Index Tracking on S&P500 with Weighted Average Value at Risk and Robust Optimization

Seungmoon Rieh

Division of Engineering Science: Mathematics, Statistics, and Finance, University of Toronto, 40 Saint George Street, Toronto, Ontario, Canada M5S 2E4

Abstract

Index tracking aims to replicate an Index fund by constructing a portfolio with a subset of stocks from that Index. For this paper, the financial risk of portfolio is quantified by the Weighted Average Value at Risk. Tracking error is defined as the closeness of portfolio risk to Index risk, namely the relative difference in the Weighted Average Value at Risk. A mathematical programming computes portfolio weights, which is tested against real financial data. Portfolio return and risk are compared against those of SPDR Exchange Traded Fund.

1. Introduction

An Index tracking refers to constructing a portfolio which replicates a certain Index fund. It is an example of passive portfolio management strategy. The biggest challenge Index tracking faces is tracking error, which arises due to stratification. In other words, because Index tracking only utilizes a subset of stocks that constitutes an Index fund, not all the stocks, tracking error is unavoidable. However, tracking error should be minimized if a decision maker wants to closely replicate the risk profile of Index fund. Also, the performance of portfolio should be considered in addition to tracking error. In short, there are two things to consider in index tracking.

1. Tracking performance: How does my portfolio perform compared to Index?
2. Tracking error: How close is portfolio risk to Index risk? Is it higher or lower?

In optimization framework, two possible approaches exist

1. Minimize tracking error while tracking performance achieves a certain percentage

$$\min\ (Tracking\ error)$$
$$s.t.\ Tracking\ performance \geq R$$

2. Maximize tracking performance while tracking error is bounded by certain percentage

$$\max\ (Tracking\ performance)$$
$$s.t.\ Tracking\ error \leq K$$

The tracking performance can be defined as

$$Tracking\ performance\ =\ Return\ of\ portfolio - Return\ of\ Index$$

The definition of tracking error varies between decision makers. For this paper, tracking error is defined as,

$$Tracking\ error\ =\ \left|\frac{Risk\ Profile_{portfolio} - Risk\ Profile_{index}}{Risk\ Profile_{index}}\right|$$

To find alternative definitions of tracking error used in other literature, please refer to Appendix D.

2. Stock Stratification

Stocks need to be stratified such that the number of selected stocks is not too large and they are well diversified enough to represent the market. Since SPDR, which is an ETF tracking S&P500, currently has around 240 stocks, selecting less than 240 stocks would make portfolio competitive against SPDR. The best way to guarantee good diversification amongst stocks is to pool them from all industry sectors in S&P500.

For this paper, the number of stocks belonging to each industry sector is calculated and about 45% of stocks in each industry sector are selected to be in the portfolio. Market capitalization is used to order importance of stocks in each industry sector.

3. Optimization Framework

A portfolio may be constructed with different objectives. It may be risk taking, risk neutral or risk averse. Or, it may consist of sub portfolios with different attitudes toward risk. For this paper, the underlying assumption is that investors are risk averse.

This paper attempts to analyse what happens at negative tail of portfolio return distribution. It is because risk-averse investors are most concerned about low probability and high impact events that can cripple their investment. Long term investors who are less concerned about extreme events thinking that their investment would bounce back eventually may not have too much interest in tail analysis. However, for short term investors, tail events would certainly be big concern.

Modelling risk is most significant part of portfolio optimization. For this paper, the Weighted Average Value at Risk is used as risk measure. For detailed discussion on different risk measures, please refer to Appendix C.

3.1. Choice of risk-aversion function

A utility function must satisfy specific properties in order to be a risk averse function. Risk averse functions also need to satisfy a set of axioms to make Weight Average Value at Risk coherent. For detailed discussion of utility functions, please refer to Appendix C.

An exponential risk aversion function is chosen to assign weights to different tail events. The strength of exponential function is that risk aversion coefficient can be chosen by the decision maker. For example, higher coefficient leads to more risk averseness.

Exponential function has the form,

$$U(x) = \gamma e^{-\delta x}$$

A specific form of utility function is chosen to satisfy three properties which makes the Weighted Average Value at Risk coherent.

1. Nominator is a positive constant multiplied by an exponential function, which is positive. Denominator is also positive. Thus, weights are always positive.

$$\emptyset(p) = \frac{ae^{a(1-p)}}{e^a - 1}, a > 0, p \in [0,1]\ interval\ at\ tail\ of\ distribution$$

2.

$$\int_0^1 \frac{ae^{a(1-p)}}{e^a - 1} dp = \int_0^1 \frac{ae^a e^{-ap}}{e^a - 1} dp = \frac{ae^a}{e^a - 1} \int_0^1 e^{-ap} dp = \frac{ae^a}{e^a - 1}\left[-\frac{1}{a} e^{-ap}\right]_0^1$$
$$= \frac{ae^a}{e^a - 1}\left[-\frac{1}{a}(e^{-a} - e^0)\right] = \frac{-e^a}{e^a - 1}[(e^{-a} - e^0)] = \frac{-e^a}{e^a - 1}[(e^{-a} - 1)]$$
$$= \frac{-1 + e^a}{e^a - 1} = 1$$

3. Clearly, derivative is less than or equal to zero

$$\emptyset'(p) = \frac{-a^2 e^{a(1-p)}}{e^a - 1}, a > 0$$

3.2. Weighted Average Value at Risk

The Weighted Average Value at Risk can be defined in the following way,

$$wAVaR_{\emptyset} = \int_{\emptyset}\int_p X^T r \emptyset'(p) dp\, d\emptyset$$

$$where\ R_{\emptyset}\ is\ threshold\ for\ each\ ordered\ statistic$$

$$wAVaR_{\emptyset} = \int_{\emptyset} X^T r \int_p \frac{ae^{a(1-p)}}{e^a - 1} dp\, d\emptyset, a > 0$$

An example is shown to illustrate how the chosen exponential function assigns different weights to different realizations at tail of distribution.

