Portfolio Resampling and Efficiency Issues

A Master Thesis Presented

by

Wei Jiao
(161425)

to

Prof. Dr. Wolfgang Härdle

Institute of Statistics and Econometrics

in partial fulfillment of the requirements

for the degree of

Master of Science

Humboldt-Universität zu Berlin

School of Business and Economics

Spandauer Str. 1

D-10178 Berlin

Berlin, December 16, 2003

Declaration of Authorship

I hereby confirm that I have authored this master thesis independently, no other than the indicated references and resources have been used. All contents, which are literally or in general matter taken out of publications or other resources, are marked as such.

Wei Jiao

Berlin, 5th January 2004

Abstract

This thesis starts with a review of the traditional portfolio theory and a discussion of its limitations. The new technique portfolio resampling is introduced, followed by two different portfolio efficiency testing methods. The final part is an empirical study of portfolio revision. A short conclusion is made at the end.

Thanks to

Contents

List of Tables

List of Figures

1 Traditional Portfolio Construction

1.1 Defining Markowitz Efficiency

Markowitz mean-variance efficiency is a cornerstone of the modern finance for asset management. Given the presumption that rational investors make investment decisions based on risky assets' expected return and risk, with risk measured as variance, a portfolio is considered mean-variance efficient if it has the minimum variance for a given level of portfolio expected return, or if it has the maximum expected return for a given level of portfolio variance.

1.2 Mathematical Notations

The expected return for asset i in the n asset universe is $\mu_i, i = 1...n$. ω_i is the weight of asset i in portfolio P. The portfolio expected return is defined as $\mu_p = \sum_i \omega_i \mu_i$

The variance σ_p^2 of portfolio P, is the double sum of the product for all ordered pairs of assets of the portfolio weight ω_i for asset i, the portfolio weight ω_j for asset j, the standard deviation σ_i for asset i, the standard deviation σ_j for asset j, and the correlation $\rho_{i,j}$ between asset i and j. In mathematical notation, $\sigma_p^2 = \sum_i \sum_j \omega_i \omega_j \sigma_i \sigma_j \rho_{i,j} = \sum_i \omega_i^2 \sigma_i^2 + 2 \sum_{i \neq j} \sigma_{ij} \omega_i \omega_j$

Expressed in matrix format: the covariance matrix of expected returns, Σ, the

portfolio weights, w, the expected returns, μ, can be written as

$$\Sigma = \begin{bmatrix} \sigma_{11} \cdots \sigma_{1n} \\ \vdots \ddots \vdots \\ \sigma_{n1} \cdots \sigma_{nn} \end{bmatrix}, \quad w = \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_n \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}$$

Portfolio risk, σ_p^2, measured as variance, and portfolio return, μ_p, are calculated from

$$\sigma_p^2 = \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_n \end{bmatrix}^\top \begin{bmatrix} \sigma_{11} \cdots \sigma_{1n} \\ \vdots \ddots \vdots \\ \sigma_{n1} \cdots \sigma_{nn} \end{bmatrix} \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_n \end{bmatrix}, \quad \mu_p = \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_n \end{bmatrix}^\top \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}$$

1.3 Efficient Frontier

There are two ways to find the efficient frontier:

- minimize portfolio variance for all portfolios ranging from minimum return to maximum return to trace out an efficient frontier; or

- maximize investors utility function for a given risk-tolerance parameters λ, and by varying λ, trace out the efficient frontier.

These two methods leads to the same efficient frontier if the utility function is quadratic or asset returns are normal distributed.

1.3.1 Minimize variance approach

Following the first approach, and including two constraints which require that the portfolio return $w^\top \mu$ equals π and that the sum of the portfolio weights equals one, the problem can be expressed as the following:

$$\begin{cases} \underset{w}{Min} \, w^\top \Sigma w \\ w^\top \mu = \pi \\ w^\top I = 1 \end{cases} \tag{1.1}$$

solving with Lagrangian

$$L = w^\top \Sigma w + \lambda_1 (\pi - w^\top \mu) + \lambda_2 (1 - w^\top I)$$

$$\begin{cases} \frac{dL}{dw} = 2\Sigma w - \lambda_1 \mu - \lambda_2 I = 0 \\ \frac{dL}{d\lambda_1} = w^\top \mu - \pi = 0 \\ \frac{dL}{d\lambda_2} = w^\top I - 1 = 0 \end{cases} \tag{1.2}$$

from the first equation above, we have $w = \frac{1}{2}\lambda_1 \Sigma^{-1}\mu + \frac{1}{2}\lambda_2 \Sigma^{-1}I$ plug it in the last two equations above, we have

$$\begin{cases} \frac{1}{2}\lambda_1 \mu^\top \Sigma^{-1}\mu + \frac{1}{2}\lambda_2 \mu^\top \Sigma^{-1}I = \pi \\ \frac{1}{2}\lambda_1 \mu^\top \Sigma^{-1}I + \frac{1}{2}\lambda_2 I^\top \Sigma^{-1}I = 1 \end{cases} \tag{1.3}$$

Defining the following terms: $a = I^\top \Sigma^{-1}I \quad b = \mu^\top \Sigma^{-1}I \quad c = \mu^\top \Sigma^{-1}\mu$ where a, b, c are constants, and rewrite the above formula

$$\begin{cases} \frac{1}{2}c\lambda_1 + \frac{1}{2}b\lambda_2 = \pi \\ \frac{1}{2}b\lambda_1 + \frac{1}{2}a\lambda_2 = 1 \end{cases} \tag{1.4}$$

solve the equations above we have the values of the two multipliers:

$$\lambda_1 = \frac{2(a\pi - b)}{ac - b^2} \qquad \lambda_2 = \frac{2(c - b\pi)}{ac - b^2} \tag{1.5}$$

plugging the two multipliers back to the expression of w, we have:

$$w(\pi) = \frac{(a\Sigma^{-1}\mu - b\Sigma^{-1}I)\pi + (c\Sigma^{-1}I - b\Sigma^{-1})\mu}{ac - b^2} \tag{1.6}$$

3
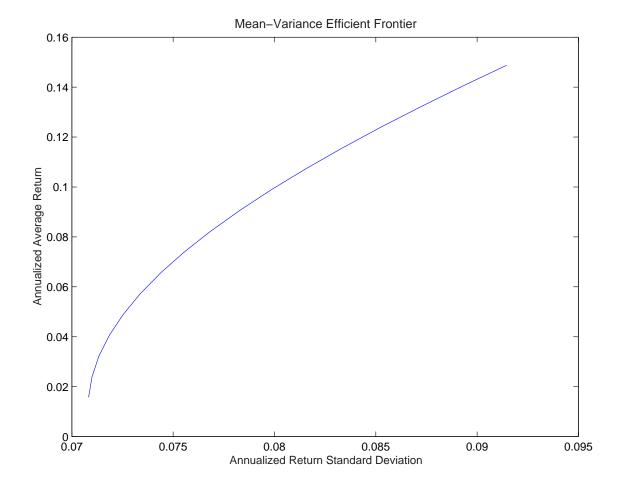
Figure 1.1: Efficient Frontier

Notice that the optimal portfolio weight vector is only a function of the absolute expected return π.

The portfolio variance is thus:

$$w^\top \Sigma w = \frac{a}{ac - b^2}\pi^2 - \frac{2b}{ac - b^2}\pi + \frac{c}{ac - b^2} \tag{1.7}$$

Therefore the portfolio with the lowest risk has co-ordinates $(\frac{1}{a}; \frac{b}{a})$

Figure 2.1 shows the mean-variance efficient frontier using parameters of data set B (explained in the Data Analysis chapter).
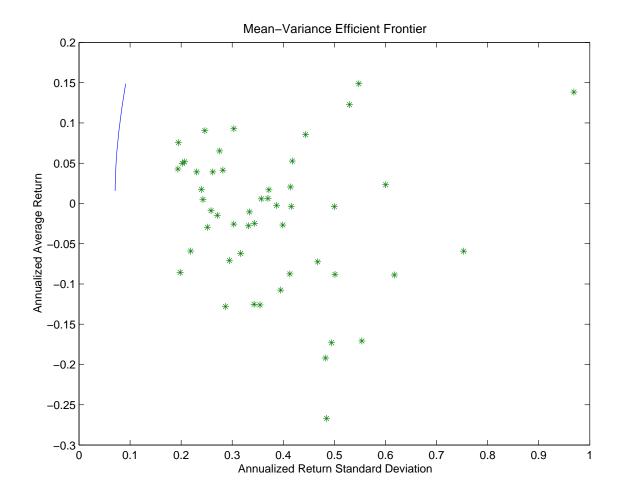
Figure 1.2: Efficient Frontier with asset points

In figure 2.2 I also added the single asset points to make the optimization effect more clearer.

In reality the asset weights can not be negative because short selling is not allowed. Figure 2.3 shows mean-variance efficient frontier with non-negative weight constraint.

Now comparing with the efficient frontier without non-negative weight constraint as showed in figure 2.2, we found out the efficient frontier with non-negative weight constraint is much longer, in another word less efficient, than the one without. The fact is the more constraints we add, the less efficient the frontier will be.
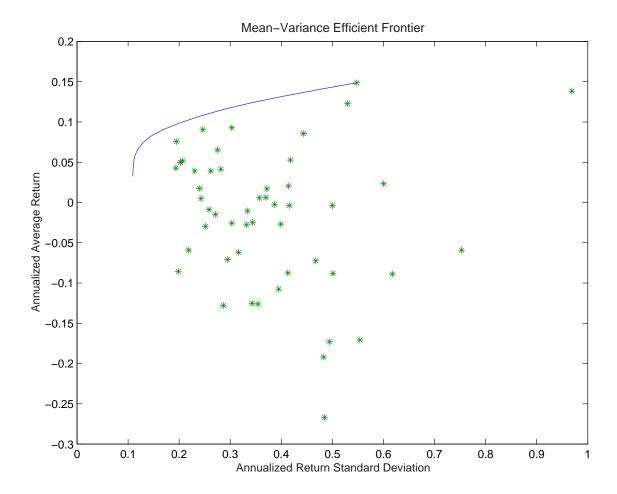
Figure 1.3: Efficient Frontier with non-negative weight constraint

1.3.2 Maximize utility approach

Given the quadratic utility function of a rational investor $Utility = \mu_p - \frac{1}{2\lambda}\sigma_p^2 = w^\top\mu - \frac{1}{2\lambda}w^\top\Sigma w$, the later approach trades off risk against return by maximizing utility for various risk-tolerance parameter λ. The higher the risk tolerance, the less weight is given to the variance (penalty) term and the more aggressive our portfolios will become.

The optimal solution is found by taking the first derivative with respect to portfolio weights, setting the term to zero and solving for the optimal weight vector, w^*:

$$\frac{dUtility}{dw} = \mu - \frac{1}{2\lambda}2\Sigma w = \mu - \frac{1}{\lambda}\Sigma w = 0 \tag{1.8}$$

$$w^* = \lambda\Sigma^{-1}\mu \tag{1.9}$$

Now we introduce general linear constraints $Aw = b$, where A denotes a matrix with m rows (equal to the number of equality constraints) and n columns (equal to the number of assets). And b is a $m \times 1$ vector of limits. We maximize: $Utility = w^\top\mu - \frac{1}{2\lambda}w^\top\Sigma w$ subject to $Aw = b$

Forming the standard Lagrangian $L = w^\top\mu - \frac{1}{2\lambda}w^\top\Sigma w - \gamma^\top(Aw - b)$, where γ is the $m \times 1$ vector of Lagrangian multipliers (one for each constraint), and taking the first derivatives with respect to the optimal weight vector and the vector of multipliers yields

$$\begin{cases} \frac{dL}{dw} = \mu - \frac{1}{\lambda}\Sigma w - A^\top\gamma = 0 & w^* = \lambda\Sigma^{-1}(\mu - A^\top\gamma) \\ \frac{dL}{d\gamma} = Aw - b = 0 & Aw = b \end{cases} \tag{1.10}$$

Inserting w^* into the lower equation above and solving the resulting equation for the Lagrange multipliers, we arrive at

$$\lambda A\Sigma^{-1}\mu - b = \lambda A\Sigma^{-1}A^\top\gamma$$

$$\gamma = \frac{A\Sigma^{-1}\mu}{A\Sigma^{-1}A^\top} - \frac{1}{\lambda}\frac{b}{A\Sigma^{-1}A^\top} \tag{1.11}$$

Substituting Equation 2.11 into Equation 2.10, we finally get the optimal solution under linear equality constraints:

$$w^* = \Sigma^{-1} A^\top (A\Sigma^{-1}A^\top)^{-1}b + \lambda\Sigma^{-1}(\mu - A^\top(A\Sigma^{-1}A^\top)^{-1}A\Sigma^{-1}\mu) \qquad (1.12)$$

According to Scherer, the optimal solution is split into a (constrained) minimum-variance portfolio and a speculative portfolio. This is know as "two-fund separation", and can be seen from the equation above, where the first term depends neither on expected returns nor on risk tolerance and is hence the minimum-risk solution - whereas the second term is sensitive to both inputs.

