good fit for the data. This can be justified by Figure B.4a - B.4k. To verify whether their joint distribution is multivariate-t with degrees of freedom 5, or $Mt_{11}(5; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, we must at least, have a look on some linear combinations of the returns series. By Theorem 2.2, any linear combinations of r_i's should follow univariate-t with degrees of freedom 5 if returns are distributed as $Mt_{11}(5; \boldsymbol{\mu}, \boldsymbol{\Sigma})$. From Figure B.5a - B.5i, we can see the QQ plots of 9 linear combinations of r_i's and the t-distribution quantiles with degrees of freedom 5. We chose the first principal component and 8 randomly selected linear combinations of r_i's. All these are some possible choices of portfolio. The plots show that these linear combinations are quite close to t-distribution with degrees of freedom 5. To have a clearer picture, we can compare the QQ plots of the same linear combinations and normal quantiles in Figure B.6a - B.6i. Therefore, assuming our data to follow the $Mt_{11}(5; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is reasonable.

It should be noted that we by no means try to say that $Mt_{11}(5; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the best choice for fitting the data. In fact, other members within the EC distribution class are possible. Our main point is to investigate the difference in the performance of portfolio schemes when we have a better fit of distributions in describing the behavior of the data.

3.3 Comparison of Portfolio Schemes

Before we actually carry out an empirical analysis using our data set, we have to define clearly what we would like to analyze. In this section, we will describe in detail how we compare different portfolio schemes.

In our study, we calculate the performances of portfolio schemes:

- with different assumptions of the distributions of returns data, i.e., $MVN_{11}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $Mt_{11}(5; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, and empirical distribution;
- before and after the Asian Financial crisis in late October of 1997;
- using different disaster levels of returns, d (or VaR values)[2]; and
- using different window sizes of historical data.

The most important thing we are investigating is whether the non-normality assumption of the distribution can lead us to a much better portfolio scheme, in terms of the returns and risk management ability.

[2]A disaster level of return with the value d means the daily return of the portfolio is d. The VaR will then be *current capital* $\times$ $|d|$.

Other than the distribution issue, from the time series plots of the returns data, we find that for all the indices, fluctuations before and after the Asian financial crisis are very different. As crises do not happen very often, we can treat the period before the crisis as the "normal" period. On the other hand, period after the crisis can be thought as the "abnormal" period. In the "abnormal" period, obviously, investors are in a much riskier position. By looking at the portfolio returns after the crisis, we can tell the abilities on managing risk in the "abnormal" period for different portfolio schemes. To be more specific, we are going to look at the cumulative returns of each scheme for the year before the crisis, October 28, 1996 to October 27, 1997, and for the year after the crisis, October 28, 1997 to October 27, 1998.

We also look at different d which can be chosen by an investor. We know that the greater the disaster level (smaller d), or the larger value of VaR, is chosen, the less risk adverse the investor is behaving. This is because it means the investor can tolerate losing more capital with the same probability. Thus we can tell how the risk attitude affects the investor's portfolio returns. Moreover, we can investigate whether some choices of d will lead to an inefficient portfolio in Baumol's sense: riskier but with lower expected return. The disaster levels of returns we have considered in this thesis range from -0.0001 to -0.035.

The last factor we will look at is the use of different window sizes of historical data in portfolio optimization. It is sometimes argued that using all the past data in forming a portfolio is inappropriate. The main argument is that the really old data did not have much influence on the behavior of the recent data. There are 2 common possible ways to adjust for the smaller influence of the older data. The first one is assigning different weights to the data at different time points. The really old data will receive smaller weights and the recent data will have larger weights when we calculate the sample means and variance-covariance matrix, and when we construct the empirical distribution. However, there are many possible weighting schemes that can be chosen and it is not obvious which one is the best. The second method is simply truncating the really old data in our series. That means we will always use the most recent data, 3 years say. This is what we call a window. So a 3-year window means that when we are finding the optimal portfolio for tomorrow, we only make use of the past data from today up to 3 years ago. This method is easier to implement compared to the first one and thus we decided to use it. The only uncertain thing is the choice of the window size. Therefore, we will use different window sizes and compare them. In this thesis, we use 1-year, 2-year, ..., up to 6-year as our window sizes. In addition, we will also construct the portfolios on the basis of using all the past data.

At this point, we can see that there are many possible portfolio schemes. When we compare them, we will look at their cumulative returns over a specific period of time. The following example will give us a clearer picture.

Example 3.1

Suppose we are considering the following portfolio scheme:

- distribution assumption: empirical distribution;
- disaster level of return: -0.001; and
- window size: 3 year.

Then we will find the optimal portfolio holding for October 28, 1996 based on the data in the period of October 28, 1993 to October 25, 1996.[3] Let the optimal portfolio be:

$$(a_{0,28/10/1996},\ a_{1,28/10/1996},\ \ldots,\ a_{11,28/10/1996}).$$

Using the realized daily returns of the indices on October 28, 1996, we can calculate the daily return of the portfolio, $r_{p,28/10/1996}$, on October 28, 1996:

$$a_{0,28/10/1996}r_0 + a_{1,28/10/1996}r_{1,28/10/1996} + \ldots + a_{11,28/10/1996}r_{11,28/10/1996},$$

where $r_{i,28/10/1996}$ is the daily returns of index i, for $i = 1, \ldots, 11$, and $r_0 = 0$. Similarly, we can calculate the daily return of the portfolio on October 29, 1996, $r_{p,29/10/1996}$, by using the data in the period of October 29, 1993 to October 28, 1996. This process is repeated until we obtain the daily return of the portfolio on October 27, 1998, $r_{p,27/10/1998}$. Now, we can see that the cumulative return of this portfolio scheme for the year before the Asian financial crisis, i.e., for the period of October 28, 1996 to October 27, 1997, is

[3]The Exchange markets were closed on October 26, 1996 and October 27, 1996.

$$r_{p,28/10/1996} + r_{p,29/10/1996} + \ldots + r_{p,27/10/1997}.$$[4]

In a similar way, the cumulative return of this portfolio scheme for the year after the Asian financial crisis, i.e., for the period of October 28, 1997 to October 27, 1998, is

$$r_{p,28/10/1997} + r_{p,29/10/1997} + \ldots + r_{p,27/10/1998}.$$

Of course, the above example is only for one of many possible portfolio schemes.

In the next section, we will present the performance of all the portfolio schemes we have considered. With the help of some graphs, we can visualize the comparisons of different portfolio schemes. More importantly, we will try to find the "best portfolio scheme".

3.4 Empirical Results and Discussion

In the last section, several factors were considered in constructing a portfolio scheme, choosing the best portfolio scheme is not an easy task. Firstly,

[4]Note that r_i's are the continuously compounded rate of returns per day. Therefore, suppose the investor had an amount of capital M_0 putting on the indices and the riskless assets before October 28, 1996, he will then have $M_0e^{r_{p,28/10/1996}}$ on October 28, 1996, $M_0e^{r_{p,28/10/1996}}e^{r_{p,29/10/1996}}$ on October 29, 1996, and so on. Eventually, on October 27, 1997, he will have

$$\begin{aligned} & M_0e^{r_{p,28/10/1996}}e^{r_{p,29/10/1996}} \ldots e^{r_{p,27/10/1997}} \\ = {} & M_0e^{(r_{p,28/10/1996}+r_{p,29/10/1996}+\cdots+r_{p,27/10/1997})}. \end{aligned}$$

Therefore, the continuously compounded rate of return for this period of time (or the cumulative return) is $r_{p,28/10/1996} + r_{p,29/10/1996} + \ldots + r_{p,27/10/1997}$.

we will present the results for the performances of different portfolio schemes. Secondly, we will use the idea of Baumol's "efficient portfolio" to find the "efficient frontier". Lastly, by defining a new return-to-risk ratio, we attempt to find the best portfolio scheme.

The way we visualize the data is by using some plots of cumulative returns against the disaster levels of returns. The cumulative returns are calculated in the way described in Example 3.1. For each window size, we will first consider the cumulative returns for the year before the Asian financial crisis. Then we will look at those for the year after the crisis. In addition, we are also interested in comparing the cumulative returns of these two years for each portfolio scheme. Thus, there will be 3 graphs for each window size. Also, in each graph, we have drawn three lines representing the returns and disaster levels of returns pairs for the 3 distribution assumptions. We call these lines **portfolio lines**. Since we have 6 different window sizes, plus the one that we always use all past data, we will have 21 plots (Figures C.1a - C.7c) in total.

In each of the plots, the $y-axis$ is the cumulative returns, which is the realized performances of different portfolio schemes. The $x-axis$ is the disaster levels of returns, d, ranging from some negative values to zero. It can be seen that the portfolio lines are very different among different windows sizes, before and after the Asian crisis, and various distribution assumptions. However, only based on these plots, it is difficult to draw a conclusion. Hence, we will use the idea of Baumol's "efficient portfolio". Recall that a rational investor should never choose a disaster level of return while he can get a higher return with a larger value of d. To find the "portfolio frontier", we simply look for the maximum cumulative returns an investor can get at each disaster level of return. Then we can construct the portfolio frontier curve. Tables 3.3 - 3.5 summarize the optimal window size, distribution assumption and the corresponding cumulative returns for different disaster levels of returns at different periods of time.

One thing we may notice is that before the Asian financial crisis, from Table 3.3, the multivariate normal distribution assumption gives us maximum cumulative returns for most of the disaster levels of returns. Moreover, when d is between -0.0001 and -0.0120, the window size of 3 years is the best choice while the 1-year window is the best for the other disaster levels of returns. However, after the crisis, from Table 3.4, the empirical distribution assumption always gives the maximum returns. For the optimal window sizes, they vary with the disaster levels of returns.

For an investor, it may be interesting to look at the cumulative returns for the period October 28, 1996 to October 27, 1998 since he would never know when a crash would come. This will capture the performances of different portfolio schemes for both "normal" and "abnormal" years. From Table 3.5, it is noted that empirical distribution assumption gives the best returns for most of the disaster levels of returns, especially for the efficient ones. In our case, the disaster levels of returns in the range of -0.0001 to -0.0120 (except for -0.0110) are considered to be efficient since the cumulative returns are increasing. In Figure 3.1, portfolio frontier curve is created by considering the returns and disaster level for the period October 28, 1996 to October 27, 1998. The efficient portfolios are those on the negative sloping part of the portfolio frontier curve. Therefore, an investor is only advised to choose a disaster level of return in the range of -0.0001 to -0.0120 other than -0.0110. If the disaster level is chosen between -0.0001 and -0.0100, a 3-year window size and empirical distribution should be used. A 5-year window and multivariate normal distribution is recommended for the disaster level of return -0.0120.

One observation should be noted is that the use of multivariate-t distribution with degrees of freedom 5 does not lead to any portfolio scheme in the efficient frontier. This can also be seen from Figures C.1a - C.7c, where we find that $Mt_{11}(5; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ assumption only leads to higher returns when the disaster level of the return is smaller than the efficient ones.

Before we attempt to find the best performing portfolio scheme, we will define a measure for the performances of different portfolio schemes. The following definition is a new return-to-risk ratio we are going to use:

Definition 3.1:

Suppose a portfolio scheme gives an cumulative return r at a disaster level of return d, its return-to-risk ratio (r/d ratio) is defined as $|\frac{r}{d}|$.

The meaning of this ratio will be the cumulative return for each unit of absolute disaster level of return. Obviously, a larger value of this ratio represents a better portfolio scheme when both return and risk are taken into account.

For the portfolio schemes in the portfolio frontier curve, we calculate their r/d ratios and summarize them in Tables 3.6 - 3.8. Again, we only consider the period October 28, 1996 to October 27, 1998 so that both "normal" and "abnormal" years are included. From Table 3.8, we find that the r/d ratio is greatest when the disaster level of return is -0.0090. Therefore, we conclude that by using a 3-year window, empirical distribution assumption, and with -0.0090 as the disaster level of return, we can obtain the best portfolio scheme for these 11 stock indices and riskless asset. Since this result has already taken the "abnormal" year into account, it also has the ability to manage the risk when some unexpected negative market movements occur. Therefore, this scheme should be advised to an investor interested in investing these assets. For the period October 28, 1996 to October 27, 1998, the portfolio holdings of this scheme are shown in Figure 3.2a - 3.2l.

Table 3.1: Sample averages of the daily returns for each index.

