Portfolio Resampling on Various Financial Models

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Science in Statistics

by

Yu-Ching Chen

2013
Abstract of the Thesis

Portfolio Resampling on Various Financial Models

by

Yu-Ching Chen

Master of Science in Statistics
University of California, Los Angeles, 2013
Professor Hongquan Xu, Chair

Portfolio resampling is a new approach to portfolio optimization. It generates a higher degree of diversification and smoothness on classic Markowitz’s model, which in turn reduces risk and enhances forecast ability. This paper examines this technology on three financial models: single index model, constant correlation model, and multigroup model. Two sets of constraints regarding short sales are also implemented on the models above.

The analysis starts with the price data of 30 stocks and a market index. On the risk-return space, a concept called efficient frontier is introduced. Efficient frontier enables us to determine a set of optimal portfolios by setting a wide range of hypothetical risk-free rates. Some of the restrictions from the original assumption are also discussed. Through Monte Carlo simulation, we are able to visualize the resampling effect. The benefits and the limitations of this technique are put into the context of statistical analysis.
The thesis of Yu-Ching Chen is approved.

Nicolas Christou

Ying Nian Wu

Hongquan Xu, Committee Chair

University of California, Los Angeles
2013
My sincere gratitude to . . .
the professors at UCLA Statistics for insightful advice,
and my family for endless support.
# Table of Contents

1 Introduction ................................................. 1

2 Data ...................................................... 3
   2.1 Probability Assumption ................................. 4

3 Financial Models ................................. 6
   3.1 Risk - Return Space ....................................... 6
      3.1.1 Portfolio Terminology .............................. 7
      3.1.2 Finding the Efficient Frontier: Optimization .... 7
      3.1.3 Risk Free Rate and Sharpe Ratio ................... 8
   3.2 Models for Optimal Portfolio ....................... 9
      3.2.1 Short Sales ........................................... 9
      3.2.2 Markowitz’s Model .................................. 9
      3.2.3 Single Index Model ................................ 10
      3.2.4 Constant Correlation Model ....................... 11
      3.2.5 Multigroup Model .................................. 11
      3.2.6 Model Performance ................................. 11

4 Resampling ............................................. 14
   4.1 Monte Carlo Simulation ................................ 14
      4.1.1 Tracing Out the Efficient Frontier ............. 15
      4.1.2 The effect of risk free rate $R_f$ ............... 15
   4.2 Results .................................................... 16
4.2.1 Smoother Transitions ......................................... 16
4.2.2 Greater Diversification ................................. 16
4.2.3 Effect on Models .............................................. 23

5 Conclusion ............................................................. 27

6 Appendix ............................................................... 29

References ............................................................... 30
LIST OF FIGURES

2.1 $R_i$ of 30 Stocks and the market index over 109 periods . . . . . . . . . 4
2.2 Distributions of stock XOM: price and log returns . . . . . . . . . 5
3.1 30 Stocks and the market index on the risk-return space . . . . . . . 6
3.2 Performance in terms of optimal portfolio returns . . . . . . . . . . . 12
3.3 Performance in terms of efficient frontiers . . . . . . . . . . . . . . . 13
4.1 Resampled Single Index Model: short sales allowed . . . . . . . . . . 17
4.2 Original Single Index Model: short sales allowed . . . . . . . . . . . 17
4.3 Resampled Single Index Model: short sales not allowed . . . . . . . 18
4.4 Original Single Index Model: short sales not allowed . . . . . . . . . 18
4.5 Resampled Constant Correlation Model: short sales allowed . . . 19
4.6 Original Constant Correlation Model: short sales allowed . . . . . 19
4.7 Resampled Constant Correlation Model: short sales not allowed . . 20
4.8 Original Constant Correlation Model: short sales not allowed . . . 20
4.9 Resampled Multigroup Model . . . . . . . . . . . . . . . . . . . . . 21
4.10 Original Multigroup Model . . . . . . . . . . . . . . . . . . . . . . 21
4.11 $m = 13^{th}$ portfolio allocation of resample and original methods.
   Single Index Model, short sales not allowed . . . . . . . . . . . . . . . 22
## List of Tables

2.1 30 stocks from 10 different industries ........................................ 3

4.1 $m = 13^{th}$ portfolio allocation under Single Index Model where short sales are not allowed: original, resampled, and their corresponding $\hat{\beta}$, ......................................................... 24

4.2 Intervals of Returns ................................................................. 25

4.3 Coefficients of Variation .......................................................... 26
CHAPTER 1

Introduction

Portfolio resampling is a sampling technique that facilitates portfolio selection. This technique applies Monte Carlo simulation to find the optimized asset allocation. Resampling aims at fixing the error-maximizing problem that stems from the original procedures. The problem arises because investors naturally seek return-maximization and risk-minimization. Hence, the algorithm is prone to error. \(^1\)

In Portfolio Resampling: Review and Critique, Sherer (2002) addresses the pros and cons of this statistical tool under Markowitz’s model for optimal portfolio selection. Markowitz’s model requires estimations of the entire \(k \times k\) variance-covariance matrix. Estimations can be inefficient when the number of asset \(k\) is large. Consequently, our ability to measure might be restricted by computational power.