1.4 Applications of Mean-Variance Optimization

The two most popular applications of Mean-Variance optimization are asset allocation and equity portfolio optimization. In both cases, the goal is to maximize expected portfolio return and minimize risk.

With asset allocation though the candidate pool is composed of large asset categories, such as domestic equities and corporate government bonds, international equities and bonds, real estate, and venture capital.

With equity portfolio optimization, a large pool of securities are included. And more complicated constraints on portfolio characteristics, industry or sector membership and trading cost restrictions are also under consideration which substantially increase the complexity of the optimization process.

The input starting points are also very different. For asset allocation optimization sample means, variances and correlations, based on monthly, quarterly, or annual historic data are the starting points. The source of equity optimization inputs can be very different. Expected and residual return for equities can be derived from some version of the Capital Asset Pricing Model or Arbitrage Pricing Theory. In

practice, portfolio managers often use α - the expected return net of systematic risk expected return as the optimization inputs.

1.5 Benchmark Relative Optimization

Markowitz model uses the absolute risk measure variance to find out the efficient portfolio, in practice however, benchmark relative portfolio optimization is widely used. This is due to the fact that investors would like to know what kind of risk their portfolios carry relative to benchmark and given the amount of relative risk how well do their portfolio perform. Thus the benchmark is becoming an important standard to evaluate the portfolio managers performance, and at the same time brings more questions to the portfolio construction process. Does the benchmark relative risk optimization bring the same result as the Markowitz absolute risk optimization, and is benchmark a good performance measure? To answer these questions above, I would like to first introduce the important concept Tracking Error.

1.5.1 Definition

The relative risk measure tracking error is defined as the standard deviation of portfolio active return (portfolio return minus benchmark return). It can be calculated either ex-ante $TE = \sqrt{w_a^\top \Sigma w_a}$ where w_a denotes the active weight vector, or ex-post $TE = \sqrt{\frac{1}{T-1} \sum_{t=1}^{T} (r_{at} - \bar{r}_a)^2}$. where r_{at} denotes the active return and $\bar{r}_a$ denotes the mean active return.

1.5.2 Tracking Error Optimization

The same procedure as minimize variance approach can be used to find the lowest tracking error for a given level of portfolio active return E. As formulated below:

$$\begin{cases} \underset{w_a}{Min}\, w_a^\top \Sigma w_a \\ w_a^\top \mu = E \\ w_a^\top I = 0 \end{cases} \tag{1.13}$$

solving with Lagrangian

$$L = w_a^\top \Sigma w_a + \lambda_1(E - w_a^\top \mu) + \lambda_2(0 - w_a^\top I)$$

$$\begin{cases} \frac{dL}{dw_a} = 2\Sigma w_a - \lambda_1 \mu - \lambda_2 I = 0 \\ \frac{dL}{d\lambda_1} = w_a^\top \mu - E = 0 \\ \frac{dL}{d\lambda_2} = w_a^\top I = 0 \end{cases} \tag{1.14}$$

from the first equation above, we have $w_a = \frac{1}{2}\lambda_1 \Sigma^{-1}\mu + \frac{1}{2}\lambda_2 \Sigma^{-1}I$ plug it in the last two equations above, we have

$$\begin{cases} \frac{1}{2}\lambda_1 \mu^\top \Sigma^{-1}\mu + \frac{1}{2}\lambda_2 \mu^\top \Sigma^{-1}I = E \\ \frac{1}{2}\lambda_1 \mu^\top \Sigma^{-1}I + \frac{1}{2}\lambda_2 I^\top \Sigma^{-1}I = 0 \end{cases} \tag{1.15}$$

Again using the terms: $a = I^\top \Sigma^{-1}I \quad b = \mu^\top \Sigma^{-1}I \quad c = \mu^\top \Sigma^{-1}\mu$ and rewrite the above formula

$$\begin{cases} \frac{1}{2}c\lambda_1 + \frac{1}{2}b\lambda_2 = E \\ \frac{1}{2}b\lambda_1 + \frac{1}{2}a\lambda_2 = 0 \end{cases} \tag{1.16}$$

solve the equations above we have the values of the two multipliers:

$$\lambda_1 = \frac{2aE}{ac - b^2} \qquad \lambda_2 = -\frac{2bE}{ac - b^2} \tag{1.17}$$

plugging the two multipliers' value to the expression of w_a, we have:

$$w_a(E) = \frac{E(a\Sigma^{-1}\mu - b\Sigma^{-1}I)}{ac - b^2} \tag{1.18}$$

Which is the optimum active weight vector given a desired level of relative return E, and the optimized tracking error

$$
\begin{aligned}
TE^2 &= \left(\frac{(a\Sigma^{-1}\mu - b\Sigma^{-1}I)E}{ac - b^2}\right)^{\top} \Sigma \left(\frac{(a\Sigma^{-1}\mu - b\Sigma^{-1}I)E}{ac - b^2}\right) \\
&= \frac{E^2}{(ac - b^2)^2}(\mu^{\top}\Sigma^{-1}a - I^{\top}\Sigma^{-1}b)(a\mu - bI) \\
&= \frac{E^2}{(ac - b^2)^2}(a^2\mu^{\top}\Sigma^{-1}\mu - abI^{\top}\Sigma^{-1}\mu - ab\mu^{\top}\Sigma^{-1}I + b^2 I^{\top}\Sigma^{-1}I) \qquad (1.19) \\
&= \frac{E^2}{(ac - b^2)^2}(a^2 c - ab^2) \\
&= \frac{aE^2}{ac - b^2}
\end{aligned}
$$

We notice from the solution above if the portfolio active return E is set to zero, the active weights vector and the tracking error will both be zero too, therefore the optimum portfolio is the benchmark itself.

In contrary to figure 2.1, the tracking error efficient frontier will be a straight line if the x axis is standard deviation instead of variance.

Another thing to notice is the upper and lower bounds for active weights are not that easy to formulate. Besides each one has to be between -1 and +1, the sum of negative active weight or the sum of positive active weight has to be between -1 and +1 too. And I couldn't include this constraint to the quadratic programming optimization function.

1.5.3 Comparing with Mean-variance Optimization

It will be interesting to find out how is the tracking error efficiency comparing with a Markowitz mean-variance efficiency in a mean-variance space. In another word, we would like to see whether tracking error efficient portfolio is also mean-variance efficient.

w_p is the portfolio weight vector, w_b the benchmark weight vector. φ is the benchmark return, and E is the portfolio active return.
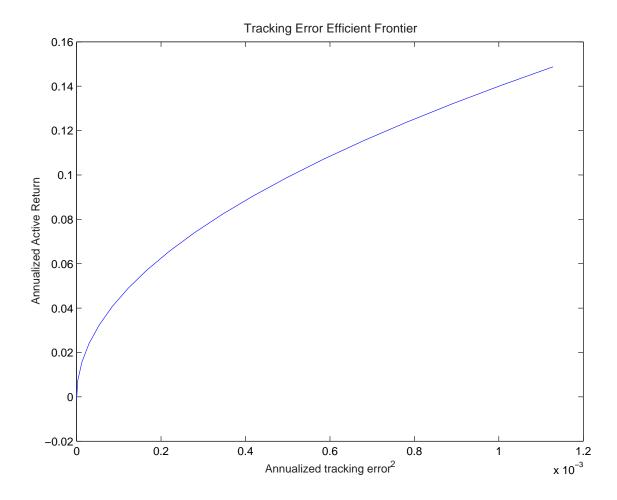
Figure 1.4: Tracking Error Efficient Frontier

$$w_a = w_p - w_b$$

$$\varphi = w_b^\top \mu$$

$$E = \pi - \varphi$$

$$\sigma_p^2 = (w_b + w_a)^\top \Sigma (w_b + w_a)$$

$$= w_b^\top \Sigma w_b + w_a^\top \Sigma w_a + 2w_b^\top \Sigma w_a$$

$$= w_b^\top \Sigma w_b + \frac{(\pi - \varphi)(a\Sigma^{-1}\mu - b\Sigma^{-1}I)^\top}{ac - b^2} \Sigma \frac{(\pi - \varphi)(a\Sigma^{-1}\mu - b\Sigma^{-1}I)}{ac - b^2}$$

$$+ 2w_b^\top \Sigma \frac{(\pi - \varphi)(a\Sigma^{-1}\mu - b\Sigma^{-1}I)}{ac - b^2}$$

$$= w_b^\top \Sigma w_b + \frac{a}{ac - b^2}\pi^2 - \frac{2a\varphi}{ac - b^2}\pi + \frac{a\varphi^2}{ac - b^2} + \frac{2w_b^\top \pi(a\mu - bI)}{ac - b^2} - \frac{2w_b^\top \varphi(a\mu - bI)}{ac - b^2}$$

$$= \frac{a}{ac - b^2}\pi^2 - \frac{2b}{ac - b^2}\pi + \frac{2b\varphi - a\varphi^2}{ac - b^2} + w_b^\top \Sigma w_b$$

$$(1.20)$$

This equation represents all the tracking error optimization portfolios located in a expected return and variance space. Comparing with equation 2.3.1, we notice these two efficient frontiers have only a difference of a constant term: $d = \frac{2b\varphi - a\varphi^2}{ac - b^2} + w_b^\top \Sigma w_b - \frac{c}{ac - b^2}$ The distance will be zero if the benchmark lies on the Markowitz efficient frontier. It also makes it clear that a tracking error optimization will not provide an optimum solution in absolute terms unless the benchmark is a mean-variance optimum portfolio, and that is seldom the case.

Even if we include tracking error as a constraint instead of as the objective function, the optimization result will still be the tracking error efficient frontier, which as showed above, is not absolute efficient.

Andrea Nardon suggests "it is very important before starting any optimization to understand where the benchmark lies in a mean-variance space and in conjunction with performance and risk targets the portfolio strategist has to choose (or help the client to choose) the most appropriate level of tracking error."

1.6 Criticism and Limitations of Mean-Variance Efficiency

1.6.1 Criticisms of Mean-Variance Efficiency

The first criticism is concerned with the assumptions of Mean-Variance efficiency. As a common knowledge, in reality, returns are not multivariate normal distributed. Investors might exhibit different utility functions other than quadratic form. And the investors might have multi-periodic investment horizon, in contrast to the Mean-Variance one period framework. Also the risk measure variance as used in mean-variance optimization, might not be proper. As the variance measures variability above and below the mean, from an investor's point of view the variance above the mean is actually not "risk". Returns below the mean or any specified level of return is much more important to an investor. Downside risk measures of variability such as semivariance $\sum_{x_i \leq \mu}(x_i - \mu)^2$ or semistandard deviation of return, the mean absolute deviation $\sum_i \mid x_i - \mu \mid$ and range measures could be good alternatives to the traditional risk measure variance or standard deviation.

Then how serious indeed are these problems on the practical use of mean-variance based portfolio construction? I will examine the questions below:

1. How well does the mean-variance framework approximate reality, where investors might have different utility functions and returns might not be normally distributed?

2. How well does the one-period solution approximate multiperiod optimality?

3. Whether, in practice, non-variance risk measures lead to significantly different efficient portfolios.

Since Markowitz mean-variance efficiency is only consistent with expected utility maximization either when asset returns are normally distributed or when investors

have quadratic utility functions. Given that in reality neither of the two assumptions are all the time true, mean-variance efficiency is not strictly consistent with expected utility maximization.

For the second question, we can divide this problem in to two separate questions.

- Does the mean-variance frontier change as the investment horizon lengthens?

