

Mean absolute deviation skewness model with
transactions costs

Author: Victor Gumbo*
Supervisor: Prof DCJ De Jongh

Submitted in partial fulfillment of the requirements for the degree of
Magister Scientiae
in the Faculty of Natural and Agricultural Sciences
UNIVERSITY OF PRETORIA

December 2002

*Email: gumbovictor@hotmail.com

DECLARATION

I, the undersigned, hereby declare that the dissertation submitted herewith for the degree Magister Scientiae to the University of Pretoria contains my own, independent work and has not been submitted for any degree at any other university.

Name: Victor Gumbo

Date: 2002-12-06



DEDICATION

For Maitumelo who missed me so much and experienced the difficulty of being mostly alone with the pregnancy of our beloved first born, Aubry Tshiamo. God bless them.



Acknowledgements

The success of this work was a result of the professional guidance of Prof De Jongh who accepted my requests and initiated the search for the bibliographic material that produced this topic of research. I gratefully appreciate and thank Prof B Swart for her professional and motherly advice and assistance. I wish the two Professors many more happy and fruitful years.

Contents

1	Introduction	4
2	Mean Variance Portfolio Theory	6
2.1	Describing the probability distributions	6
2.2	Expected return of a portfolio	8
2.3	Variance of a portfolio	8
3	Review of large-scale portfolio optimization	10
3.1	Problems with Markowitz' model	12
4	Mean absolute deviation model	15
4.1	Compound L_1 risk function	15
4.2	Mean absolute deviation model	23
4.2.1	Method for solving $[P_{MAD}]$	26
5	Utility Theory and Risk Aversion	29
5.1	The von Neumann-Morgenstern utility theory	30
5.2	Economic properties of utility functions	32
5.3	Higher-order derivatives of the utility function	33
6	Skewness and portfolio analysis	34
6.1	A mean-variance skewness model	37
7	Mean-absolute-deviation-skewness model with transaction costs	40
7.1	An algorithm for the solution of the problem	46
8	Back-Testing MADS	49
9	Conclusion	52

1 Introduction

In [4], the authors propose a portfolio optimization model under concave transaction costs employing "absolute deviation" as a measure of risk as outlined in [3]. It is further shown in [3], by applying the model to historical data of NIKKEI 225, that the "mean-absolute deviation model" (MAD) removes most of the difficulties associated with H. Markowitz' mean-variance (MV) model which leads to tedious quadratic programming. The MAD model leads to a linear programming instead of a quadratic one, thus enabling us to solve a large-scale programme of more than 1 000 stocks at a faster and efficient way.

The standard Markowitz MV model [6] bases itself upon the assumptions that: (1) the distribution of the rate of return is multivariate normal, or (2) the utility of the investor is a quadratic function of the rate of return; assumptions which do not necessarily hold in practice.

Investors prefer a positively skewed distribution to a negative one, if the expected value and the variance are the same. Moreover, some investors prefer a distribution with larger skewness at the expense of larger variance, meaning to say utility functions of investors are not quadratic. In the late 50's, Samuelson [10] suggested the importance of the third moment in portfolio optimization. As recent as 1992, Maghrebi [13] tested the skewness preference and persistence hypothesis based on the extended CAPM which incorporates the effect of the third moment of the rate of return using the data of Tokyo Stock Exchange. He reported that investors have a preference for positive skewness in their portfolios and that it is not rejected that positively skewed assets in one period are likely to remain positively skewed in the next period.

In 1995 [2] the authors proposed a mean-variance-skewness (MVS) portfolio optimization model as a natural extension of the classical MV model to the situation where the third order term is not negligible. Earlier on in 1993 [1] the authors had proposed the mean-absolute-deviation-skewness (MADS) model and demonstrated that this model generates a portfolio with a large third moment very quickly.



It is the intention of this project to extend further the MADS model to take into consideration transaction costs. We will assume under MADS that the amount of investment is below the critical point α_j where the transactions cost function is a well specified concave function. We also assume that short-selling (borrowing) is not allowed. We will derive a fairly large-scale non-linearly constrained minimization problem using special techniques. Fortunately, due to ongoing advances in programming softwares, we are now able to solve some large-scale programming problems within a short space of time, and it is our hope that the system proposed in [1] can also be adjusted to solve our problem.

2 Mean Variance Portfolio Theory

As a first example of a portfolio problem, we consider standard mean-variance optimization. Historically, this was one of the earliest problems considered, and it is important because mean-variance analysis provides a basis for the derivation of the equilibrium model known variously as the capital asset pricing model (CAPM), Sharpe-Lintner model, Black model, and the two-factor model. Mean-variance analysis is fully consistent with expected utility maximization but only under special circumstances as will be seen later.

2.1 Describing the probability distributions

In the standard mean-variance portfolio problem, the treatment of risk is limiting in that it takes the variance (or equivalently the standard deviation) of portfolio returns as an adequate risk measure. The question we pose is "how else can one best describe the uncertainty of portfolio rates of return"? In principle, one could list all possible outcomes for the portfolio over a given period. If each outcome results in a payoff such as a R1 profit or rate of return, then this payoff value is the "random variable" in question. A list assigning a probability to all possible values of the random variable is called the "probability distribution" of the random variable.

The reward for holding a portfolio is typically measured by the expected rate of return across all possible scenarios (next section). Actually, the expected value or mean is not the only candidate for the central value of a probability distribution. Other candidates are the median and the mode.

The median is defined as the outcome value that exceeds the outcome value for half the population and is exceeded by the other half. Whereas the expected rate of return is a weighted average of the outcomes, the weights being the probabilities, the median is based on the rank order of the outcomes and takes into account only the order of the outcome values rather than the values themselves. The median differs significantly from the mean in cases where the expected value is dominated by extreme values. An example is the income (or wealth) distribution in a population. A relatively small number of households command a disproportionate share of total income (wealth). The mean income is "pulled up" by these extreme values which makes it non-

representative. The median is free of this effect, since it equals the income level that is exceeded by half the population, regardless of by how much.

Finally, a third candidate for the measure of central value is the mode, which is the most likely value of the distribution or the outcome with the highest probability. However, the expected value is by far the most widely used measure of central or average tendency.

Lets turn to the characterization of the risk implied by the nature of the probability distribution of returns. The idea is to describe the likelihood and magnitude of "surprises" (deviations from the mean) with as small a set of statistics as is needed for accuracy. The easiest way to accomplish this is to answer a set of questions in order of their informational value and to stop at the point where additional questions would not affect our notion of the risk-return trade-off.

The first question is "what is a typical deviation from the expected value?" A natural answer would be, "The expected deviation from the expected value is —." Unfortunately this answer is meaningless because it is necessarily zero: positive deviations from the mean are offset exactly by negative deviations.

There are two ways of getting around this problem. The first is to use the expected "squared deviation" from the mean, which is simply the variance of the probability distribution. The second is to use the expected "absolute" value of the deviation. This is known as MAD (mean-absolute-deviation) which will be seen later and forms the backbone of this present work.

2.2 Expected return of a portfolio

We assume the standard assumption that the class of potentially optimal portfolios are those with greatest expected return for a given level of variance, and, simultaneously, the smallest variance for a given expected return, and no short-selling. The return on a portfolio of assets is a weighted average of the return on the individual assets, the weight applied to each return being a fraction of the portfolio invested in that asset. If x_i is the amount invested in the i th asset, and R_p the return on the portfolio, we have:

$$R_p = \sum_{i=1}^n x_i R_i \quad (1)$$

where R_i is the return on the asset i and $i = 1, \dots, n$

The expected return therefore becomes a weighted average of the expected returns on the individual assets:

$$\begin{aligned} E(R_p) &= E\left[\sum_{i=1}^n x_i R_i\right] \\ &= \sum_{i=1}^n x_i E(R_i) \\ &= \sum_{i=1}^n x_i r_i \end{aligned} \quad (2)$$

where $r_i = E(R_i)$.

2.3 Variance of a portfolio

The variance of a portfolio is given by:

$$\begin{aligned} \sigma_p^2 &= E[R_p - r_p]^2 \\ &= \sum_{i=1}^n x_i^2 \sigma_i^2 + \sum_{i=1}^n \sum_{j=1}^n x_i x_j \sigma_{ij} \\ &= \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij} x_i x_j \end{aligned} \quad (3)$$

where $\sum_{i=1}^n x_i^2 \sigma_i^2$ captures the variance terms and the remainder captures the covariance terms.

Let M_0 be the investor's total fund and ρ be the minimal rate of return he requires. His objective is therefore to minimize risk while getting back at least his minimal expected return. Thus Markowitz employed the standard deviation as a measure of risk to solve the classical MV problem (We shall consider n assets right through):

$$\begin{aligned}
 \text{Minimize} \quad & E\left[\left(\sum_{j=1}^n R_j x_j - E\left[\sum_{j=1}^n R_j x_j\right]\right)^2\right] \\
 \text{Subject to} \quad & E\left[\sum_{j=1}^n R_j x_j\right] \geq \rho M_0 \\
 & \sum_{j=1}^n x_j = M_0 \\
 & x_j \geq 0
 \end{aligned} \tag{4}$$

Since $\sigma_{ij} = \text{cov}[R_i, R_j]$, our problem reduces to:

$$\begin{aligned}
 \text{Minimize} \quad & \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij} x_i x_j \\
 \text{Subject to} \quad & \sum_{j=1}^n r_j x_j \geq \rho M_0 \\
 & \sum_{j=1}^n x_j = M_0 \\
 & x_j \geq 0
 \end{aligned} \tag{5}$$

Among the factors that discredited the application of Markowitz' model was the computational burden associated with it. For a portfolio of only n stocks, we need to calculate $n(n-1)/2$ standard deviations. Thus for S&P500 an amount of 124 750 is required.

3 Review of large-scale portfolio optimization

From now onwards we normalize our problem by considering the x_j 's as fractions of the total fund, rather than the amount, corresponding to each asset j . Thus we want to solve:

$$\begin{aligned}
 \text{Minimize} \quad & \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij} x_i x_j \\
 \text{Subject to} \quad & \sum_{i=1}^n r_j x_j \geq \rho \\
 & \sum_{j=1}^n x_j = 1 \\
 & x_j \geq 0
 \end{aligned} \tag{6}$$

If n is small, we can use many standard algorithms to solve the above. When n is over a few thousands, the problem becomes more complex mainly due to the fact that the σ_{ij} 's are usually non-zero for all i, j and that the number of arithmetic operations to solve the programme depends on the number of non-zero coefficients contained in the model as well as the number of variables.