The goal of this research is to explore portfolio resampling on three financial models: single index model, constant correlation model, and multigroup model. All three models reduce the estimation required in Markowitz’s model. Single index model approaches the optimization with linear regression. Constant correlation model and multigroup model simplify the variance-covariance matrix under different assumptions. We also address the effect of short sales. We are interested in knowing whether the technique offers similar preferable criteria under those

\(^1\)Michaud (1989) and Nawrocki (1996) addressed this problem.
models. The portfolio performance in terms of asset return is also included in the discussion.

The analysis stresses more on the models than data itself. Nevertheless, we begin with a wide range of data collection to avoid bias. We also compare the parametric assumption and empirical observation, which leads to the Monte Carlo simulation. Several portfolio terminologies will be explained in addition to the models. We introduce risk - return space for its pertinence to the two-dimensional optimization. The optimization is performed by setting a wide range of risk-free rates. Elton et al. (1978) describes this procedure in Simple Criteria for Optimal Portfolio Selection: Tracing Out the Efficient Frontier.

With an underlying probability assumption and estimated parameters, re-sampling, as a Monte Carlo simulation, generates different asset allocations. We then examine several subsequent statistical criteria. Overall performances of each model are compared under the resampling and original scheme. Finally, a study of the results allows us to make further inference from this technique.
We begin with 10 different industries. In each industry, we pick 3 large companies from the Forbes 500 list of 2013. Consequently, \( k = 30 \) stocks are chosen with longer history of stock prices and therefore more data available. S&P 500 (code:\( ^\wedge \text{GSPC} \)) is also included as the market index. It is based on the market capitalizations of 500 leading companies in the U.S.. Below are the stocks and the industries they belong to:

<table>
<thead>
<tr>
<th>Oil and Gas</th>
<th>Variety Stores</th>
<th>Machinery</th>
<th>Insurance</th>
<th>Computers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exxon Mobile</td>
<td>Wal-Mart</td>
<td>General Electric</td>
<td>Berkshire Hathaway</td>
<td>HP</td>
</tr>
<tr>
<td>Chevron</td>
<td>Kroger</td>
<td>Caterpillar</td>
<td>AIG</td>
<td>IBM</td>
</tr>
<tr>
<td>Conocophillips</td>
<td>Costco</td>
<td>United Technologies</td>
<td>Allstate</td>
<td>Dell</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Telecom</th>
<th>Banks</th>
<th>Health Care</th>
<th>Software</th>
<th>Drugs</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT&amp;T</td>
<td>Bank of America</td>
<td>UnitedHealth</td>
<td>Apple</td>
<td>Pfizer</td>
</tr>
<tr>
<td>Verizon</td>
<td>J.P. Morgan</td>
<td>WellPoint</td>
<td>Microsoft</td>
<td>Johnson &amp; Johnson</td>
</tr>
<tr>
<td>Sprint Nextel</td>
<td>Citigroup</td>
<td>Aetna</td>
<td>Oracle</td>
<td>Merck</td>
</tr>
</tbody>
</table>

Table 2.1: 30 stocks from 10 different industries

The adjusted stock price data is acquired from Yahoo! Finance. For statistical analysis in R, the function \texttt{getReturns()} in package \texttt{stockPortfolio} is capable of data gathering. The stock code, time period, as well as time range need to be specified.
2.1 Probability Assumption

Each stock has a lifespan of 110 monthly price data $S_i, i = 1, \ldots, 110$. The price data can be converted to monthly return

$$R_i = \frac{S_{t+1} - S_t + D}{S_t}$$

where $S_{t+1}$ stands for ending investment value, $S_t$ stands for beginning investment value, and $D$ stands for cash dividends. We adopt adjusted stock price which takes into account of dividends, so $D$ can be ignored. As a result, the only two values of concern are the beginning investment value and the ending investment value, or $S_i$ and $S_{i+1}$. Note that we lose one observation in the process, which makes the data length $T = 109$. Figure 2.1 captures the $R_i$ of 30 stocks and the market index over that time period.

![Figure 2.1](image)

Figure 2.1: $R_i$ of 30 Stocks and the market index over 109 periods

It is common to assume that the returns follow Gaussian distribution under
the Black-Scholes framework. Suppose $A$ is the continuously compounded annual rate of return and is assumed to be normally distributed, then we have $S_t/S_0 = e^{At}$. As a result, $S_t/S_0$ is lognormally distributed.

![Distribution of XOM Price](image1.png)

![Distribution of XOM Log Returns](image2.png)

**Figure 2.2: Distributions of stock XOM: price and log returns**

Taking Exxon Mobile (code: XOM) as an example, we present the distribution of both the stock prices and log returns in Figure 2.2. The prices are distributed in a rather random manner in the first panel. In the second panel, however, the distribution of log returns appears to be bell-shaped. A normal curve is fitted in support of our probability assumption. We also observe that there is a slight inclination of 'fat-tail' in the distribution. The extreme values occasionally occur as potential violation of the assumption. However, the normality assumption remains fairly reasonable in the context of resampling.
CHAPTER 3

Financial Models

3.1 Risk - Return Space

Every set of historical returns $R_i$ can be summarized by its expected value and standard deviation, which is frequently referred to as risk. The expected returns $E(R_i)$ and risk $\sigma_i$ form the risk-return space. Figure 3.1 is the risk-return space that contains our sample of 31 assets. $E(R_i)$ is on the y-axis and $\sigma_i$ on the x-axis. All 31 assets are included in the plots.