- Does repeatedly investing in one-period-efficient portfolios result in multiperiod-efficient portfolios?

The first question is relative easy to answer. Assuming homoskedastic, zero serial correlated and normally distributed assets returns, portfolio returns and variance are proportional to the time horizon. Which means the curvature of the efficient frontier should be unchanged across different time period, and all investors will chose the same portfolio irrespective of the time horizon.

To answer the second question, According to Scherer, under fairly strict assumptions, repeatedly investing in one-period-efficient portfolios will also result in multiperiod-efficient portfolios if:

- investors have constant relative risk-aversion (wealth level does not change optimal allocations) and only possess financial wealth;

- asset returns are not autocorrelated (investment opportunities are not time-varying)-ie, period returns are not forecastable;

- there is no uncertainty about estimated parameters.

- portfolio returns are not path-dependent due to intermediate cash-flows (no cash infusion and/or withdrawals)

- there are no transaction costs

Most of these assumption, especially the last two, are very unrealistic as investment opportunities are time-varying and transaction costs are unavoidable. I would say in reality repeatedly investing in one-period-efficient portfolios will result in incomparable or multiperiod inefficient portfolios.

Now to the problem of appropriate risk measure. As pointed out by Michaud, the returns of diversified equity portfolios, equity indexes, and other assets are often approximately symmetric over periods of institutional interest, efficiency based on nonvariance risk measures may be nearly equivalent to mean-variance efficiency, for symmetric returns downside risk contains same information as variance. Bond returns and fixed-income indexes are less symmetric than equities classes. Options do not have return distributions that are approximately symmetric. In addition, the return distribution of diversified equity portfolios becomes increasingly asymmetric over a long-enough period. Consequently, the variance measure for defining portfolio risk is not appropriate. For many applications of institutional interest, however, a variance-based efficient frontier is often little different (and even less often statistically significantly different) from frontiers that use other measures of risk, which makes variance still an acceptable or even in most cases more convenient measure of risk.

1.6.2 The Fundamental Limitations of Mean-Variance Efficiency

As pointed out by Michaud, the most serious problems in practical application of mean-variance efficiency are instability and ambiguity. By instability and ambiguity, we mean small changes in input will often lead to large changes in the optimized portfolio. Another problem with mean-variance optimized portfolios is that they do not make investment sense and do not have investment value.

2 Data Analysis

Dow Jones Euro stoxx50 monthly return data from February 1993 to September 2003 were downloaded from Thomson Financial Datastream. I named it data set A, which includes altogether 128 months' data. The constituents of the the index are those listed in September 2003.

One problem with the data set A is that some of the index constituents' were not listed back to the early 90's. Stocks whose historical data are partially missing include: AVENTIS (from 02.1993), BNP PARIBAS (from 11.93), DAIMLER-CHRYSLER (from 11.98), DEUTSCHE TELEKOM (from 12.96), ENEL (from 11.99), ENI (from 12.95), FRANCE TELECOM (from 11.97), MUNCH.RUCK. (from 02.96), TELECOM ITAL.MOBL. (from 08.95).

This makes it impossible to calculate the covariance matrix with all real numbers directly. I write a Matlab function myself, which is called "covariance", using the maximum available data to get the all real number covariance matrix. The function works as the following: take two columns (two time series) from the data matrix and compare the length of the available data, use the starting point of the shorter one as the starting point for both to calculate the covariance of the two time series. The code of the function is attached in Appendix.

Even with this improved way to calculate covariance, data set A still has the problem of reliability and integrity. As some of the means and variances are from different time period, and are thus not comparable. I setup another data set B with monthly returns starting December 1999 ending September 2003. There are only 46

months's data available, but without any missing value.

In order to decide which data set is more suitable for my following portfolio optimization and portfolio resampling analysis, I will first do a statistic analysis of the two data sets respectively. Since data set B covers the whole bear market period in the past few years, It is also very interesting to do a comparison.

2.1 Descriptive Statistics

The following table shows the mean as the measure of location, standard deviation as the measure of dispersion for the two data sets respectively. With A representing the monthly return data set from February 1993 to September 2003, and B the monthly return data set from December 1999 to September 2003.

Table 2.1: Data Analysis: Descriptive Statistics

No.	Titel	Mean(A)	Mean(B)	STD(A)	STD(B)
1	ABN AMRO HOLDING	0.0123	-0.0021	0.0898	0.0993
2	AEGON	0.0167	-0.0142	0.1141	0.1599
3	AHOLD KON.	0.0105	-0.0074	0.1177	0.1783
4	AIR LIQUIDE	0.0062	0.0042	0.0543	0.0584
5	ALCATEL	0.0122	0.0115	0.1954	0.2797
6	ALLIANZ (XET)	0.0060	-0.0144	0.1073	0.1427
7	GENERALI	0.0076	-0.0052	0.0806	0.0914
8	AVENTIS	0.0100	-0.0007	0.0819	0.0746
9	AXA	0.0124	-0.0060	0.1105	0.1349
10	BASF (XET)	0.0138	0.0034	0.0765	0.0813
11	BAYER (XET)	0.0071	-0.0073	0.0898	0.1192

continued on next page

No.	Titel	Mean(A)	Mean(B)	STD(A)	STD(B)
12	BBV ARGENTARIA	0.0196	-0.0009	0.1008	0.0963
13	SANTANDER CTL.HISPANO	0.0170	0.0005	0.1033	0.1030
14	BNP PARIBAS	0.0108	0.0054	0.0950	0.0794
15	CARREFOUR	0.0153	-0.0107	0.0821	0.0828
16	DAIMLERCHRYSLER (XET)	-0.0059	-0.0104	0.1013	0.0990
17	DEUTSCHE BANK (XET)	0.0084	0.0005	0.0930	0.1068
18	DEUTSCHE TELEKOM (XET)	0.0049	-0.0160	0.1278	0.1393
19	E ON (XET)	0.0095	0.0004	0.0637	0.0701
20	ENDESA	0.0105	-0.0021	0.0777	0.0875
21	ENEL	-0.0071	-0.0071	0.0572	0.0572
22	ENI	0.0128	0.0063	0.0671	0.0562
23	FORTIS (AMS)	0.0123	-0.0105	0.0880	0.1023
24	FRANCE TELECOM	0.0139	-0.0049	0.1912	0.2174
25	DANONE	0.0066	0.0033	0.0668	0.0665
26	SOCIETE GENERALE	0.0127	0.0077	0.1004	0.0875
27	IBERDROLA	0.0128	0.0043	0.0686	0.0597
28	ING GROEP CERTS.	0.0150	-0.0022	0.0949	0.1152
29	L'OREAL	0.0149	0.0033	0.0821	0.0756
30	LAFARGE	0.0070	-0.0023	0.0830	0.0958
31	LVMH	0.0145	0.0071	0.1109	0.1281
32	MUNCH.RUCK. (XET)	0.0102	-0.0073	0.1304	0.1447
33	NOKIA	0.0445	0.0019	0.1492	0.1733
34	PHILIPS ELTN.KON	0.0260	0.0102	0.1258	0.1529
35	REPSOL YPF	0.0105	-0.0025	0.0705	0.0726

No.	Titel	Mean(A)	Mean(B)	STD(A)	STD(B)
36	ROYAL DUTCH PTL.	0.0090	-0.0049	0.0620	0.0631
37	RWE (XET)	0.0044	-0.0059	0.0730	0.0850
38	SAINT GOBAIN	0.0096	0.0044	0.0954	0.1206
39	SAN PAOLO IMI	0.0096	-0.0002	0.1058	0.1117
40	SANOFI - SYNTHELABO	0.0159	0.0075	0.0727	0.0711
41	SIEMENS (XET)	0.0150	0.0124	0.1170	0.1582
42	SUEZ	0.0041	-0.0090	0.0905	0.1140
43	TELECOM ITALIA	0.0139	-0.0003	0.1528	0.1443
44	TELEFONICA	0.0187	0.0017	0.1004	0.1195
45	TELECOM ITAL.MOBL.	0.0201	-0.0003	0.1066	0.1200
46	TOTAL SA	0.0131	0.0036	0.0673	0.0559
47	UNICREDITO ITALIANO	0.0131	0.0015	0.1012	0.0692
48	UNILEVER CERTS.	0.0093	-0.0012	0.0718	0.0782
49	VIVENDI UNIVERSAL	0.0017	-0.0222	0.1054	0.1399
50	VOLKSWAGEN (XET)	0.0155	0.0014	0.1009	0.1073

In order to make the comparison between the two data sets clearer, I made a graphic of the means and standard deviations for the 50 constituents. From figure 3.1 we see, the mean returns of data set A dating from February 1993 to September 2003 are generally higher than that of the data set B dating from December 1999 to September 2003, and the standard deviations of data set A are generally lower than that of data set B. This is coherent with the fact that starting 2000 the world capital markets have experienced a very volatile bear market.

Since Interquartile Range is more robust to outliers as a measure of dispersion, here I showed two boxplots for data set A and data set B to make the comparison

of volatility among single titles more obvious.

From figure 3.2 we see, during the period 02.1993 to 09.2003, No.33 (NOKIA), No.43 (TELECOM ITALIA), No.24 (FRANCE TELECOM), No.5 (ALCATEL) and No.34 (PHILIPS ELTN.KON) have relatively wide dispersion (broader interquartile range), while No.21 (ENEL), No.4 (AIR LIQUIDE), No.19 (E ON), No.46 (TOTAL SA), No.10 (BASF) have relatively low level of dispersion (narrow interquartiel range). From figure 3.3 we see during the period 12.1999 to 09.2003, index component No.5 (ALCATEL) has extremely wide dispersion followed by No.24 (FRANCE TELECOM), No.41 (SIEMENS), No.18 (DEUTSCHE TELEKOM) and No.33 (NOKIA), while No.47 (UNICREDITO ITALIANO) No.21 (ENEL) No.27 (IBERDROLA) No.14 (BNP PARIBAS) No.35 (REPSOL YPF) have relative low level of dispersion. The result is coherent to the fact that telecommunication stocks performed very volatile during the last four years.

2.2 Normal Distribution Test

To do simulations of asset returns, I need to know the corresponding distribution, whether it is reasonable to suppose the returns are normal distributed. Here I have chosen Lilliefors goodness of fit to a normal distribution test.

The Lilliefors test evaluates the null hypothesis H_0 that input data vector X in the population has a normal distribution with unspecified mean and variance, against the alternative H_1 that X in the population does not have a normal distribution. This test compares the empirical distribution of X with a normal distribution having the same mean and variance as X. The parameters of the normal distribution are estimated from X rather than specified in advance.