$$Assume\ a = 1$$

$$There\ are\ five\ observations\ at\ tail\ \ R_{0.2} > R_{0.4} > R_{0.6} > R_{0.8} > R_{1.0}$$

$$\emptyset_{0.2} = \emptyset(0) - \emptyset(0.2) = \frac{e^{1-0}}{e^1 - 1} - \frac{e^{1-0.2}}{e^1 - 1} = 1.582 - 1.295 = 0.287$$

$$\emptyset_{0.4} = \emptyset(0.2) - \emptyset(0.4) = \frac{e^{1-0.2}}{e^1 - 1} - \frac{e^{1-0.4}}{e^1 - 1} = 1.295 - 1.06 = 0.235$$
$$\emptyset_{0.6} = \emptyset(0.4) - \emptyset(0.6) = \frac{e^{1-0.4}}{e^1 - 1} - \frac{e^{1-0.6}}{e^1 - 1} = 1.06 - 0.868 = 0.192$$
$$\emptyset_{0.8} = \emptyset(0.6) - \emptyset(0.8) = \frac{e^{1-0.6}}{e^1 - 1} - \frac{e^{1-0.8}}{e^1 - 1} = 0.868 - 0.711 = 0.157$$
$$\emptyset_{1.0} = \emptyset(0.8) - \emptyset(1.0) = \frac{e^{1-0.8}}{e^1 - 1} - \frac{e^{1-1.0}}{e^1 - 1} = 0.711 - 0.582 = 0.129$$

$$wAVaR_{\emptyset} = \emptyset_{0.2}R_{0.2} + \emptyset_{0.4}R_{0.4} + \emptyset_{0.6}R_{0.6} + \emptyset_{0.8}R_{0.8} + \emptyset_{1.0}R_{1.0}$$

3.3. Linearizing risk measure

Rockafellar, T and Uryasev S. [27] came up with a method to linearly approximate the Average Value at Risk. The same technique with slight modification can be applied to approximate Weighted Average Value at Risk as well.

$$wAVaR_{\emptyset} = \int_{\emptyset}\int_{p} X^T r \emptyset'(p) dp \, d\emptyset$$

$$wAVaR_{\emptyset.\alpha} = \int_{\emptyset}\int_{p} R_{\emptyset}\emptyset'(p) dp \, d\emptyset - \int_{\emptyset}\int_{p} [R_{\emptyset} - X^T r]^{+} \emptyset'(p) dp \, d\emptyset$$

$$wAVaR_{\emptyset,\alpha} = \sum_{p=1}^{j} R_{\emptyset}\emptyset_p - \sum_{p=1}^{j} [R_{\emptyset} - X^T r]^{+} \emptyset_p$$

It is very important that the weights must be applied after ordering observations at tail.

3.4. Optimization framework

Stochastic programming model can be written in the following way,

$$\max \boldsymbol{x}^T \boldsymbol{r}$$

$$\left| \frac{\mathrm{wAVaR}_{\alpha}^{\mathbf{x}} - \mathrm{wAVaR}_{\mathrm{a}}^{\mathbf{X}}}{\mathrm{wAVaR}_{\mathrm{a}}^{\mathbf{X}}} \right| \leq K$$

$$\boldsymbol{x}^T \mathbf{1} = 1$$

$$\boldsymbol{ub} \leq \boldsymbol{x} < \boldsymbol{lb}$$

$$\boldsymbol{x} = weights\ for\ portfolio$$
$$\boldsymbol{X} = weights\ for\ Index$$
$$r_i = expected\ return\ for\ stock\ i$$
$$r_j = vector\ of\ stock\ returns\ for\ j^{th}\ simulation$$

Uncertainty in vector of stock returns in the objective function can be dealt with by introducing a new variable z,

$$\max \boldsymbol{x}^T \boldsymbol{r}$$

$$\max \sum_{i=1}^{n} x_i r_i$$

$$\min(- \sum_{i=1}^{n} x_i r_i)$$

$$\min z$$
$$z + \sum_{i=1}^{n} x_i r_i \geq 0$$

Tracking error can be written using the definition introduced earlier,

$$\left| \frac{\mathrm{wAVaR}^{\mathbf{x}} - \mathrm{wAVaR}^{\mathbf{X}}}{\mathrm{wAVaR}^{\mathbf{X}}} \right| \leq K$$

$\mathrm{wAVaR}^{\mathbf{X}}$ can be estimated using Monte Carlo simulation. For example, one data point for $\mathrm{wAVaR}^{\mathbf{X}}$ is computed by multiplying simulated return vector by the weights. A large number of data points can be computed, which leads to a distribution. Once a distribution is created, a confidence intervals can be calculated to provide both point forecast and range estimate of $\mathrm{wAVaR}^{\mathbf{X}}$.

Incorporating ideas from Rockafellar, T and Uryasev S. [27] the Weighted Average Value at Risk is written as,

$$\left| \frac{\sum_{j=1}^{J} \emptyset_j R_j - \sum_{j=1}^{J} \left[\emptyset_j R_j - \emptyset_j \boldsymbol{x}^T \boldsymbol{r}_j \right]^+ - \mathrm{wAVaR}^{\mathbf{X}}}{\mathrm{wAVaR}^{\mathbf{X}}} \right| \leq K$$

Wang, M., Xu, C., Xu, F., and Xue, H. [35] came up with a method to rewrite the constraint for the Average Value at Risk. A similar technique is used to rewrite above expression. However, unlike Average Value at Risk where portfolio returns are unordered statistics, here the realization at the tail are ordered statistics. Thus the model becomes,

$$\left| \frac{\sum_{j=1}^{J} \emptyset_j R_j - \sum_{j=1}^{J} p_j - \mathrm{wAVaR}^{\mathbf{X}}}{\mathrm{wAVaR}^{\mathbf{X}}} \right| \leq K$$
$$p_j \geq \ \emptyset_j R_j - \emptyset_j \boldsymbol{x}^T \boldsymbol{r}_j \ \forall j$$
$$\boldsymbol{x}^T \boldsymbol{r}_j \geq R_{j+1} \ \forall j - 1$$

$$R_j \leq 0.8 * R_{j+1}\, \forall j-1$$
$$-c \leq R_1 \leq c\, \forall j$$
$$p_j \geq 0\, \forall j$$
$$R_1 = value\ at\ risk$$

The following two constraints will force observations at tail to be ordered statistics.

$$\boldsymbol{x}^T \boldsymbol{r}_j \geq R_{j+1}\, \forall j-1$$
$$R_j \leq 0.8 * R_{j+1}\, \forall j-1$$

For example, $R_1 \geq x^T r_1$ and $x^T r_1 \geq R_2$ establishes that $R_1 \geq R_2$. Moreover, $R_2 \geq x^T r_2$ leads to $x^T r_1 \geq R_2 \geq x^T r_2$. Therefore, $x^T r_j$ becomes ordered statistics. The constraint $R_j \leq 0.8 * R_{j+1}$ prevents drastic decrease of threshold value, which may lead to infeasible solutions.