![Figure 3.1: 30 Stocks and the market index on the risk-return space](image)

Note that the dots in Figure 3.1 are the investment opportunity sets. They are all feasible pairs of $E(R_i)$ and $\sigma_i$ from all portfolios resulting from different
values of asset allocations. The boundary to the left is the portfolio possibilities curve. That leads us to the introduction of portfolio terminology below.

### 3.1.1 Portfolio Terminology

Suppose a portfolio consists of $k$ assets. Let $E(R_i)$ and $\sigma_i$ be the expected return and standard deviation of stock $i$, $i = 1, 2, ..., k$. and $\sigma_{ij}$ be the covariance between stocks $i$ and $j$. Asset allocation $w_i$ is the proportion of each of the assets in the portfolio, $\sum_{i=1}^{k} w_i = 1$. The expected return of portfolio is therefore

$$E(R_p) = \sum_{i=1}^{k} w_i E(R_i) = \mathbf{w}' \bar{\mathbf{R}}, \quad (3.1)$$

where $\mathbf{w}' = (w_1, w_2, ..., w_k)$, and $\bar{\mathbf{R}} = (E(R_1), E(R_2), ..., E(R_k))$. The variance is defined as

$$\sigma^2_p = \sum_{i=1}^{k} \sum_{j=1}^{k} w_i w_j \sigma_{ij} = \mathbf{w}' \Sigma \mathbf{w} \quad (3.2)$$

where $\Sigma$ is the symmetric, positive definite $k \times k$ variance-covariance matrix of the returns of the $k$ assets as shown below:

$$\Sigma = \begin{pmatrix}
\sigma^2_1 & \sigma_{12} & \cdots & \sigma_{1k} \\
\sigma_{21} & \sigma^2_2 & \cdots & \sigma_{2k} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{k1} & \sigma_{k2} & \cdots & \sigma^2_k
\end{pmatrix} \quad (3.3)$$

### 3.1.2 Finding the Efficient Frontier: Optimization

The efficient frontier is the upper part of the portfolio possibilities curve. It is efficient because the Finance Axioms state that investors:

- prefer more to less: higher expected value $E(R_i)$ is preferred.
- are risk-averse: lower standard deviation $\sigma_i$ is preferred.
The most efficient risk-return tradeoff is therefore characterized by the upward curve in the risk-return space. The curve, or the efficient frontier, starts with minimum variance portfolio to the left and ends with maximum return portfolio to the right. Minimum variance portfolio, as its name suggests, is the portfolio that provides the lowest variance, and therefore the lowest standard deviation, among all possible portfolios of risky assets. Maximum portfolio offers the highest return as well as the highest risk.

3.1.3 Risk Free Rate and Sharpe Ratio

An asset that has a certain future return is called risk-free asset. Treasury Bills are considered to be one because they are backed by the U.S. Government. The return of risk free rate is denoted $R_f$, usually with a lower rate of return. Analytical complexity may arise when $R_f$ becomes higher than some of the assets, which further influences our resampling procedure.

Since the risk free rate is often the minimum rate of return, a rational investor will require their asset to generate a return above that base level. Risk premium $E(R_i) - R_f$ represents the difference and is frequently referred to as the excess return.

When an investor faces the efficient frontier, the combinations of the risk-free asset with the risky portfolio $p$ lie on the capital allocation line (CAL):

$$E(R_p) = R_f + \sigma_p \left( \frac{E(R_i) - R_f}{\sigma_i} \right)$$

$$= R_f + \sigma_p SR_i$$

where $E(R_p)$ and $\sigma_p$ are the expected return and the standard deviation of the portfolio returns. Sharpe ratio $SR_i$ is defined as the excess return per unit
of risk

\[ SR_i = \frac{E(R_i) - R_f}{\sigma_i} \]  

(3.6)

From equation (3.2), it is clear that \( R_f \) is the y-intercept of CAL, and \( SR_i \) is the slope of CAL. One can also regard \( SR_i \) as the market price of risk.

### 3.2 Models for Optimal Portfolio

#### 3.2.1 Short Sales

Short sales are characterized by negative asset allocation. Not allowing short sales adds another constraint in Kuhn-Tucker conditions where the asset allocation \( w_i > 0 \). \( w_{im} \) denotes the \( k \times 1 \) vector of the \( m \)th portfolio along the frontier for the \( i \)th resampling. This results in truncated distribution (at 0) that affects the algorithm as a whole. We therefore examine resampling for both allowing and disallowing short sales under the models separately.