Perold and Markowitz obtained an alternative representation of a quadratic programme by using a multi-factor model. Assume:

$$R_i = \alpha_i + \beta_{i1} F_1 + \dots + \beta_{iK} F_K + \epsilon_i, i = 1, \dots, n \tag{7}$$

where F_k is the k -th random factor; α_i, β_{ij} are constants, ϵ_i is a random disturbance with mean zero and $cov[F_k, \epsilon_i] = 0, k = 1, \dots, K$

Theorem 3.1 *Let $\sigma_i^2 = E[\epsilon_i^2], f_{rs} = cov[F_r, F_s]$, then the above relation leads to the following expression:*

$$\sum_{i=1}^n \sum_{j=1}^n \sigma_{ij} x_i x_j = \sum_{i=1}^n \sigma_i^2 x_i^2 + \sum_{i=1}^n \sum_{j=1}^n \sum_{r=1}^K \sum_{s=1}^K f_{rs} \beta_{ir} \beta_{js} x_i x_j \tag{8}$$

Proof

$$\begin{aligned}
\sum_{i=1}^n \sum_{j=1}^n \sigma_{ij} x_i x_j &= \sum_{i=1}^n [\sigma_{i1} x_i x_1 + \sigma_{i2} x_i x_2 + \dots + \sigma_{in} x_i x_n] \\
&= (\sigma_1^2 x_1^2 + \sigma_{12} x_1 x_2 + \dots + \sigma_{1n} x_1 x_n) + \\
&\quad (\sigma_{21} x_2 x_1 + \sigma_2^2 x_2^2 + \dots + \sigma_{2n} x_2 x_n) \\
&\quad + \dots + (\sigma_{n1} x_n x_1 + \sigma_{n2} x_n x_2 + \dots + \sigma_n^2 x_n^2) \\
&= (\sigma_1^2 x_1^2 + \sigma_2^2 x_2^2 + \dots + \sigma_n^2 x_n^2) + \\
&\quad [\sigma_{12} x_1 x_2 + \sigma_{13} x_1 x_3 + \dots \\
&\quad + \sigma_{1n} x_1 x_n + \sigma_{21} x_2 x_1 + \dots + \sigma_{2n} x_2 x_n + \sigma_{31} x_3 x_1 + \dots \\
&\quad + \dots + \sigma_{n(n-1)} x_n x_{n-1}] \\
&= \sum_{i=1}^n \sigma_i^2 x_i^2 + \sum_{i=1}^n \sum_{j=1}^n \sum_{r=1}^K \sum_{s=1}^K \text{cov}[F_r, F_s] \beta_{ir} \beta_{js} x_i x_j \\
&= \sum_{i=1}^n \sigma_i^2 x_i^2 + \sum_{i=1}^n \sum_{j=1}^n \sum_{r=1}^K \sum_{s=1}^K f_{rs} \beta_{ir} \beta_{js} x_i x_j \blacksquare \tag{9}
\end{aligned}$$

If we let $y_k = \sum_{j=1}^n \beta_{jk} x_j$, $k = 1, \dots, K$, our problem reduces to:

$$\begin{aligned}
\text{Minimize} \quad & \sum_{i=1}^n \sigma_i^2 x_i^2 + \sum_{r=1}^K \sum_{s=1}^K f_{rs} y_r y_s \\
\text{Subject to} \quad & \sum_{j=1}^n r_j x_j \geq \rho \\
& \sum_{j=1}^n \beta_{jk} x_j - y_k = 0 \\
& \sum_{j=1}^n x_j = 1 \\
& x_j \geq 0
\end{aligned} \tag{10}$$

Because $K \ll n$, this programme can be efficiently solved by using, for example, the sparse matrix techniques developed by Pang and Perold. Taking into account transaction costs $c_j(x_j)$ we get the programme:

$$\begin{aligned}
\text{Minimize} \quad & \sum_{i=1}^n \sigma_i^2 x_i^2 + \sum_{r=1}^K \sum_{s=1}^K f_{rs} y_r y_s \\
\text{Subject to} \quad & \sum_{j=1}^n (r_j x_j - c_j(x_j)) \geq \rho \\
& \sum_{j=1}^n \beta_{jk} x_j - y_k = 0 \\
& \sum_{j=1}^n x_j = 1 \\
& x_j \geq 0 \\
& \sum_{j=1}^n a_{ij} x_j \geq b_i \\
& l_j \leq x_j \leq u_j
\end{aligned} \tag{11}$$

(12)

where the last two correspond to institutional constraints. The model appears nice from a theoretical viewpoint, but it has not found the doors open in the practitioners' homes.

3.1 Problems with Markowitz' model

Markowitz' model itself was not used extensively by practitioners as a tool for optimizing a large-scale portfolio. According to one fund manager of a leading security company in Japan, the problems containing more than 200 variables are rarely solved in practice because of the following:

Computational burden. To build a model, we have to calculate $n(n-1)/2$ constants σ_{ij} 's through historical data or through some future projection. We would not be surprised if practitioners felt that this computation is quite a tedious task. Furthermore, solving a large-scale dense quadratic programming problem like the one just seen where almost all σ_{ij} 's are non-zero is

very difficult if n is over, say 500. This computational difficulty can be substantially alleviated through the use of the factor (index) models (Sharpe 1963, Perold 1984) and sparse matrix techniques (Pang 1980, Perold 1984), but it is still not easy to obtain an optimal solution of a large-scale quadratic programming problem on a real-time basis.

Investor perception. Many practitioners were not fully convinced by the validity of the standard deviation as a measure of risk. They were certainly unhappy to have small or negative profit, but as we know, they feel happy to have larger profits. This means that the investors' perception against risk is not asymmetric around the mean. Unfortunately, recent studies on stock prices in Tokyo Stock Market revealed that most R_j 's are not normally nor symmetrically distributed. Thus we need to consider the third moment of the distribution in addition to the first and the second. So we can view Markowitz' model as a great motivation to the main task we wish to solve.

Transaction/Management cost and cut-off effect. An optimal solution \mathbf{x}^* of a large-scale quadratic programming problem seen in the previous section usually contains many non-zero elements. In fact at least 100-200 components of the solution are expected to be positive when n is over 1000. This means that an investor has to purchase many different stocks, most of which are just a fraction of a percent of the total fund. This is very inconvenient in practice since we have to pay significant amount of transaction costs to buy many different stocks by a small amount. On the other hand we may not be able to purchase small amounts of stock below minimum transaction units. Thus we have to round the numbers to the integer multiples of this minimal unit or else we have to solve an integer quadratic programming problem which is intractable when n is larger than 20, for example. Thus we are forced to eliminate stocks with smaller weight to get around the difficulty. But then this cut-off process distorts the portfolio to an extent that the resulting standard deviation is considerably larger than the one obtained through an exact model. In summary, even though Perold's "Optimizer" is widely used by practitioners to solve problems of the type above when $c_j(x_j) = 0$, (1) we have to introduce as many as one hundred factors to obtain a good statistical fitting, making the process time consuming and tedious. It has been reported that it takes several hours to solve the above programme for n bigger than a few thousands.

- (2) it contains n quadratic terms, making it difficult to solve for n bigger than a few thousands, at least until recently.
- (3) the optimal solution usually contains many positive x variables which then requires us to spend time eliminating a portion of variables to organize a manageable size of assets.
- (4) many investors were not convinced of the validity of the quadratic risk function.

In [15], the author proposed an approximation scheme which can particularly take care of the transaction costs and the constraints associated with minimal transaction units. Also, he demonstrated that this scheme is very effective for the model consisting of up to 500 stocks. In brief, large-scale portfolio optimization using Markowitz' model has been considered impractical not only because of the reasons above but also because of the complications inherent in the implementation of the solution. To address this problem of large-scale optimization, the absolute deviation model has been recently proposed as a substitute to the standard deviation, as a measure of risk. This is the objective of the next section.

4 Mean absolute deviation model

The purpose of this section is to introduce a portfolio optimization model L_1 or better known as MAD, that removes most of the difficulties associated with the classical Harry Markowitz' model.

4.1 Compound L_1 risk function

Let

$$\omega_\alpha(x_1, \dots, x_n) = E\left[\left|\sum_{j=1}^n R_j x_j - E\left[\sum_{j=1}^n R_j x_j\right]\right|_-\right] - \alpha E\left[\left|\sum_{j=1}^n R_j x_j - E\left[\sum_{j=1}^n R_j x_j\right]\right|_+\right] \quad (13)$$

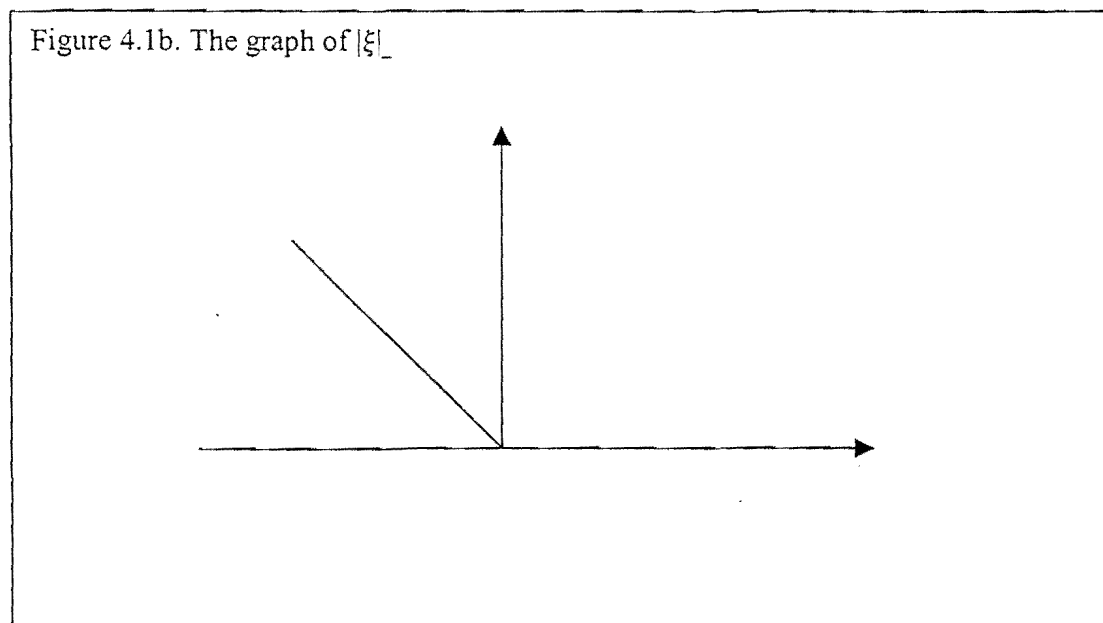
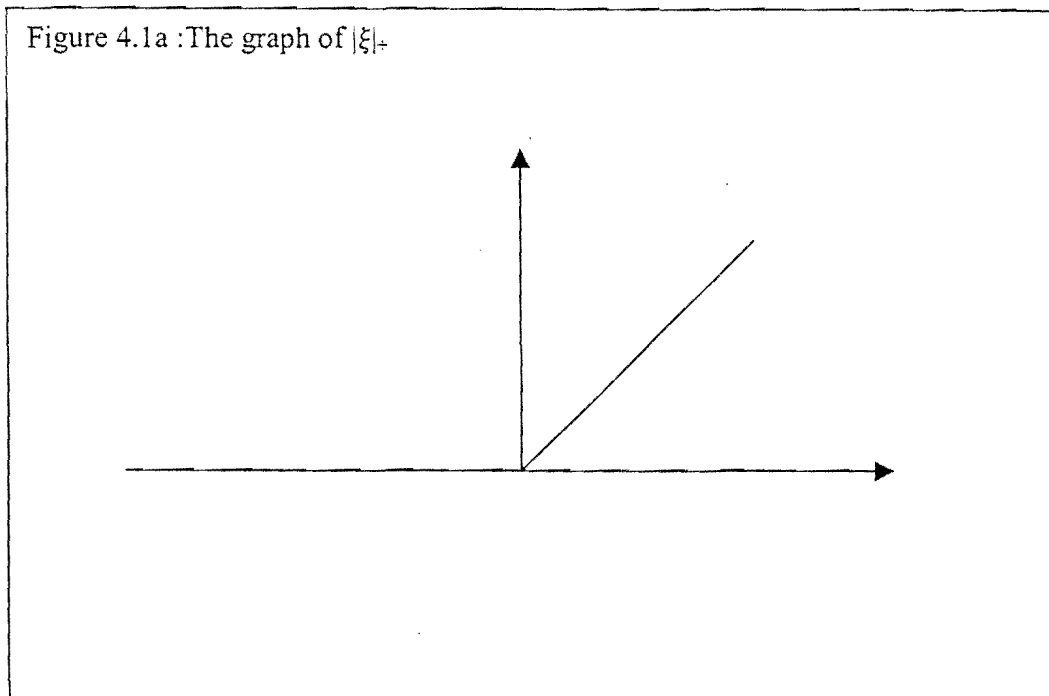
where α is a positive parameter representing the degree of risk aversion of an investor and

$$|\xi|_+ = \begin{cases} \xi & ; \xi \geq 0 \\ 0 & ; \xi < 0 \end{cases} \quad (14)$$

$$|\xi|_- = \begin{cases} 0 & ; \xi \geq 0 \\ -\xi & ; \xi < 0 \end{cases} \quad (15)$$

See corresponding graphs on next page.

Table of graphs.