The absolute values can be written as two different constraints.

$$\frac{\sum_{j=1}^{J} \emptyset_j R_j - \sum_{j=1}^{J} p_j - \mathrm{wAVaR}_{\mathrm{a}}^{\mathbf{x}}}{\mathrm{wAVaR}_{\mathrm{a}}^{\mathbf{x}}} \leq K$$
$$\frac{\sum_{j=1}^{J} \emptyset_j R_j - \sum_{j=1}^{J} p_j - \mathrm{wAVaR}_{\mathrm{a}}^{\mathbf{x}}}{\mathrm{wAVaR}_{\mathrm{a}}^{\mathbf{x}}} \geq -\mathrm{K}$$
$$p_j \geq \emptyset_j \boldsymbol{x}^T \boldsymbol{r}_j - \emptyset_j R_j\, \forall j$$
$$\boldsymbol{x}^T \boldsymbol{r}_j > R_{j+1}\, \forall j-1$$
$$R_j \leq 0.8 * R_{j+1}\, \forall j-1$$
$$-c \leq R_1 \leq c\, \forall j$$
$$p_j \geq 0\, \forall j$$

Above constraints can be reordered as,

$$\sum_{j=1}^{J} \emptyset_j R_j - \sum_{j=1}^{J} p_j \leq (K+1)\mathrm{wAVaR}_{\mathrm{a}}^{\mathbf{x}}$$
$$\sum_{j=1}^{J} \emptyset_j R_j - \sum_{j=1}^{J} p_j \geq (-K+1)\mathrm{wAVaR}_{\mathrm{a}}^{\mathbf{x}}$$
$$p_j \geq \emptyset_j \boldsymbol{x}^T \boldsymbol{r}_j - \emptyset_j R_j\, \forall j$$
$$\boldsymbol{x}^T \boldsymbol{r}_j \geq R_{j+1}\, \forall j-1$$
$$R_j \leq 0.8 * R_{j+1}\, \forall j-1$$
$$-c \leq R_1 \leq c\, \forall j$$
$$p_j \geq 0\, \forall j$$

The presence of uncertainties in constraints should be taken care of. The uncertainty in estimated Weighted Average Value at Risk can be dealt with by introducing a new variable b,

$$\sum_{j=1}^{J} \emptyset_j R_j - \sum_{j=1}^{J} p_j \le (K+1)\mathrm{wAVaR}_{\mathrm{a}}^{\mathbf{X}}$$
$$\sum_{j=1}^{J} \emptyset_j R_j - \sum_{j=1}^{J} p_j \ge (-K+1)\mathrm{wAVaR}_{\mathrm{a}}^{\mathbf{X}}$$

$$\sum_{j=1}^{J} \emptyset_j R_j - \sum_{j=1}^{J} p_j \le (K+1)(\overline{\mathrm{wAVaR}_{\mathrm{a}}^{\mathbf{X}}} - b)$$
$$\sum_{j=1}^{J} \emptyset_j R_j - \sum_{j=1}^{J} p_j \ge (-K+1)(\overline{\mathrm{wAVaR}_{\mathrm{a}}^{\mathbf{X}}} - b)$$
$$-l \le b \le l$$

Consider these two constraints. The situation is not as simple as previous case where uncertainty involved only one constant. Here, uncertainty lies on the vector of stock returns which also happens to be coefficients of decision variable x. Bertsimas, D. and Thiele, A. [3] introduced techniques for dealing with such situation.

$$p_j \ge \emptyset_j \boldsymbol{x}^T \boldsymbol{r}_j - \emptyset_j R_j \ \forall j$$
$$\boldsymbol{x}^T \boldsymbol{r}_j \ge R_{j+1} \ \forall j - 1$$

Above expressions are reordered such that all uncertainties are on the left side.

$$\emptyset_j \boldsymbol{x}^T \boldsymbol{r}_j \le p_j + \emptyset_j R_j \ \forall j$$
$$-\boldsymbol{x}^T \boldsymbol{r}_j \le R_{j+1} \ \ \forall j - 1$$

Rewrite constraints with coefficient under uncertainty in A

$$\min Ax \le B$$

Assume that decision maker knows the range of values to forecast

$$[\bar{r}_{ij} - \widehat{r_{ij}}, \bar{r}_{ij} + \widehat{r_{ij}}]$$
$$z_{ij} = \frac{r_{ij} - \bar{r}_{ij}}{\widehat{r_{ij}}}$$
$$z_{ij} \ \in [-1,1]$$
$$\sum_{i=1}^{n} |z_{ij}| \le \ \Gamma_j, \forall j$$
$$\Gamma_{\mathrm{j}} \ \in [0, n]$$

Γ_j is accumulated difference between the point forecast and the realized value. Intuitively, some realization will be above estimate and some will be below estimate so they will cancel out. Thus, there is no need to assume extremely large number for Γ_j