Formulated in a mathematical way: We test the sample distribution $F_n(x)$, where n is the sample size, against the theoretical distribution $F_0(x) = \Phi(\frac{x-\bar{x}}{s})$ where $\bar{x}$
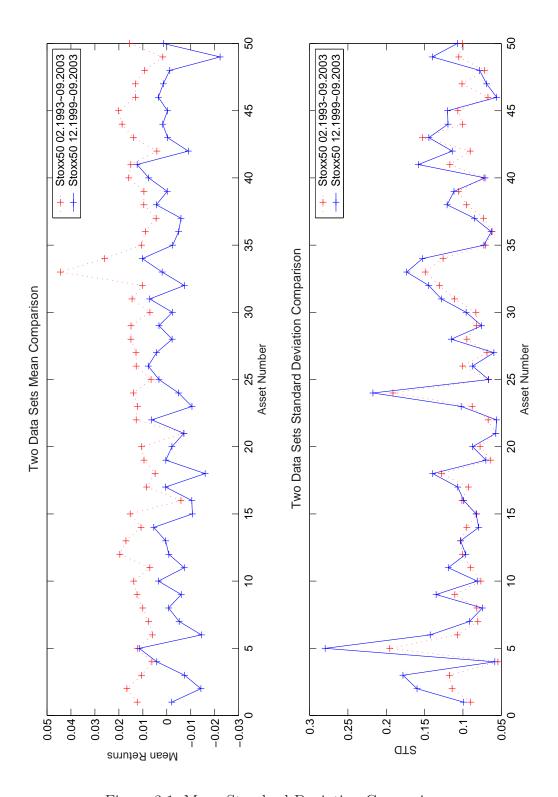
Figure 2.1: Mean-Standard Deviation Comparison
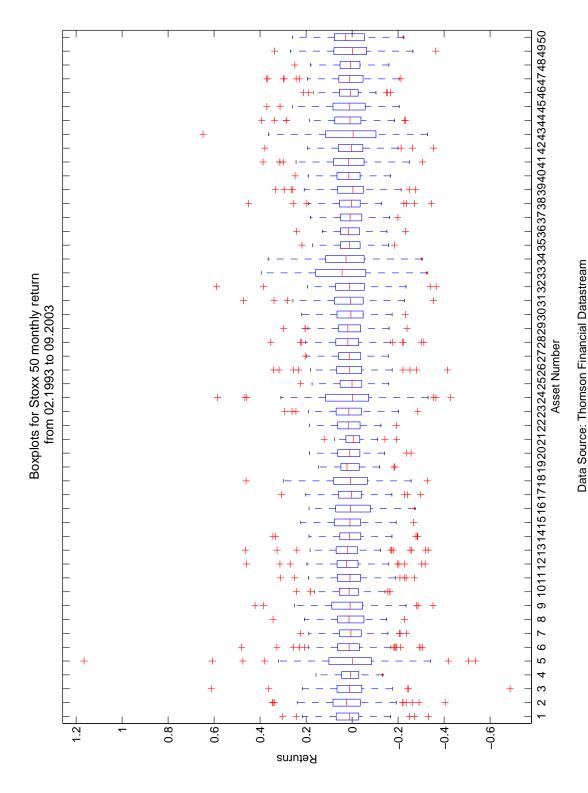
Figure 2.2: Boxplot of Data Set A

23
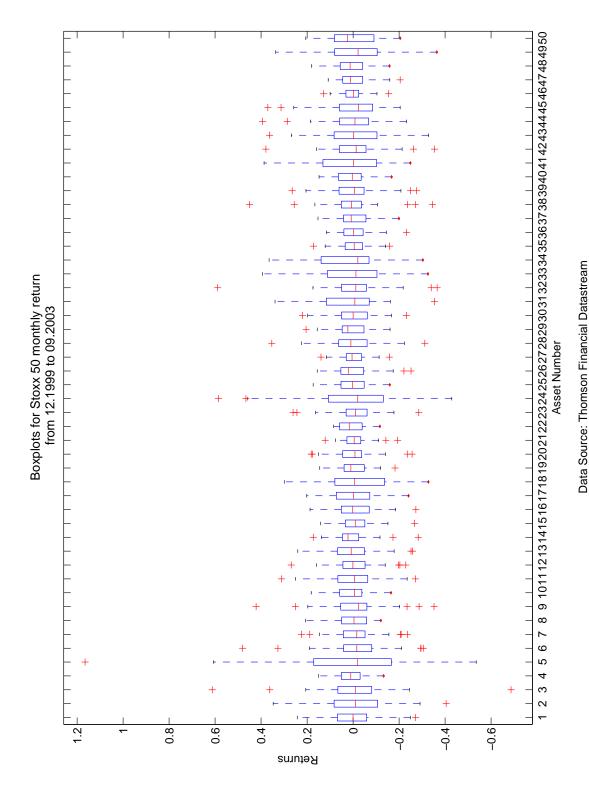
Figure 2.3: Boxplot of Data Set B

and s are estimated mean and variance from the sample X. The test statistic is:

$$D_n = max_x |F_n(x) - F_0(x)| = max_x |F_n(x) - \Phi(\frac{x - \bar{x}}{s})| \qquad (2.1)$$

D_n is the biggest absolute vertical distance between empirical and hypothetical distribution function. Under the null hypothesis, the distribution function of D_n only depends on n not on $F_0(x)$. To determine the the test statistic D_n, we have to consider the empirical discrete distribution function is a stair function. The distance of $F_n(x)$ to $F_0(x)$ therefore has to be calculated not only from the lower but also from the upper jump point. As showed below:

$$D_n^1 = max_{x_i} |F_n(x_{i-1}) - F_0(x_i)|$$
$$\qquad (2.2)$$
$$D_n^2 = max_{x_i} |F_n(x_i) - F_0(x_i)|$$

The maximum distance is then $D_n = max(D_n^1, D_n^2)$. If the observed distribution is coherent with the hypothetical distribution, the distance between F_n and F_0 will be very small and is randomly decided. For test statistic $Z_n = D_n n^{\frac{1}{2}}$ there is a Lillefors table with critical quantile value for normal distribution. So the null hypothesis H_0 will be rejected at significance level α if $Z_n > L_{n,1-\alpha}$ where $L_{n,1-\alpha}$ is the Lillefors critical value for significant level α.

The result of the hypothesis test H is 1 if we can reject the hypothesis that X has a normal distribution, or 0 if we cannot reject that hypothesis. We reject the hypothesis if the test is significant at the 5 percent level.

Other parameters are also included in the testing result table below. P is the p-value of the test, obtained by linear interpolation in a set of table created by Lilliefors. LSTAT is the value of the test statistic. CV is the critical value for determining whether to reject the null hypothesis. If the value of LSTAT is outside the range of the Lilliefors table, P is returned as NaN but H indicates whether to reject the hypothesis.

The results show in table 3.2 for data set A, 16 stocks out of 50 are rejected the hypothesis that they have normal distributions at the 5 percent significant level. For the other 34 stocks Lilliefors test can not reject the normal distributions hypothesis at 5 percent significant level. For data set B, the result is even better. As show in table 3.3 Normal distribution hypothesis are rejected to only 7 out of 50 stocks at 5 percent significant level.

Based on the test results, I decided to use normal distribution to simulate stock returns in the portfolio resampling part.

Table 2.2: Lilliefors goodness of fit to a normal distribution test: Data Set A

No.	Titel	H	P	LSTAT	CV
1	ABN AMRO HOLDING	1.0000	0.0301	0.0879	0.0783
2	AEGON	1.0000	0.0365	0.0848	0.0783
3	AHOLD KON.	1.0000	NaN	0.1250	0.0783
4	AIR LIQUIDE	0	NaN	0.0475	0.0783
5	ALCATEL	1.0000	0.0269	0.0894	0.0783
6	ALLIANZ (XET)	1.0000	NaN	0.1174	0.0783
7	GENERALI	0	NaN	0.0516	0.0783
8	AVENTIS	0	0.1730	0.0668	0.0786
9	AXA	1.0000	0.0491	0.0788	0.0783
10	BASF (XET)	0	NaN	0.0587	0.0783
11	BAYER (XET)	1.0000	NaN	0.0990	0.0783
12	BBV ARGENTARIA	1.0000	NaN	0.1117	0.0783
13	SANTANDER CTL.HISPANO	1.0000	NaN	0.1055	0.0783
14	BNP PARIBAS	0	0.0612	0.0799	0.0816
15	CARREFOUR	0	NaN	0.0528	0.0783

No.	Titel	H	P	LSTAT	CV
16	DAIMLERCHRYSLER (XET)	0	NaN	0.0644	0.1163
17	DEUTSCHE BANK (XET)	0	0.1622	0.0672	0.0783
18	DEUTSCHE TELEKOM (XET)	0	0.0681	0.0952	0.0984
19	E ON (XET)	1.0000	0.0127	0.0962	0.0783
20	ENDESA	0	NaN	0.0611	0.0783
21	ENEL	0	0.1293	0.1155	0.1306
22	ENI	0	NaN	0.0560	0.0919
23	FORTIS (AMS)	0	NaN	0.0529	0.0783
24	FRANCE TELECOM	0	NaN	0.0831	0.1059
25	DANONE	0	NaN	0.0586	0.0783
26	SOCIETE GENERALE	1.0000	0.0154	0.0949	0.0783
27	IBERDROLA	0	NaN	0.0497	0.0783
28	ING GROEP CERTS.	1.0000	0.0123	0.0964	0.0783
29	L'OREAL	0	0.1705	0.0667	0.0783
30	LAFARGE	0	NaN	0.0437	0.0783
31	LVMH	0	0.1678	0.0669	0.0783
32	MUNCH.RUCK. (XET)	1.0000	0.0190	0.1105	0.0929
33	NOKIA	0	NaN	0.0411	0.0783
34	PHILIPS ELTN.KON	0	NaN	0.0505	0.0783
35	REPSOL YPF	0	NaN	0.0437	0.0783
36	ROYAL DUTCH PTL.	0	NaN	0.0597	0.0783
37	RWE (XET)	0	NaN	0.0561	0.0783
38	SAINT GOBAIN	1.0000	NaN	0.1019	0.0783
39	SAN PAOLO IMI	0	0.0756	0.0746	0.0783

No.	Titel	H	P	LSTAT	CV
40	SANOFI-SYNTHELABO	0	NaN	0.0401	0.0783
41	SIEMENS (XET)	1.0000	0.0300	0.0879	0.0783
42	SUEZ	0	0.1671	0.0669	0.0783
43	TELECOM ITALIA	0	NaN	0.0563	0.0783
44	TELEFONICA	0	0.1974	0.0652	0.0783
45	TELECOM ITAL.MOBL.	0	NaN	0.0686	0.0900
46	TOTAL SA	0	NaN	0.0567	0.0783
47	UNICREDITO ITALIANO	1.0000	NaN	0.1009	0.0783
48	UNILEVER CERTS.	0	NaN	0.0459	0.0783
49	VIVENDI UNIVERSAL	0	NaN	0.0540	0.0783
50	VOLKSWAGEN (XET)	0	NaN	0.0618	0.0783
Sum		16			

Table 2.3: Lilliefors goodness of fit to a normal distribution test: Data Set B

No.	Titel	H	P	LSTAT	CV
1	ABN AMRO HOLDING	0	NaN	0.0928	0.1306
2	AEGON	0	NaN	0.0841	0.1306
3	AHOLD KON.	1.0000	0.0229	0.1523	0.1306
4	AIR LIQUIDE	0	NaN	0.0948	0.1306
5	ALCATEL	1.0000	0.0308	0.1460	0.1306
6	ALLIANZ (XET)	0	0.1219	0.1163	0.1306
7	GENERALI	0	0.0740	0.1249	0.1306
8	AVENTIS	0	NaN	0.1021	0.1306

No.	Titel	H	P	LSTAT	CV
9	AXA	0	0.0568	0.1290	0.1306
10	BASF (XET)	0	NaN	0.0751	0.1306
11	BAYER (XET)	0	NaN	0.0968	0.1306
12	BBV ARGENTARIA	0	NaN	0.0979	0.1306
13	SANTANDER CTL.HISPANO	1.0000	0.0486	0.1318	0.1306
14	BNP PARIBAS	1.0000	0.0363	0.1416	0.1306
15	CARREFOUR	0	NaN	0.0972	0.1306
16	DAIMLERCHRYSLER (XET)	0	NaN	0.0657	0.1306
17	DEUTSCHE BANK (XET)	0	NaN	0.0636	0.1306
18	DEUTSCHE TELEKOM (XET)	0	0.1676	0.1116	0.1306
19	E ON (XET)	0	NaN	0.0833	0.1306
20	ENDESA	1.0000	0.0324	0.1447	0.1306
21	ENEL	0	0.1293	0.1155	0.1306
22	ENI	0	0.1276	0.1157	0.1306
23	FORTIS (AMS)	0	0.0711	0.1256	0.1306
24	FRANCE TELECOM	0	NaN	0.0920	0.1306
25	DANONE	0	NaN	0.0650	0.1306
26	SOCIETE GENERALE	0	0.1260	0.1159	0.1306
27	IBERDROLA	0	NaN	0.0942	0.1306
28	ING GROEP CERTS.	0	0.1386	0.1145	0.1306
29	L'OREAL	0	0.0877	0.1216	0.1306
30	LAFARGE	0	NaN	0.0777	0.1306
31	LVMH	0	NaN	0.1060	0.1306
32	MUNCH.RUCK. (XET)	1.0000	0.0419	0.1371	0.1306

No.	Titel	H	P	LSTAT	CV
33	NOKIA	0	NaN	0.0711	0.1306
34	PHILIPS ELTN.KON	0	NaN	0.1001	0.1306
35	REPSOL YPF	0	NaN	0.0808	0.1306
36	ROYAL DUTCH PTL.	0	NaN	0.0994	0.1306
37	RWE (XET)	0	NaN	0.0876	0.1306
38	SAINT GOBAIN	1.0000	NaN	0.2048	0.1306
39	SAN PAOLO IMI	0	NaN	0.1012	0.1306
40	SANOFI-SYNTHELABO	0	NaN	0.0828	0.1306
41	SIEMENS (XET)	0	NaN	0.0706	0.1306
42	SUEZ	0	0.1439	0.1139	0.1306
43	TELECOM ITALIA	0	NaN	0.0698	0.1306
44	TELEFONICA	0	NaN	0.0812	0.1306
45	TELECOM ITAL.MOBL.	0	NaN	0.1057	0.1306
46	TOTAL SA	0	NaN	0.0998	0.1306
47	UNICREDITO ITALIANO	0	NaN	0.0917	0.1306
48	UNILEVER CERTS.	0	NaN	0.1059	0.1306
49	VIVENDI UNIVERSAL	0	NaN	0.0635	0.1306
50	VOLKSWAGEN (XET)	0	0.1362	0.1147	0.1306
Sum		7			

3 Resampled Efficient Frontier

3.1 Estimation Error

3.1.1 Estimation Error Definition

Estimation Error is defined as the difference between the estimated distribution parameters and the true parameters when samples are not large enough. The impact of estimation error on portfolio optimization could be very serious.