Theorem 4.1 If (R_1, \dots, R_n) are multi-variate normally distributed with mean (μ_1, \dots, μ_n) and variance-covariance matrix $\Sigma \equiv (\sigma_{ij})$ then

$$\begin{aligned}\omega_\alpha(x_1, \dots, x_n) &= \frac{1 - \alpha}{\sqrt{(2\pi)}} \left\{ \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij} x_i x_j \right\}^{\frac{1}{2}} \\ &= \frac{1 - \alpha}{\sqrt{(2\pi)}} \sigma\end{aligned}\quad (16)$$

Proof. Perold shows that the random variable $Y = \sum_{j=1}^n R_j x_j$ is normally distributed with

$$(\mu, \sigma^2) \equiv \left(\sum_{j=1}^n \mu_j x_j, \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij} x_i x_j \right) \quad (17)$$

Therefore

$$\begin{aligned}\omega_\alpha(x_1, \dots, x_n) &= E\left[\left| \sum_{j=1}^n R_j x_j - E\left[\sum_{j=1}^n R_j x_j \right] \right|_- \right] - \alpha E\left[\left| \sum_{j=1}^n R_j x_j - E\left[\sum_{j=1}^n R_j x_j \right] \right|_+ \right] \\ &= E\left[|Y - E[Y]|_- \right] - \alpha E\left[|Y - E[Y]|_+ \right] \\ &= 1/\sqrt{2\pi\sigma^2} \left[\int_{-\infty}^0 u \exp\left(-\frac{u^2}{2\sigma^2}\right) du - \alpha \int_0^{\infty} u \exp\left(-\frac{u^2}{2\sigma^2}\right) du \right] \\ &= \sigma(1 - \alpha)/(2\pi)^{1/2} \blacksquare\end{aligned}$$

which shows that minimizing variance is the same as minimizing ω_α if $\alpha < 1$ and (R_1, \dots, R_n) are multi-variate and normally distributed.

Our next step is to try and represent ω_α using historical data from some future projection. We shall assume also that the expected value of the random variable can be approximated by the average derived from these data.

Let r_{jt} be the realization of random variable R_j during time period $t = 1, \dots, T$. In particular let

$$\begin{aligned}r_j &\equiv E[R_j] \\ &= \frac{1}{T} \sum_{t=1}^T r_{jt}\end{aligned}\quad (18)$$

Then

$$\begin{aligned}
 r(x_1, \dots, x_n) &= E\left[\sum_{j=1}^n R_j x_j\right] \\
 &= \sum_{j=1}^n r_j x_j
 \end{aligned} \tag{19}$$

and

$$\begin{aligned}
 \omega_\alpha(x_1, \dots, x_n) &= E\left[|\sum R_j x_j - E[\sum R_j x_j]|_{-}\right] - \alpha E\left[|\sum R_j x_j - E[\sum R_j x_j]|_{+}\right] \\
 &= E\left[|\sum R_j x_j - \sum E[R_j] x_j|_{-}\right] - \alpha E\left[|\sum R_j x_j - \sum E[R_j] x_j|_{+}\right] \\
 &= \frac{1}{T} \sum_{t=1}^T \left\{ \left| \sum_{j=1}^n (r_{jt} - r_j) x_j \right|_{-} - \alpha \left| \sum_{j=1}^n (r_{jt} - r_j) x_j \right|_{+} \right\}
 \end{aligned} \tag{20}$$

If we let

$$\xi_t = \sum_{j=1}^n (r_{jt} - r_j) x_j \tag{21}$$

and

$$g_\alpha(\xi) \equiv |\xi|_{-} - \alpha |\xi|_{+} \tag{22}$$

Thus

$$\begin{aligned}
 \omega_\alpha &= \frac{1}{T} \sum_{t=1}^T \{|\xi_t|_{-} - \alpha |\xi_t|_{+}\} \\
 &= \frac{1}{T} \sum_{t=1}^T g_\alpha(\xi_t)
 \end{aligned} \tag{23}$$

For $\alpha = -1$ we have the absolute value L_1 risk while $\alpha = 0$ is associated with the investor who cares about "below average" returns, and $\alpha > 0$ is

associated with the investor whose "below average" returns are compensated by some "above average" returns and specifically for $\alpha > 1$ the investor is viewed as prone to risk. Refer to tables of graphs below and on the next page for the corresponding graphs of $g(\cdot)$ for some different values of α .

Figure 4.2a) Graph of $g_{-1}(\xi)$

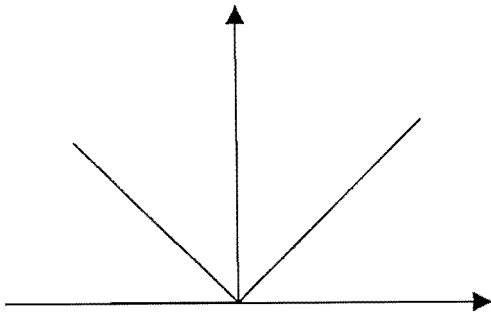
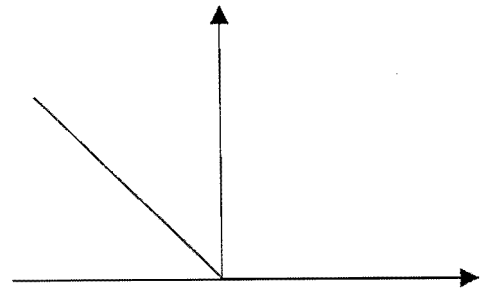


Figure 4.2b) Graph of $g_0(\xi)$



Going back to our portfolio problem, now replacing the quadratic risk function by compound L_1 risk function ω_α , and letting $a_{jt} = r_{jt} - r_j, j = 1, \dots, n, t = 1, \dots, T$ we get the programme:

$$\begin{aligned}
 [P_\alpha] : \text{Minimize} \quad & \sum_{t=1}^T \left\{ \left| \sum_{j=1}^n a_{jt} x_j \right|_- - \alpha \left| \sum_{j=1}^n a_{jt} x_j \right|_+ \right\} \\
 \text{Subject to} \quad & \sum_{j=1}^n r_j x_j \geq \rho \\
 & \sum_{j=1}^n x_j = 1 \\
 & x_j \geq 0
 \end{aligned} \tag{24}$$

We show that the class P_α of such portfolio optimization problems have the same optimal solution for all $\alpha \in (0, 1) \cup (1, +\infty)$

Figure 4.2c) Graph of $g_{1/2}(\xi)$

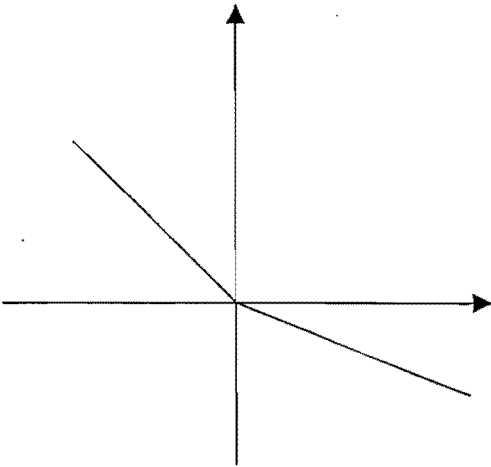
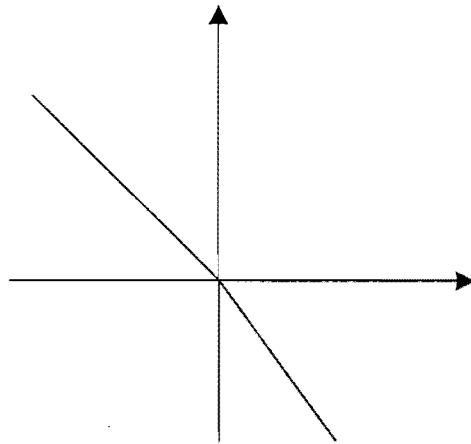


Figure 4.2b) Graph of $g_2(\xi)$



Theorem 4.2 *The class of optimization programme $[P_\alpha]$ have the same optimal solution for all $\alpha \in (0, 1)$. Also they have the same optimal solution for all $\alpha > 1$.*

Proof

Note that

$$\begin{aligned} |\xi|_+ &= \frac{1}{2}(|\xi| + \xi) \\ |\xi|_- &= \frac{1}{2}(|\xi| - \xi) \end{aligned}$$

Thus

$$\begin{aligned} \omega_\alpha(\cdot) &= \frac{1}{T} \sum_{t=1}^T \{|\xi_t|_- - \alpha|\xi_t|_+\} \\ &= \frac{1}{T} \sum_{t=1}^T \frac{1}{2} \{(|\xi_t| - \xi_t) - \alpha(|\xi_t| + \xi_t)\} \\ &= \frac{1}{T} \sum_{t=1}^T \left(\frac{1-\alpha}{2} |\xi_t| - \frac{1+\alpha}{2} \xi_t \right) \tag{25} \\ &= \frac{1}{T} \left\{ \frac{1-\alpha}{2} \sum_{t=1}^T \left| \sum_{j=1}^n a_{jt} x_j \right| - \frac{1+\alpha}{2} \sum_{t=1}^T \sum_{j=1}^n a_{jt} x_j \right\} \\ &= \frac{1-\alpha}{2T} \sum_{t=1}^T \left| \sum_{j=1}^n a_{jt} x_j \right| \end{aligned}$$

since

$$\begin{aligned} \sum_{t=1}^T \sum_{j=1}^n a_{jt} x_j &= \sum_{j=1}^n \sum_{t=1}^T (r_{jt} - r_j) x_j \\ &= \sum_{j=1}^n x_j (T r_j - T r_j) \tag{26} \\ &= 0 \end{aligned}$$

Thus for fixed T , minimizing ω_α is the same as minimizing $\sum_{t=1}^T \left| \sum_{j=1}^n a_{jt} x_j \right|, \forall \alpha < 1$. Also for fixed T it is equivalent to maximizing $\sum_{t=1}^T \left| \sum_{j=1}^n a_{jt} x_j \right|, \forall \alpha > 1$. ■

So our portfolio problem becomes split into two according to alpha.
For $\alpha < 1$ we get the programme:

$$\begin{aligned}
 [P_{\alpha_1}] \text{Minimize} \quad & \sum_{t=1}^T \left| \sum_{j=1}^n a_{jt} x_j \right| \\
 \text{Subject to} \quad & \sum_{j=1}^n r_j x_j \geq \rho \\
 & \sum_{j=1}^n x_j = 1 \\
 & x_j \geq 0
 \end{aligned} \tag{27}$$

For $\alpha > 1$:

$$\begin{aligned}
 [P_{\alpha_2}] \text{Maximize} \quad & \sum_{t=1}^T \left| \sum_{j=1}^n a_{jt} x_j \right| \\
 \text{Subject to} \quad & \sum_{j=1}^n r_j x_j \geq \rho \\
 & \sum_{j=1}^n x_j = 1 \\
 & x_j \geq 0
 \end{aligned} \tag{28}$$

The second programme cannot be converted into a linear programme. Note that the objective function is convex.

Theorem 4.3 *There exists an optimal solution \mathbf{x}_j^* of $[P_{\alpha_2}]$ for which at most two indices j satisfy $\mathbf{x}_j^* > 0$.*

Proof Since the objective function of the programme is convex, there exists an optimal solution among extreme points of the feasible region, which has the stated property. ■

We go back to $[P_{\alpha_1}]$, letting $y_t = \sum_{j=1}^n a_{jt}x_j$, we can convert it to:

$$\begin{aligned}
\text{Minimize} \quad & \sum_{t=1}^T y_t \\
\text{Subject to} \quad & y_t - \sum_{j=1}^n a_{jt}x_j \geq 0 \\
& y_t + \sum_{j=1}^n a_{jt}x_j \geq 0 \\
& \sum_{j=1}^n r_j x_j \geq \rho \\
& \sum_{j=1}^n x_j = 1; x_j \geq 0, j = 1, \dots, n; t = 1, \dots, T
\end{aligned} \tag{29}$$

whose dual is

$$\begin{aligned}
\text{Maximize} \quad & z_1 + z_2 \\
\text{Subject to} \quad & 2 \sum_{t=1}^T a_{jt} \xi_t + r_j z_1 + z_2 \leq 0 \\
& 0 \leq \xi_t \leq 1; z_1 \geq 0
\end{aligned} \tag{30}$$

which is easier to solve than $[P_{\alpha_1}]$.