Then,

$$Ax + \min \sum_{i=1}^{n} \widehat{r_{ij}}\, x_i z_{ij} \le B$$
$$\min \sum_{i=1}^{n} \widehat{r_{ij}}\, x_i z_{ij} =$$
$$-\max \sum_{i=1}^{n} \widehat{r_{ij}}\, |x_i| z_{ij}$$
$$\sum_{i=1}^{n} z_{ij} \le \Gamma_j \;\; \forall j$$
$$0 \le z_{ij} \le 1 \; \forall i$$

$$Ax - \Gamma_j \varphi_j - \sum_{i=1}^{n} \omega_{ij} \le B \; \forall j$$
$$\varphi_j + \omega_{ij} \ge \widehat{r_{ij}} y_j \; \forall i,j$$
$$-y_j \le x_i \le y_j \; \forall i$$
$$\varphi_j \ge 0 \; \forall j$$
$$\omega_{ij} \ge 0 \; \forall i,j$$

$$\emptyset_j \boldsymbol{x}^T \overline{\boldsymbol{r}_j} - \Gamma_j \varphi_j - \sum_{i=1}^{n} \omega_{ij} \le p_j + \emptyset_j R_j \; \forall j$$
$$\boldsymbol{x}^T \overline{\boldsymbol{r}_j} - \Gamma_j \varphi_j - \sum_{i=1}^{n} \omega_{ij} \le R_{j+1} \; \forall j - 1$$
$$\varphi_j + \omega_{ij} \ge \widehat{r_{ij}} y_j \; \forall i,j$$
$$-y_j \le x_i \le y_j \; \forall i$$
$$\varphi_j \ge 0 \; \forall j$$
$$\omega_{ij} \ge 0 \; \forall i,j$$

Simulating $\overline{\boldsymbol{r}_j}$ and $\widehat{r_{ij}}$ are very important. They represent stock returns at the tail. Also, R_1 needs to be carefully chosen, which represents the Value at Risk.

After putting everything together the model becomes,

$$\min_{x_i, p_j, \omega_{ij}, \varphi_j, y_j, a, b, z} z$$
$$z + \sum_{i=1}^{n} x_i r_i \geq 0$$
$$\sum_{j=1}^{J} \emptyset_j R_j - \sum_{j=1}^{J} p_j \leq (K+1)(\overline{\mathrm{wAVaR}_{\mathrm{a}}^{\mathbf{X}}} - b)$$
$$\sum_{j=1}^{J} \emptyset_j R_j - \sum_{j=1}^{J} p_j \geq (-K+1)(\overline{\mathrm{wAVaR}_{\mathrm{a}}^{\mathbf{X}}} - b)$$
$$-l \leq b \leq l$$
$$\emptyset_j \boldsymbol{x}^T \overline{\boldsymbol{r}_J} - \Gamma_j \varphi_j - \sum_{i=1}^{n} \omega_{ij} \leq p_j + \emptyset_j R_j \ \forall j$$
$$\boldsymbol{x}^T \overline{\boldsymbol{r}_J} - \Gamma_j \varphi_j - \sum_{i=1}^{n} \omega_{ij} \leq R_{j+1} \ \forall j - 1$$
$$R_j \leq 0.8 * R_{j+1} \ \forall j - 1$$
$$\varphi_j + \omega_{ij} \geq \widehat{r_{ij}} y_j \ \forall i, j$$
$$-y_j \leq x_i \leq y_j \ \forall i$$
$$\varphi_j \geq 0 \ \forall j$$
$$\omega_{ij} \geq 0 \ \forall i, j$$
$$-c \leq R_1 \leq c$$
$$p_j \geq 0 \ \forall j$$
$$\boldsymbol{x}^T \boldsymbol{1} = 1$$
$$\boldsymbol{ub} \leq \boldsymbol{x} < \boldsymbol{lb}$$

Before converting above model to linear equations, the decision variables and constants need to be separated.

$$\min_{x_i, p_j, \omega_{ij}, \varphi_j, y_j, R, a, b, z} z$$
$$-z - \sum_{i=1}^{n} x_i r_i \leq 0$$
$$(K+1)b - \sum_{j=1}^{J} p_j + \sum_{j=1}^{J} \emptyset_j R_j \leq (K+1)\overline{\mathrm{wAVaR}_{\mathrm{a}}^{\mathbf{X}}}$$
$$(K-1)b + \sum_{j=1}^{J} p_j - \sum_{j=1}^{J} \emptyset_j R_j \leq (K-1)\overline{\mathrm{wAVaR}_{\mathrm{a}}^{\mathbf{X}}}$$
$$-l \leq b \leq l$$
$$\emptyset_j \boldsymbol{x}^T \overline{\boldsymbol{r}_J} - \Gamma_j \varphi_j - \sum_{i=1}^{n} \omega_{ij} - p_j - \emptyset_j R_j \leq 0 \ \forall j$$

$$\boldsymbol{x}^T\overline{\boldsymbol{r}_j} - \Gamma_j\varphi_j - \sum_{i=1}^{n}\omega_{ij} + R_{j+1} \leq 0\ \forall j-1$$
$$R_j - 0.8 * R_{j+1} \leq 0\ \forall j-1$$
$$-\varphi_j - \omega_{ij} + \widehat{r_{ij}}y_i \leq 0\ \forall i,j$$
$$x_i - y_i \leq 0\ \forall i$$
$$-x_i - y_i \leq 0\ \forall i$$
$$\varphi_j \geq 0\ \forall j$$
$$\omega_{ij} \geq 0\ \forall i,j$$
$$-c \leq R_1 \leq c$$
$$p_j \geq 0\ \forall j$$
$$\boldsymbol{x}^T\mathbf{1} = 1$$
$$\boldsymbol{ub} \leq \boldsymbol{x} < \boldsymbol{lb}$$