#### 3.2.2 Markowitz’s Model

Markowitz’s model attempts to find the best market price of risk, or, in other words, the largest Sharpe ratio. The solution

\[
\max SR_i = \frac{E(R_i) - R_f}{\sigma_i}
\]  

(3.7)

is subject to \( \sum_{i=1}^{k} w_i = 1 \). We then define \( \mathbf{R} = (E(R_1) - R_f, E(R_2) - R_f, ..., E(R_k) - R_f) \). The solution involves solving for

\[
\mathbf{Z} = \Sigma^{-1} \mathbf{R}
\]  

(3.8)
where $\mathbf{Z} = (z_1, z_2, \ldots, z_k)$. And then rescale $z_i$’s to meet the constraint that sum of asset allocation should be 1. Namely,

$$w_i = \frac{z_i}{\sum_{i=1}^{k} z_i} \quad (3.9)$$

Note that the number of estimates in $\Sigma$ is $k(k-1)/2$. The number gets large as $k$ increases. When multiple dimension of calculation is required, especially in resampling, the calculation can be lengthy.

### 3.2.3 Single Index Model

Single index model requires smaller number of calculation compares to Markowitz procedure with number of estimates $= k(k-1)/2$. Consequently, the estimation of the covariance matrix is simplified and the analysis of security risk premiums is enhanced. Single index model is implemented by regression:

$$R_i(t) = \alpha_i + \beta_i R_M(t) + e_i(t)$$

Where $R_i$ = excess return of a security; $\alpha_i$ = Security’s expected excess return when the market return is zero(intercept); $R_M$ = excess return of the market; $\beta_i = \text{cov}(R_i, R_M)/\sigma^2_M$; $\beta_i R_M$ = systematic risk; $e_i$ = Idiosyncratic risk.

The goal is to find $\beta_i$. This model is based on several assumptions from Capital Asset Pricing Model (CAPM). In particular, the market is assumed to be in competitive equilibrium, which suggests that the market portfolio is the collection of all existing risky asset. The regression therefore measures the proportion of risk contribution from the security with respect to the market index. Hence, the higher the regression coefficient $\beta_i$, the more extra return is gained.
3.2.4 Constant Correlation Model

One of the popular correlation structures that simplifies the calculation of variance-
covariance matrix $\Sigma$ is constant correlation, $\rho$, such that

$$
\Sigma = \begin{pmatrix}
\sigma_1^2 & \rho \sigma_1 \sigma_2 & \cdots & \rho \sigma_1 \sigma_k \\
\rho \sigma_1 \sigma_2 & \sigma_2^2 & \cdots & \rho \sigma_2 \sigma_k \\
\vdots & \vdots & \ddots & \vdots \\
\rho \sigma_1 \sigma_k & \rho \sigma_2 \sigma_k & \cdots & \sigma_k^2
\end{pmatrix}
$$

(3.10)

This structure also reduces the estimation to $k$ of $\sigma_i$ and an average correlation $\rho$. However, assuming all pairs of stocks have the same correlation can be an oversimplification in many cases.

3.2.5 Multigroup Model

Multigroup model can be seen as a slight release on the constant correlation assumption. Here, the correlation is assumed to be level only within an industry. The model still reduce the number of correlation to be estimated from $k(k-1)/2$ to $k'(k'-1)/2$, where $k'$ is the number of industries, $k' < k$.

3.2.6 Model Performance

Apart from simplifying calculation, the main purpose of any financial model is, after all, generating return. The financial world has yet to conclude the superiority among different models above. However, we may at least examine those models with our particular data set.

Figure 3.2 presents all four models introduced, and we separate the cases either short sales are allowed or not for single index model and constant correlation model. The optimal portfolio allocation is determined by the six approaches,
Figure 3.2: Performance in terms of optimal portfolio returns

and we plot the performance over the observational period. Meanwhile, the equal allocation \((w_i = 1/k \text{ for all } i)\) is presented for comparison, as well as the market index S&P 500. In terms of returns,

\[
\text{Markowitz} > CCM_{SS} \approx SIM_{NSS} > MGM > SIM_{SS} > CCM_{NSS} > \text{Equal} > \text{S&P500}
\]

where SS is coded for short sales allowed and NSS otherwise. The other way to examine the performance is to take into consideration of risk, or the fluctuation. That can be summarized in risk-return space in terms of efficient frontier. A rational investor is indifference on any combination on the curve. However, the dominance principle states that the combination is superior in the upper-left: the expected return is higher and the risk is lower, both preferable to a rational
investor. As can be seen in Figure 3.3:

Markowitz > SIM_{SS} > MGM > CCM_{SS} > SIM_{NSS} > CCM_{NSS}

Note that two different factors affect the ranking from each graphs. Figure 3.2 include a particular portfolio for each model, where Figure 3.3 measures a set of portfolios for each. Also, the former presents the relative position of the returns whereas the latter gives precise summary statistics $E(R_i)$ and $\sigma_i$. 