4.2 Mean absolute deviation model

A special case is when $\alpha = -1$ (see table of graphs) Letting $\omega_{\alpha=-1} = W(x)$, we denote the absolute deviation function by:

$$W(x) = E\left[\left|\sum_{j=1}^n R_j x_j - E\left[\sum_{j=1}^n R_j x_j\right]\right|\right] \tag{31}$$

Theorem 4.4 *If (R_1, \dots, R_n) are multi-variate normally distributed, then $W(x) = \sqrt{\frac{2}{\pi}}\sigma(x)$.*

Note that this is a special case of Theorem 4.1. Thus our problem reduces to the linear programme:

$$\begin{aligned}
\text{Minimize} \quad & W(x) = E\left[\left|\sum_{j=1}^n R_j x_j - E\left[\sum_{j=1}^n R_j x_j\right]\right|\right] \\
\text{Subject to} \quad & \sum_{j=1}^n E(R_j) x_j \geq \rho \\
& \sum_{j=1}^n x_j = 1 \\
& x_j \geq 0
\end{aligned} \tag{32}$$

Let r_{jt} be the realization of the random variable R_j during a time period $t = 1, \dots, T$ and let

$$\begin{aligned}
r_j &= E[R_j] \\
&= \frac{1}{T} \sum_{t=1}^T r_{jt}
\end{aligned} \tag{33}$$

Hence we can approximate $W(x)$ by:

$$\begin{aligned}
W(x) &= E\left[\left|\sum_{j=1}^n R_j x_j - E\left[\sum_{j=1}^n R_j x_j\right]\right|\right] \\
&= E\left[\left|\sum_{j=1}^n R_j x_j - \left[\sum_{j=1}^n E[R_j] x_j\right]\right|\right] \\
&= E\left[\left|\sum_{j=1}^n R_j x_j - \left[\sum_{j=1}^n r_j x_j\right]\right|\right] \\
&= \frac{1}{T} \sum_{t=1}^T \left| \left(\sum_{j=1}^n r_{tj} x_j - \sum_{j=1}^n r_j x_j\right) \right| \\
&= \frac{1}{T} \sum_{t=1}^T \left| \sum_{j=1}^n x_j (r_{tj} - r_j) \right|
\end{aligned} \tag{34}$$

Let $a_{jt} = r_{jt} - r_j, j = 1, \dots, n; t = 1, \dots, T$
Then our portfolio problem becomes:

$$\begin{aligned}
 \text{Minimize} \quad & W(x) = \frac{1}{T} \sum_{t=1}^T \left| \sum_{j=1}^n a_{jt} x_j \right| \\
 \text{Subject to} \quad & \sum_{j=1}^n r_j x_j \geq \rho \\
 & \sum_{j=1}^n x_j = 1 \\
 & x_j \geq 0
 \end{aligned} \tag{35}$$

Let $y_t = \left| \sum_{j=1}^n a_{jt} x_j \right|$. Thus we have the standard linear programme:

$$\begin{aligned}
 \text{Minimize} \quad & W(x) = \frac{1}{T} \sum_{t=1}^T y_t \\
 \text{Subject to} \quad & \sum_{j=1}^n r_j x_j \geq \rho \\
 & \sum_{j=1}^n x_j = 1 \\
 & y_t + \sum_{j=1}^n a_{jt} x_j \geq 0 \\
 & y_t - \sum_{j=1}^n a_{jt} x_j \geq 0 \\
 & x_j \geq 0
 \end{aligned} \tag{36}$$

We can certainly add transaction costs and linear institutional constraints to obtain:

$$\begin{aligned}
[P_{MAD}] \text{Minimize} \quad & \sum_{t=1}^T y_t \\
\text{Subject to} \quad & \sum_{j=1}^n \{r_j x_j - c_j(x_j)\} \geq \rho \\
& y_t - \sum_{j=1}^n a_{jt} x_j \geq 0 \\
& y_t + \sum_{j=1}^n a_{jt} x_j \geq 0 \\
& \sum_{j=1}^n x_j = 1; x_j \geq 0 \\
& \sum_{j=1}^n \hat{a}_{ij} x_j \geq b_i; j, i = 1, \dots, n \\
& l_j \leq x_j \leq u_j
\end{aligned} \tag{37}$$

In [3] it is shown that the portfolios generated by the above programme are easy to construct because of the following advantages of the mean absolute deviation model :

- (1) It can be solved much faster than its counterparts, the MV model, since it is a linear programme (when $c_j(x_j)$ is linear)
- (2) Its optimal solution contains no more than $(T+2)$ assets with positive weights
- (3) It can incorporate all the other features like transaction and institutional constraints (we look at how to deal with the non-linear $c_j(x_j)$ next).

4.2.1 Method for solving $[P_{MAD}]$

We look at how we can tackle our programme with the inclusion of the non-linear costs function $c_j(x_j)$. We let

$$\begin{aligned}
F = \{(\mathbf{x}, \mathbf{y}) : & y_t - \sum_{j=1}^n a_{jt} x_j \geq 0; y_t + \sum_{j=1}^n a_{jt} x_j \geq 0; \sum_{j=1}^n x_j = 1; \\
& x_j \geq 0; \sum_{j=1}^n \hat{a}_{ij} x_j \geq b_i; l_j \leq x_j \leq u_j; y_t \geq 0\}
\end{aligned} \tag{38}$$

Thus we want to solve

$$\begin{aligned}
& \text{Minimize} && \sum_{t=1}^n y_t \\
& \text{Subject to} && \sum_{j=1}^n \{r_j x_j - c_j(x_j)\} \geq \rho \\
& && (\mathbf{x}, \mathbf{y}) \in F \\
& && \sum_{j=1}^n x_j = 1
\end{aligned} \tag{39}$$

Approximate the concave non-decreasing transaction costs function by a linear under-estimating function $\delta_{j0}x_j$ and solve the programme:

$$\begin{aligned}
& [P_{MAD_0}] \text{Minimize} && \sum_{t=1}^n y_t \\
& \text{Subject to} && \sum_{j=1}^n \{r_j x_j - \delta_{j0} x_j\} \geq \rho \\
& && (\mathbf{x}, \mathbf{y}) \in F \\
& && \sum_{j=1}^n x_j = 1, l_j \leq x_j \leq u_j
\end{aligned} \tag{40}$$

Let \mathbf{x}^* be the optimal solution to the above programme. If $\sum \{c_j(x_j^*) - \delta_{j0}x_j^*\} < \epsilon$ then $\delta_{j0}x_j$ is a good approximation of $c_j(x_j)$ and the problem is solved.

Suppose its not a good approximation. Divide $[l_j, u_j]$ into two equal parts $[l_j, \frac{l_j+u_j}{2}]$ and $[\frac{l_j+u_j}{2}, u_j]$ and solve two linear sub-programmes by under-estimating $c_j(x_j)$ by two linear functions $\delta_{j1}x_j$ and $\delta_{j2}x_j$ and solve:

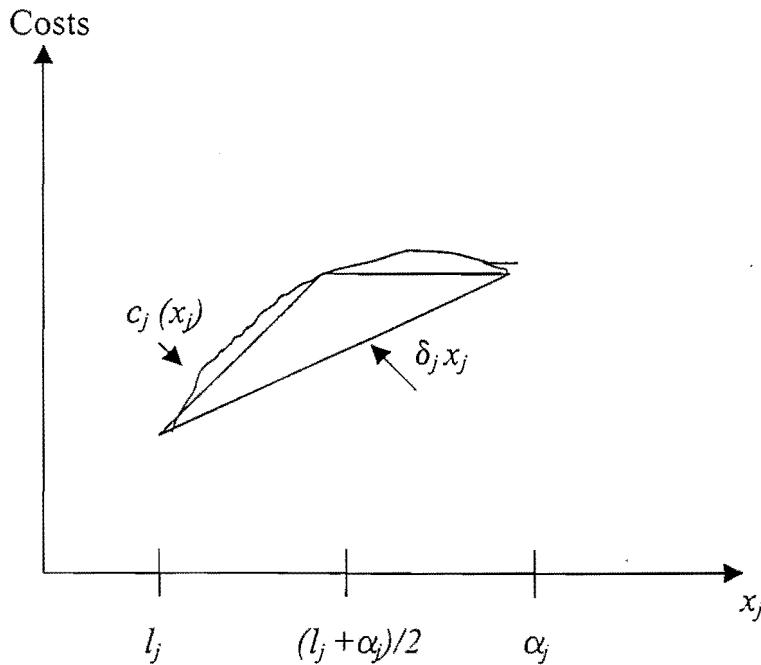
$$\begin{aligned}
& [P_{MAD_1}] \text{Minimize} && \sum_{t=1}^T y_t \\
& \text{Subject to} && \sum_{j=1}^n \{r_j x_j - \delta_{j1} x_j\} \geq \rho \\
& && (\mathbf{x}, \mathbf{y}) \in F \\
& && \sum_{j=1}^n x_j = 1, l_j \leq x_j \leq \frac{l_j + u_j}{2}
\end{aligned} \tag{41}$$

and

$$\begin{aligned}
 [P_{MAD_2}] \text{Minimize} \quad & \sum_{t=1}^T y_t \\
 \text{Subject to} \quad & \sum_{j=1}^n \{r_j x_j - \delta_{j2} x_j\} \geq \rho \\
 & (\mathbf{x}, \mathbf{y}) \in F \\
 & \sum_{j=1}^n x_j = 1, \frac{l_j + u_j}{2} < x_j < u_j \quad (42)
 \end{aligned}$$

Let $(\mathbf{x}_1^*, \mathbf{x}_2^*)$ be the optimal solution to (P_{MAD_1}, P_{MAD_2}) . If $\sum_{j=1}^n \{c_j(x_j^*) - \delta_{ji} x_j^*\} < \epsilon, i = 1, 2$ then $\delta_{ji} x_j$ is a good approximation of our costs function and we compare our solutions to get the ultimate solution to our problem. If otherwise, we continue the process by increasing the number of iterations and fathoming sub-programmes accordingly.

Linear under-estimation of the cost function



5 Utility Theory and Risk Aversion

"Utility" is used as a measure of "happiness" to compare competing investment portfolios based on the expected return and risk of those portfolios. The utility score is then used as a means of ranking portfolios:- higher utility scores imply higher expected return, and lower utility scores imply higher volatility.

The standard consumer's allocation problem is for consumers to choose the most preferred complex in the feasible set. They wish to maximize utility or "happiness" subject to their budget constraints. Extending utility maximization to situations involving risk, the investor expects the greatest of happiness with minimal or no risk. Put in short, the investor wishes to maximize expected utility subject to minimal risk.

What do we mean by risk aversion? A decision maker with a von Neumann-Morgenstern utility function is said to be "risk averse" (at a particular wealth level) if he is unwilling to accept every actuarially fair and immediately resolved gamble with only wealth consequences, that is, those that leave consumption good prices unchanged. If the decision maker is risk averse at all (relevant) wealth levels, he is globally risk averse.

Lets examine a little bit the rationale behind the contention that investors are risk averse. Recognition of risk aversion as central in investment decisions goes back at least to 1733. Daniel Bernoulli, one of a famous Swiss family of distinguished mathematicians spent the years 1725 through 1733 in St Petersburg where he analyzed the following coin-toss game. To enter the game one pays an entry fee. Thereafter a coin is tossed until the first head appears. The number of tails, denoted by n , that appears until the first head is tossed is used to compute the pay-off, \$R to the participant, as :

$$R(n) = 2^n \quad (43)$$

The probability of no tails before the first head ($n = 0$) is $1/2$ and the corresponding pay-off is $2^0 = \$1$. The probability of one tail and then heads ($n = 1$) is $1/2 \times 1/4$ with pay-off $2^1 = \$2$. The probability of two tails and then heads ($n = 2$) is $1/2 \times 1/2 \times 1/2$ etc.