Above model can be rewritten as following linear equations,

$$min\ \boldsymbol{x^T0} + \boldsymbol{p^T0} + \boldsymbol{\omega^T0} + \boldsymbol{\varphi}^T\mathbf{0} + \boldsymbol{y^T0} + \boldsymbol{R^T0} + 0b + 1z$$
$$-\boldsymbol{x^Tr} + \boldsymbol{p^T0} + \boldsymbol{\omega^T0} + \boldsymbol{\varphi^T0} + \boldsymbol{y^T0} + \boldsymbol{R^T0} + 0b - 1z \leq 0$$
$$\boldsymbol{x^T0} - \boldsymbol{p^T1} + \boldsymbol{\omega^T0} + \boldsymbol{\varphi^T0} + \boldsymbol{y^T0} + \boldsymbol{R^Tw} + (K+1)b + 0z \leq (K+1)\overline{\mathrm{wAVaR}_{\mathrm{a}}^{\mathbf{X}}}$$
$$\boldsymbol{x^T0} + \boldsymbol{p^T1} + \boldsymbol{\omega^T0} + \boldsymbol{\varphi^T0} + \boldsymbol{y^T0} - \boldsymbol{R^Tw} + (K-1)b + 0z \leq (K-1)\overline{\mathrm{wAVaR}_{\mathrm{a}}^{\mathbf{X}}}$$
$$(\boldsymbol{wx})^{\boldsymbol{T}}r_{\boldsymbol{j}} - \boldsymbol{p^T1} - \boldsymbol{\omega^T1} - \boldsymbol{\varphi}^T\boldsymbol{\Gamma_j} + \boldsymbol{y^T0} - \boldsymbol{R^Tw_j} + 0b + 0z \leq 0\ \forall j$$
$$\boldsymbol{x^T}r_{\boldsymbol{j}} - \boldsymbol{p^T0} - \boldsymbol{\omega^T1} - \boldsymbol{\varphi}^T\boldsymbol{\Gamma_j} + \boldsymbol{y^T0} + R_1 0 + \boldsymbol{R_{+1}^T1} + 0b + 0z \leq 0\ \forall j-1$$
$$\boldsymbol{x^T}0 - \boldsymbol{p^T0} - \boldsymbol{\omega^T0} - \boldsymbol{\varphi}^T\mathbf{0} + \boldsymbol{y^T0} + R_j 1 - R_{j+1}0.8 + \boldsymbol{R}_{!\boldsymbol{j},!\boldsymbol{j+1}}0 + 0b + 0z \leq 0\ \forall j-1$$
$$\boldsymbol{x^T0} + \boldsymbol{p^T0} - \boldsymbol{\omega^T1} - \boldsymbol{\varphi}^T\mathbf{1} + \boldsymbol{y^T}\widehat{\boldsymbol{r_{ij}}} + \boldsymbol{R^T0} + 0b + 0z \leq 0\ \forall i,j$$
$$\boldsymbol{x^T1} + \boldsymbol{p^T0} + \boldsymbol{\omega^T0} + \boldsymbol{\varphi}^T\mathbf{0} - \boldsymbol{y^T1} + \boldsymbol{R^T0} + 0b + 0z \leq 0\ \forall i$$
$$-\boldsymbol{x^T1} + \boldsymbol{p^T0} + \boldsymbol{\omega^T0} + \boldsymbol{\varphi}^T\mathbf{0} - \boldsymbol{y^T1} + \boldsymbol{R^T0} + 0b + 0z \leq 0\ \forall i$$
$$lb = [\mathbf{0.0002}, \mathbf{0}, \mathbf{0}, \mathbf{0}, -\boldsymbol{Inf}, (c, -\boldsymbol{Inf}), -l, -Inf]^T$$
$$ub = [\mathbf{0.02}, \boldsymbol{Inf}, \boldsymbol{Inf}, \boldsymbol{Inf}, \boldsymbol{Inf}, (c, \boldsymbol{Inf}), l, Inf]^T$$
$$\boldsymbol{x^T1} + \boldsymbol{p^T0} + \boldsymbol{\omega^T0} + \boldsymbol{\varphi}^T\mathbf{0} + \boldsymbol{y^T0} + \boldsymbol{R^T0} + 0b + 0z = 1$$

3.5. Distribution

Ma, T. and Serota, R.A. [18] has studied distributions that best model stock return and volatility from S&P500. Central findings from their investigation was that the Generalized Inverse Gamma function multiplied by the Standard Normal variables, which turned out to be generalized Student's t-distribution, produced best fit for stock returns from S&P500. Therefore, for this paper, Student's t numbers are used to generate random returns for individual stocks. Low degree of freedom (for example, 3) is chosen to assume fat tailed distribution.

4. Data Generation

Historical stock prices from Yahoo Finance are used to calculate daily average return of stocks. Historical composition of S&P500 with industry sector information obtained from Rotman Bloomberg is used to stratify stocks. Using ten years of data, return and risk are computed on a yearly basis.

In order to stratify stocks, S&P500's historical constituents and their industry sector are needed. However, this information is very confidential and S&P500 does not provide such information for free. Unfortunately, for this paper, present data as of November 2014 is used for analysis. That is, assumption is made such that historical constituents and their industry sector would be the same as November 2014. Only justification for this assumption is inaccessibility of data. Thus, anyone with access to historical S&P500's constituents and sector information should use appropriate data for analysis. Data used for this paper is obtained from Bloomberg subscription of Rotman Commerce at University of Toronto.

5. Result

Because simulations are used to compute expected return of stocks, the weights found from mathematical programming will be slightly different each time the model is computed. This will lead to a slight difference in portfolio return and $wAVaR$. For analysis purpose, the model is computed ten times for each year from 2004 to 2013 and the performance of model is calculated on average.

Following table summarizes the test results using real data.

	Daily return			$wAVaR$			Tracking Error	
Year	SPDR	P/F	S%P500	SPDR	P/F	S%P500	SPDR	P/F
2004	0.0007	0.00062	0.0007	-0.0148	-0.01766	-0.0148	0.0049	0.19493
2005	0.0006	0.0008	0.0005	-0.0152	-0.01914	-0.0154	0.0103	0.24278
2006	0.0002	0.00018	0.0001	-0.0321	-0.03755	-0.0324	0.0098	0.15763
2007	0.0007	0.00085	0.0006	-0.0235	-0.02778	-0.0237	0.0119	0.16955
2008	0.0011	0.00173	0.001	-0.0314	-0.04021	-0.0325	0.0335	0.23687
2009	-0.0018	-0.00221	-0.0019	-0.0597	-0.07476	-0.0602	0.0077	0.24219
2010	-0.0002	-0.00031	-0.0002	-0.0237	-0.02626	-0.0242	0.0211	0.08535
2011	0.0005	0.0005	0.0004	-0.0119	-0.01539	-0.0116	0.0289	0.32744
2012	0.0004	0.00081	0.0004	-0.0127	-0.01585	-0.0121	0.0504	0.30641
2013	0.0002	0.00081	0.0001	-0.0138	-0.01521	-0.0132	0.0465	0.15382

The model outperforms SPDR for five periods and underperforms for five periods in terms of return. A strength of the model is that when it outperforms SPDR, the magnitude is higher than the magnitude when it underperforms SPDR. However, the model always underperforms SPDR in terms of $wAVaR$. One notable thing is the striking closeness of $wAVaR$ between SPDR and S&P500.