$\bar{r}_1$	$\bar{r}_2$	$\bar{r}_3$	$\bar{r}_4$	$\bar{r}_5$	$\bar{r}_6$
0.000591	0.000582	0.000776	0.000565	-0.000291	-0.000362

$\bar{r}_7$	$\bar{r}_8$	$\bar{r}_9$	$\bar{r}_{10}$	$\bar{r}_{11}$
0.000302	0.000448	0.000546	0.000368	0.000311

Table 3.2: Sample correlation matrix of the daily returns.

	r_1	r_2	r_3	r_4	r_5	r_6
r_1	1.0000	0.9510	0.8165	0.1277	0.1115	0.0872
r_2	0.9510	1.0000	0.7405	0.1468	0.1110	0.0872
r_3	0.8165	0.7405	1.0000	0.1492	0.1274	0.0940
r_4	0.1277	0.1468	0.1492	1.0000	0.2666	0.1286
r_5	0.1115	0.1110	0.1274	0.2666	1.0000	0.0876
r_6	0.0872	0.0872	0.0940	0.1286	0.0876	1.0000
r_7	0.0872	0.1098	0.1204	0.4470	0.3195	0.1556
r_8	0.3505	0.3562	0.3479	0.3031	0.2737	0.1202
r_9	0.2731	0.2862	0.2805	0.4037	0.2861	0.1115
r_{10}	0.3413	0.3538	0.3240	0.2779	0.2485	0.0784
r_{11}	0.6502	0.6488	0.6172	0.2385	0.1795	0.0877

	r_7	r_8	r_9	r_{10}	r_{11}
r_1	0.0872	0.3505	0.2731	0.3413	0.6502
r_2	0.1098	0.3562	0.2862	0.3538	0.6488
r_3	0.1204	0.3479	0.2805	0.3240	0.6172
r_4	0.4470	0.3031	0.4037	0.2779	0.2385
r_5	0.3195	0.2737	0.2861	0.2485	0.1795
r_6	0.1556	0.1202	0.1115	0.0784	0.0877
r_7	1.0000	0.2982	0.4103	0.2624	0.2257
r_8	0.2982	1.0000	0.5493	0.6768	0.3777
r_9	0.4103	0.5493	1.0000	0.6239	0.3459
r_{10}	0.2624	0.6768	0.6239	1.0000	0.3649
r_{11}	0.2257	0.3777	0.3459	0.3649	1.0000

Table 3.3: Summary of maximum cumulative returns for each disaster level (d) in the period Oct 28, 1996 to Oct 27, 1997.

d	*Window size*	*Distribution**	*Maximum returns*
-0.0001	3	MVN	0.0025
-0.0005	3	MVN	0.0126
-0.0010	3	MVN	0.0253
-0.0030	3	MVN	0.0759
-0.0050	3	MVN	0.1265
-0.0070	3	MVN	0.1771
-0.0090	3	MVN	0.2227
-0.0100	3	Emp	0.2227
-0.0110	3	Mt5	0.2259
-0.0120	3	Mt5	0.2236
-0.0130	1	MVN	0.2216
-0.0140	1	MVN	0.2394
-0.0150	1	MVN	0.2452
-0.0160	1	Emp	0.2478
-0.0170	1	Emp	0.2496
-0.0180	1	MVN	0.2531
-0.0200	1	MVN/Emp	0.2567

* MVN: Multivariate normal distribution, Mt5: Multivariate-t distribution, Emp: Empirical distribution.

Table 3.4: Summary of maximum cumulative returns for each disaster level of return (d) in the period Oct 28, 1997 to Oct 27, 1998.

d	*Window size*	*Distribution*	*Maximum returns*
-0.0001	4	Emp	0.0012
-0.0005	4	Emp	0.0062
-0.0010	4	Emp	0.0123
-0.0030	4	Emp	0.0370
-0.0050	4	Emp	0.0619
-0.0070	4	Emp	0.0860
-0.0090	3	Emp	0.1116
-0.0100	5	Emp	0.1305
-0.0110	5	Emp	0.1399
-0.0120	5	Emp	0.1715
-0.0130	5	Emp	0.1654
-0.0140	2	Emp	0.1745
-0.0150	2	Emp	0.1794
-0.0160	2	Emp	0.1926
-0.0170	2	Emp	0.1944
-0.0180	2	Emp	0.1968
-0.0200	2	Emp	0.1739

Table 3.5: Summary of maximum cumulative returns for each disaster level of return (d) in the period Oct 28, 1996 to Oct 27, 1998.

d	*Window size*	*Distribution*	*Maximum returns*
-0.0001	3	Emp	0.0036
-0.0005	3	Emp	0.0179
-0.0010	3	Emp	0.0358
-0.0030	3	Emp	0.1075
-0.0050	3	Emp	0.1799
-0.0070	3	Emp	0.2494
-0.0090	3	Emp	0.3327
-0.0100	3	Emp	0.3399
-0.0110	5	Emp	0.3389
-0.0120	5	MVN	0.3721
-0.0130	2	Emp	0.3680
-0.0140	5	Emp	0.3668
-0.0150	5	Mt5	0.3665
-0.0160	2	Emp	0.3572
-0.0170	2	Emp	0.3545
-0.0180	2	Emp	0.3560
-0.0200	5	Mt5	0.3383

Table 3.6: Summary of return-to-risk ratio (r/d ratio) for each disaster level of return (d) in the period Oct 28, 1996 to Oct 27, 1997.

d	*Window size*	*Distribution*	*r/d ratio*
-0.0001	3	MVN	25.2968
-0.0005	3	MVN	25.2964
-0.0010	3	MVN	25.3014
-0.0030	3	MVN	25.2994
-0.0050	3	MVN	25.2923
-0.0070	3	MVN	25.2976
-0.0090	3	MVN	24.7484
-0.0100	3	Emp	22.2659
-0.0110	3	Mt5	20.5364
-0.0120	3	Mt5	18.6309
-0.0130	1	MVN	17.0448
-0.0140	1	MVN	17.1023
-0.0150	1	MVN	16.3499
-0.0160	1	Emp	15.4901
-0.0170	1	Emp	14.6837
-0.0180	1	MVN	14.0627
-0.0200	1	MVN/Emp	12.8360

Table 3.7: Summary of return-to-risk ratio (r/d ratio) for each disaster level of return (d) in the period Oct 28, 1997 to Oct 27, 1998.

d	*Window size*	*Distribution*	r/d *ratio*
-0.0001	4	Emp	12.3326
-0.0005	4	Emp	12.3420
-0.0010	4	Emp	12.3021
-0.0030	4	Emp	12.3201
-0.0050	4	Emp	12.3861
-0.0070	4	Emp	12.2920
-0.0090	3	Emp	12.4054
-0.0100	5	Emp	13.0524
-0.0110	5	Emp	12.7203
-0.0120	5	Emp	14.2933
-0.0130	5	Emp	12.7253
-0.0140	2	Emp	12.4651
-0.0150	2	Emp	11.9611
-0.0160	2	Emp	12.0400
-0.0170	2	Emp	11.4367
-0.0180	2	Emp	10.9316
-0.0200	2	Emp	8.6942

Table 3.8: Summary of return-to-risk ratio (r/d ratio) for each disaster level of return (d) in the period Oct 28, 1996 to Oct 27, 1998.

d	*Window size*	*Distribution*	*r/d ratio*
-0.0001	3	Emp	35.6745
-0.0005	3	Emp	35.8006
-0.0010	3	Emp	35.7771
-0.0030	3	Emp	35.8360
-0.0050	3	Emp	35.9815
-0.0070	3	Emp	35.6246
-0.0090	3	Emp	36.9694
-0.0100	3	Emp	33.9917
-0.0110	5	Emp	30.8121
-0.0120	5	MVN	31.0108
-0.0130	2	Emp	28.3107
-0.0140	5	Emp	26.1991
-0.0150	5	Mt5	24.4358
-0.0160	2	Emp	22.3242
-0.0170	2	Emp	20.8531
-0.0180	2	Emp	19.7753
-0.0200	5	Mt5	16.9146

Figure 3.1: Portfolio frontier curve for the period 28/10/96 to 27/10/98

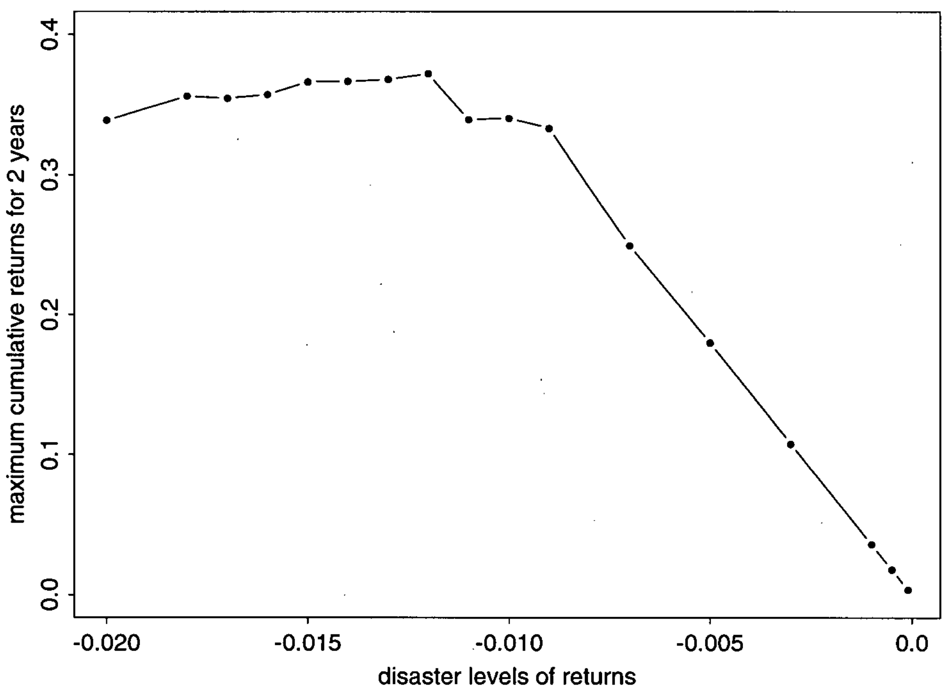


Figure 3.2: Optimal portfolio scheme curves from 28/10/96 to 27/10/98

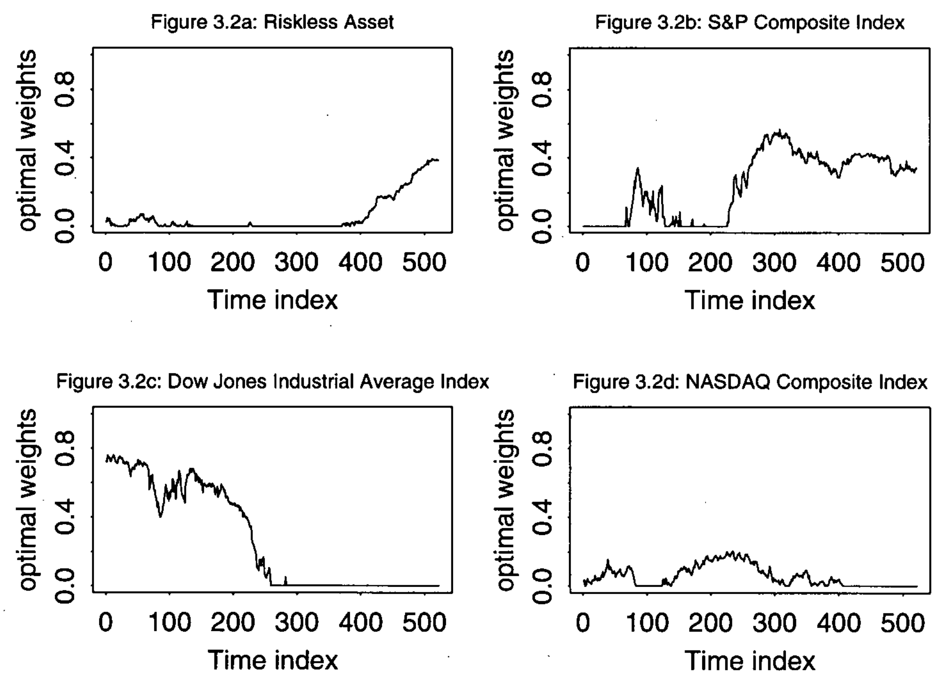


Figure 3.2: Optimal portfolio scheme curves from 28/10/96 to 27/10/98 (Cont'd)

Figure 3.2e: Hong Kong Heng Seng Index

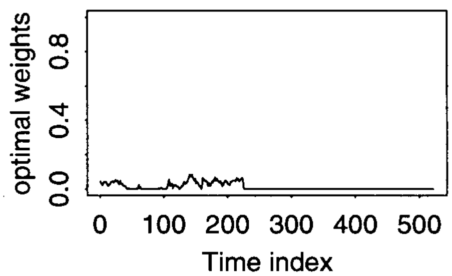


Figure 3.2f: Tokyo Nikkei 225 Composite Index

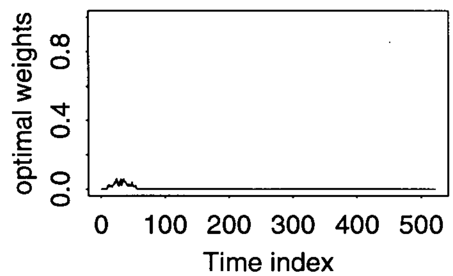


Figure 3.2g: Seoul Composite Index

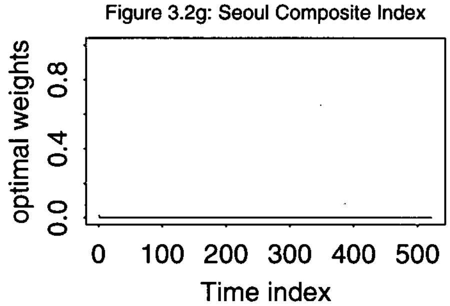


Figure 3.2h: Australia All Ordinaries Index

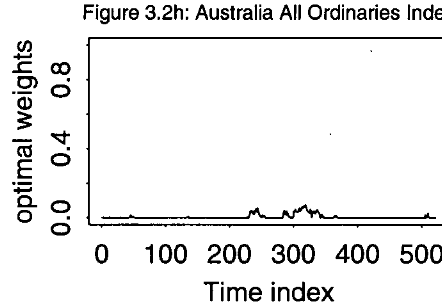


Figure 3.2: Optimal portfolio scheme curves from 28/10/96 to 27/10/98 (Cont'd)