Thus the expected pay-off is therefore:

$$\begin{aligned}
 E(R) &= \sum_{n=0}^{\infty} Pr(n)R(n) \\
 &= 1/2 + 1/2 + \dots \\
 &= \infty
 \end{aligned} \tag{44}$$

The evaluation of this game is called the "St Petersburg Paradox". Although the expected pay-off is infinite, participants obviously will be willing to purchase tickets to play the game only at a finite, and possibly quite modest, entry fee.

Bernoulli resolved the paradox by noting that the investors do not assign the same value per dollar to all pay-offs. Specifically, the greater their wealth, the less their "appreciation" for each extra dollar. We can make this insight mathematically precise by assigning a welfare or "utility value" to any level of investor wealth. Our utility should increase as wealth is higher, but each extra dollar of wealth should increase utility by progressively smaller amounts. The utility function here refers not to investors' satisfaction with alternative portfolio choices but only to the subjective welfare they derive from different levels of wealth. Modern economists would say that investors exhibit "decreasing marginal utility" from an additional pay-off dollar.

Von-Neumann and Morgenstern adapted this approach to investment theory in a complete axiomatic system in 1946, details of which we shall not bother look into in this report.

5.1 The von Neumann-Morgenstern utility theory

Lets assume that we have a risk-less asset with return R_0 and n risky assets with stochastic returns R_i . Then the overall return on the portfolio is $R_p(\mathbf{x}) = R(\mathbf{x})$. Let $U : \mathfrak{R} \mapsto \mathfrak{R}$ be a utility function.

Definition 5.1 (optimal efficient portfolio) *A portfolio \mathbf{x}^* is called optimal relative to the N-M utility $U : \mathfrak{R} \mapsto \mathfrak{R}$ if it is a solution of the optimization problem*

$$\begin{aligned}
 & \text{Maximize } E\{U[R(\mathbf{x})]\} \\
 & \text{Subject to } \sum_{i=1}^n x_i = 1
 \end{aligned} \tag{45}$$

But $R(\mathbf{x}) := \sum_{i=1}^n x_i R_i$. Thus the problem reduces to (max = maximize):

$$\begin{aligned}
 \max E\{U[\sum_{i=1}^n x_i R_i]\} &= \max E\{U[R_0 + \sum_{i=1}^n x_i R_i - R_0]\} \\
 &= \max E\{U[R_0 + \sum_{i=1}^n x_i R_i - R_0 \sum_{i=1}^n x_i]\} \\
 &= \max E\{U[R_0 + \sum_{i=1}^n x_i (R_i - R_0)]\} \quad (46)
 \end{aligned}$$

Proposition 5.1 \mathbf{x}^* is optimal relative to U iff

$$E\{U'[R(\mathbf{x}^*)](R_i - R_0)\} = 0 \quad (47)$$

Proof

By forming the Langrangian

$$L = E[U(\cdot)] + \lambda(1 - \sum_{i=1}^n x_i) \quad (48)$$

we get the first order conditions

$$\frac{\partial L}{\partial x_i} = E[U'(\cdot)R_i] - \lambda = 0 \quad (49)$$

$$\frac{\partial L}{\partial \lambda} = 1 - \sum_{i=1}^n x_i = 0 \quad (50)$$

and since we have a riskless asset in our portfolio, it is convenient to use the second condition to write the first as:

$$E\{U'[(\sum_{i=1}^n x_i R_i)](R_i - R_0)\} = 0 \quad (51)$$

$$E\{U'[R(\mathbf{x}^*)](R_i - R_0)\} = 0 \quad (52)$$

and if \mathbf{x}^* is an optimal solution, $\sum_{i=1}^n x_i R_i = \mathbf{x}^*$, then $E\{U'[R(\mathbf{x}^*)](R_i - R_0)\} = 0$. The other direction follows. ■

5.2 Economic properties of utility functions

1. Utility increases as end of period wealth increases, thus $U'(\cdot) > 0$.
2. The second property takes into account the investor's taste for risk
 - i) if the investor is averse to risk then $U''(\cdot) < 0$, i.e the investor rejects fair gamble.
 - ii) if the investor is risk neutral then $U''(\cdot) = 0$, i.e he is indifferent to fair gamble.
 - iii) otherwise $U''(\cdot) > 0$ and the investor selects a fair gamble.
3. The third is related to the question: if the investor's wealth increases, will more or less wealth be invested in risky assets?
 - i) if the investor increases the amount in risky assets as wealth increases, then he exhibits decreasing absolute risk aversion.
 - ii) if the investor's investments remain unchanged as wealth increases then he manifests constant absolute risk aversion.
 - iii) otherwise, he exhibits increasing absolute risk aversion. Absolute risk aversion is measured by the function:

$$A(w) = \frac{-U''(w)}{U'(w)} \quad (53)$$

4. The last property seeks to answer the question: how does the percentage of wealth invested in risky assets change as wealth changes? This is relative risk aversion. Its measure is the function:

$$R(w) = \frac{-wU'''(w)}{U''(w)} = -wA(w) \quad (54)$$

The class of NM utility functions is important in risk theory and its known as HARA (hyperbolic absolute risk aversion) or LRT (linear risk tolerance) class of utility functions. Its characterized by

$$U(w) = \frac{1-\gamma}{\gamma} \left(\frac{aw}{1-\gamma} + b \right)^\gamma, b > 0 \quad (55)$$

with absolute risk tolerance function

$$T(w) = \frac{1}{A(w)} = \frac{w}{1-\gamma} + ba \quad (56)$$

which is linear as the name suggests. We get different special functions each corresponding to some γ and/or b value(s). We shall not occupy ourselves analyzing them.

5.3 Higher-order derivatives of the utility function

Theorem 5.1 *If investors are consistent in their first m preferences (each of the first m derivatives of U is uniformly positive, negative, or zero) over an unbounded positive domain of w , then the derivatives must alternate in signs, i.e*

$$(-1)^i U^i(w) < 0, i = 1, \dots, m \quad (57)$$

Proof We prove this result by induction. Define $f_n(w) \equiv (-1)^n U^n(w)$ and assume that $f_i(w) < 0$ for $i = 1, \dots, n$. Using the mean value theorem we get

$$\begin{aligned} f_{n-1}(w_2) &= f_{n-1}(w_1) + f'_{n-1}(w^*)(w_2 - w_1) \\ &= f_{n-1}(w_1) - f_n(w^*)(w_2 - w_1) \end{aligned} \quad (58)$$

for some w^* in $[w_1, w_2]$. Now assume $(-1)^i U^i(w) < 0, i = 1, \dots, m$ is false for $n + 1$; that is

$$f_{n+1}(\cdot) = -f'_n(\cdot) \geq 0 \quad (59)$$

Then

$$f_n(w^*) \leq f_n(w_1) \quad (60)$$

thus

$$f_{n-1}(w_2) \geq f_{n-1}(w_1) - f_n(w_1)(w_2 - w_1) \quad (61)$$

Now choose $w_2 > w_1 + \frac{f_{n-1}(w_1)}{f_n(w_1)}$. This choice is possible since the ratio is positive and the domain of interest is unbounded above. Substituting gives us

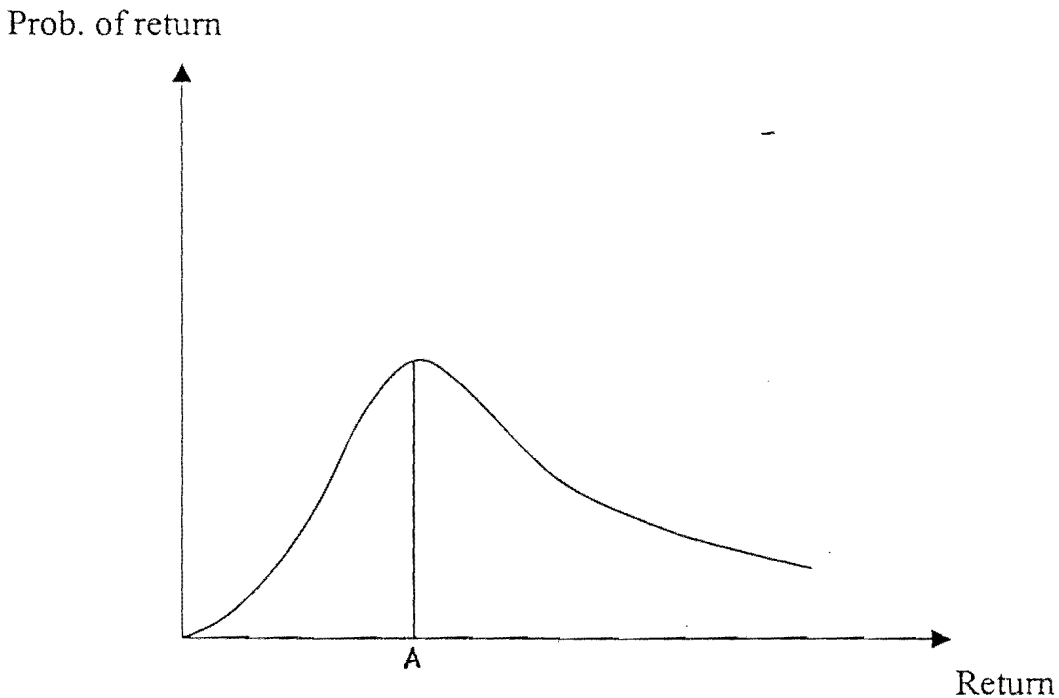
$$f_{n-1}(w_2) > f_{n-1}(w_1) - f_n(w_1) \frac{f_{n-1}(w_1)}{f_n(w_1)} = 0 \quad (62)$$

which contradicts our assumption. ■

6 Skewness and portfolio analysis

Elton and Gruber in [8] referring to the importance of skewness in portfolio analysis said. "...this developmental work has not been done. Thus practical portfolio analysis in three moments must await development of a set of analytical techniques for estimating and solving problems involving skewness". In this respect, a number of authors have proposed selecting portfolios on the basis of the first three moments of return distributions rather than the first two. Skewness is a measure of the asymmetry of a distribution. The normal distribution has zero skewness since the shape of the distribution above the mode is a mirror's image of the shape below the mode. The log-normal distribution in the diagram below has positive skewness. Point A indicates the mode. The log-normal has more observations above this value than below. It is said to be skewed towards high values or exhibit positive skewness. Researchers in skewness believe investors should prefer positive skewness. All else constant, they should prefer portfolios with a larger probability of very large payoffs. This is not only logical but also consistent with the empirical evidence that investors are risk averse and looking for higher returns.