Figure 3.3: Performance in terms of efficient frontiers
CHAPTER 4

Resampling

4.1 Monte Carlo Simulation

The resampling is performed as an application of Monte Carlo Simulation. We estimate the mean and standard deviation of the monthly returns of each of the 30 stocks. And then we sample from a multivariate normal distribution with the estimated mean and standard deviation. This Monte Carlo simulation generates $T = 109$ monthly log returns.

The algorithm is comparable to resampling under Markowitz’s model described in Scherer’s paper. [3] The procedure is recapitulated below:

Step 1 Estimate $\hat{\beta}$ for single index model and variance-covariance matrix for constant correlation model and multigroup model. Also estimate the mean vector of historical returns.

Step 2 Resample from the returns by taking $T = 109$ draws from the return distribution. Estimate $\hat{\beta}$ and variance-covariance matrix from the sampled series.

Step 3 With the input derived in Step 2, trace out efficient frontiers by setting a range of risk free rates. Note that this is an increment of one dimension in the computation.
Step 4 After repeating Steps 2 and 3 many times \((n = 50\) in this paper), calculate the average portfolio weights for each return point.

4.1.1 Tracing Out the Efficient Frontier

One critical mission above is to find the efficient frontier. We attempt to find \(m = 25\) equally distributed return points along the frontier.

The task is performed by setting various risk free rates \(R_f\). Each \(R_f\) generates an optimal portfolio. By collecting many \(R_f\), we are able to trace out the efficient frontier between the minimum variance portfolio and the maximum return portfolio.

Note that the upper bound, or maximum return portfolios, only exists under the constraint that short sales are not allowed. Therefore, we apply the same maximum return portfolio for the same model, regardless of the constraint on short sales. Repeating the procedure for \(n = 50\) times, we obtain a new set of optimization inputs \((\Sigma_j, \mu_j), j = 1, \ldots, 50\). Each of the \(m = 25\) resampled allocation is defined as the average weight over \(n = 50\) simulated results:

\[
\bar{w}_{resampled} = \frac{1}{n} \sum_{i=1}^{n} w_{im}'
\]  

(4.1)

Compare to original optimal inputs, \((\Sigma_0, \mu_0)\), where \(\Sigma_j\) is a \(k \times k\) variance-covariance matrix and \(\mu_j\) is a \(k \times 1\) return vector.

4.1.2 The effect of risk free rate \(R_f\)

In the process of tracing out the efficient frontier, we set a wide range of \(R_f\). Graphically, the high risk free rate on the y-axis of the risk-return space \((\sigma = 0)\) initiate the capital allocation line (CAL), which is tangent to the optimized portfolio curve. The increase in \(R_f\) pushes the CAL upward from the left and
therefore restricts the options of asset allocation. This can be seen numerically
because the slope of CAL is exactly the Sharpe ratio \( SR_i = \frac{E(R_i) - R_f}{\sigma_i} \). Higher \( R_f \)
reduces possible \( E(R_i) \), and therefore less diversification. For single index model,
a potential problem of negative \( \beta \) may arise when the \( R_f \) is higher than \( E(R_i) \). On
the other hand, lowering \( R_f \) eventually allows CAL to pass through the leftmost
point on the curve. That means the frontier converges to the minimum variance
portfolio as \( R_f \to -\infty \). Since the actual \( R_f \) is at the record low\(^1\), the upper
bound of hypothetical \( R_f \) should not be too high.

4.2 Results

4.2.1 Smoother Transitions

The resampling results can be visualized from Figure 4.1 to 4.10, in comparison
to the original portfolio allocation of each models. Resampling creates smoother
transitions. Sudden shifts between each of the \( m = 25 \) samples are largely reduced
compared to the original models. The number of asset \( k = 30 \) is also large enough
to contribute to the smoothness even without resampling. The seemingly straight
lines in the resample increase the predictability due to smoothness. The technique
appears to perform well on all models.

4.2.2 Greater Diversification

The resampling technique also generates greater diversification along the efficient
frontier. Greater diversification generally implies lower risk. A mathematical
proof is provided in the appendix.

\(^1\)1-month Treasury Yield Curve Rate as of 05/20/13 is 0.01% according to the U.S. Depart-
ment of the Treasury.
Figure 4.1: Resampled Single Index Model: short sales allowed

Figure 4.2: Original Single Index Model: short sales allowed
Figure 4.3: Resampled Single Index Model: short sales not allowed

Figure 4.4: Original Single Index Model: short sales not allowed
Figure 4.5: Resampled Constant Correlation Model: short sales allowed

Figure 4.6: Original Constant Correlation Model: short sales allowed
Figure 4.7: Resampled Constant Correlation Model: short sales not allowed

Figure 4.8: Original Constant Correlation Model: short sales not allowed
Figure 4.9: Resampled Multigroup Model

Figure 4.10: Original Multigroup Model
Figure 4.11: \( m = 13^{th} \) portfolio allocation of resample and original methods. Single Index Model, short sales not allowed.

In particular, the 10\textsuperscript{th} and the 29\textsuperscript{th} stocks, or BRK-A and MFK, under original method have allocations of 0.1981 and 0.2161, respectively. The two stocks contribute to 41\% of the allocation. On the contrary, those are only 0.0789 and 0.0745 in the resample, albeit being the largest two allocation. This sign of less centralized distribution can also be visualized in Figure 4.11.