Figure 3.2i: London FTSE 100 Index

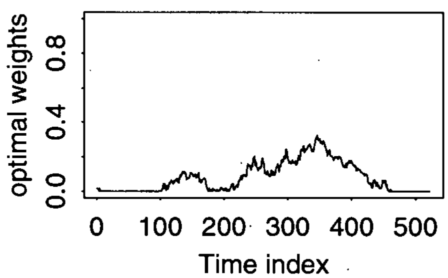


Figure 3.2j: Frankfurt DAX Composite Index

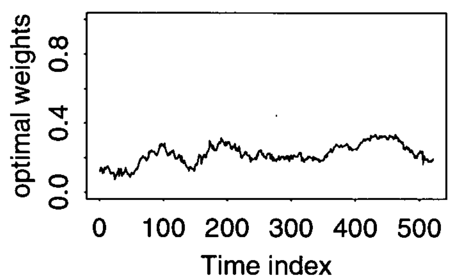


Figure 3.2k: Paris CAC 40 Composite Index

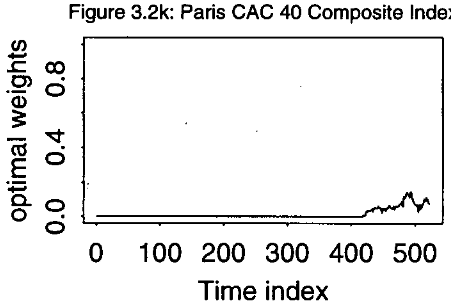
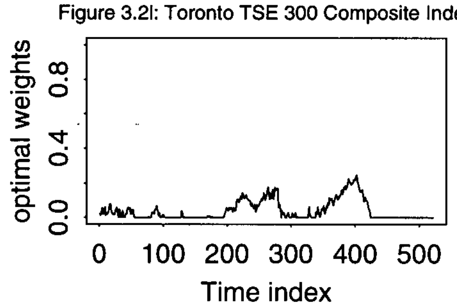


Figure 3.2l: Toronto TSE 300 Composite Index



Chapter 4

Further Research and Discussion

4.1 Introduction

The main advantage of the method in choosing a portfolio, developed in the previous section, is that it is simple and easy to implement in reality. In addition, the use of cutting plane method is also computationally efficient. Our experience is that the computer time of finding the optimal portfolio is within seconds. The details of computational aspects are given in section 4.2.

However, if we really like to use the suggested model in practice, it may seem to be too simple. Moreover, we may want to consider some more technical issues when we are looking for the optimal portfolio scheme. We will discuss all these in section 4.3, along with some directions so that we can actually apply our study to the real world.

Although the VaR approach (or the chance constrained programming approach) is popular in risk management for many firms, there are criticisms for it. We will have a brief discussion in section 4.4.

4.2 Computational Aspects

In the algorithm described in Chapter 2, basically, we have to solve a linear programming problem in each iteration. In the first iteration, we have a linear objective function, and two linear constraints. After that, for each additional iteration, we will include one more linear constraint to the linear programming problem. To solve these linear programming problems, we use the simplex method.

The approach we used in this thesis has been implemented using C programming language. In the more than one thousand lines of code we have written, six files of source code from the Numerical Recipes (Press et al (1992))are used. One of them is the C subroutine for the simplex method, which is in the file "simplx.c". The other five files we used contain the supporting functions used in "simplx.c". They are "simp1.c", "simp2.c", "simp3.c", "nrutil.c" and "nrutil.h". Among them, "nrutil.c" and "nrutil.h" are the ANSI C version of the Numerical Recipes utility files.

Our program was written in the way that when the disaster level of return, the distribution assumption, and the window size are inputted, the optimal portfolio holdings and the realized returns for the period from October 28, 1996 to October 27, 1998 will be calculated. According to our experience, the computer time needed for such calculation is around 30 to 60 minutes in a Sun Enterprise 450 computer running Solaris 2.6. Therefore, to find the optimal portfolio holding for a particular date, the computer time required is within seconds.

4.3 Beyond the model

Transaction Cost

One very important thing we have not considered in our model is the transaction cost needed when the portfolio holdings are changed. In our study, since the optimization is done everyday and the portfolio holdings will be revised daily, transaction costs may be of concern. Even though in the recent years, electronic trading through the Internet has reduced the transaction costs a lot,[1] we may still want to consider it.

[1] Based on the commissions charged by online brokers nowadays, even in our case where portfolio holdings are revised daily, total transaction cost involved is still small compared to the rate of return obtained by the optimal portfolio scheme. For the two-year time horizon in our example, the net return rate is at least 29% while the gross return rate is 33%.

To include the transaction cost in the process of selecting the optimal portfolio, two modifications of our model can be made. One is simply incorporating the transaction cost into the objective function. For our case in the last chapter, suppose the transaction costs for a unit change in the holding of the assets are $c_0, c_1, \ldots, c_k$ respectively. If our current portfolio has the weights vector $(a_0^0, a_1^0, \ldots, a_k^0)$, then to find the optimal portfolio in the next time period, we solve the following maximization problem:

$$\text{Max}_{\{\mathbf{a}_{(\mathbf{0})}\}} a_0 r_0 + \sum_{i=1}^{k} a_i \mu_i - \sum_{i=0}^{k} c_i |a_i - a_i^0|$$

$$\text{s.t. } a_0 r_0 + \mathbf{a}'\boldsymbol{\mu} + F^{-1}(\alpha)\sqrt{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}} \geq d,$$

$$\sum_{i=0}^{k} a_i = 1,$$

$$\text{and } a_i \geq 0, \;\; i = 0, \ldots, k.$$

The difficulty in solving the above problem is that the objective function is no longer a simple linear function. In fact, such an objective function is piecewise linear. We can perform a 2-stage optimization by first maximizing over a partition of domain where the objective function is locally linear in each sub-domain. This part can utilize the cutting plane method. Then we can optimize over all the sub-domain to find the optimal portfolio. The trade-off is that it requires much more computer time.

The second modification we can make for our model is the case when an investor has a limited amount of transaction cost to spend for changing the assets holding. This can be done by adding the following constraint in order to restrict the transaction cost to be less than an amount C:

$$\sum_{i=0}^{k} c_i |a_i - a_i^0| \leq C.$$

Some remarks on the use of empirical distribution

In the last chapter, we used the empirical distribution to approximate the theoretical critical value of the distribution of the linear combination of returns data. Recall that we randomly selected 100 linear combinations of the returns data and constructed their empirical distributions. Then we used the simple average of the critical points of these 100 generated empirical distributions as an estimate of the theoretical critical value. These sample critical points are supposed to be the same theoretically. We found that the variance of these sample critical values are around 0.2% (or less) of their average. Therefore, it is reasonable to believe that all linear combinations of the returns data have approximately the same distribution form. Certainly, further testing can be done to varify its appropriateness.

Another point to take note is the use of sample mean as an estimate of the theoretical critical point. We used it because it is simple, and reasonable.

In fact, other estimation methods are also possible.

4.4 Discussion on the Use of VaR Approach

The main criticism for the use of VaR as a risk measure in the context of the chance constrained programming model, which we discussed in the previous chapter, is that it gives the same penalty to different levels of violation of the chance constraint. In our model, the original probabilistic constraint is

$$\Pr\{\sum_{i=0}^{k} a_i r_i \geq d\} \geq 1 - \alpha.$$

Such a constraint is, in fact, only requiring the inequality $\sum_{i=0}^{k} a_i r_i \geq d$ to be true for $100 \times (1-\alpha)$ times out of 100. However, it cannot distinguish between a serious violation of the inequality (i.e., $\sum_{i=0}^{k} a_i r_i \ll d$) and a slight violation of the inequality (i.e., $d - \sum_{i=0}^{k} a_i r_i$ is a very small positive number).

Certainly, it may not be an important issue if the violation is not so serious. But in the case of a serious violation, a huge loss will occur.

We admit that this is the disadvantage of using VaR as the risk measure because it treat huge and small losses equally which does not seem to make economic sense. However, we are not only choosing the optimal portfolio at one time point, where in that case, it is true that we may end up with losing a large portion of the original capital. Instead, we repeat the optimization process in selecting our portfolio for a period of time,[2] plus the choice of optimal portfolio scheme is over several different values of disaster levels of return.[3] Therefore, the disaster level of return (or the VaR value) is carefully chosen, and it will be extremely unlikely to lose a large portion of the original capital. For further study, we can also manipulate the value of disaster probability, α, when we choose the optimal portfolio scheme.

In addition, we would like to point out that in our generalized VaR approach to portfolio selection, the distribution assumption of the data is crucial. The reason is that an inappropriate distribution assumption may underestimate our risk exposure (or the VaR value), which may lead us to a huge loss. On the other hand, it may also overestimate our risk exposure. Then, our choice of portfolio may be too conservative, which means it tends to give us low portfolio return. The way we use the empirical distribution in our optimal portfolio scheme suggests a method to address the distribution assumption problem by the use of historical data.

In summary, we believe that the use of VaR with various disaster levels of returns in our model is reasonable and appropriate.

[2]There are 600 times of optimization in the 2 years period.

[3]In fact, what we did is a sensitivity analysis over the values of d.

Chapter 5

Conclusion

The traditional VaR approach to portfolio selection problem often assumes that the joint distribution of the returns data is multivariate normal. This assumption is mainly implemented for handling linear combinations of the returns data. The class of multivariate normal distribution has the desirable property that any linear combination also has a univariate normal distribution. In fact, this similar property can be found in the class of EC distributions, which contains multivariate normal distribution as a subclass.

In our generalized VaR approach, firstly, we extended the distribution assumption to the class of EC distribution family. The second generalization we made to our portfolio selection model was that we allowed for time-varying distributional form of the data set. Therefore, we were able to capture any change in the behavior of the data that might be due to some unexpected environmental changes. For this extension, the empirical distribution was used.

In chapter 3, we analyzed some stock index returns data in the context of selecting the optimal portfolio over a period of time. The results showed that empirical distribution had an overall good performance compared to multivariate normal and multivariate-t distributions. More, it had the best performance in the "abnormal" year, which meant its ability of managing the risk was higher. This was mainly because empirical distribution can capture the time-dependency character of our returns data.

In addition, we also found that the number of historical data values being used (window size) was an important factor in solving our investor's problem.

Finally, we would like to emphasize that our study gave an idea on how to approach a portfolio selection problem with VaR as the risk measure. To apply our model and method in practice, we can consider other factors such as the transaction cost in order to obtain a better and more accurate solution.

Appendix A

Proof of Quasi-concavity of Constraint Functions

Lemma A.1:

$\theta(\mathbf{a}) = \sqrt{\mathbf{a}'\mathbf{\Sigma}\mathbf{a}}$ is a convex function.

Proof. Let $\lambda \in (0, 1)$, and $\bar{\lambda} = 1 - \lambda$. To show the convexity of $\theta(\mathbf{a})$ is the same as showing

$$\sqrt{(\lambda\mathbf{a} + \bar{\lambda}\mathbf{b})'\mathbf{\Sigma}(\lambda\mathbf{a} + \bar{\lambda}\mathbf{b})} \leq \lambda\sqrt{\mathbf{a}'\mathbf{\Sigma}\mathbf{a}} + \bar{\lambda}\sqrt{\mathbf{b}'\mathbf{\Sigma}\mathbf{b}}$$

$$\Leftrightarrow \lambda^2\mathbf{a}'\mathbf{\Sigma}\mathbf{a} + \bar{\lambda}^2\mathbf{b}'\mathbf{\Sigma}\mathbf{b} + 2\lambda\bar{\lambda}\mathbf{a}'\mathbf{\Sigma}\mathbf{b} \leq \lambda^2\mathbf{a}'\mathbf{\Sigma}\mathbf{a} + \bar{\lambda}^2\mathbf{b}'\mathbf{\Sigma}\mathbf{b} + 2\lambda\bar{\lambda}\sqrt{\mathbf{a}'\mathbf{\Sigma}\mathbf{a}}\sqrt{\mathbf{b}'\mathbf{\Sigma}\mathbf{b}}$$

$$\Leftrightarrow \mathbf{a}'\mathbf{\Sigma}\mathbf{b} \leq \sqrt{\mathbf{a}'\mathbf{\Sigma}\mathbf{a}}\sqrt{\mathbf{b}'\mathbf{\Sigma}\mathbf{b}}.$$

Since $\mathbf{\Sigma}$ is positive definite, we can write $\mathbf{\Sigma} = \mathbf{\Sigma^{1/2}\Sigma^{1/2}}$.