Figure 6.1: The log-normal distribution



Skewness usually means the third central moment divided by the cube of the standard deviation, i.e if R is a random variable then:

$$\kappa(R) = \frac{E[R - E[R]]^3}{E[(R - E(R))\frac{2}{3}]} \quad (63)$$

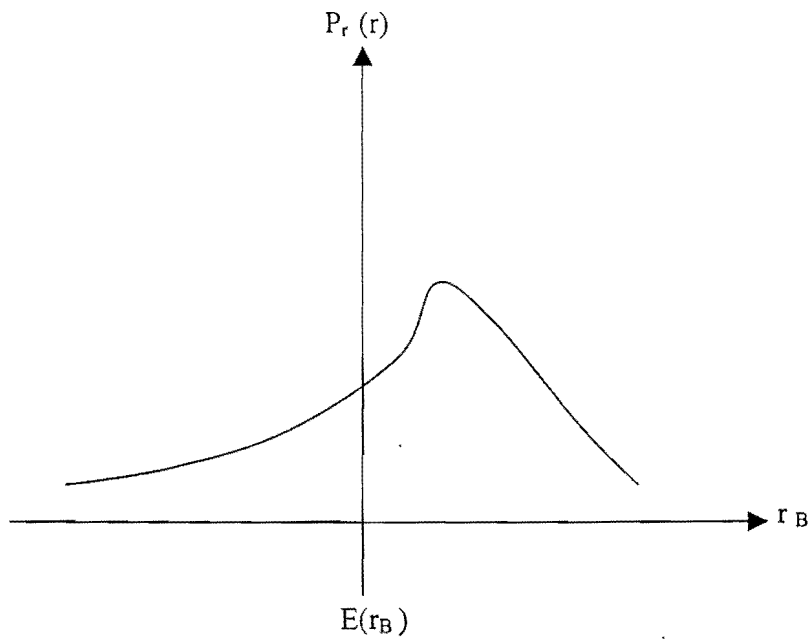
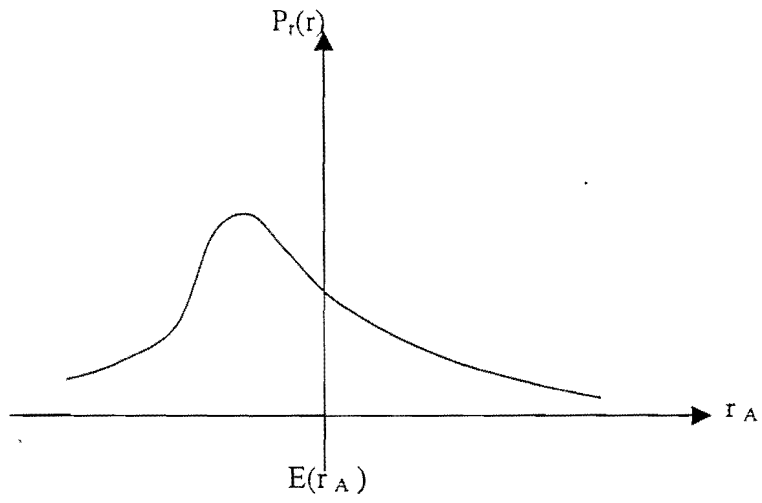
but in general it is the un-normalized third central moment :

$$m^3 = E[R - E(R)]^3 \quad (64)$$

Although the variance measures the average squared deviation from the expected value, it does not provide a full description of risk. To see why, consider two log-normal distributions for rates of return on a portfolio (See tables of graphs on next page). A and B are probability distributions with identical expected values and variances. The graphs show that variances are identical because probability distribution B is a mirror image of A . What is the empirical difference between A and B ? A is characterized by more likely but small losses and less likely but extreme gains. This pattern is reversed in B . The difference is important. When we talk about risk, we really mean "bad surprises". The "bad surprises" in A , although they are more likely, are small (and limited) in magnitude. The ones in B could be extreme, indeed unbounded! A risk averse investor will prefer A to B on these grounds; hence it is worthwhile to quantify this characteristic. The asymmetry of the distribution is called "skewness", and is measured by the third central moment m^3 seen above in this section.

Cubing the deviations from the expected value preserves their signs, which allows us to distinguish good from bad surprises. Because this procedure gives greater weight to larger deviations, it causes the "long tail" of the distribution to dominate the measure of skewness. Thus the skewness of the distribution will be positive for a right-skewed distribution such as A and negative for a left-skewed distribution like B . To summarize, or rather to introduce, the first moment represents the reward. The second and higher moments characterize the uncertainty of the reward. All the even moments (variance, m^4 , etc) represent the likelihood of extreme values. Larger values for these moments indicate greater uncertainty. The odd moments (m^3 , m^5 , etc) represent measures of symmetry. Positive numbers are associated with positive skewness and hence are desirable.

Log-normal distributions for rates of return on a portfolio



6.1 A mean-variance skewness model

Let R be the random variable representing rate of return of the assets S_j and let x_j be the fraction of the fund to be invested in asset S_j . The rate of return of the portfolio $\mathbf{x} = (x_1, \dots, x_n)$ is given by:

$$R(\mathbf{x}) = \sum_{j=1}^n R_j x_j \quad (65)$$

Let $U = U(R(\mathbf{x}))$ be the investor's utility function. Then, as seen in the previous section, his portfolio optimization problem is to maximize his expected happiness subject to his budget constraints. In other words, he wishes to:

$$\begin{aligned} & \text{Maximize} && E[U(R(\mathbf{x}))] \\ & \text{Subject to} && \sum_{j=1}^n x_j = 1 \\ & && x_j \geq 0 \end{aligned} \quad (66)$$

Assume that $U(\cdot)$ can be approximated by the third order Taylor's expansion around the mean $r(\mathbf{x})$ of $R(\mathbf{x})$. Thus

$$\begin{aligned} U(R(\mathbf{x})) &= U(r(\mathbf{x})) + U'(r(\mathbf{x}))[R(\mathbf{x}) - r(\mathbf{x})] + \\ &+ \frac{1}{2}U''(r(\mathbf{x}))[R(\mathbf{x}) - r(\mathbf{x})]^2 + \frac{1}{6}U'''(r(\mathbf{x}))[R(\mathbf{x}) - r(\mathbf{x})]^3 \end{aligned} \quad (67)$$

Thus applying $E(\cdot)$ we get:

$$E[U(R(\mathbf{x}))] = U(r(\mathbf{x})) + \frac{1}{2}U''(r(\mathbf{x}))E[(R(\mathbf{x}) - r(\mathbf{x}))^2] + \frac{1}{6}U'''(r(\mathbf{x}))E[(R(\mathbf{x}) - r(\mathbf{x}))^3] \quad (68)$$

since $E[R(\mathbf{x})] = r(\mathbf{x})$.

So the maximal value of the expected utility can be obtained by solving

$$\begin{aligned}
 & \text{Maximize} && E[U(R(x))] \\
 & \text{Subject to} && E[R(x)] = r \\
 & && \sum_{j=1}^n x_j = 1; x_j \geq 0
 \end{aligned} \tag{69}$$

Maximizing the objective function here is simply reduced to maximizing only the last term $E[(R(x)-r(x))^3]$ since $U(r(x))$ is a constant and $\frac{1}{2}U''(r(x))E[(R(x)-r(x))^2] < 0$. Therefore formulating our MVS problem we wish to:

$$\begin{aligned}
 & \text{Maximize} && E[(R(x) - r(x))^3] \\
 & \text{Subject to} && E[(R(x) - r(x))^2] = \sigma^2 \\
 & && E[R(X)] = r \\
 & && \sum_{j=1}^n x_j = 1; x_j \geq 0
 \end{aligned} \tag{70}$$

which can also be written in the following way letting:

$$\begin{aligned}
 V[R(x)] &= E[(R(x) - r(x))^2] \\
 \gamma[R(x)] &= E[(R(x) - r(x))^3]
 \end{aligned} \tag{71}$$

where r, σ are given parameters:

$$\begin{aligned}
 & \text{Maximize} && \gamma[R(x)] \\
 & \text{Subject to} && E[R(x)] = r \\
 & && V[R(x)] = \sigma^2 \\
 & && \sum_{j=1}^n x_j = 1 \\
 & && x_j \geq 0
 \end{aligned} \tag{72}$$

Let $\mathbf{x}^*(r, \sigma)$ be an optimal solution of this parametric programme and let $\gamma^*(r, \sigma)$ be the associated maximal value of $\gamma[R(x)]$. Then we have:

$$E[U(R(x))] = U(r(x)) + \frac{1}{2}U''(r(x))\sigma^2 + \frac{1}{6}U'''(r(x))\gamma^*(r, \sigma) \tag{73}$$

Thus we could be able to obtain an approximate optimal value of $E[U(R(x))]$ if we can parametrically calculate the "efficient surface" $\gamma^*(r, \sigma)$. Let

$$\begin{aligned}
 r_j &= E[R_j] \\
 \sigma_{ij} &= E[(R_i - r_i)(R_j - r_j)] \\
 \gamma_{ijk} &= E[(R_i - r_i)(R_j - r_j)(R_k - r_k)]
 \end{aligned} \tag{74}$$

Thus our MVS problem can now be written as:

$$\begin{aligned}
 \text{Maximize} \quad & \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \gamma_{ijk} x_i x_j x_k \\
 \text{Subject to} \quad & \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij} x_i x_j = \sigma^2 \\
 & \sum_{j=1}^n r_j x_j = r \\
 & \sum_{j=1}^n x_j = 1; x_j \geq 0
 \end{aligned} \tag{75}$$

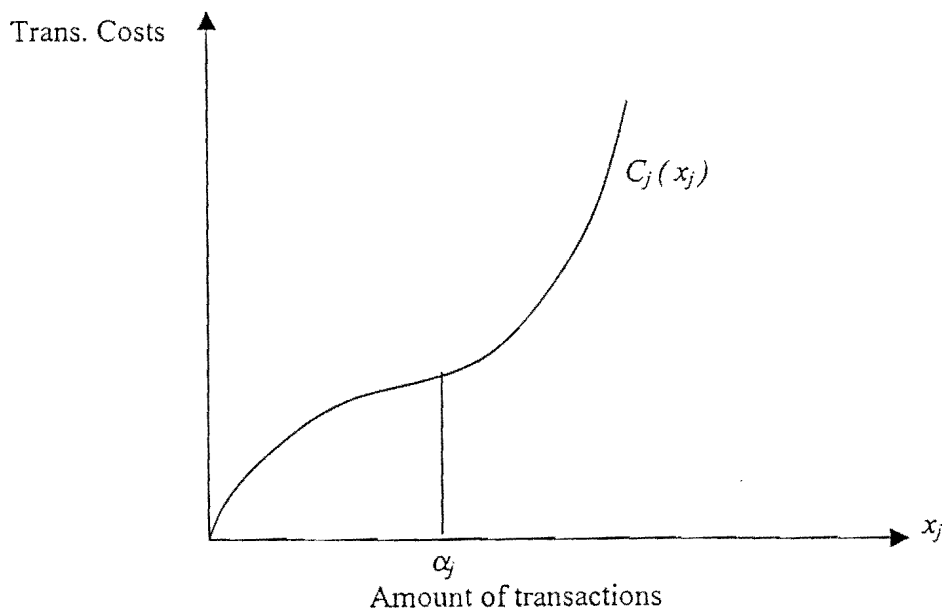
Unfortunately both forms of our MVS problem are typical non-concave maximization problems whose global maximum cannot be calculated by state-of-the-art non-linear programming algorithms. Also it is virtually impossible to collect the γ_{ijk} when n is over one thousand, not to mention the σ_{ij} . Hence we need to introduce some kind of approximation to convert an intractable problem into a tractable one.

7 Mean-absolute-deviation-skewness model with transaction costs

In the standard portfolio analysis, it is assumed that the investor is risk averse and that his utility is a function of the mean and variance of the return of the portfolio, or can be approximated as such. It turns out that the third moment plays an important role if the distribution of the rate of return of the assets is asymmetric around the mean. As mentioned earlier, an investor would prefer a portfolio with larger third moment if the mean and the variance are the same. In this section we propose a portfolio with a large third moment under the constraints of the first and the second moments, and concave transaction costs. We wish to formulate and propose a solution to a programme with among the constraints, a strictly concave function.

An investor has to pay a certain amount of fees when investing (or dis-investing). Let x_j be the amount of investment in asset j . Let $c_j(x_j)$ be the transaction costs associated with investing in asset j . $c_j(x_j)$ is non-decreasing and concave up to some point α_j . The total transaction costs is therefore $\sum_{j=1}^n c_j(x_j)$. See graph below.

Transaction costs function



We now proceed to approximate the third moment and introduce our model.