Table 4.1 also reveals the relationship between asset allocation and corresponding \( \hat{\beta} \), the slope coefficient of the regression. Generally, the assets that
enter the allocation are those with lower $\hat{\beta}$, mostly below 1. The 10th and the 29th appear to be the lowest two with $\hat{\beta} = 0.39$ and 0.49. Since $\hat{\beta}$ represents how a security responds to market forces, more stocks with lower $\hat{\beta}$ in the original allocation appear to be less influenced by market fluctuation. Resampling, on the other hand, allows more assets with higher beta and therefore less diversifiable risk.

### 4.2.3 Effect on Models

Table 4.2 compares the returns under different models over original and resampled technique. It is clear that resampling reduces the size of the intervals of for all five models. In other words, the resampling offers a less risky set of selection, regardless of the model. One reason might be the sampling error that feeds through the allocation. [3]

$\chi^2$ test is frequently used to determine whether or not the resampled and original portfolios are statistically equivalent. However, the normal assumption is violated when short sales are not allowed. The long-only constraint forces the allocation to be non-negative. Therefore, the distribution is truncated at 0.

As an alternative, we display the coefficient of variation $CV = \sigma/\mu$ in Table 4.3 to examine the dispersion of the allocation distribution. In particular, single index model disallowing short sales appears to have higher CVs, where 26 out of 30 allocations are above 1. CVs above one are rare in other models. Resampling seems to generate the largest dispersion effect here.
<table>
<thead>
<tr>
<th>Stock Code</th>
<th>Original</th>
<th>Resampled</th>
<th>Original $\hat{\beta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>XOM</td>
<td>0.1554</td>
<td>0.0705</td>
<td>0.53</td>
</tr>
<tr>
<td>CVX</td>
<td>0.0822</td>
<td>0.0382</td>
<td>0.71</td>
</tr>
<tr>
<td>COP</td>
<td>0</td>
<td>0.0309</td>
<td>0.99</td>
</tr>
<tr>
<td>WMT</td>
<td>0.1681</td>
<td>0.0638</td>
<td>0.42</td>
</tr>
<tr>
<td>KR</td>
<td>0.0389</td>
<td>0.0314</td>
<td>0.58</td>
</tr>
<tr>
<td>COST</td>
<td>0.0355</td>
<td>0.0524</td>
<td>0.76</td>
</tr>
<tr>
<td>GE</td>
<td>0</td>
<td>0.0176</td>
<td>1.36</td>
</tr>
<tr>
<td>CAT</td>
<td>0</td>
<td>0.0242</td>
<td>1.61</td>
</tr>
<tr>
<td>UTX</td>
<td>0.0127</td>
<td>0.0647</td>
<td>0.92</td>
</tr>
<tr>
<td>BRK-A</td>
<td>0.1981</td>
<td>0.0789</td>
<td>0.39</td>
</tr>
<tr>
<td>AIG</td>
<td>0</td>
<td>0.0015</td>
<td>2.86</td>
</tr>
<tr>
<td>ALL</td>
<td>0</td>
<td>0.0269</td>
<td>1.23</td>
</tr>
<tr>
<td>HPQ</td>
<td>0</td>
<td>0.0254</td>
<td>1.32</td>
</tr>
<tr>
<td>IBM</td>
<td>0</td>
<td>0.0369</td>
<td>1.03</td>
</tr>
<tr>
<td>DELL</td>
<td>0</td>
<td>0.0165</td>
<td>1.28</td>
</tr>
<tr>
<td>T</td>
<td>0</td>
<td>0.0425</td>
<td>0.93</td>
</tr>
<tr>
<td>VZ</td>
<td>0</td>
<td>0.0377</td>
<td>0.85</td>
</tr>
<tr>
<td>S</td>
<td>0</td>
<td>0.0068</td>
<td>1.38</td>
</tr>
<tr>
<td>BAC</td>
<td>0</td>
<td>0.0081</td>
<td>1.65</td>
</tr>
<tr>
<td>JPM</td>
<td>0</td>
<td>0.0239</td>
<td>1.35</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0.0064</td>
<td>2.14</td>
</tr>
<tr>
<td>UNH</td>
<td>0.0408</td>
<td>0.0260</td>
<td>0.68</td>
</tr>
<tr>
<td>WLP</td>
<td>0.0404</td>
<td>0.0279</td>
<td>0.71</td>
</tr>
<tr>
<td>AET</td>
<td>0.0037</td>
<td>0.0184</td>
<td>1.02</td>
</tr>
<tr>
<td>AAPL</td>
<td>0.0080</td>
<td>0.0259</td>
<td>1.40</td>
</tr>
<tr>
<td>MSFT</td>
<td>0</td>
<td>0.0347</td>
<td>1.06</td>
</tr>
<tr>
<td>ORCL</td>
<td>0</td>
<td>0.0236</td>
<td>1.12</td>
</tr>
<tr>
<td>PFE</td>
<td>0</td>
<td>0.0410</td>
<td>0.66</td>
</tr>
<tr>
<td>JNJ</td>
<td>0.2161</td>
<td>0.0745</td>
<td>0.49</td>
</tr>
<tr>
<td>MRK</td>
<td>0</td>
<td>0.0228</td>
<td>0.74</td>
</tr>
</tbody>
</table>