By Cauchy-Schwarz inequality,

$$\mathbf{a}'\mathbf{\Sigma}\mathbf{b} = (\mathbf{\Sigma^{1/2}a})'(\mathbf{\Sigma^{1/2}b}) \leq \sqrt{\mathbf{a}'\mathbf{\Sigma}\mathbf{a}}\sqrt{\mathbf{b}'\mathbf{\Sigma}\mathbf{b}}.$$

Q.E.D.

Recall

$$g_j(\mathbf{a_{(0)}}) = a_0 r_0 + \mathbf{a}'\boldsymbol{\mu} + F^{-1}(\alpha_j)\sqrt{\mathbf{a}'\mathbf{\Sigma}\mathbf{a}},\ j = 1, \ldots, m,$$

where $F^{-1}(\alpha_j)$'s are all negative. Applying Lemma A.1, $g_j(\mathbf{a_{(0)}})$ can be expressed in the following form:

linear function - convex function,

which is a concave function. Hence, $g_j(\mathbf{a_{(0)}})$ is also a quasi-concave function for $j = 1, \ldots, m$.

Appendix B

Returns Data

Figure B.1: Time Series Plots of returns data

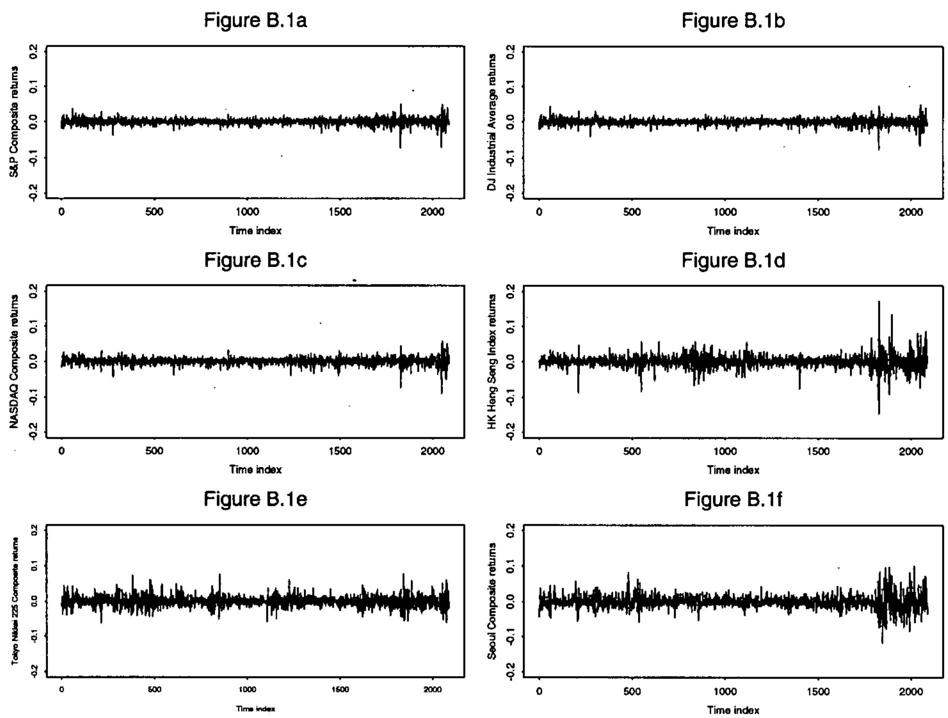


Figure B.1: Time Series Plots of returns data (Cont'd)