As pointed out by Scherer, portfolio optimization suffers from error maximization. "The optimizer tends to pick those assets with very attractive features (high return and low risk and/or correlation) and tends to short or deselect those with the worst features. These are exactly the cases where estimation error is likely to be highest, hence maximizing the impact of estimation error on portfolio weights. The quadratic programming optimization algorithm takes point estimates as inputs and treats them as if they were known with certainty (which they are not) will react to tiny differences in returns that are well within measurement error." This is exactly the reason that mean-variance optimized portfolios suffer from instability and ambiguity.

A Monte Carlo measure called portfolio resampling can be used to illustrate the effect of estimation error. And it works like this: Suppose what we got are the true distribution parameters covariance matrix Σ_0, and the mean return vector μ_0, we generate a random sample based on the same distribution with n observations as the original sample. Repeating this procedure t times. Each time we got a new

set of optimization input which goes from Σ_1, μ_1 to Σ_t, μ_t. For each of these inputs we can calculate a new efficient frontier represented by m efficient portfolios with the corresponding allocation vectors $w_1...w_m$. But we use each set of allocation vectors $w_i, i = 1...m$ back to the original variance-covariance matrix Σ_0 and the mean return vector μ_0 and get a new efficient frontier which plot below the original efficient frontier. This is because any weight vector optimal for $\Sigma_i, \mu_i, i = 1...t$ can not be optimal for Σ_0, μ_0 The result of the resampling procedure is that estimation error in the inputs parameters is transformed as the uncertainty of the optimal weight vector.

3.1.2 Visualising Estimation Error

I chose data set B to do resampling and to show the effects of estimation error caused by both variance and mean, by variance alone and by mean alone.

Below is a table of input data for portfolio resampling. It includes a partial covariance matrix and a mean return vector for constituents of Stoxx50.

Table 3.1: Partial Covariance Matrix: Data Set B

	Titel						mean
ABN AMRO HOLDING	0.0099	0.0108	0.0073	0.0021	0.0180	...	-0.0021
AEGON	0.0108	0.0256	0.0139	0.0048	0.0255	...	-0.0142
AHOLD KON.	0.0073	0.0139	0.0318	0.0016	0.0114	...	-0.0074
AI LIQUIDE	0.0021	0.0048	0.0016	0.0034	0.0030	...	0.0042
ALCATEL	0.0180	0.0255	0.0114	0.0030	0.0782	...	0.0115
ALLIANZ (XET)	0.0078	0.0181	0.0119	0.0039	0.0167	...	-0.0144
GENERALI	0.0050	0.0086	0.0018	0.0026	0.0132	...	-0.0052
AVENTIS	0.0008	0.0025	0.0048	0.0008	0.0033	...	-0.0007
AXA	0.0095	0.0174	0.0093	0.0035	0.0258	...	-0.0060
BASF (XET)	0.0047	0.0089	0.0038	0.0023	0.0084	...	0.0034

continued on next page

32

Titel					mean		
BAYER (XET)	0.0065	0.0123	0.0130	0.0033	0.0107	...	-0.0073
BBV ARGENTARIA	0.0074	0.0110	0.0055	0.0019	0.0194	...	-0.0009
SANTANDER CTL.HISPANO	0.0084	0.0114	0.0059	0.0025	0.0189	...	0.0005
BNP PARIBAS	0.0061	0.0074	0.0040	0.0019	0.0137	...	0.0054
CARREFOUR	0.0039	0.0054	0.0044	0.0012	0.0094	...	-0.0107
DAIMLERCHRYSLER (XET)	0.0040	0.0080	0.0047	0.0017	0.0104	...	-0.0104
DEUTSCHE BANK (XET)	0.0061	0.0072	0.0078	0.0024	0.0123	...	0.0005
DEUTSCHE TELEKOM (XET)	0.0049	0.0066	0.0063	0.0006	0.0207	...	-0.0160
E ON (XET)	0.0018	0.0053	0.0052	0.0007	0.0018	...	0.0004
ENDESA	0.0062	0.0080	0.0065	0.0009	0.0137	...	-0.0021
ENEL	0.0029	0.0033	0.0034	0.0005	0.0058	...	-0.0071
ENI	0.0025	0.0025	0.0044	0.0010	0.0020	...	0.0063
FORTIS (AMS)	0.0068	0.0131	0.0075	0.0026	0.0136	...	-0.0105
FRANCE TELECOM	0.0074	0.0124	0.0078	-0.0004	0.0438	...	-0.0049
DANONE	0.0027	0.0051	0.0029	0.0019	0.0025	...	0.0033
SOCIETE GENERALE	0.0070	0.0104	0.0066	0.0025	0.0144	...	0.0077
IBERDROLA	0.0015	0.0015	0.0013	-0.0002	0.0002	...	0.0043
ING GROEP CERTS.	0.0078	0.0156	0.0114	0.0036	0.0156	...	-0.0022
L'OREAL	0.0017	0.0047	0.0038	0.0020	0.0011	...	0.0033
LAFARGE	0.0045	0.0097	0.0052	0.0029	0.0049	...	-0.0023
LVMH	0.0077	0.0126	0.0062	0.0030	0.0235	...	0.0071
MUNCH.RUCK. (XET)	0.0061	0.0156	0.0113	0.0040	0.0115	...	-0.0073
NOKIA	0.0053	0.0104	0.0067	0.0034	0.0231	...	0.0019
PHILIPS ELTN.KON	0.0082	0.0125	0.0081	0.0018	0.0296	...	0.0102
REPSOL YPF	0.0023	0.0036	0.0041	0.0005	0.0054	...	-0.0025
ROYAL DUTCH PTL.	0.0034	0.0039	0.0045	0.0012	0.0052	...	-0.0049
RWE (XET)	0.0035	0.0073	0.0087	0.0014	0.0061	...	-0.0059
SAINT GOBAIN	0.0077	0.0127	0.0090	0.0038	0.0129	...	0.0044
SAN PAOLO IMI	0.0077	0.0110	0.0082	0.0023	0.0180	...	-0.0002
SANOFI-SYNTHELABO	0.0000	0.0018	0.0027	0.0010	-0.0002	...	0.0075

Titel					mean		
SIEMENS (XET)	0.0078	0.0119	0.0045	0.0024	0.0328	...	0.0124
SUEZ	0.0053	0.0121	0.0138	0.0020	0.0131	...	-0.0090
TELECOM ITALIA	0.0062	0.0070	0.0044	0.0012	0.0237	...	-0.0003
TELEFONICA	0.0052	0.0064	0.0030	0.0001	0.0237	...	0.0017
TELECOM ITAL.MOBL.	0.0038	0.0055	0.0046	0.0011	0.0168	...	-0.0003
TOTAL SA	0.0019	0.0026	0.0035	0.0004	0.0023	...	0.0036
UNICREDITO ITALIANO	0.0045	0.0065	0.0045	0.0017	0.0069	...	0.0015
UNILEVER CERTS.	0.0017	0.0044	0.0011	0.0018	0.0006	...	-0.0012
VIVENDI UNIVERSAL	0.0050	0.0083	0.0104	0.0007	0.0188	...	-0.0222
VOLKSWAGEN (XET)	0.0044	0.0096	0.0069	0.0018	0.0119	...	0.0014

Figure 4.1 shows the estimation error effect of variance and mean. As discussed by Scherer, the problem gets worse as the number of assets rises because this increases the chance of outliers. The simulated mean-variance efficient frontier is not necessarily consistent with efficient frontier intuition and may not monotonically increase in expected return with increasing risk as in our case. Since the weight vector optimal for simulated input parameters is not optimal for the original inputs parameters.

we can also distinguish the impact of the uncertainty due to estimation errors in means from that due to estimation errors in variance. To measure the estimation error in means, we resample still from the original covariance matrix Σ_0 and mean vector μ_0, but we optimize with the resampled means $\mu_i, i = 1...n$ and the original covariance matrix Σ_0. The result is showed in figure 4.2. To measure the estimation error in variance, we just do the opposite. Optimize with the resampled covariance matrix and the original mean vector. The effect is showed in figure 4.3.

We noticed the dispersion of risk-return points is considerably reduced when estimation error is confined to variances. And small estimation error in means can
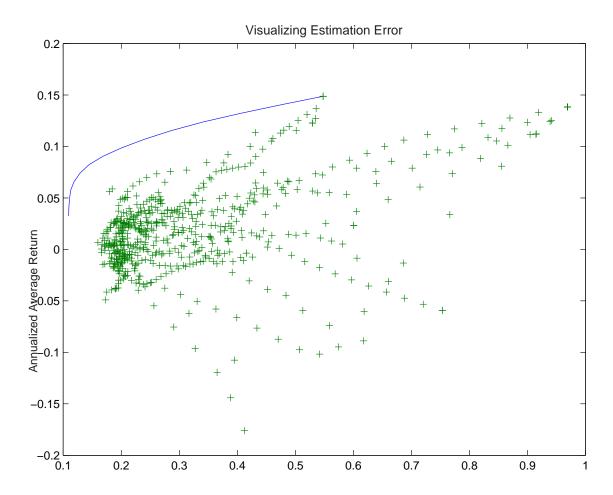
Figure 3.1: Estimation Error Effect

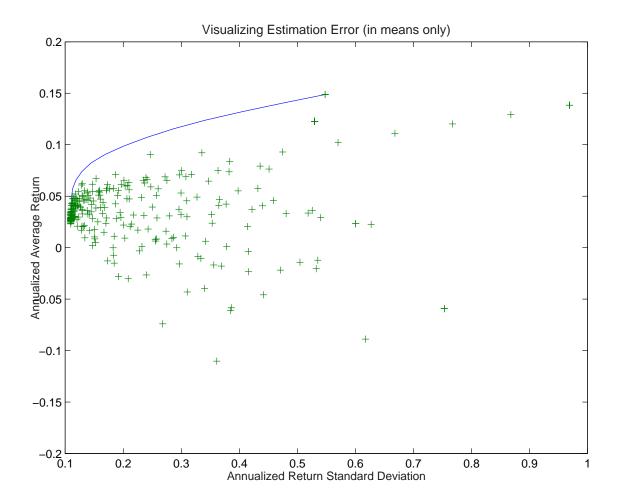Figure 3.2: Estimation Error Caused by Mean

cause the efficient frontier shift considerably.

3.2 Resampled Efficient Frontier

3.2.1 Michaud's Methodology

As pointed out in the earlier section, the quadratic programming optimization algorithm is too sensitive to the quality of input parameters. The result is it maximizes the estimation error problems. Resampled Efficiency, a new concept introduced to
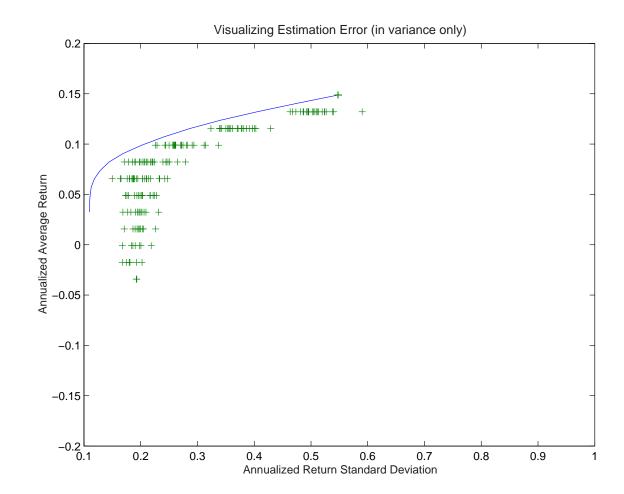
Figure 3.3: Estimation Error Caused by Variance

the asset management world by Michaud, dealt with the estimation error problem.