Table 4.1: $m = 13^{th}$ portfolio allocation under Single Index Model where short sales are not allowed: original, resampled, and their corresponding $\hat{\beta}$,
<table>
<thead>
<tr>
<th></th>
<th>Original</th>
<th>Resampled</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIM-SS</td>
<td>(3.89, 8.44)</td>
<td>(5.37, 6.16)</td>
</tr>
<tr>
<td>SIM-NSS</td>
<td>(4.38, 6.68)</td>
<td>(5.48, 6.52)</td>
</tr>
<tr>
<td>CCM-SS</td>
<td>(3.98, 9.65)</td>
<td>(5.37, 6.30)</td>
</tr>
<tr>
<td>CCM-NSS</td>
<td>(4.83, 8.15)</td>
<td>(5.38, 6.40)</td>
</tr>
<tr>
<td>MGM-SS</td>
<td>(3.91, 8.65)</td>
<td>(5.36, 6.16)</td>
</tr>
</tbody>
</table>

Table 4.2: Intervals of Returns
<table>
<thead>
<tr>
<th>Stock Code</th>
<th>SIM-SS</th>
<th>SIM-NSS</th>
<th>CCM-SS</th>
<th>CCM-NSS</th>
<th>MGM-SS</th>
</tr>
</thead>
<tbody>
<tr>
<td>XOM</td>
<td>0.1454</td>
<td>0.9318</td>
<td>0.1669</td>
<td>0.1690</td>
<td>0.2757</td>
</tr>
<tr>
<td>CVX</td>
<td>0.1877</td>
<td>1.3839</td>
<td>0.1805</td>
<td>0.1950</td>
<td>0.2491</td>
</tr>
<tr>
<td>COP</td>
<td>0.2192</td>
<td>1.2269</td>
<td>0.1797</td>
<td>0.2149</td>
<td>0.3785</td>
</tr>
<tr>
<td>WMT</td>
<td>0.1824</td>
<td>1.0738</td>
<td>0.1940</td>
<td>0.1732</td>
<td>0.2128</td>
</tr>
<tr>
<td>KR</td>
<td>0.2257</td>
<td>1.2027</td>
<td>0.2039</td>
<td>0.1639</td>
<td>0.3441</td>
</tr>
<tr>
<td>COST</td>
<td>0.1611</td>
<td>0.9201</td>
<td>0.1792</td>
<td>0.1642</td>
<td>0.2131</td>
</tr>
<tr>
<td>GE</td>
<td>0.2225</td>
<td>1.3866</td>
<td>0.2051</td>
<td>0.3520</td>
<td>0.4508</td>
</tr>
<tr>
<td>CAT</td>
<td>0.1611</td>
<td>1.2095</td>
<td>0.1798</td>
<td>0.1850</td>
<td>0.4043</td>
</tr>
<tr>
<td>UTX</td>
<td>0.1635</td>
<td>1.1354</td>
<td>0.1743</td>
<td>0.1819</td>
<td>0.2544</td>
</tr>
<tr>
<td>BRK-A</td>
<td>0.1683</td>
<td>0.8660</td>
<td>0.1597</td>
<td>0.1310</td>
<td>0.3107</td>
</tr>
<tr>
<td>AIG</td>
<td>0.5497</td>
<td>1.4258</td>
<td>0.8125</td>
<td>1.0683</td>
<td>1.3623</td>
</tr>
<tr>
<td>ALL</td>
<td>0.2528</td>
<td>1.0615</td>
<td>0.2153</td>
<td>0.2713</td>
<td>0.4309</td>
</tr>
<tr>
<td>HPQ</td>
<td>0.1999</td>
<td>1.0444</td>
<td>0.2224</td>
<td>0.2948</td>
<td>0.3275</td>
</tr>
<tr>
<td>IBM</td>
<td>0.2069</td>
<td>1.1130</td>
<td>0.1553</td>
<td>0.2832</td>
<td>0.3075</td>
</tr>
<tr>
<td>DELL</td>
<td>0.2603</td>
<td>1.2277</td>
<td>0.2458</td>
<td>0.3228</td>
<td>0.4587</td>
</tr>
<tr>
<td>T</td>
<td>0.2140</td>
<td>0.9781</td>
<td>0.2154</td>
<td>0.1968</td>
<td>0.3280</td>
</tr>
<tr>
<td>VZ</td>
<td>0.2047</td>
<td>1.0603</td>
<td>0.2016</td>
<td>0.1913</td>
<td>0.4089</td>
</tr>
<tr>
<td>S</td>
<td>0.3270</td>
<td>1.3021</td>
<td>0.4184</td>
<td>0.4143</td>
<td>0.7648</td>
</tr>
<tr>
<td>BAC</td>
<td>0.2868</td>
<td>1.3775</td>
<td>0.3321</td>
<td>0.2786</td>
<td>0.6138</td>
</tr>
<tr>
<td>JPM</td>
<td>0.2134</td>
<td>1.0513</td>
<td>0.2460</td>
<td>0.2069</td>
<td>0.3988</td>
</tr>
<tr>
<td>C</td>
<td>0.4418</td>
<td>1.3773</td>
<td>0.4932</td>
<td>0.4121</td>
<td>0.8165</td>
</tr>
<tr>
<td>UNH</td>
<td>0.2171</td>
<td>1.2103</td>
<td>0.1721</td>
<td>0.2126</td>
<td>0.3069</td>
</tr>
<tr>
<td>WLP</td>
<td>0.2197</td>
<td>1.1356</td>
<td>0.1700</td>
<td>0.2530</td>
<td>0.3151</td>
</tr>
<tr>
<td>AET</td>
<td>0.2212</td>
<td>1.3060</td>
<td>0.1846</td>
<td>0.2206</td>
<td>0.3489</td>
</tr>
<tr>
<td>AAPL</td>
<td>0.1611</td>
<td>0.9854</td>
<td>0.1757</td>
<td>0.2011</td>
<td>0.3855</td>
</tr>
<tr>
<td>MSFT</td>
<td>0.2570</td>
<td>1.1368</td>
<td>0.1997</td>
<td>0.1819</td>
<td>0.3970</td>
</tr>
<tr>
<td>ORCL</td>
<td>0.2462</td>
<td>1.0329</td>
<td>0.2314</td>
<td>0.1835</td>
<td>0.4077</td>
</tr>
<tr>
<td>PFE</td>
<td>0.2480</td>
<td>1.1554</td>
<td>0.1882</td>
<td>0.2467</td>
<td>0.3037</td>
</tr>
<tr>
<td>JNJ</td>
<td>0.1426</td>
<td>1.2103</td>
<td>0.1607</td>
<td>0.1649</td>
<td>0.2116</td>
</tr>
<tr>
<td>MRK</td>
<td>0.2195</td>
<td>1.3493</td>
<td>0.2813</td>
<td>0.2750</td>
<td>0.3544</td>
</tr>
</tbody>
</table>