Let

$$g(u) = \begin{cases} 0 & : u \geq 0 \\ u^3 & : u < 0 \end{cases} \quad (76)$$

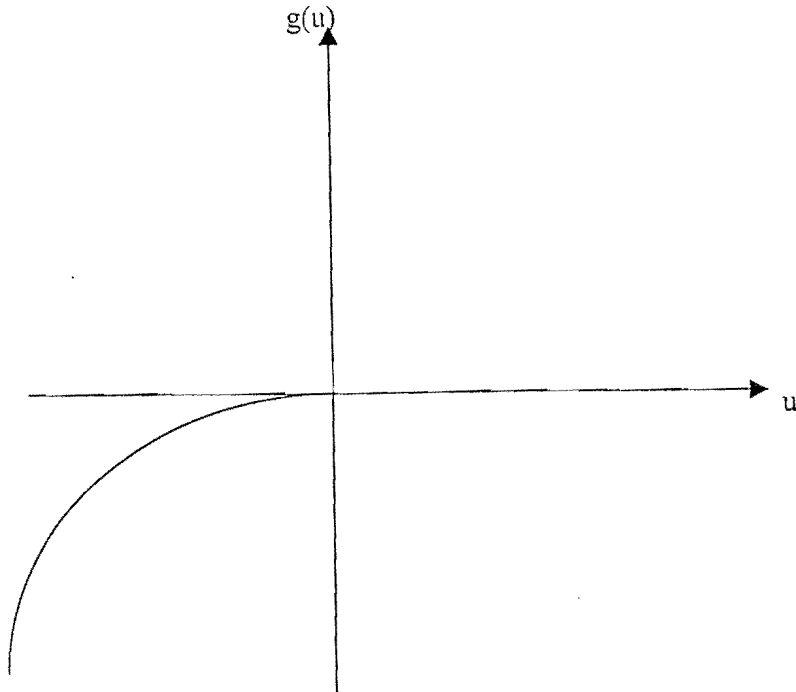
See graph below. We define the lower semi-third moment of a random variable R by:

$$\gamma_-(R) = E[g(R - E[R])] \quad (77)$$

Instead of optimizing the third moment in the previous section, we will optimize the lower semi-third moment of the rate of return of the portfolio.

$$\begin{aligned} \text{Minimize} & \quad \gamma_-(R(x)) \\ \text{Subject to} & \quad E[R(x)] = r \\ & \quad V[R(x)] = \sigma^2 \\ & \quad \sum_{j=1}^n x_j = 1 \\ & \quad x_j \geq 0 \end{aligned} \quad (78)$$

Figure 7.2



We wish to replace the non-convex constraint in terms of variance by "absolute deviation" as seen earlier on.

$$W[R] = E[|R - E[R]|] \quad (79)$$

Then we obtain the programme:

$$\begin{aligned} \text{Minimize} \quad & \gamma_-(R(x)) \\ \text{Subject to} \quad & E[R(x)] = r \\ & W[R(x)] \leq \omega \\ & \sum_{j=1}^n x_j = 1 \\ & x_j \geq 0 \end{aligned} \quad (80)$$

where ω is some specified risk.

If \mathbf{x}^* is an optimal solution to the above programme, then the portfolio \mathbf{x}^* is expected to have a shorter tail to the left of the mean. Hence it will be expected to have a relatively big positive third moment.

But

$$R(x) = \sum_{j=1}^n R_j x_j \quad (81)$$

and

$$\begin{aligned} \gamma_-(R) &= E[g(R - E(R))] \\ &= E[g(\sum_{j=1}^n R_j x_j - E(\sum_{j=1}^n R_j x_j))] \\ &= E[g(\sum_{j=1}^n R_j x_j - \sum_{j=1}^n r_j x_j)] \\ &= E[g(\sum_{j=1}^n (R_j x_j - r_j x_j))] \\ &= E[g(\sum_{j=1}^n (R_j - r_j) x_j)] \end{aligned} \quad (82)$$

We wish to replace the non-convex constraint in terms of variance by "absolute deviation" as seen earlier on.

$$W[R] = E[|R - E[R]|] \quad (79)$$

Then we obtain the programme:

$$\begin{aligned} \text{Minimize} \quad & \gamma_-(R(x)) \\ \text{Subject to} \quad & E[R(x)] = r \\ & W[R(x)] \leq \omega \\ & \sum_{j=1}^n x_j = 1 \\ & x_j \geq 0 \end{aligned} \quad (80)$$

where ω is some specified risk.

If \mathbf{x}^* is an optimal solution to the above programme, then the portfolio \mathbf{x}^* is expected to have a shorter tail to the left of the mean. Hence it will be expected to have a relatively big positive third moment.

But

$$R(x) = \sum_{j=1}^n R_j x_j \quad (81)$$

and

$$\begin{aligned} \gamma_-(R) &= E[g(R - E(R))] \\ &= E[g(\sum_{j=1}^n R_j x_j - E(\sum_{j=1}^n R_j x_j))] \\ &= E[g(\sum_{j=1}^n R_j x_j - \sum_{j=1}^n r_j x_j)] \\ &= E[g(\sum_{j=1}^n (R_j x_j - r_j x_j))] \\ &= E[g(\sum_{j=1}^n (R_j - r_j) x_j)] \end{aligned} \quad (82)$$

So that our problem becomes:

$$\begin{aligned}
 \text{Minimize} \quad & E[g(\sum_{j=1}^n (R_j - r_j)x_j)] \\
 \text{Subject to} \quad & W(\sum_{j=1}^n R_j x_j) \leq \omega \\
 & \sum_{j=1}^n r_j x_j = r \\
 & \sum_{j=1}^n x_j = 1 \\
 & x_j \geq 0
 \end{aligned} \tag{83}$$

This is a non-linear programming problem because the preference function is concave. We replace $g(\cdot)$ by a piecewise linear concave function $G(\cdot)$ where:

$$G(u) = -|u - \rho_1|_- - \alpha|u - \rho_2|_-, \alpha > 0, \rho_1 > \rho_2 \tag{84}$$

and

$$|v|_- = \begin{cases} 0 & ;v \geq 0 \\ -v & ;v < 0 \end{cases} \tag{85}$$

Thus we obtain the problem:

$$\begin{aligned}
 \text{Minimize} \quad & E[|\sum_{j=1}^n R_j x_j - \rho_1|_-] + \alpha E[|\sum_{j=1}^n R_j x_j - \rho_2|_-] \\
 \text{Subject to} \quad & W(\sum_{j=1}^n R_j x_j) \leq \omega \\
 & \sum_{j=1}^n r_j x_j = r \\
 & \sum_{j=1}^n x_j = 1; x_j \geq 0
 \end{aligned} \tag{86}$$

Recall that:

$$E\left[\left|\sum_{j=1}^n R_j x_j - \rho_1\right|_{-}\right] = \frac{1}{T-1} \sum_{t=1}^T \left|\sum_{j=1}^n r_{jt} x_j - \rho_1\right|_{-} \quad (87)$$

$$E\left[\left|\sum_{j=1}^n R_j x_j - \rho_2\right|_{-}\right] = \frac{1}{T-1} \sum_{t=1}^T \left|\sum_{j=1}^n r_{jt} x_j - \rho_2\right|_{-} \quad (88)$$

Thus

$$\begin{aligned} E\left[g\left(\sum_{j=1}^n R_j x_j\right)\right] &= E\left[G\left(\sum_{j=1}^n R_j x_j\right)\right] \\ &= E\left[\left|\sum_{j=1}^n R_j x_j - \rho_1\right|_{-}\right] + \alpha E\left[\left|\sum_{j=1}^n R_j x_j - \rho_2\right|_{-}\right] \\ &= \frac{1}{T-1} \sum_{t=1}^T \left|\sum_{j=1}^n r_{jt} x_j - \rho_1\right|_{-} + \alpha \frac{1}{T-1} \sum_{t=1}^T \left|\sum_{j=1}^n r_{jt} x_j - \rho_2\right|_{-} \\ &= \frac{1}{T-1} \sum_{t=1}^T (u_t + \alpha v_t) \end{aligned} \quad (89)$$

where

$$u_t = \left|\sum_{j=1}^n r_{jt} x_j - \rho_1\right|_{-} \quad (90)$$

$$v_t = \left|\sum_{j=1}^n r_{jt} x_j - \rho_2\right|_{-} \quad (91)$$

On the other hand

$$\begin{aligned} W\left(\sum_{j=1}^n R_j x_j\right) &= E\left[\left|\sum_{j=1}^n R_j x_j - \sum_{j=1}^n r_j x_j\right|\right] \\ &= \frac{1}{T-1} \sum_{t=1}^T \left|\sum_{j=1}^n r_{jt} x_j - \sum_{j=1}^n r_j x_j\right| \\ &= \frac{1}{T-1} \sum_{t=1}^T \left|\sum_{j=1}^n (r_{jt} - r_j) x_j\right| \\ &= \frac{1}{T-1} \sum_{t=1}^T \xi_t \\ &\leq \omega \end{aligned} \quad (92)$$

where $|\sum_{j=1}^n (r_{jt} - r_j)x_j| = \xi_t$

Now if $c_j(x_j)$ is a concave function representing the expected transaction costs associated with the investment then our expected return with transaction costs is transformed to:

$$\sum_{j=1}^n [r_j x_j - c_j(x_j)] = r \quad (93)$$

where again r is the given parameter of rate of return.

Substituting accordingly we get the non-linear programming problem:

$$\begin{aligned} \text{Minimize} \quad & \frac{1}{T-1} \sum_{t=1}^T (u_t + \alpha v_t) \\ \text{Subject to} \quad & \frac{1}{T-1} \sum_{t=1}^T \xi_t \leq \omega \\ & \sum_{j=1}^n [r_j x_j - c_j(x_j)] = r \\ & u_t + \sum_{j=1}^n r_{jt} x_j \geq \rho_1 \\ & v_t + \sum_{j=1}^n r_{jt} x_j \geq \rho_2 \\ & \sum_{j=1}^n x_j = 1, u_t \geq 0, v_t \geq 0, \xi_t \geq 0, 0 \leq x_j \leq \alpha_j, t = 1, \dots, T, \end{aligned} \quad (94)$$

Which can be transformed into:

$$\begin{aligned} \text{Minimize} \quad & \sum_{t=1}^T y_t \\ \text{Subject to} \quad & F = \{(\mathbf{x}, \mathbf{y}) : \frac{1}{T-1} \sum_{t=1}^T \xi_t \leq \omega; u_t + \sum_{j=1}^n r_{jt} x_j \geq \rho_1; v_t + \sum_{j=1}^n r_{jt} x_j \geq \rho_2; \\ & \sum_{j=1}^n x_j = 1; u_t \geq 0, v_t \geq 0; \xi_t \geq 0, 0 \leq x_j \leq \alpha_j\} \\ & \sum_{j=1}^n [r_j x_j - c_j(x_j)] = r \\ & y_t \geq 0, 0 \leq x_j \leq \alpha_j \end{aligned} \quad (95)$$

where $y_t = u_t + \alpha v_t$.

This program is non-linear because of the transactions cost function, and thus it will not be possible to apply directly conventional methods to solve it. We suggest the following method to tackle it. It should be noted however that if a solution is not encountered early, the method can be tedious but a computer programme should iron out this hurdle.