Portfolios on the resampled frontier are composed of assets weight vectors which are the average of the mean-variance efficient portfolios weight vectors given a certain level of portfolio return. This procedure guaranties that after averaging, the weight vector still sum up to one. But this procedure has no economic justification, and the resampled efficient portfolio is not mean-variance efficient any more by definition.

The procedure can be summarized as follows:

First we run a standard mean-variance optimization. The efficient frontier composed of portfolios varying from the minimum-variance to the maximum return portfolio. Dividing the difference between the minimum and maximum return into m ranks.

The resampled weight for a portfolio of rank m (portfolio number m along the frontier) is given by

$$\bar{w}_m^{resampled} = \frac{1}{n}\sum_{i=1}^{n} w_{im} \qquad (3.1)$$

where w_{im} denotes the weight vector of the mth portfolio along the frontier for the ith resampling.

Step 1 Estimate the variance-covariance matrix and the mean vector of the historical inputs. (Alternatively, the inputs can be prespecified.)

Step 2 Resample, using the inputs created in Step 1, taking T draws from the input distribution; the number of draws, T, reflects the degree of uncertainty in the inputs. Calculate a new variance-covariance matrix from the sampled series. Estimation error will result in different variance-covariance matrices and mean vector from those in Step 1
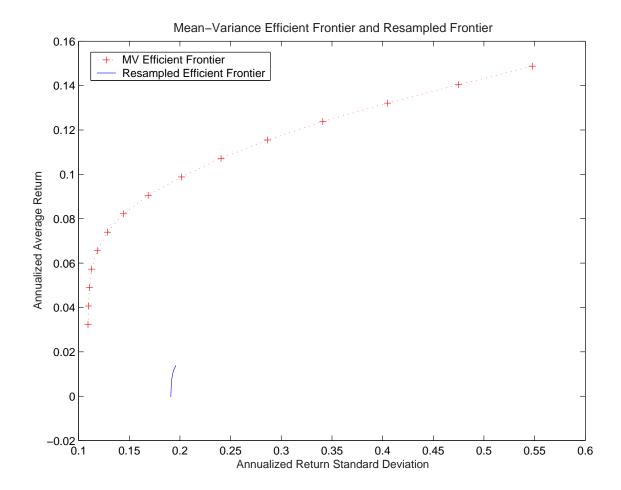
Figure 3.4: Resampled Frontier-Michaud's method

Step 3 Calculate an efficient frontier for the inputs derived in Step 2. Record the optimal portfolio weights for m equally distributed return points along the frontier.

Step 4 Repeat Steps 2 to 3 many times. Calculate average portfolio weights for each return point. Evaluate a frontier of averaged portfolios with the variance-covariance matrix from Step 1 to plot the resampled frontier.

3.2.2 Improved Resampled Frontier

Figure 4.4 shows the mean-variance efficient frontier and the resampled frontier based on data set B. The curve of resampled frontier is remarkably short comparing with mean-variance efficient frontier. Especially in the high return area, there is no point of resampled portfolio at all. Why is it so?

After considered it carefully, I can only see two explanations. One is due to the number of assets. In my case is 50. With the number of assets increases the estimation error problem is getting worse (as showed in figure 4.1) which means the resampled frontier get less efficient (even further away from the mean-variance efficient frontier). The second reason and probably the main reason is due to the methodology itself. If we want to get return level comparable resampled frontier, in my opinion we should take the average of the resampled portfolio weights whose corresponding resampled return (transposed weight vector multiply the original mean return vector) belongs to the same return rank. Not the average of the resampled portfolio weights whose simulated portfolio return (transposed weight vector multiply the simulated mean return vector) belongs to the same rank. And that is actually Michaud's method to get the resampled frontier. So it is not surprising that we see a extremely shortened resampled frontier due to the average of different levels portfolio return.

I redid the resampled frontier with my method of both data sets. As showed in figure 4.5 and figure 4.6 now the resampled frontier is much more comparable to the mean-variance efficient frontier.

It is very interesting to compare the two graphics. In figure 4.5 based on data set A, the efficient frontier annualized return ranges from 5% to 55%, with the annualized standard deviation ranging from 16% to 55%. In figure 4.6 based on data set B, the efficient frontier annualized return ranges from 3% to 15% with the annualized standard deviation ranging from 10% to 55%. A clearly lower return at
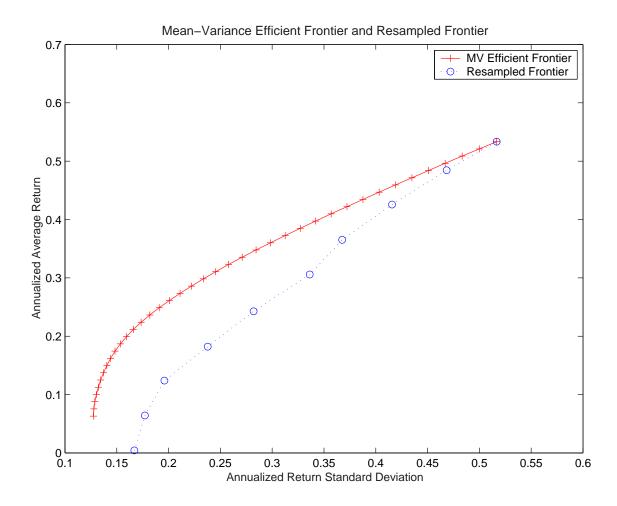
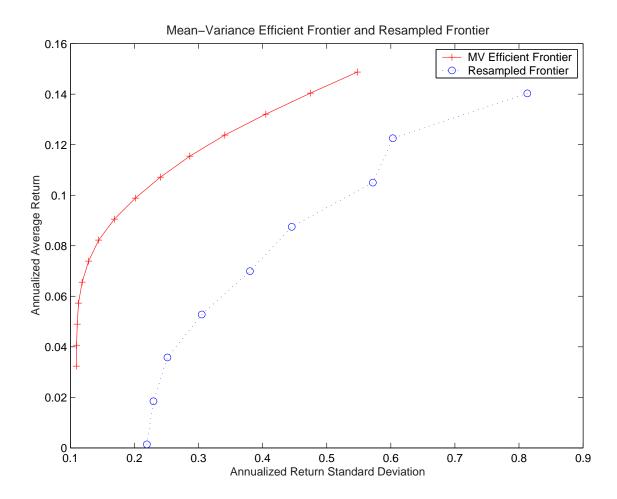Figure 3.5: Resampled Frontier of Data Set A-improved method

Figure 3.6: Resampled Frontier of Data Set B-improved method

comparable risk. This is coherent with the fact that data set B has a lower mean return due to the bear market in the last three years.

The resampled frontier based on data set A is much closer to the efficient frontier, with a maximum standard deviation distance of approximately 10%. On the contrary, the resampled frontier based on data set B is further away from the efficient frontier, with a maximum standard deviation distance of approximately 35%. This shows the estimation error problem is more serious with data set B due to the relatively short time series.

3.2.3 Pros and Cons of Resampled Frontier

As can be foresee, resampled portfolios show a higher diversification, with more assets entering the solutions than mean-variance efficient portfolios. They exhibit less sudden shifts in allocations, giving smooth transition as return requirements change. Thus makes it more desirable for practitioners.

But one thing can not be neglect is the "lucky draws" problem with resampled portfolios. Due to the averaging procedure, one or two heavy allocation in one asset could influence the averaging allocation to that asset greatly.

3.3 Portfolio Revision

Portfolio revision is a very practical problem in the investment management field. When to do a revision, and how to do a revision to maximize portfolio return given a curtain level of portfolio risk are decisions almost every portfolio manager has to make.

After we have chosen a portfolio efficiency measure, whether it is mean-variance or resampled efficiency, as the next step, we have to decide whether the portfolio needs revision to be efficient. Since not all portfolios need revision, some are close to the

43

efficient frontier and are statistically indistinguishable from efficiency. As showed earlier in figure 4.1, wide range of portfolios are statistically equivalent to the efficient frontier. The level of variability is high and illustrates the instability and ambiguity of traditional mean-variance optimization for investment management. Here we need a statistical inference procedure to transform the statistical equivalence region into a sample acceptance region to control the type I error.

3.3.1 Sample Acceptance Region

As introduced by Michaud, an intuitive way to approximate the sample acceptance region from the statistical equivalence region is to find an area under the efficient frontier that includes, on average, $100(1 - \alpha)\%$ of resampled portfolios. The procedure works as the following: "Divide the area under the efficient frontier into mutually exclusive column rectangles that include all the simulated portfolios. Define the base of the rectangle as the minimum return point that contains $100(1-\alpha)\%$ of the simulated portfolios in the rectangle. The curve connection the midpoint of the base of the rectangles contains approximately $100(1 - \alpha)\%$ of the simulated portfolios under the curve. This curve is an estimate of the lower boundary of a $100(1 - \alpha)\%$ sample acceptance region. The test for MV efficiency at the 90% acceptance level proceeds by determining whether the risk and return of a candidate portfolio is within the sample acceptance region. If the portfolio is within the sample acceptance region, no revisions may be required; if the candidate portfolio is outside the region, it probably requires revision."

Here in figure 4.7 and figure 4.8 I showed graphics of the 12500 resampling points for data set A and B respectively. With data set A the resampling points are closer to the efficient frontier, while with data set B the resampling points are further away and not so concentrated along the efficient frontier as with data set A.

Figure 4.9 and 4.10 show the 80%, 90% and 95% sample acceptance region and
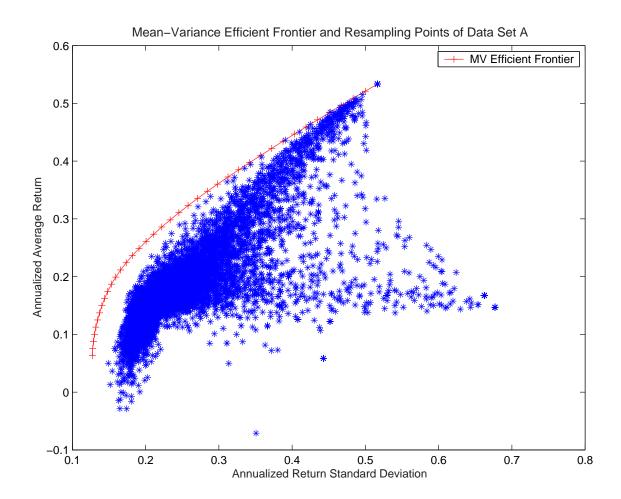
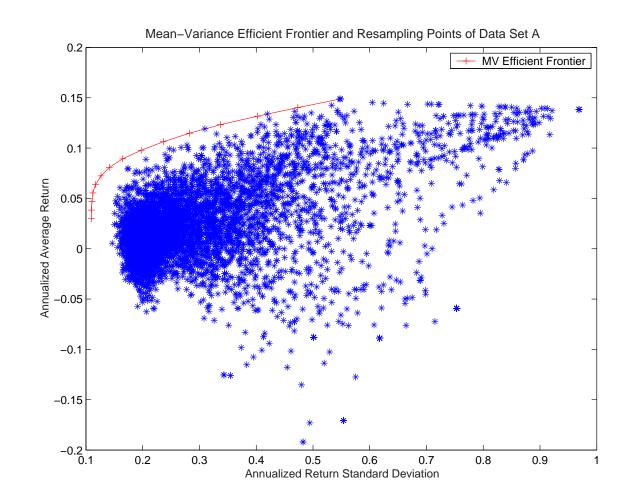Figure 3.7: Resampling Data Set A

Figure 3.8: Resampling Data Set B

efficient frontiers based on data set A and data set B.

Both figures show a too broad acceptance region to tell whether the portfolio is really efficient, especially in the high return high variance area. Again this is, in my opinion, due to the estimation error associated with big number of assets.