Table 4.3: Coefficients of Variation
CHAPTER 5

Conclusion

We conduct the analysis of resampling on single index model, constant correlation model, and multigroup model. The simulation-optimization process is established by Monte Carlo simulation and tracing out efficient frontiers on a series of hypothetical risk-free rates $R_f$.

The resampling results appear to be preferable in two ways: smoother transitions and greater diversification. Those are consistent among all models, including Markowitz’s model that is discussed in Sherer’s paper (2002). The smoothness from resampling may improve the predictability of each model. Also, error-maximization can be rectified due to a wider range of asset selection. As another benefit, the time required for resampling computation is effectively reduced due to the simplification nature of the three models.

Nevertheless, the performance of the resampled allocation seems to be limited compared to the original portfolio selection. This can be seen by a smaller range of return intervals across all models. Besides, it remains unclear why resampling has greater dispersion particularly on single index model when short sales are not allowed. Different regression technique may be applied in order to deal with the truncated distribution.

A potential improvement can be the assumption on the constant risk free rate $R_f$. In reality, $R_f$ changes over time. A constant risk free rate can be a strong assumption as the data spans over a long period of time. However, that requires
$R_f$ to be a random variable or a function of time. In both cases, the complexity of models increases. Another direction of further study can be cross-validation. Cross-validation takes advantage of different sets of observation and are capable of verifying the forecast.

Finally, the stocks with more data available are companies that operate better in nature, or it would have been out of the market. A sample of well-perform stocks might result in selection bias. A careful examination of methodology is required. Different samples of data and industries can also be used. Many ascending firms from, for example, technology industries are excluded from the analysis. The inclusion of them can be closer to the real picture of the latest market. However, the number of observation $T$ may be too short to generate convincing result.
In modern portfolio theory, the variance of a portfolio with $N$ assets is given by

$$\sigma_p^2 = \sum_{i=1}^{N} w_i^2 \sigma_i^2 + 2 \sum_{i=1}^{N} \sum_{j > 1}^{N} Cov(R_i, R_j)$$

For simplicity, equal weight is assumed here. Note that

$$\frac{2}{N} = \frac{N - 1}{N} \times \frac{1}{N(N - 1)/2}$$

That allows

$$\sigma_p^2 = \frac{1}{N} \bar{\sigma}^2 + (1 - \frac{1}{N}) \left[ \frac{1}{N(N - 1)/2} \sum_{i=1}^{N} \sum_{j > 1}^{N} Cov(R_i, R_j) \right]$$

$$= \frac{1}{N} \bar{\sigma}^2 + (1 - \frac{1}{N}) Cov(R_i, R_j)$$

As a result,

$$\lim_{N \to \infty} \sigma_p^2 = Cov(R_i, R_j)$$

$\bar{\sigma}^2$ is eliminated when $N$ is large. The risk is therefore diversified.
REFERENCES