7.1 An algorithm for the solution of the problem

Replace $c_j(x_j)$ by an underestimating linear function $\delta_j x_j$ and solve the standard linear programme P_0 by the simplex method, taking into account that we require higher returns than expected:

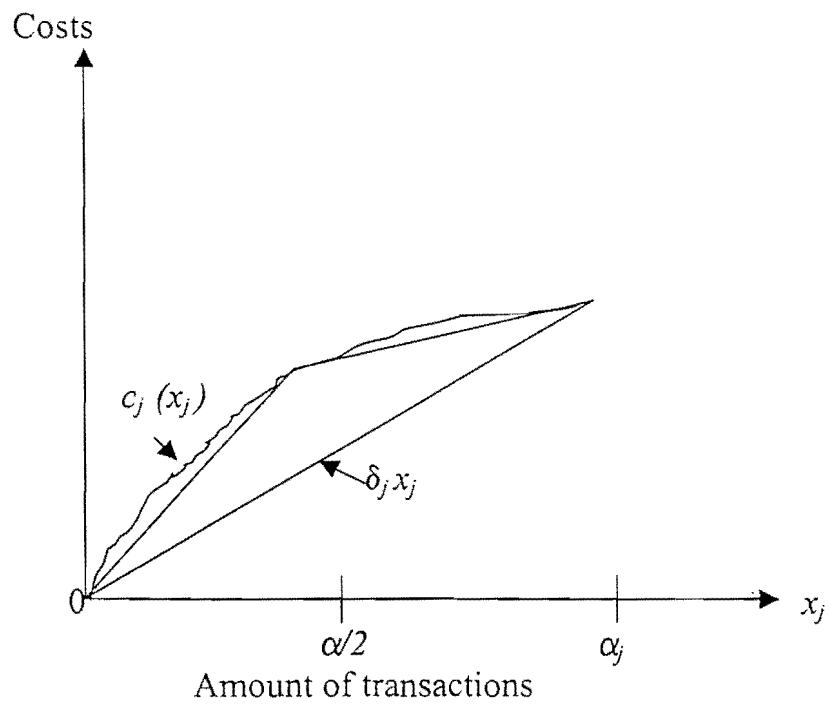
$$\begin{aligned}
 [P_0] : \text{Minimize} \quad & \sum_{t=1}^T y_t \\
 \text{Subject to} \quad & (\mathbf{x}, \mathbf{y}) \in \mathbf{F} \\
 & \sum_{j=1}^n [r_j x_j - \delta_j x_j] \geq r \\
 & 0 \leq x_j \leq \alpha_j; y_t \geq 0
 \end{aligned} \tag{96}$$

Let \mathbf{x}_0^* be the optimal solution to P_0 . If $\sum_{j=1}^n [c_j(x_0^*) - \delta_j x_0^*] < \epsilon$ then $\delta_j x_j$ is a good approximation of $c_j(x_j)$ with error less than a chosen ϵ and \mathbf{x}_0^* is the optimal solution to our problem.

Now, suppose that $\sum_{j=1}^n [c_j(x_0^*) - \delta_j x_0^*] < \epsilon$ is not true. Subdivide the interval $[0; \alpha_j]$ into two equal intervals $[0, \frac{\alpha_j}{2}]$ and $[\frac{\alpha_j}{2}, \alpha_j]$ Solve two linear sub-programmes P_1 and P_2 :

Graph showing iterated under-estimations of the concave transactions cost function.

Figure 7.3: Linear under-estimation of the costs function





$$\begin{aligned} [P_1] : \text{Minimize} \quad & \sum_{t=1}^T y_t \\ \text{Subject to} \quad & (\mathbf{x}, \mathbf{y}) \in \mathbf{F} \\ & \sum_{j=1}^n [r_j x_j - \delta_j x_j] \geq r \\ & 0 \leq x_j \leq \frac{\alpha_j}{2}; y_t \geq 0 \end{aligned} \quad (97)$$

$$\begin{aligned} [P_2] : \text{Minimize} \quad & \sum_{t=1}^T y_t \\ \text{Subject to} \quad & (\mathbf{x}, \mathbf{y}) \in \mathbf{F} \\ & \sum_{j=1}^n [r_j x_j - \delta_j x_j] \geq r \\ & \frac{\alpha_j}{2} \leq x_j \leq \alpha_j; y_t \geq 0 \end{aligned} \quad (98)$$

For P_1 and P_2 ; approximate $c_j(x_j)$ each by an underestimating linear function $\delta_{jh}x_j$; $h = 1, 2$ and solve the programmes:

$$\begin{aligned} [P_1] : \text{Minimize} \quad & \sum_{t=1}^T y_t \\ \text{Subject to} \quad & (\mathbf{x}, \mathbf{y}) \in \mathbf{F} \\ & \sum_{j=1}^n [r_j x_j - \delta_{j1} x_j] \geq r \\ & 0 \leq x_j \leq \frac{\alpha_j}{2}; y_t \geq 0 \end{aligned} \quad (99)$$

$$\begin{aligned} [P_2] : \text{Minimize} \quad & \sum_{t=1}^T y_t \\ \text{Subject to} \quad & (\mathbf{x}, \mathbf{y}) \in \mathbf{F} \\ & \sum_{j=1}^n [r_j x_j - \delta_{j2} x_j] \geq r \\ & \frac{\alpha_j}{2} \leq x_j \leq \alpha_j; y_t \geq 0 \end{aligned} \quad (100)$$

Let $\{\mathbf{x}_1^*; \mathbf{x}_2^*\}$ be an optimal solution to $\{P_1; P_2\}$. If $\sum [c_j(x_j^*) - \delta_{jh}x_j^*] \leq \epsilon$, then $\delta_{jh}x_j$ is an approximation of $c_j(x_j)$ with error less than ϵ . The solution is therefore $S = \min\{\mathbf{x}_1^*; \mathbf{x}_2^*\}$ because we are dealing with a minimization problem.

It may happen that some of the resultant sub-programmes have no solutions, in which case they are fathomed. If $\delta_{jh}x_j$ is not an approximation of $c_j(x_j)$ then repeat the process with more iterations as shown on the graph.

8 Back-Testing MADS

In [1] the authors did some numerical experiments of MADS (without transaction costs ofcourse). They used historical data of the Tokyo Stock Exchange with the aim of checking whether this model actually generates a portfolio with large skewness.

They prepared three sets of data D_1, D_2 and D_3 , all of which consist(ed) of 36 data representing the rate of return of 224 stocks for 36 months. D_1 covered three years from 1984 to 1986, while D_2 and D_3 covered 36 months from 1985 to 1987, 1986 to 1988, respectively.

They then first solved the MAD model for $r = 2.0, 2.5, 3.0\%$ per month and calculated the minimal absolute deviation.

$$\underline{w}(r) = \min \left\{ W \left(\sum_{j=1}^n R_j x_j \right) : \sum_{j=1}^n r_j x_j = r, \sum_{j=1}^n x_j = 1, x_j \geq 0 \right\} \quad (101)$$

The skewness was negative for all r . They then proceeded to conduct preliminary experiments to solve the MADS for $w = 1.10\underline{w}(r)$ and found that the maximal value of the skewness is attained when the parameters in the objective function of the MADS model are chosen as follows: $(\alpha, \rho_1, \rho_2) = (1.0, r - 1.0, r - 2.0)$. Then they fixed the value of these parameters at this level throughout subsequent experiments. Let $P(r, w)$ be the portfolio corresponding to an optimal solution of the MADS for fixed value of (r, w) . Also, let $\kappa(r, w)$ be the skewness of the distribution of the portfolio $P(r, w)$. Refer to the corresponding table for clarity on this issue. The table shows the value of $\kappa(r, w)$ for data sets $D_1 - D_3$. We see from this that $\kappa(r, w)$ increases as w increases. Particularly the skewness of the portfolio associ-

	r	$\underline{w}(r)$	Standard Dev.	Skewness
D_1	2.0	0.8350	2.048	-1.252
	2.5	0.8400	1.969	-2.226
	3.0	0.4270	1.961	-3.060
D_2	2.0	0.8440	1.756	-0.286
	2.5	0.7990	1.705	-1.121
	3.0	0.8770	1.866	-1.618
D_3	2.0	1.5930	2.840	-0.693
	2.5	1.2520	2.442	-0.848
	3.0	1.0710	2.255	-1.220

Table 1: Results showing negative skewness for all r

ated with $w = 1.50 \times \underline{w}(r)$ is always positive, which is contrary to the MAD portfolio where skewness is always negative.

With the three portfolios

$$\begin{aligned}
 P_1 &\equiv P(r, \underline{w}(r)) \\
 P_2 &\equiv P(r, 1.20\underline{w}(r)) \\
 P_3 &\equiv P(r, \infty)
 \end{aligned} \tag{102}$$

for $r = 3.01\%$ per month, they discovered that the distribution associated with the MAD portfolio P_1 has a larger tail to the left of the mean and hence has a large negative skewness. The distribution associated with P_3 has a long tail to the right of the mean. Also, it has a large absolute deviation since the constraint on the absolute deviation is completely relaxed. The distribution of P_2 lies between the other two. In particular, the skewness is much larger than P_1 while the absolute deviation is slightly larger than P_1 . Even though it appears most investors would prefer P_2, P_3 to P_1 , the preference also is dependent on the functional form of the utility function. Their experiments were carried out using the SUN 4/280 system which produced results in less than one minute for $T = 36$ and $n = 225$!



r	$w/\bar{w}(r)$	<i>Skewness</i>		
		D_1	D_2	D_3
2	1.0	-2.252	-0.287	-0.693
	1.1	0.230	0.652	0.543
	1.2	0.271	1.187	1.217
	1.5	0.846	1.802	
2.5	1.0	-2.226	-1.121	-0.848
	1.1	-0.120	0.148	0.323
	1.2	0.206	0.707	0.850
	1.5	0.537	1.520	1.637
3	1.0	-3.060	-1.618	-1.220
	1.1	-1.405	-0.740	-0.902
	1.2	-0.421	-0.076	0.346
	1.5	0.351	1.522	1.292

Table 2: Values of $\kappa(r, w)$ for data sets D_1 to D_3

Note: Value for D_3 when $r = 2, w/\bar{w}(r) = 1.5$ not available.

9 Conclusion

We have seen how various techniques have come into play to reduce some very difficult optimization problems to more manageable levels. The advances in computational technologies have helped a lot in the running of some mathematical models and programmes, unfortunately these softwares do not come cheap.

We have here suggested a model that appears easy to solve, and which can be used as a practical tool to generate a portfolio with larger if not maximal skewness under the constraints on the first and second moments of the distribution. Moreover, this model has the advantage that it also admits the incorporation of other institutional linear constraints. However, it remains the ambitions of the author to try to produce a computer programme and, with the availability of data, put to practical test the benefits herein justified.

It has been noted that we only considered investments up to some point α_j , point of inflection. The reason being that from this point up the transaction costs start to unbearably sky-rocket making investment worthless beyond this point.

References

- [1] Konno H., Shirakawa H. and Yamazaki H. *A mean-absolute deviation-skewness portfolio optimization model*. Annals of Operations Research, Vol 45 (1993) 205-220.
- [2] Konno H. and Suzuki K. *A mean variance skewness portfolio optimization model*. J. of Oper. Research Society, Japan Vol 38 (1995) 173-187.
- [3] Konno H. and Yamazaki H. *Mean-absolute deviation portfolio optimization model and its application to Tokyo stock market*. Management Science, Vol 37, (1991) 519-531.
- [4] Konno H. and Wijayanayake A. *Mean-absolute deviation portfolio optimization model under transaction costs*. J. Oper Res. Soc. Japan, Vol 42 (1999) 422-435.
- [5] Konno H. and Wijayanayake A. *Portfolio optimization problem under concave transaction costs and minimal transaction units constraints*. Math. Prog. Ser B 89: (2001) 233-250.
- [6] Markowitz H. *Portfolio Selection: Efficient divesification of investments*. 1959, John Wiley & Sons.
- [7] Chvatal V. *Linear programming*. 1983, Freeman and Co.
- [8] Elton J. E. and Gruber M. J. *Morden Portfolio Theory and Investment Analysis*. 4th ed. 1991 John Wiley and Sons.
- [9] Ingersoll Jr. J. E. *Theory of Financial Decision Making*. 1987, Rowman and Littlefield.
- [10] Samuelson P. A. *The fundamental approximation theorem of portfolio analysis in terms of means, variances and higher moments*. Rev Econ Studies Vol 25 (1958) 65-86.
- [11] Konno H and Suzuki K. *A fast algorithm for solving large-scale mean-variance models by compact factorization of covariance matrices*. J. Oper. Res. Soc. Japan, Vol 35 (1992), 93-104.



- [12] Konno H, Suzuki T and Kobayashi D. *A branch and bound algorithm for solving mean-risk-skewness model*. Optimization Methods Software Vol 10 1998, 297-317
- [13] Maghrebi N. *On Skewness Preference and Persistence Hypothesis*. Japan Financial Review, Vol 15 (1992), 17-35.
- [14] Konno, H *Piecewise linear risk functions and portfolio optimization*. J. of Operations Research Soc. of Japan, Vol 33 (1990), 139-156
- [15] Perold A. *Large-scale portfolio optimization*, Management Science, Vol 30 (1984), 1143-1160