6. Conclusion

A robust optimization model with the Weighted Average Value at Risk was performed to compute weights for a portfolio that attempted to outperform SPDR. The model managed to achieve higher return than SPDR when tested against real data. However, the model always had higher risk profile than SPDR meaning it underperformed SPDR in terms of risk measured as Weighted Average Value at Risk. Therefore, it is conclusive that the model did not outperform SPRD overall.

7. References

[1] Adam, A., Houkari, M., and Laurent, J.P. (2008) Spectral risk measure and portfolio selection. *Science Direct*. Vol.32, pp.1870-1882.

[2] Adjei, F. (2009) Diversification, performance, and performance persistence in exchange-traded funds. *International Review of Applied Financial Issues and Economics*. Vol.1, No.1, p.4.

[3] Bertsimas, D. and Thiele, A. (2006) Robust and data-driven optimization: modern decision-making under uncertainty. *Tutorials on operations research*. Vol.4, p.122-195.

[4] Canakgoz, N.A. and Beasley, J.E. (2008) Mixed-integer programming approaches for index tracking and enhanced indexation. *European Journal of Operational Research*. Vol.196, p.384-399.

[5] Charupat, N. and Miu, P. (2012) Recent developments in exchange-traded fund literature. *Managerial Finance*. Vol.39, No.5, p.427-443.

[6] Chen, C and Kwon, R.H. (2012). Robust portfolio selection for index tracking. *Computers & Operations Research*. Vol.39, p.829-837.

[7] Chen, G. (2013) Based on Monte Carlo simulation investment portfolio VaR risk analysis. *Journal of Convergence Information Technology*. Vol.8, No.9, p.1071.

[8] Chen, L. and Huang, L (2009) Portfolio optimization of equity mutual funds with fuzzy return rate and risks. *Expert Systems with Applications*. Vol.36, p.3720-3727.

[9] Corielli, F. and Marcellino M. (2005) Factor based index tracking. *Journal of Banking and Finance*. Vol.30, No.8, p.2215-2233.

[10] Dose, C. and Cincotti, S. (2005) Clustering of financial time series with application to index and enhanced index tracking portfolio. *Physica A.* Vol.355, p.145-151.

[11] Down, K and Cotter, J. (2007) Spectral risk measures and the choice of risk aversion function. Ireland: UCD Geary Institute, University College Dublin.

[12] Down, K and Cotter, J. (2007) Expected spectral risk measures. Ireland: UCD Geary Institute, University College Dublin.

[13] Down, K and Cotter, J. (2007) Extreme spectral risk measures: an application to future clearinghouse margin requirements. Ireland: UCD Geary Institute, University College Dublin.

[14] Gaivoronski, A.A., Krylov, S., and van der Wijst, N. (2005) Optimal portfolio selection and dynamic benchmark tracking. *European Journal of Operational Research*. Vol.163, No.1, p.115-131.

[15] Garcia, F., Guijarro, F., and Moya, I. (2011) The curvature of the tracking frontier: a new criterion for the partial index tracking problem. *Mathematical and Computer Modelling*. Vol.54, p.1781-1784.

[16] Imeni, S.A. and Najafi, A,A. (2013) Fuzzy index tracking portfolio selection model based on Value-at-Risk. *International journal of Research in Industrial Engineering*. Vol.2, No.4, p.41-57.

[17] Jianhe L., Xin, W., and Keyi, C. (2012) Index tracking with factor model: evidence from shanghai 50. *Journal of Convergence Information Technology*. Vol.7, No.19, p.347-355.\

[18] Ma, T. and Serota, R.A. (2014) A model for stock returns and volatility. *Physica A*. Vol.398, p.89-115.

[19] Maringer, D and Oyewumi, O. (2007) Index tracking with constrained portfolios. *Intelligent systems in accounting, finance, and management.* Vol.15, p.57-71.

[20] Mezali, H. and Beasley, J.E. (2012) Index tracking with fixed and variable transaction costs. *Optim Lett*. Vol.8, p.61-80.

[21] Mezali, H. and Beasley, J.E. (2013) Quintile regression for index tracking and enhanced indexation. *Journal of the Operational Research Society*. Vol.64, p.1676-1692.
[22] Montfort, K., Visser, E., and Draat, L. (2008) Index tracking by means of optimized sampling. *The Journal of Portfolio Management*. Vol.34, p.143-151.
[23] Oh, K. J., Kim, T. Y, and Min, S. (2005) Using genetic algorithm to support portfolio optimization for index fund management. *Expert Systems with Application.* Vol.28, p.271-379.
[24] Okay, N. and Akman, U. (2003) Index tracking with constant aggregation. *Applied Economics Letter*s. Vol.10, p.913-916.
[25] Rachev, S., Stoyanov, S., and Fabozzi, F. (2008) Advanced Stochastic Models, Risk Assessment, and Portfolio Optimization. Hoboken, New Jersey: John Wiley & Sons, Inc.
[26] Rachev, T., Stoyanov, S. V., and Fabozzi, F. J. (2011) A Probability Metrics Approach to Financial Risk Measures. Hoboken, New Jersey: John Wiley & Sons, Inc.
[27] Rockafellar, T and Uryasev S. (2002) Optimization of conditional value at risk. *The Journal of Risk*. Vol.2, No. 3.
[28] Rudolf, M., Wolter, H., and Zimmerman, H. (1998) A linear model for tracking error minimization. *Journal of Banking & Finance*. Vol.23, p.85-103.
[29] Ruiz-Torrubiano, R and Suarez, A. (2009) A hybrid optimization approach to index tracking. *Ann Oper Res*. Vol.166, p.57-71.
[30] Satchell, Stephen. (2010) Optimizing Optimization. Boston, Massachusetts: Academic Press.
[31] Sharifzadeh, M. (2012) An analytical performance comparison of exchange-traded funds with index funds: 2002 – 2010. *Journal of Asset Management*. Vol.13, No.3, p.196-209.
[32] Stoyan, S.J. and Kwon, R.H. (2010) A two-stage stochastic mixed-integer programming approach to the index tracking problem. *Optim Eng*. Vol.11, p.247-275.
[33] Szego, G. (2002) Measures of risk. *Journal of Banking & Finance*. Vol. 26, p.1253-1272.
[34] Varsei, M., Shams, N., and Fahimnia, B. (2013) A Heuristic approach to the index tracking problem: a case study of the Tehran exchange price index. *Asian Academy of Management Journal*. Vol.18, No.1, p.19-34.
[35] Wang, M., Xu, C., Xu, F., and Xue, H. (2011) A mixed 0-1 LP for index tracking problem with CVaR risk constraints. *Springer Science + Business Media*. Vol.196, p.591-609.
[36] Wigan, David (2012) Tracking error conundrum. *The Journal of Risk*. V.1, p.12.