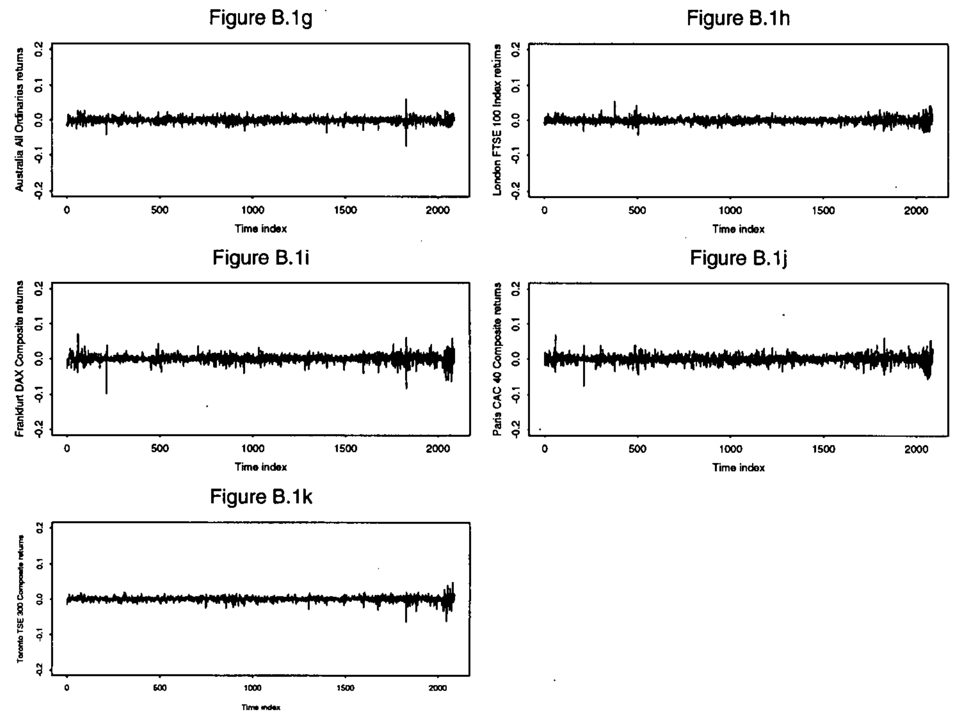


Figure B.2: QQ Plots with normal probabilities

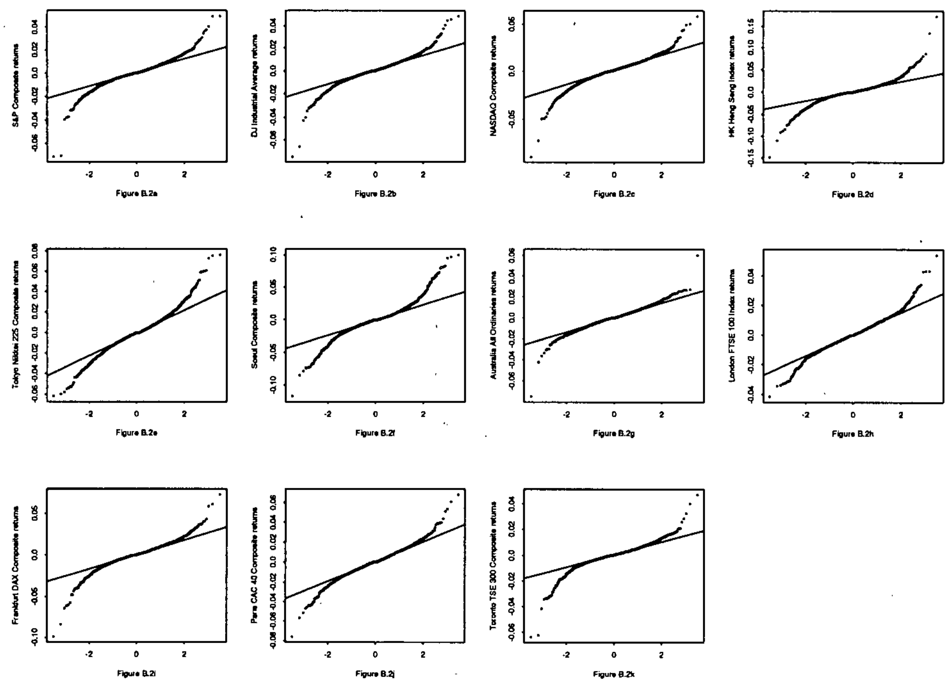


Figure B.3: Histograms of returns data

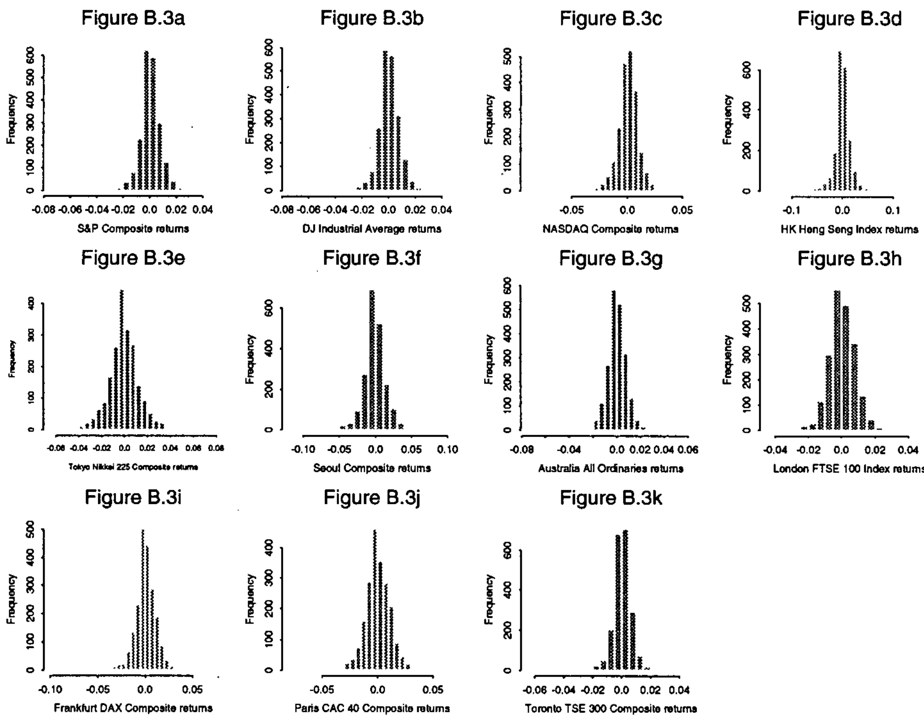


Figure B.4: QQ Plots with t(5) probabilities

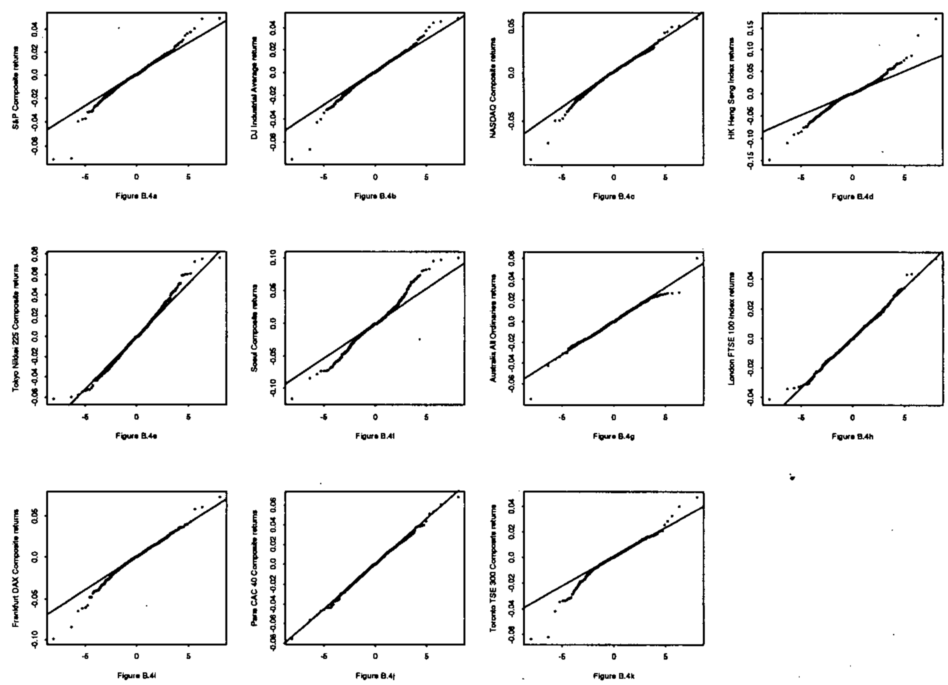


Figure B.5: QQ Plots of $t(5)$ probabilities - Linear Combinations

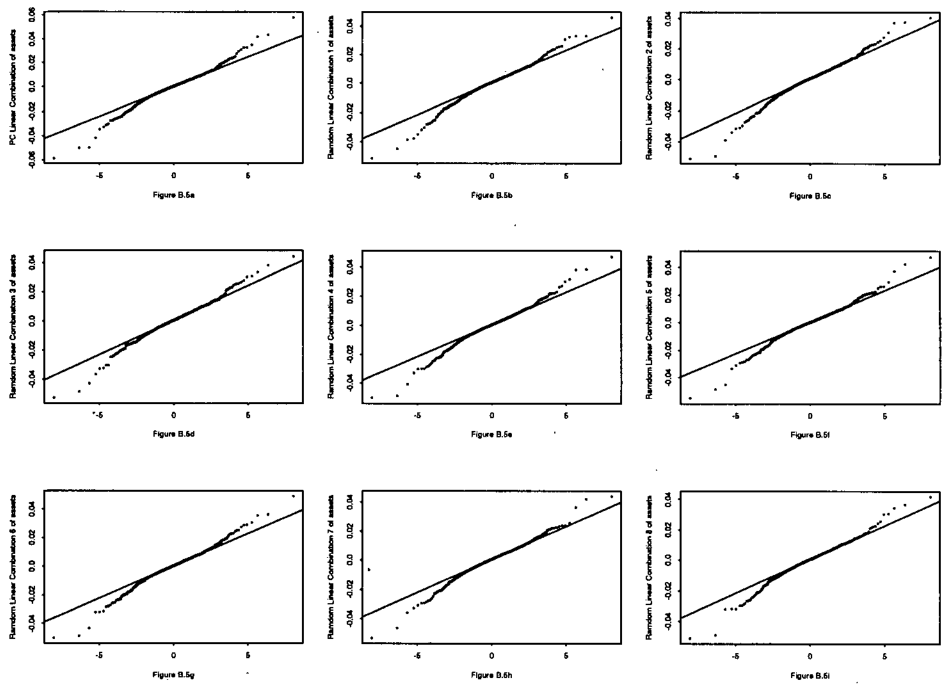
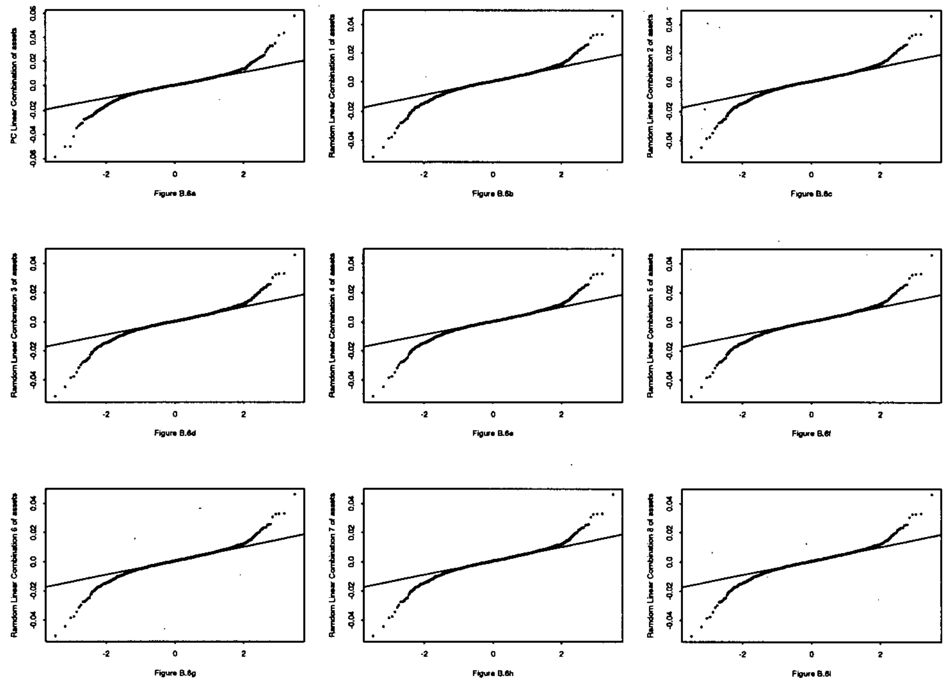


Figure B.6: QQ Plots of normal probabilities - Linear Combinations



Appendix C

Portfolio Lines

Figure C.1: Portfolio lines for various distribution assumptions, using 6-year window

Figure C.1a: Portfolio lines for the period 28/10/96 to 27/10/97

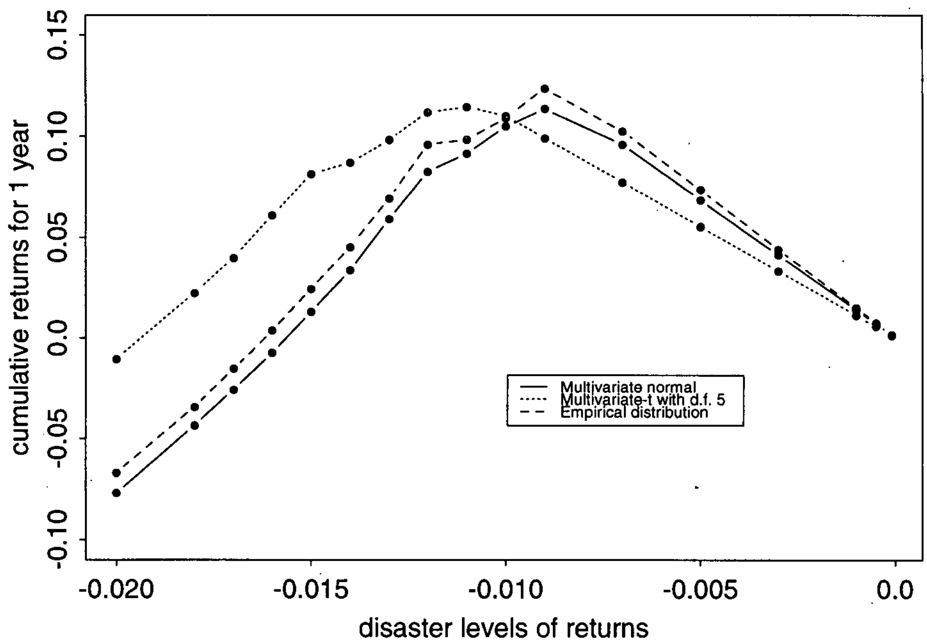


Figure C.1: Portfolio lines for various distribution assumptions, using 6-year window (Cont'd)

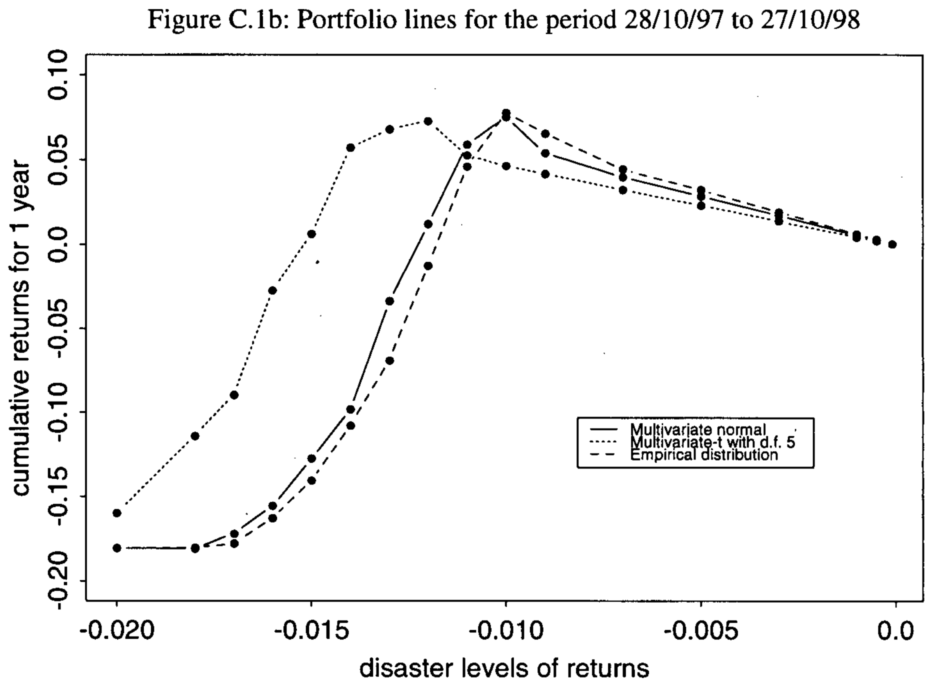


Figure C.1: Portfolio lines for various distribution assumptions, using 6-year window (Cont'd)

Figure C.1c: Portfolio lines for the period 28/10/96 to 27/10/98

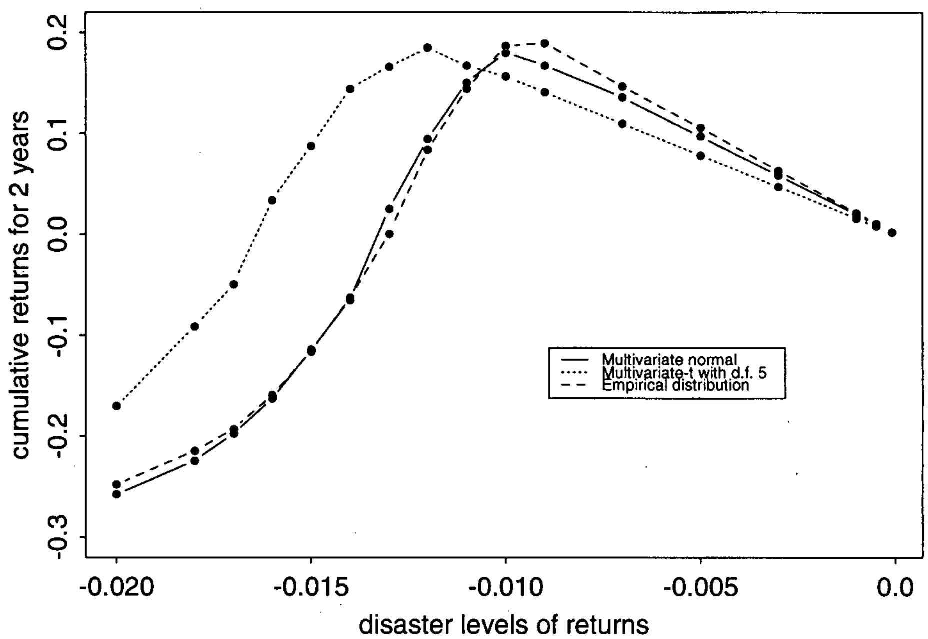


Figure C.2: Portfolio lines for various distribution assumptions, using 5-year window

Figure C.2a: Portfolio lines for the period 28/10/96 to 27/10/97

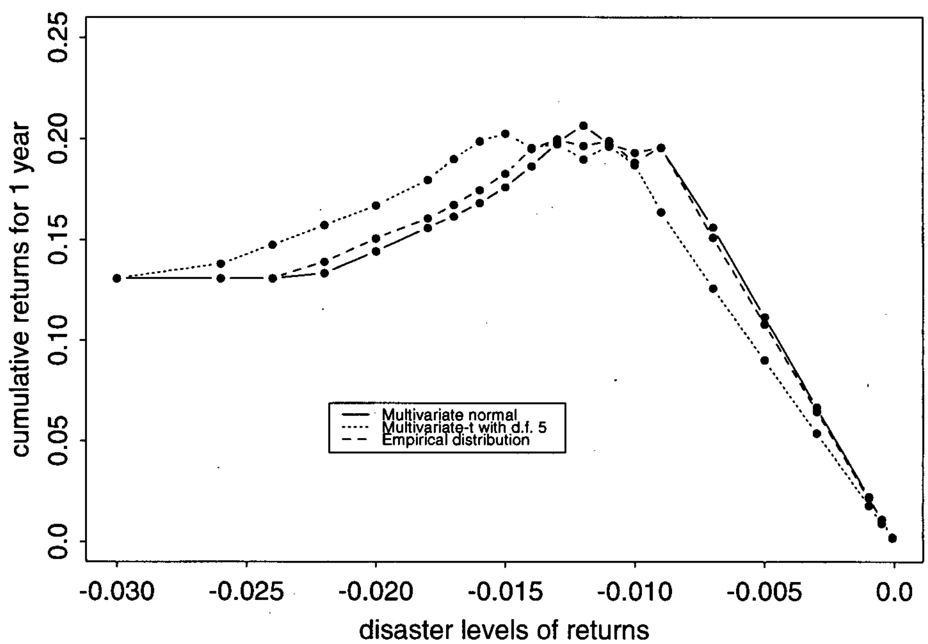


Figure C.2: Portfolio lines for various distribution assumptions, using 5-year window (Cont'd)