But with data set A the problem is not so worse as with data set B where most of the sample acceptance region line stay below zero, which means any portfolio return above zero is efficient and do not need revision. This difference between two data sets is perhaps because data set B has relatively low mean returns and high variances and also the time series are too short to make a reliable estimation of covariance matrix.

3.3.2 Confidence Regions for Resampled Portfolios

In reality the problem often arises is whether a given portfolio is statistically equivalent to an efficient portfolio which satisfies client risk objectives and constraints. Even if the current portfolio is consistent with mean-variance efficiency, but not consistent to the target efficient portfolio, it may still need revision.

In this sections, resampled frontier will represent the portfolio efficiency. This choice is based on two reasons. First, a resampled efficient portfolio is a sample mean vector, and the statistical properties of the sample mean vector are statistically convenient. Second, comparing to mean-variance efficiency, resampled efficiency has more practical investment value.

The judgement of the efficiency of a portfolio is then based on how near it is to the target resampled efficient portfolio. A distance function is required to define the confidence region.

Suppose W is the weight vector of the testing portfolio, W_0 is the weight vector of the target resampled efficient portfolio, S is the covariance matrix of historic return. The test statistic of the distance between portfolio W and W_0 is defined as
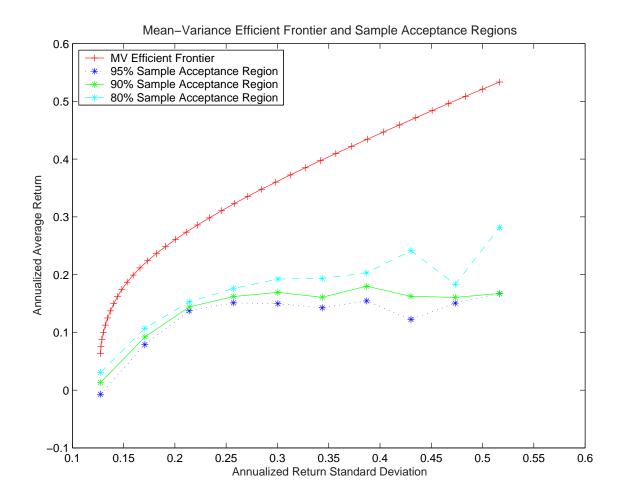
Figure 3.9: Sample-Acceptance-Regions Data Set A
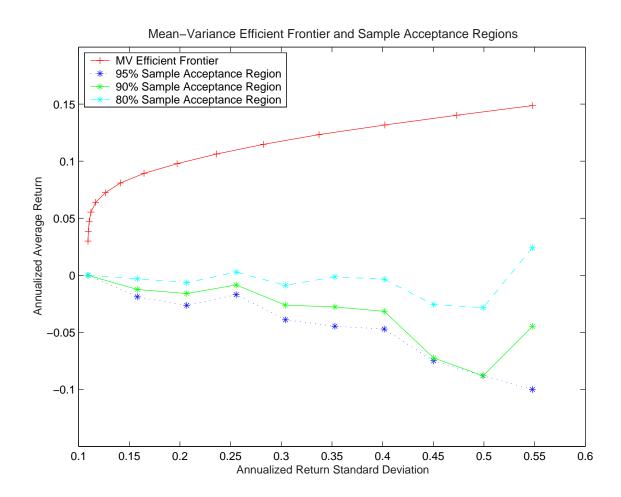
Figure 3.10: Sample-Acceptance-Regions Data Set B

the relative variance.

$$(W - W_0)'S(W - W_0) \leq C \qquad (3.2)$$

The simulation procedure is used to find the constant C which is the test statistic with $100(1 - \alpha)\%$ confidence level.

Choose an equal weight portfolio's variance as the starting point and find the portfolio weight on the resampled efficient frontier which has the same variance as the target portfolio weight. I calculated the value of C is 0.00027542.

One problem with this methodology is that the risk level of the resampled efficient portfolio dramatically affects the shape of the confidence region. The lower the risk level, the denser and compacter the confidence region, as can be foresee from the simulation graphic 4.7.

3.4 An Empirical Study of Portfolio Revision

In this section I would like to do an empirical study of portfolio revision. Given a certain portfolio efficiency judgement rule, based on Euro Stoxx50 historical data, I would like to calculate the portfolio performance and compare the result among different rules.

The study is composed of six parts.

1. Forecasting process.

2. Simulation and resampling process

3. Finding sample acceptance region

4. Finding resampled frontier

5. Portfolio revision

6. Performance calculation and comparison

Periodically fund managers make forecast of next period's assets returns and will make portfolio revision decisions accordingly. The correlation of the forecasts and the ex-post results is quite low - around 0.1 on a monthly basis. I used a simple linear equation to generate next period forecast $Forecast_t = \beta return_t + \mu_t$ where μ_t is a normal distributed random number with standard deviation and mean equal to the corresponding time series. This forecasting process can of course be improved later on.

The data simulation process is the same as those used before. We generate normal distributed time series with the same mean, standard deviation and length as the historical plus forecasted next period data. With the estimated parameters of the simulated data set we do a mean-variance optimization and use the optimal weight vector back to the original parameters. So that we have a resampled data set. This process was repeated 200 times.

With the resampled data set we could find the sample acceptance region with Michaud's method mentioned before. And also the resampled frontier. In this empirical study, I will use the resampled frontier instead of the mean-variance efficient frontier to do portfolio revision due to the more desirable nature of the resampled frontier for practical uses as mentioned before.

If the portfolio is outside of the sample acceptance region, the next step is to do a revision. Here I just tried to find out the weight vector of the portfolio on the resampled frontier with the same variance as the portfolio to be revised.

The last step is performance calculating. We multiply assets historical monthly return with each period's asset weight and get the period's portfolio return. When there is a revision we multiply the absolute value of weight changes with a transaction cost of 0.3%, and this value is deduct from the corresponding return of that period. finally we add each period's return by 1 and calculate a cumulative product of all

and find out the 12 months portfolio return.

The above procedure was repeated 12 times, and I try to find out the performance of the portfolio from October 2002 to September 2003. But unfortunately even if I use the sample acceptance region of 60% there was no portfolio to be revised after running it five times. When I use a sample acceptance region of 90% the result is the same. Actually as can be foresee, this revision rule makes 10 out of 100 portfolio really need a revision given the sample acceptance region of 90%. So I can't give a comparison table here. But the programm code is anyway included in Appendix. The frame work should still be usefully after improving the efficiency testing rule.

3.5 Conclusion

Although both of the portfolio efficiency test procedures are intuitive, due to the large dispersion of data, it is difficult to reject the null hypothesis that the portfolio is efficient. Whether it is that the portfolio is statistically equivalent to efficient portfolio or it is that the the the portfolio is equivalent to the target portfolio. The power of both test are therefore low and unfortunately can not be used in practice in my opinion.

A Matlab Program Codes

A.1 Covariance Matrix with NaN Entries

```
function covM=covariance(A)


%when the original data array A includes NaN
%this function utilize the maximum available data
%to caculate the covariance matrix


col=size(A,2);
covA=zeros(col,col);
for i=1:col
    for j=i:col
        compare=[sum(isnan(A(:,i)));sum(isnan(A(:,j)))];
        cov12=cov(A(max(compare)+1:end,i),A(max(compare)+1:end,j));
        covA(i,j)=cov12(1,2);
    end
end


covM=covA'+covA-diag(diag(covA));
```

A.2 Statistic Analysis

```
A= load('return.dat');
B= load('return_99.dat');


%descriptive statistics
meanA=nanmean(A);
meanB=mean(B);
iqrA=iqr(A);
iqrB=iqr(B);
stdA=nanstd(A);
stdB=std(B);
output=[(1:50)' meanA' meanB' stdA' stdB' iqrA' iqB'];


%Graphical Descriptions


%boxplot
boxplot(A);
boxplot(B);


%mean-std plot
x=(1:50);
subplot(2,1,1);
plot(x,meanA,'r:+',x,meanB,'b-+');
legend('Stoxx50 02.1993~09.2003','Stoxx50 12.1999~09.2003');
title('Two Data Sets Mean Comparison','FontSize',11);
xlabel('Asset Number');
ylabel('Mean Returns');
```

```
subplot(2,1,2);

plot(x,stdA,'r:+',x,stdB,'b-+');

legend('Stoxx50 02.1993~09.2003','Stoxx50 12.1999~09.2003');

title('Two Data Sets Standard Deviation Comparison','FontSize',11);

xlabel('Asset Number');

ylabel('STD');
```

A.3 Normality Test

```
%Lilifors normality test
for n=1:50
    [h p l c] = lillietest(A(:,n));
    Result1(n,:)=[h p l c];
end


for n=1:50
    [h p l c] = lillietest(B(:,n));
    Result2(n,:)=[h p l c];
end
```

A.4 Optimization

```
B = load('return_99.dat');
meanB=mean(B);
stdB=std(B);
```

```
covB=cov(B);

maxB=max(meanB');

minB=min(meanB');

options = (optimset('LargeScale','off');


%without none negative weight constraints

i=1

for n=minB:(maxB-minB)/50:maxB

[w,fval]=quadprog(2.*covB,zeros(50,1),[],[],[meanB;ones(1,50)],[n;1]);

    Result(i,:)=[fval n];

    i=i+1;

end

m=find(Result(:,1)==min(Result(:,1)));

Result(1:m-1,:)=[]

x1=sqrt(12.*Result(:,1));

y1=12.*Result(:,2);

x2=sqrt(12).*stdB;

y2=12.*meanB;

plot(x1,y1); %here single asset points can be added to the graphic

title('Mean-Variance Efficient Frontier','FontSize',11);

xlabel('Annualized Return Standard Deviation');

ylabel('Annualized Average Return');


%with none-negative weight constraints

i=1

for n=minB:(maxB-minB)/50:maxB

[w,fval] = quadprog(2.*covB,zeros(50,1),[],[],[meanB;ones(1,50)],
```

```
                    [n;1],zeros(50,1),ones(50,1),[],[],options);
    Result(i,:)=[fval n];
    i=i+1;
end m=find(Result(:,1)==min(Result(:,1)));
Result(1:m-1,:)=[]
x1=sqrt(12.*Result(:,1));
y1=12.*Result(:,2);
x2=sqrt(12).*stdB;
y2=12.*meanB;
plot(x1,y1,x2,y2,'*');
title('Mean-Variance Efficient Frontier','FontSize',11);
xlabel('Annualized Return Standard Deviation');
ylabel('Annualized Average Return');


%tracking error optimization weight constraint couldn't be formulized
i=1
Result=[]
for n=minB:(maxB-minB)/50:maxB
    [w,fval] = quadprog(2.*covB,zeros(50,1),[],[],
                        [meanB;ones(1,50)],[n;0]);
    weight(:,i)=w;
    Result(i,:)=[fval n];
    i=i+1;
end


m=find(Result(:,1)==min(Result(:,1)));
Result(1:m-1,:)=[];
```

```
x1=12.*Result(:,1);
y1=12.*Result(:,2);

plot(x1,y1);
title('Tracking Error Efficient Frontier','FontSize',11);
xlabel('Annualized tracking error^2');
ylabel('Annualized Active Return');
```

A.5 Estimation Error

```
B = load('return_99.dat');
meanB=mean(B);
covB=cov(B);
stdB=std(B);
maxB=max(meanB');
minB=min(meanB');

options = optimset('LargeScale','off');

%calculating efficient frontier
i=1
for n=minB:(maxB-minB)/50:maxB
[w,fval] = quadprog(2.*covB,zeros(50,1),[],[],[meanB;ones(1,50)],
                    [n;1],zeros(50,1),ones(50,1),[],[],options);
    Result(i,:)=[fval n];
    i=i+1;
end
```

```
m=find(Result(:,1)==min(Result(:,1)));

Result(1:m-1,:)=[];

x1=sqrt(12.*Result(:,1));

y1=12.*Result(:,2);

efB=[x1 y1]

save efB


%estimation error data generating

total=[];

for j=1:200

    sim=ones(46,1)*meanB+randn(46,50).*(ones(46,1)*stdB);

    meanS=mean(sim);

    covS=cov(sim);

    maxS=max(meanS');

    minS=min(meanS');

    i=1;

    Result=[];


for n=minB:(maxB-minB)/25:maxB

[w,fval] = quadprog(2.*covS,zeros(50,1),[],[],[meanB;ones(1,50)],

                    [n;1],zeros(50,1),ones(50,1),[],[],options);

%change covS to covB or meanS to meanB find out the

%estimation error effect of mean and variance respectively

    fval=w'*covB*w

    r=meanB*w

    Result(i,:)=[fval r];

    i=i+1;
```

```
end
m=find(Result(:,1)==min(Result(:,1)));
Result(1:m-1,:)=[];
rowend=size(total,1);
total((rowend+1):(rowend+size(Result,1)),:)=Result;
end


x2=sqrt(12.*total(:,1));
y2=12.*total(:,2);


plot(x1,y1,x2,y2,'+'); %here single asset points can be added to the graphic
title('Visualizing Estimation Error','FontSize',11);
xlabel('Annualized Return Standard Deviation');
ylabel('Annualized Average Return');
```