Appendix A. Stock Stratification Methods

In this section, various methods for stratifying a subset of stocks from Index fund in the context of Index tracking employed by other researches are reviewed.

A.1. Weights from Index

Gaivoronski, A.A., Krylov, S., and van der Wijst, N. [14] suggests stratifying stocks in order of their weights in the Index fund. Justification of this approach is that stocks consisting larger part of Index fund are likely to represent the market well. However, this approach does not consider sector diversification, which is a significant limitation.

A.2. Beta of individual stocks and the Benchmark

Gaivoronski, A.A., Krylov, S., and van der Wijst, N. [14] suggests another method using the beta. The logic is that closer the beta gets to 1, more correlated each stock is to the Index fund, thus it's likely that the stock will represent the market very well. However, historical correlations do not necessarily guide future correlations.

Beta for stock i is given as:

$$\beta_i(t) = \frac{cov(r_i(t), R_m(t))}{Var(R_m(t))}$$

$r_i(t) = rate\ of\ return\ of\ stock\ i\ at\ time\ t$
$R_m(t) = rate\ of\ return\ of\ index\ at\ time\ t$
$Order\ stocks\ based\ on\ |\beta_i - 1|$

A.3. Correlation distance

Dose and Cincotti [10] introduces a quantity called correlation distance defined as following,

$$d(x, y) = \sqrt{2(1 - c(x, y))}$$

$$x = (x_1, \dots, x_T)$$

$$y = (y_1, \dots, y_T)$$

$$c(x, y) = correlation\ coefficient\ between\ x\ and\ y$$

This quantity is very similar to beta. Thus, it shares the same weakness that the beta approach has.

A.4. Percentage difference

Dose and Cincotti [10] introduces another method, which is the percentage difference between prices of stocks. However, historical stock prices don't necessarily guide future stock prices.

$$d(x, y) = \min(d_1, d_2)$$

$$d_1 = \min(\frac{1}{T} \sum_{t=1}^{T} \left(\frac{x_t - ay_t}{x_t}\right)^2)$$

$$d_2 = \min(\frac{1}{T} \sum_{t=1}^{T} \left(\frac{x_t - ay_t}{ay_t}\right)^2)$$

A.5. Using the result of the optimization framework

Varsei, Shams, and Fahimnia [34] come up with an idea of adjusting stratification based on the result of optimization framework. Their algorithm is summarized below.

(1) Objective function can be rewritten $min \sum (AX - I)^2$
(2) Select $K + L$ stocks out of N stocks that have the highest positive correlation with the

index where $L = 0, 1, 2, \dots, 10$

(3) Use $X = (A^T A)^\wedge 2 A^T I$ to calculate weights for each portfolio

(4) Calculate TE for each portfolio

(5) Select portfolio that has minimum TE

This approach has a number of problems. First of all, the trial and error type of approach is very time consuming and there seems to be too many combinations to try. Second, picking stocks based on correlation will lead to same problem discussed in earlier approaches.

A.6. Combination of sector, market cap, trading amount, and beta

Oh, Kim, and Min [23] come up with a quite sophisticated method to stratify stocks. Detailed method is outlined below

$N = number\ of\ stocks\ in\ index$

$K = number\ of\ stocks\ in\ portfolio$

$J = number\ of\ industry\ sectors$

$I_j = number\ of\ stocks\ in\ sector\ j$

$r_{i,j}(t) = rate\ of\ return\ of\ stock\ i\ in\ sector\ j\ at\ time\ t$

$A_{i,j}(t) = trading\ amount\ of\ stock\ i\ in\ sector\ j\ at\ time\ t$

$Trading\ amount = multiply\ stock\ price\ by\ its\ trading\ volume$

$M_{i,j}(t) = market\ capitalization\ of\ stock\ i\ in\ sector\ j\ at\ time\ t$

$Market\ capitalization = multiply\ stock\ price\ by\ number\ of\ shares\ outstanding$

$R_m(t) = rate\ of\ return\ of\ index\ at\ time\ t$

$B_{i,j} = \sqrt{\frac{\frac{1}{T-2}\sum_t \left(r_{i,j}(t) - \bar{r}_{i,j}\right)^2}{\sum_t (R_m(t) - \bar{R}_m)^2}}$

The priority P is defined as:

$$P_{i,j} = v_1 * B_{i,j}^{-1} + v_2 * \bar{A}_{i,j} + v_3 * \bar{M}_{i,j}$$

$where\ v_1, v_2, v_3\ are\ normalizing\ factors\ (their\ numbers\ are\ chosen\ such\ that\ each\ of\ the\ three\ terms\ have\ equal\ contribution\ to\ the\ value\ of\ P_{i,j})$

Algorithm

(1) Select sector j with largest market capitalization

(2) Within sector j, choose stock i that has the highest $P_{i,j}$ value

(3) Remove stock i and recalculate market capitalization for sector j

(4) Repeat the process until desire number of stocks are selected

This approach is very through considering industry sectors, beta, market capitalization, and trading amount. It may be pointed out that historical quantities such as beta and trading amount might not be good indication since history doesn’t necessarily guide future. However, since sector diversification is taken care of, this method should yield a good stratification.

Appendix B. Modelling uncertainty (Forecasting stock returns)

B.1. Historical Model

This approach does not impose any distributional assumptions, rather it is constructed from historical data. Underlying assumption is that the past trends will continue in the future, which may not necessarily true.