Figure C.2b: Portfolio lines for the period 28/10/97 to 27/10/98

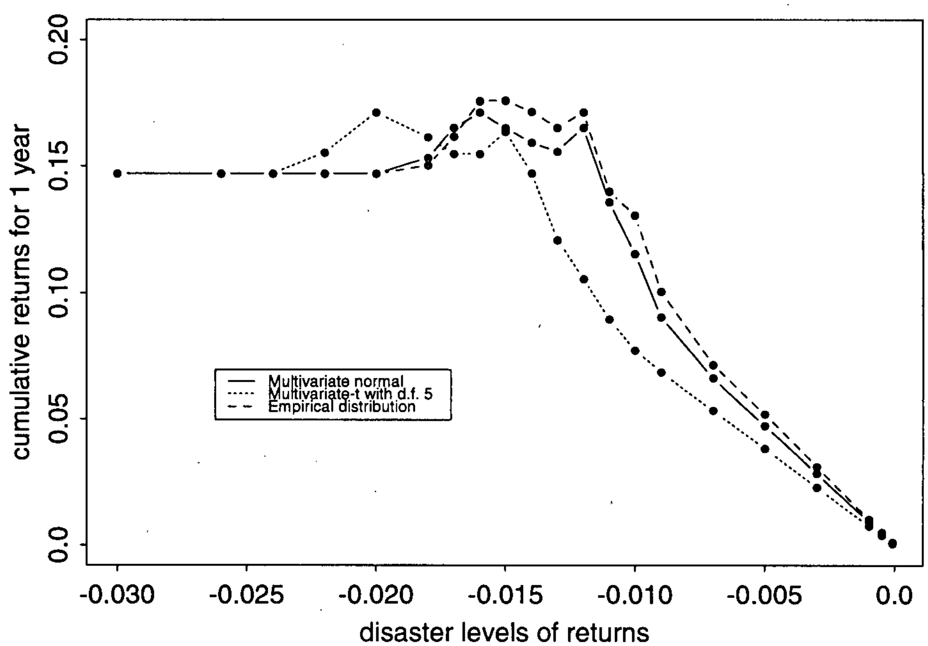


Figure C.2: Portfolio lines for various distribution assumptions, using 5-year window (Cont'd)

Figure C.2c: Portfolio lines for the period 28/10/96 to 27/10/98

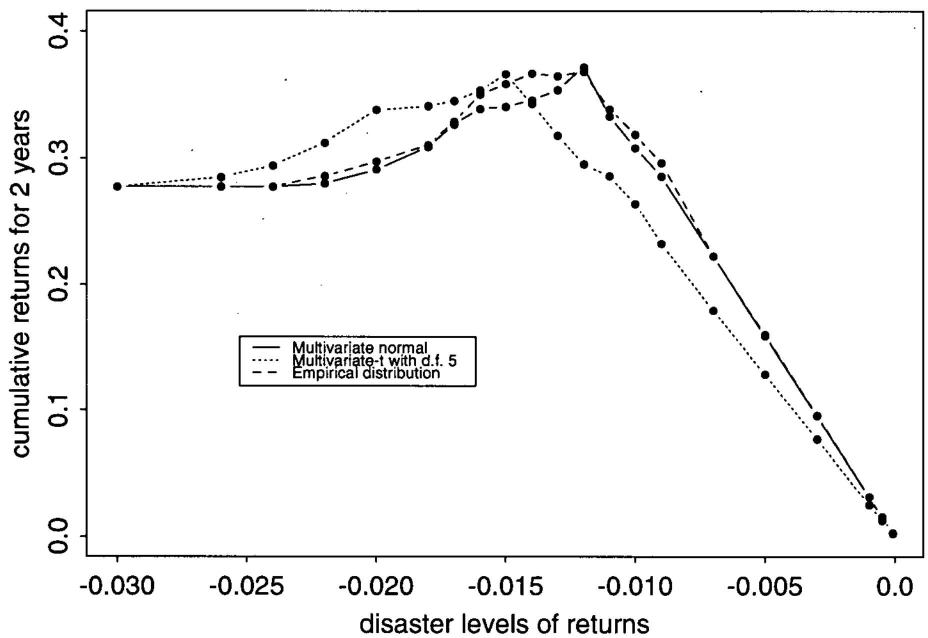


Figure C.3: Portfolio lines for various distribution assumptions, using 4-year window

Figure C.3a: Portfolio lines for the period 28/10/96 to 27/10/97

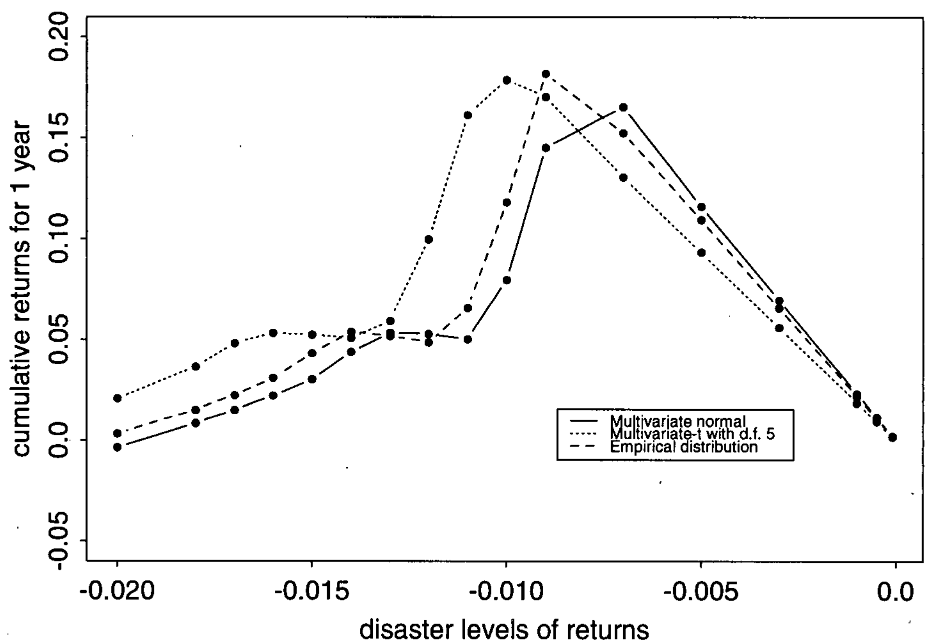


Figure C.3: Portfolio lines for various distribution assumptions, using 4-year window (Cont'd)

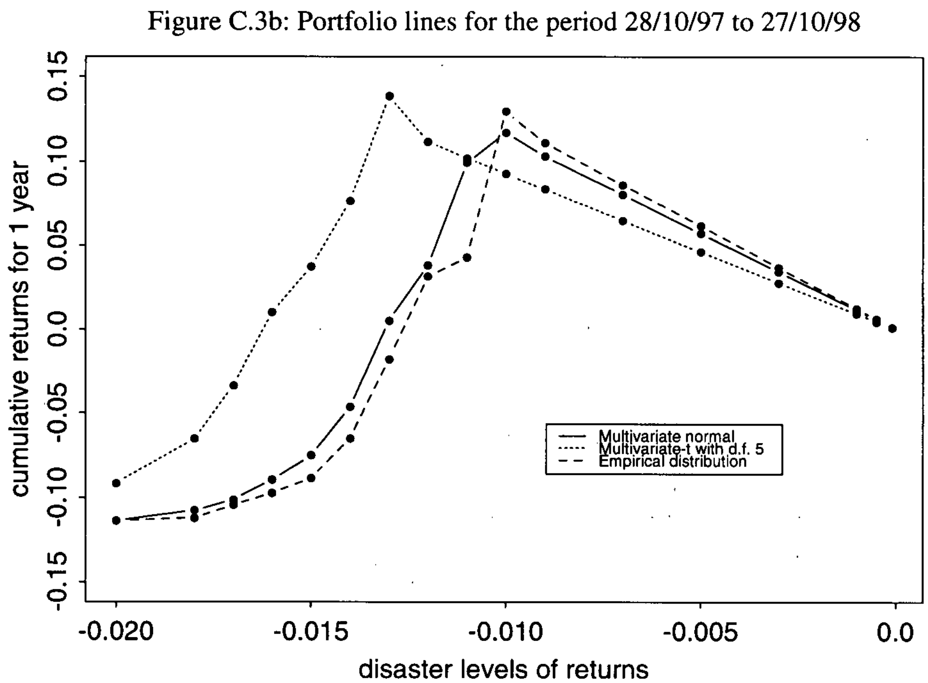


Figure C.3: Portfolio lines for various distribution assumptions, using 4-year window (Cont'd)

Figure C.3c: Portfolio lines for the period 28/10/96 to 27/10/98

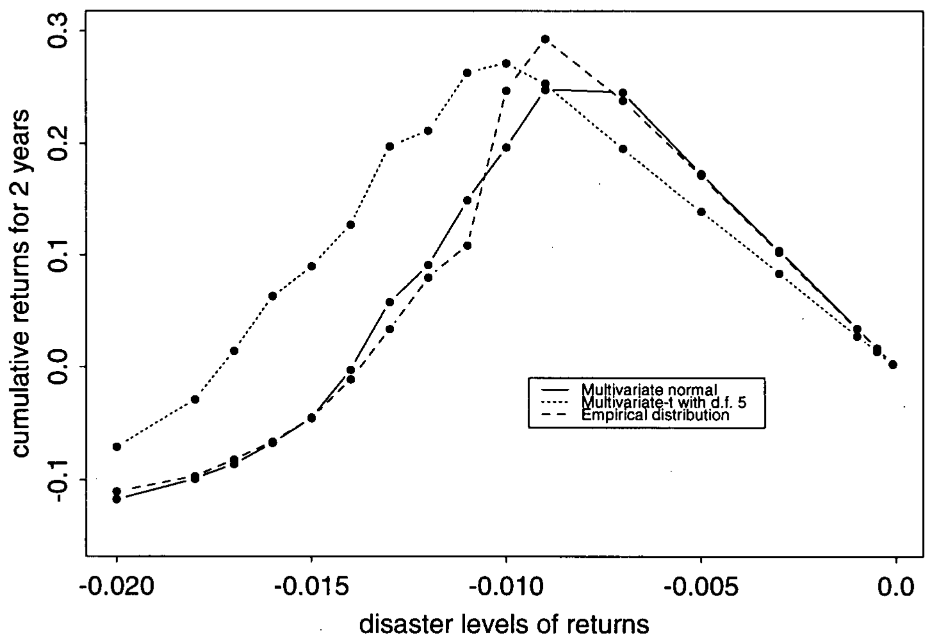


Figure C.4: Portfolio lines for various distribution assumptions, using 3-year window

Figure C.4a: Portfolio lines for the period 28/10/96 to 27/10/97

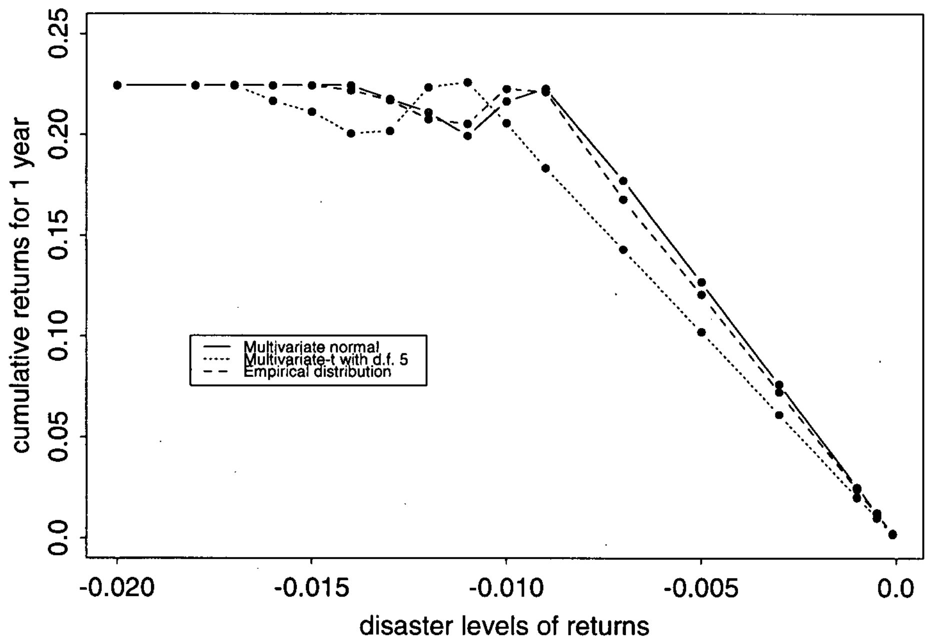


Figure C.4: Portfolio lines for various distribution assumptions, using 3-year window (Cont'd)

Figure C.4b: Portfolio lines for the period 28/10/97 to 27/10/98

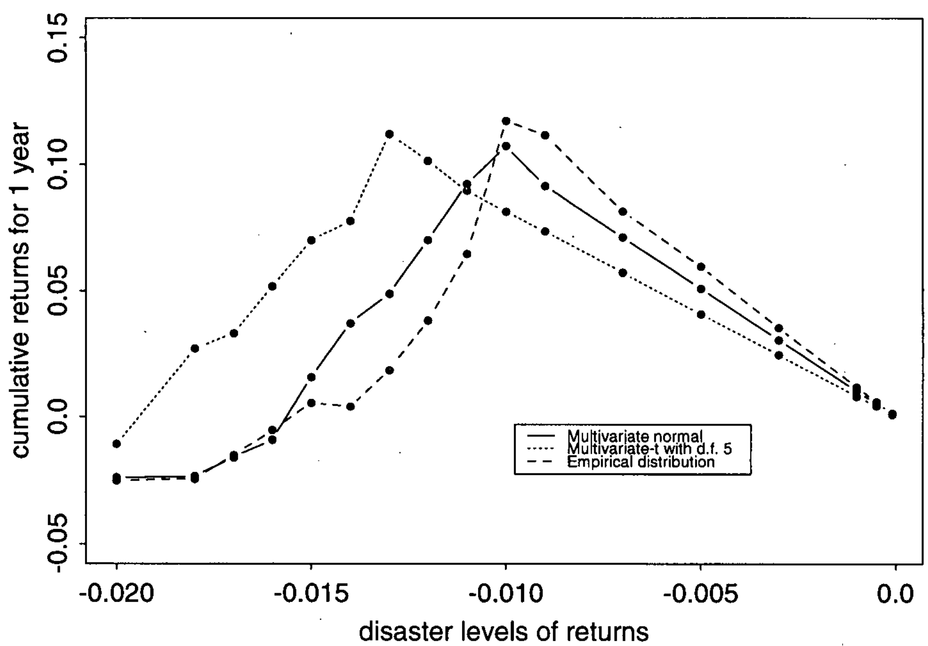


Figure C.4: Portfolio lines for various distribution assumptions, using 3-year window (Cont'd)