A.6 Simulation

```
A = load('return.dat');
meanA=nanmean(A);
stdA=nanstd(A);
covA=covariance(A);
maxA=max(meanA');
minA=min(meanA');
options = optimset('LargeScale','off');


%efficient frontier generating based on parameters of
%data set A with none-negative weight constraints
```

```
i=1

for n=minA:(maxA-minA)/49:maxA

   [w,fval] =

      quadprog(2.*covA,zeros(50,1),[],[],[meanA;ones(1,50)],

              [n;1],zeros(50,1),ones(50,1),[],[],options);

      EF(i,:)=[fval n];

      i=i+1;

end


%delete the inefficient data points

m=find(EF(:,1)==min(EF(:,1)));

EF(1:m-1,:)=[]


%annualize data

x1=sqrt(12.*EF(:,1));

y1=12.*EF(:,2);


save EFdataA


%simulation process

m=500; i=1;

for j=1:m

   sim=ones(size(A,1),1)*meanA+randn(size(A,1),50).*(ones(size(A,1),1)*stdA);

   meanS=mean(sim);

   covS=cov(sim);

   maxS=max(meanS');

   minS=min(meanS');
```

```
for n=minS:(maxS-minS)/24:maxS

    w=quadprog(2.*covS,zeros(50,1),[],[],[meanS;ones(1,50)],

                [n;1],zeros(50,1),ones(50,1),[],[],options);

    Weight(:,i)=w;

    Return(i,:)=meanA*w;

    Variance(i,:)=w'*covA*w;

    i=i+1;

end

%delete the inefficient data points

o=find(Variance((i-25):(i-1),1)==min(Variance((i-25):(i-1),1)));

Variance((i-25):(i+o-27),:)=[];

Return((i-25):(i+o-27),:)=[];

Weight(:,(i-25):(i+o-27))=[];

i=size(Return,1)+1;

end


save simulationA
```

A.7 Sample Acceptance Region

```
load('simulationA.mat');


c=1;
%the lowest and highest standard deviation of efficient frontier
minV=sqrt(12*min(EF(:,1)));
maxV=sqrt(12*max(EF(:,1)));
Variance=sqrt(12*Variance);
for n=minV:(maxV-minV)/9:maxV
```

```
    s=1;
    out=[];
    for k=1:size(Variance,1)
        if (n<=Variance(k,:))&(Variance(k,:)<(n+(maxV-minV)/9))
            out(s,:)=Return(k);
            s=s+1;
        end
    end
    if ~isempty(out)
        OUT(c,1)=prctile(out,5);
        OUT(c,2)=prctile(out,10);
        OUT(c,3)=prctile(out,20);
    end
    c=c+1;
end


x2=(minV:(maxV-minV)/9:maxV)'; y2=12.*OUT(:,1);
x3=(minV:(maxV-minV)/9:maxV)'; y3=12.*OUT(:,2);
x4=(minV:(maxV-minV)/9:maxV)'; y4=12.*OUT(:,3);


%add the oringinal simulation points
% x5=Variance
% y5=12*Return


%draw graphic of sample acceptance lines
plot(x1,y1,'r-+',x2,y2,'b:*',x3,y3,'g-*',x4,y4,'c--*');
legend('MV Efficient Frontier','95% Sample Acceptance Region',
```

```
        '90% Sample Acceptance Region','80% Sample Acceptance Region');
title('Mean-Variance Efficient Frontier and Sample Acceptance
Regions','FontSize',11);
xlabel('Annualized Return Standard Deviation');
ylabel('Annualized Average Return');
```

A.8 Resampling

A.8.1 Michaud's Method

```
load('variables-B');


%resampling Michaud's method

  total=[];
  W=zeros(50,26);
  m=200;
  for j=1:m
  sim=ones(46,1)*meanB+randn(46,50).*(ones(46,1)*stdB);
  meanS=mean(sim);
  covS=cov(sim);
  i=1;
  weight=[];
  for n=minB:(maxB-minB)/25:maxB
      [w,fval] = quadprog(2.*covS,zeros(50,1),[],[],[meanS;ones(1,50)],
                          [n;1],zeros(50,1),ones(50,1),[],[],options);
```

```
        weight(:,i)=w;

        i=i+1;

    end

    W=W+weight;

end W=W./m;


for k=1:26

    fval=W(:,k)'*covB*W(:,k);

    r=meanB*W(:,k);

    Result(k,:)=[fval r];

end

m=find(Result(:,1)==min(Result(:,1)));

Result(1:m-1,:)=[];

x2=sqrt(12.*Result(:,1));

y2=12.*Result(:,2);


%draw graphic

plot(x1,y1,'r:+',x2,y2,'b-');

legend('MV Efficient Frontier','Resampled Efficient Frontier');

title('Mean-Variance Efficient Frontier and Resampled Efficient

        Frontier','FontSize',11);


xlabel('Annualized Return Standard Deviation');

ylabel('Annualized Average Return');
```

A.8.2 Improved Method

```
load('simulationA');
```

```
%resampling my method
c=1;
minR=min(Return);
maxR=max(Return);
W=zeros(50,25);
for n=minR:(maxR-minR)/10:maxR
    s=0;
    for k=1:size(Return,1)
            if (n<=Return(k,:))&(Return(k,:)<(n+(maxR-minR)/24))
                W(:,c)=W(:,c)+Weight(:,k);
                s=s+1;
            end
     end
    %taking average
    W(:,c)=W(:,c)/s;
    fval=W(:,c)'*covA*W(:,c);
    R=meanA*W(:,c);
    Result(c,:)=[fval R];
    c=c+1;
end
%delete inefficient data points
m=find(Result(:,1)==min(Result(:,1)));
Result(1:m-1,:)=[];
x2=sqrt(12.*Result(:,1));
y2=12.*Result(:,2);
```

```
%draw graphic
plot(x1,y1,'r-+',x2,y2,'b:o');
legend('MV Efficient Frontier','Resampled Frontier');
title('Mean-Variance Efficient Frontier and
        Resampled Frontier','FontSize',11);
xlabel('Annualized Return Standard Deviation');
ylabel('Annualized Average Return');


save resampleA
```

A.9 Revision

```
A = load('return.dat');


wp0=1/10*ones(10,1);
for j=1:12
    %forecasting the monthly returns
    his=A(1:(115+j),:);
    stdH=nanstd(his);
    forecast=0.1*select((116+j),:)+randn(1,50).*stdH
    %forecasting process can be improved later
    fore(1:(115+j),:)=his;
    fore((116+j),:)=forecast;
    meanF=nanmean(fore);
    minF=min(meanF');
    maxF=max(meanF');
    covF=covariance(fore);
```

```matlab
stdF=nanstd(fore);

location=wp0'*covF*wp0;


%simulation process to find out the sample acceptance region
m=200;
i=1;
Weight=[];
Return=[];
Variance=[]
for k=1:m
    sim=ones(116+j,1)*meanF+randn(116+j,50).*(ones(116+j,1)*stdF);
    meanS=mean(sim);
    covS=cov(sim);
    maxS=max(meanS');
    minS=min(meanS');
        for n=minS:(maxS-minS)/24:maxS
        w=quadprog(2.*covS,zeros(50,1),[],[],[meanS;ones(1,50)],
                   [n;1],zeros(50,1),ones(50,1));
        Weight(:,i)=w;              %m*25 simulated portfolio weight vector
        Return(i,:)=meanF*w;        %m*25 simulated portfolio return
        Variance(i,:)=w'*covF*w;  %m*25 simulated portfolio variance
        i=i+1;
        end
    %delete the inefficient data points
     s=find(Variance(i-25:i-1,1)==min(Variance(i-25:i-1,1)));
     Return(i-25:i+s-27,:)=[];
     Variance(i-25:i+s-27,:)=[];
```

```
        Weight(:,i-25:i+s-27)=[];

         i=size(Return,1)+1;

 end


%find the sample acceptance region

c=1;

r=[];

minV=min(Variance);

maxV=max(Variance);

for a=1:size(Variance,1)

     if ((location-(maxV-minV)/25)<=Variance(a,:))&

          (Variance(a,:)<(location+(maxV-minV)/25));

     %bandwidth can be changed

          r(c,:)=Return(a);

          c=c+1;

     end

end

if ~isempty(r)

     OUT=prctile(r,40);

else

     weightE(:,j)=wp0;

     delta_weight(:,j)=abs(weightE(:,j)-wp0);

     continue;

end


%decide whether the portfolio is outside the

%sample acceptance region and need a revision
```

```matlab
if OUT<=meanF*wp0

    weightE(:,j)=wp0;

    delta_weight(:,j)=abs(weightE(:,j)-wp0);

    continue;

end


%find out same variance portfolios on resampled frontier
h=1;
minR=min(Return);
maxR=max(Return);
weight=[];
for o=minR:(maxR-minR)/24:maxR
 s=0;
 W=zeros(50,1);
 for b=1:size(Return,1)
     if (o<=Return(b,:))&(Return(b,:)<(o+(maxR-minR)/24))
         W=W+Weight(:,b);
         s=s+1;
     end
 end
weight(:,h)=W/s;
fval=weight(:,h)'*covF*weight(:,h);
R=meanF*weight(:,h);
out(h,:)=[fval R];
h=h+1;
end
q=find(out(:,1)==min(out(:,1)));
```

```
    out(1:q-1,:)=[];
    weight(:,1:q-1)=[];



    %find out the corresponding weight vector of the resampled portfolio
    for p=1:size(out,1)
        if (out(p,1)<=wp0'*covF*wp0)&(out(p+1,1)>wp0'*covF*wp0)
         weightE(:,j)=weight(:,p);
        end
    end
    if size(weightE,2)<j
        weightE(:,j)=weight(:,p);
    end
    delta_weight(:,j)=abs(weightE(:,j)-wp0);
    wp0=weightE(:,j);
end

%portfolio performance calculation
Rp=cumprod(diag(A(117:128,:)*weightE)-(sum(delta_weight)*0.003)'+1)-1
```

A.10 Confidence Region

```
load('resampleA');
wp0=1/50*ones(50,1);
minR=min(Return);
maxR=max(Return);
```

```
location=wp0'*covA*wp0;
n = interp1(Result(:,1),Result(:,2),location);
    s=0;
    W0=zeros(50,1);
    for k=1:size(Return,1)
            if (n<=Return(k,:))&(Return(k,:)<(n+(maxR-minR)/24))
                W0=W0+Weight(:,k);
                s=s+1;
            end
      end
    %taking average
    W0=W0/s;


    for a=1:size(Weight,2)
    constant(a)=(Weight(:,a)-W0)'*covA*(Weight(:,a)-W0);
    end
C=prctile(constant,10)
```

Bibliography

[1] F. Fabozzi, F. Gupta, and M.H. Markowitz. The legacy of modern portfolio theory. *The Journal of Investing*, 2002.

[2] R. Grinold and R. Kahn. *Active Portfolio Management*. McGraw-Hill, 2000.

[3] M. H. Markowitz. Portfolio selection. *The Journal of Finance*, VII, 1953.

[4] O.R. Michaud. *Efficient Asset Management*. Harvard Business School Press, 1998.

[5] A. Nardon. Absolute and relative optimisations. 2002.

[6] R. Roll. A mean/variance analysis of tracking error. *The Journal of Portfolio Management*, 2002.

[7] B. Rönz. Computergestützte statistik i. 2001.

[8] B. Scherer. *Portfolio Construction and Risk Budgeting*. Risk Books, 2002.