Figure C.4c: Portfolio lines for the period 28/10/96 to 27/10/98

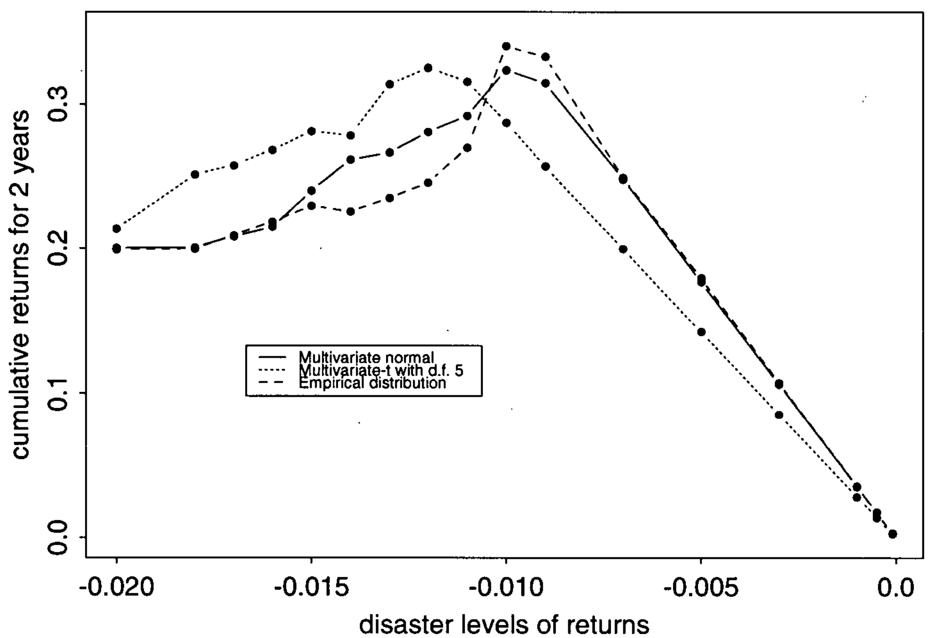


Figure C.5: Portfolio lines for various distribution assumptions, using 2-year window

Figure C.5a: Portfolio lines for the period 28/10/96 to 27/10/97

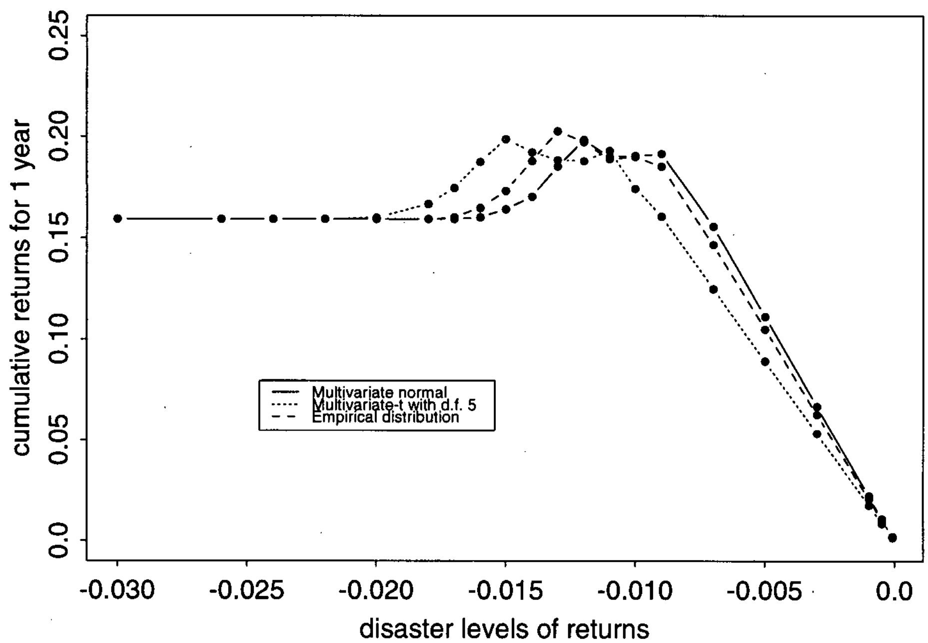


Figure C.5: Portfolio lines for various distribution assumptions, using 2-year window (Cont'd)

Figure C.5b: Portfolio lines for the period 28/10/97 to 27/10/98

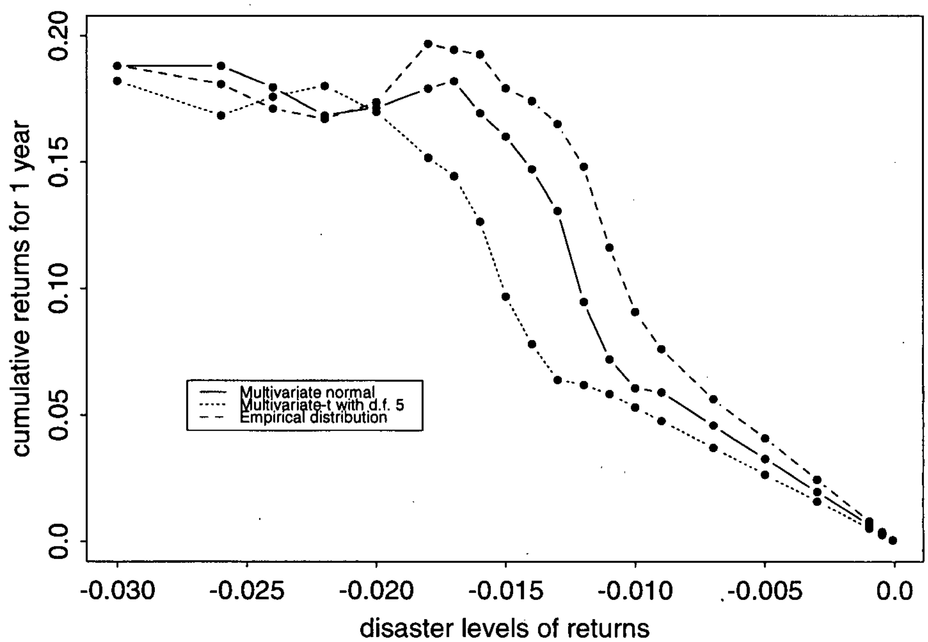


Figure C.5: Portfolio lines for various distribution assumptions, using 2-year window (Cont'd)

Figure C.5c: Portfolio lines for the period 28/10/96 to 27/10/98

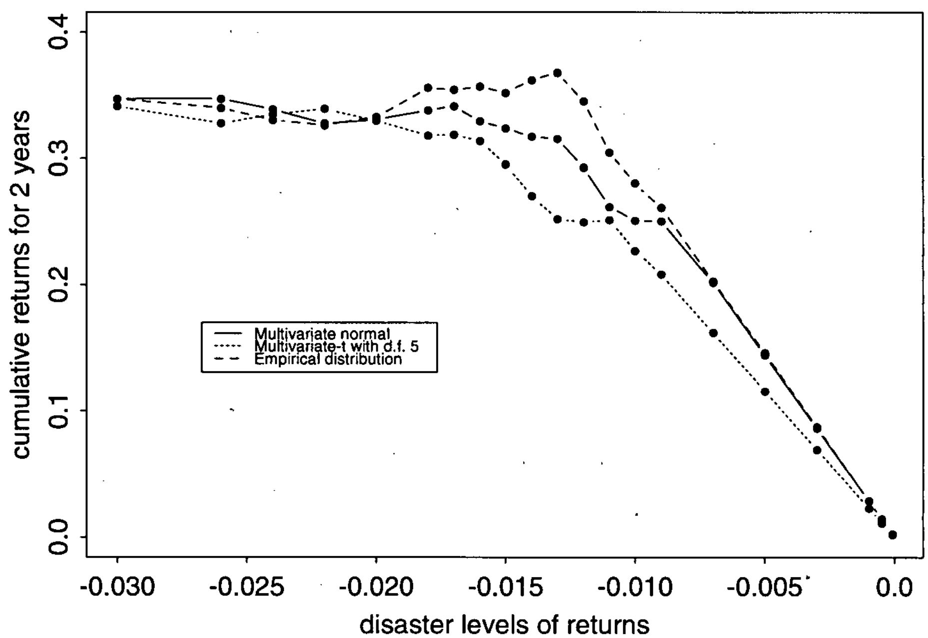


Figure C.6: Portfolio lines for various distribution assumptions, using 1-year window

Figure C.6a: Portfolio lines for the period 28/10/96 to 27/10/97

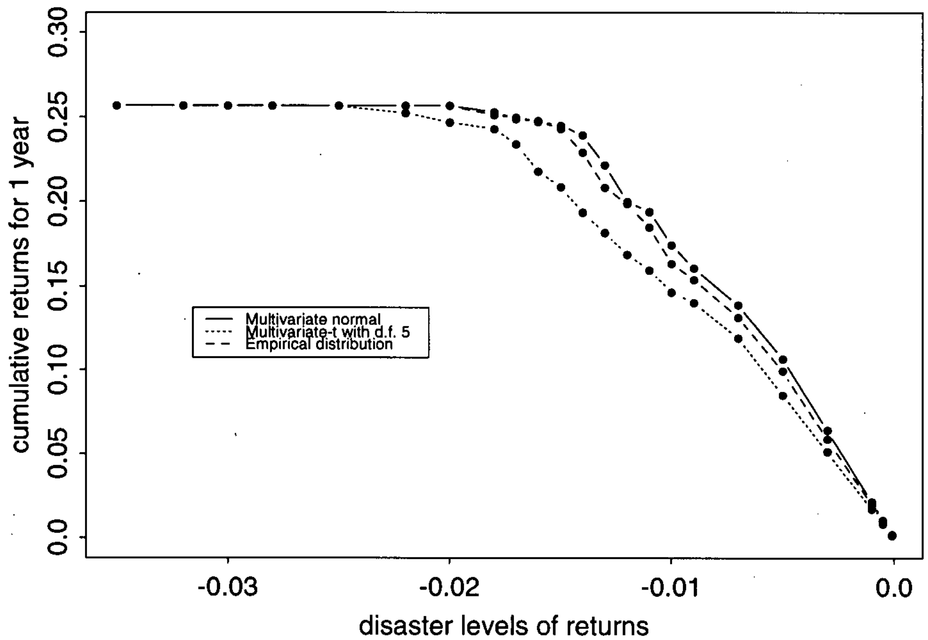


Figure C.6: Portfolio lines for various distribution assumptions, using 1-year window (Cont'd)

Figure C.6b: Portfolio lines for the period 28/10/97 to 27/10/98

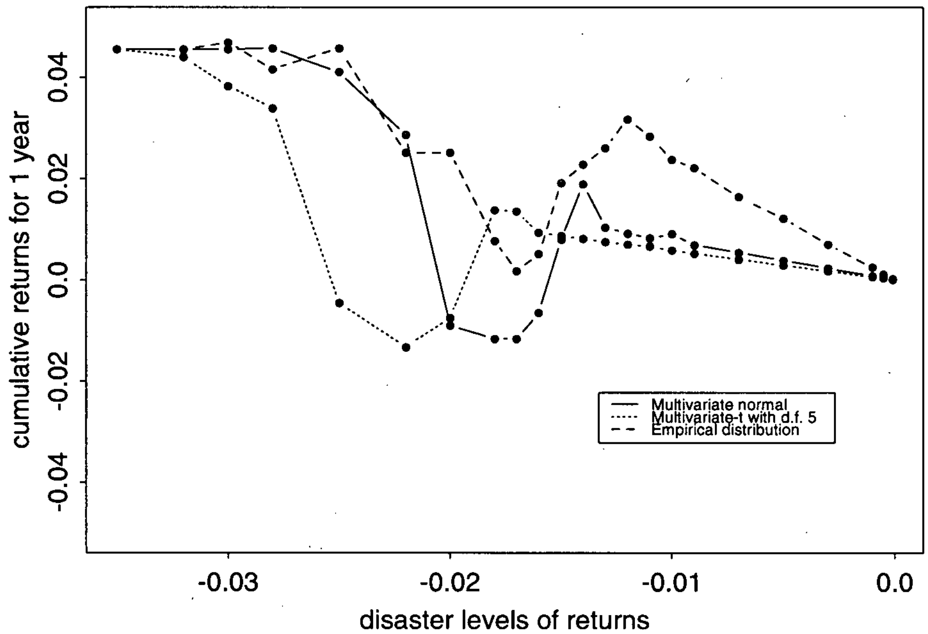


Figure C.6: Portfolio lines for various distribution assumptions, using 1-year window (Cont'd)

Figure C.6c: Portfolio lines for the period 28/10/96 to 27/10/98

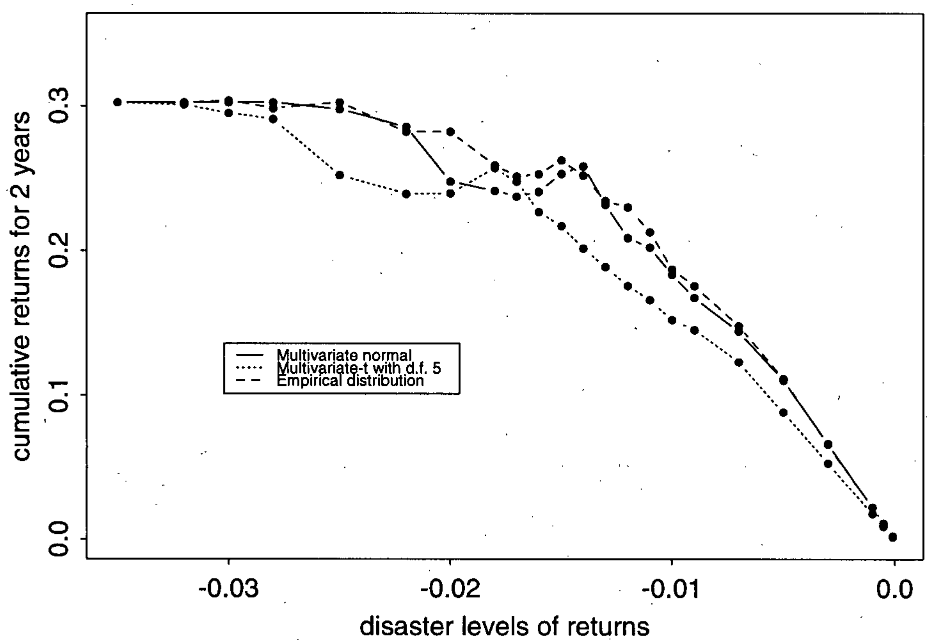


Figure C.7: Portfolio lines for various distribution assumptions, using all data

Figure C.7a: Portfolio lines for the period 28/10/96 to 27/10/97

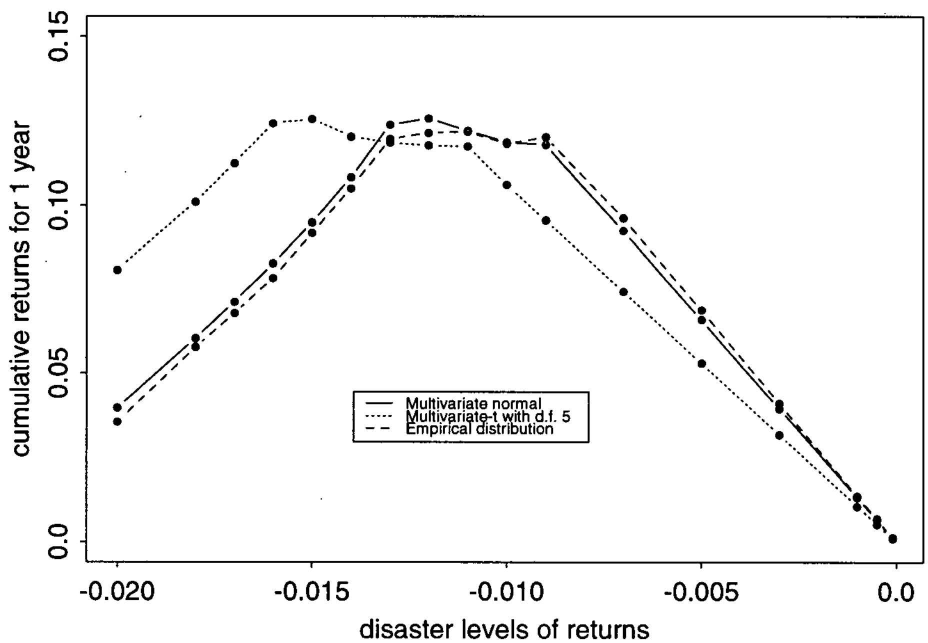


Figure C.7: Portfolio lines for various distribution assumptions, using all data (Cont'd)

Figure C.7b: Portfolio lines for the period 28/10/97 to 27/10/98

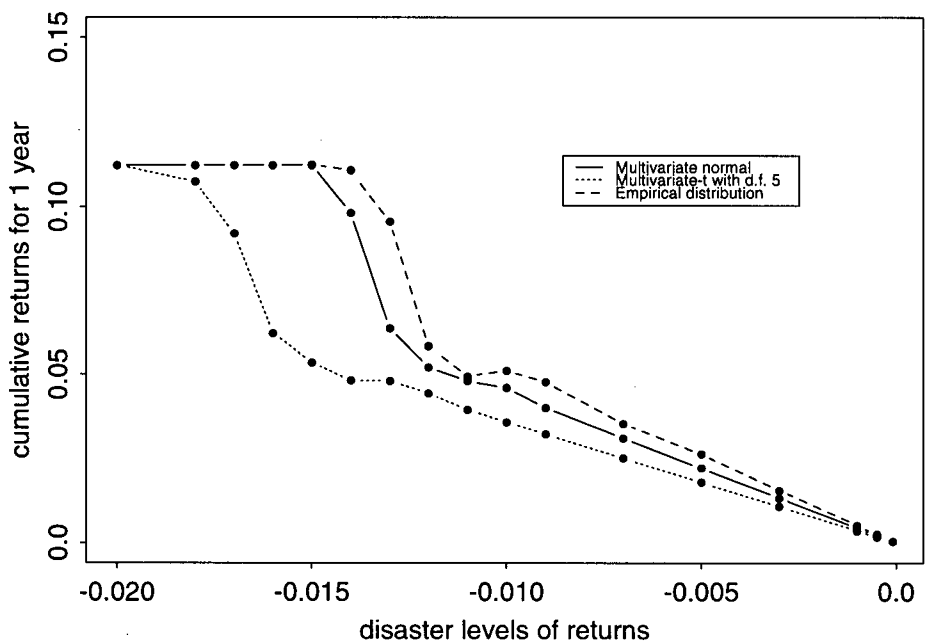
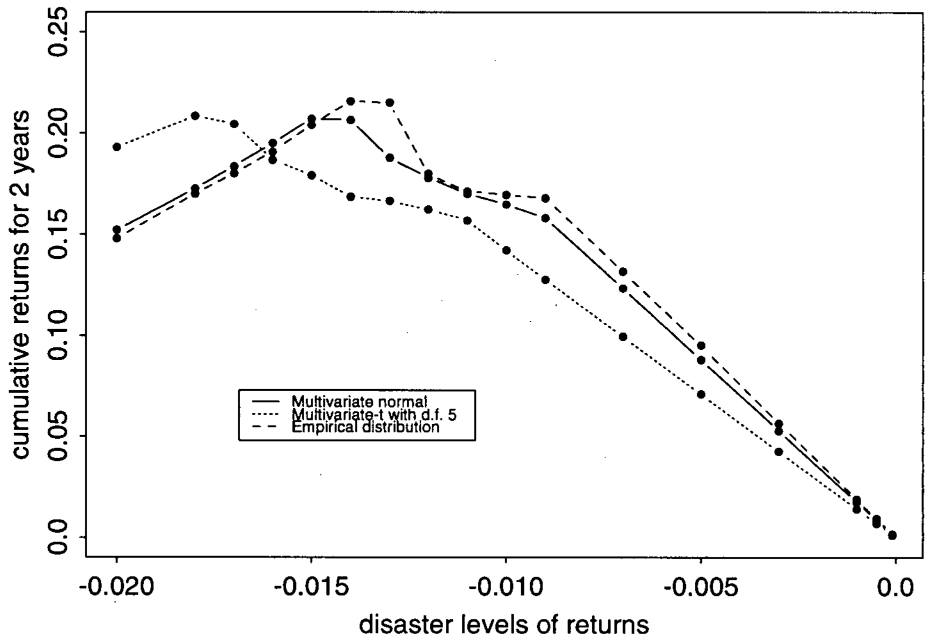


Figure C.7: Portfolio lines for various distribution assumptions, using all data (Cont'd)

Figure C.7c: Portfolio lines for the period 28/10/96 to 27/10/98



Bibliography

[1] Agnew, N. H., R. A. Agnew, J. Rasmussen and K. R. Smith (1969). An Application of Chance Constrained Programming to Portfolio Selection in a Casualty Insurance Firm. *Management Science,* **15**, B512-B520.

[2] Avriel, M., W. E. Diewert, S. Schaible and W. T. Ziemba (1981). Introduction to Concave and Generalized Concave Functions, in: Schaible S. and W. T. Ziemba (eds.), *Generalized Concavity in Optimization and Economics*, New York, London: Academic Press, 21-50.

[3] Baumol, W. J. (1963). An Expected Gain-confidence Limit Criterion for Portfolio Selection. *Management Science,* **10**, 174-182.

[4] Brockett, P. L., A. Charnes, W. W. Cooper, K. H. Kwon and T. W. Ruefli (1992). Chance Constrained Programming Approach to Empirical Analyses of Mutual Fund Investment Strategies. *Decision Sciences,* **23**, 385-408.

[5] Charnes, A. and W. W. Cooper (1959). Chance-Constrained Programming. *Management Science,* **6**, 73-79.

[6] Charnes, A. and W. W. Cooper (1963). Deterministic Equivalents for Optimizing and Satisficing under Chance Constraints. *Operations Research,* **2**, 18-39.

[7] Constantinides, G. M. and A. G. Malliaris (1995). Portfolio Theory, in: Jarrow, R. A., V. Maksimovic and W. T. Ziemba (eds.), *Handbooks in Operations Research and Management Science, Vol. 9*, Amsterdam, New York: Elsevier, 1-30.

[8] Fang, K. T., S. Kotz and K. W. Ng (1989). *Symmetric Multivariate and Related Distributions.* London, New York: Chapman and Hall.

[9] Fang, K. T. and Y. Zhang (1990). *Generalized Multivariate Analysis.* New York: Springer-Verlag.

[10] Gupta, A. K. and T. Varga (1993). *Elliptically Contoured Models in Statistics.* Dordrecht, Boston, London: Kluwer Academic Publishers.

[11] Hillier F. S. and G. J. Lieberman (1990). *Introduction to Operations Research (Fifth Edition).* New York: McGraw-Hill.

[12] J. P. Morgan and Reuters (1996). *RiskMetrics - Technical Document (Fourth Edition).* J. P. Morgan's Web page on the Internet.

[13] Kallberg, J. G. and W. T. Ziemba (1981). Generalized Concave Functions in Stochastic Programming and Portfolio Theory, in: Schaible S. and W. T. Ziemba (eds.), *Generalized Concavity in Optimization and Economics,* New York, London: Academic Press, 719-767.

[14] Kataoka, S. (1963). A Stochastic Programming Model. *Econometrica,* **31**, 181-196.

[15] Kelley, J. E. (1960). The Cutting-Plane Method for Solving Convex Programs. *SIAM* **8**, 703-712.

[16] Lintner, J. (1965). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *The Review of Economics and Statistics,* **47**, 13-37.

[17] Markowitz, H. (1952). Portfolio Selection. *Journal of Finance,* **7**, 77-91.

[18] Markowitz, H. (1959). *Portfolio Selection.* New Haven, London: Yale University Press.

[19] Press, W. H., S. A. Teukolsky, W. T. Vetterling and B. P. Flannery (1992). *Numerical Recipes in C: The Art of Scientific Computing (Second Edition).* Cambridge University Press.

[20] Pyle, D. H. and S. J. Turnovsky (1970). Safety-First and Expected Utility Maximization in Mean-Standard Deviation Portfolio Analysis. *The Review of Economics and Statistics,* **52**, 75-81.

[21] Roy, A. D. (1952). Safety First and the Holding of Assets. *Econometrica,* **20**, 431-449.

[22] Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk. *The Journal of Finance,* **19**, 425-442.

[23] Telser, L. (1955-6). Safety First and Hedging. *Review of Economic Studies,* **23**, 1-16.

[24] Veinott, A. F. (1967). The Supporting Hyperplane Method for Unimodal Programming. *Operations Research,* **15**, 147-152.

[25] Weintraub, A. and J. Vera (1991). A Cutting Plane Approach for Chance Constrained Linear Programs. *Operations Research,* **39**, 776-785.

[26] Zangwill, W. I. (1969). *Nonlinear Programming: A Unified Approach.* Englewood Cliffs, N.J.: Prentice Hall.