B.2. Hybrid

Different weights are assigned to the observations by which the more recent observations get a higher weight (Exponentially declining weights attached to historical returns, starting from the current time and going back in time)

$$Weight\ for\ i^{th}\ observation: \theta_i = c^* \lambda^{t-i}, 0 < \lambda < 1, c = \frac{1-\lambda}{1-\lambda^k}$$
$$k = number\ of\ observed\ returns\ on\ a\ given\ asset$$
$$t = current\ time$$

In this approach, observations far back in the past have less impact on the portfolio risk at the present time. However, it should be noted that this method still depends on the past trend.

B.3. Normal distribution

This approach assumes that on the basis of Central Limit Theorem, the distribution of the daily log-return is approximately normal. However, short period log-returns are not independent and autocorrelations are observed experimentally. Also when returns are summed up, the sum is often dominated by big outliers; not symmetrically negligible. Thus, if a distributional model is used, non-normal heavy tail distribution would be better choice than the normal distribution.

Appendix C. Financial risk

C.1. Coherent risk measures axioms

Before discussing financial risk models, it's worth to examine axioms for coherent risk measure. If a risk model is coherent, then it means that the risk measure makes financial sense.

$$Given\ a\ portfolio\ return\ X, denote\ corresponding\ distribution\ function\ F_X$$

$$(H1) Positive\ homogeneity: for\ every\ random\ portfolio\ return\ X\ and\ real\ value$$
$$\lambda > 0, \rho(\lambda X) = \lambda \rho(X)$$

: Scaling the return of portfolio by a positive factor scales the risk by the same factor. (If the investment in a position doubles, so does the risk of the portfolio)

$$(H2) Translation\ invariance: for\ every\ random\ portfolio\ return\ X\ and\ real\ value\ \alpha,$$
$$\rho(X + \alpha) = \rho(X) - \alpha$$

: Adding cash to a position reduces its risk by the amount of cash added

$$(H3) Monotonicity: for\ every\ random\ portfolio\ return\ X\ and\ Y such\ that\ X \geq Y,$$
$$\rho(X) \leq \rho(Y)$$

: If an investment has random return X that is not less than the return Y of another investment then the risk of X is not greater than the risk of Y

$(H4) Sub - additivity: for\ every\ random\ portfolio\ return\ X\ and\ Y,$
$\rho(X+Y) \leq \rho(X) + \rho(Y)$
: The risk of portfolio is not greater than the sum of the risks of individual returns
: The positive homogeneity and sub-additivity imply that the functional is convex.

C.2 Measure of dispersion

Variance/Standard Deviation
$\sigma_X = \sqrt{E(X-EX)^2}, \quad \sigma_X^2 = E(X-EX)^2$

Mean-Absolute Deviation
$MAD_X = E|X-EX|$

Variance/Standard Deviation/Mean-Absolute Deviation penalize symmetrically both the negative and the positive deviation from the mean. This can be overcome by only considering downside deviation. Those risk measures do not incorporate risk averseness either. It can also be overcome by applying risk averse function on random variable before computing moments. However, the biggest weakness of these risk measures is that they are not coherent under non-normal distributions.

C.3. Value at Risk

If we order observations of portfolio returns, then a^{th} observation would correspond to Value at Risk at tail probability ϵ. Value at risk tries to capture what the portfolio return would be when portfolio performs poorly. Mathematically, it is defined as

Value at Risk at confidence level $(1-\epsilon)100 = -\inf\{x | P(X \leq x) \geq \epsilon\}$

Value at Risk has some severe weakness to be a good risk measure. First if all, Value at Risk is not informative about losses beyond the confidence level. Knowing the portfolio return at a^{th} observation does not tell anything about portfolio returns beyond a^{th} observation. Moreover, it fails sub-additivity axiom of coherent risk measure in general distributions. That is, Value at Risk of portfolio may be greater than the sum of $VaRs$ of constituents

$$VaR_{\in}(X+Y) > VaR_{\in}(X) + VaR_{\in}(Y)$$

This may lead to conclusion that there is no diversification effect in portfolio and irrational decision to concentrate only into few positions.

Value at Risk also is not suited for investors who are risk averse since it places all its weights on a single quantile that corresponds to a chosen confidence level, which is extremely risk-taking behaviour.

C.4. Average Value at Risk

The Average Value at Risk is the average of observations which are larger than Value at Risk at tail probability ϵ

$$AVaR_{\in}(X) = \frac{1}{\epsilon}\int_0^{\epsilon} VaR_p\,(X)\,dp$$

Average value of risk is well defined for real-valued random variable with finite mean. It assigns equal weights for all quantile that are outside of confidence level. Thus, assumption is that investors are risk-neutral for the portfolio returns at tail.

C.5. Weighted Average Value at Risk

This risk measure maintains investor's risk-averseness by applying a risk-averse function for portfolio returns at tail, making it a weight average rather than simple average. It can be defined as following,

$$wAVaR_{\emptyset}(X) = \int_0^1 VaR_p(X)\emptyset(p)dp$$

$$\emptyset(p) = risk\ aversion\ function, p \in [0,1]$$

The choice of risk-aversion (or utility) function is very important because not every utility function makes Weighted Average Value at Risk coherent. In order for $wAVaR$ to be a coherent risk measure, function $\emptyset(p)$ must satisfy following three properties. [25]

$$positive: \emptyset(p) \geq\ 0, p \in [0,1]$$

$$Normed: \int_0^1 \emptyset(p)dp = 1$$

$$Nonincreasing: \emptyset(p_1) \geq\ \emptyset(p_2),\ p_1 \leq p_2$$

$$weight\ corresponding\ to\ interval\ p_2 - p_1 =\ \Delta p\ is\ \emptyset(p_1)\Delta p$$

The third condition is key: the weights attached to higher losses should be no less than the weights attached to lower losses

C.6. Utility Functions

In order to find suitable utility function for Weighted Average Value at Risk, some properties of utility function needs to be examined. A utility function is non-satiable when

$$u(x) \leq u(y)\ if\ x \leq y$$

The above represents non-decreasing property. Financial implication of non-decreasing property is that investors prefer more to less.

Another property to look for is risk-averseness is,

$$u(px_1 + (1-p)x_2) \geq pu(x_1) + (1-p)u(x_2), \forall x_1, x_2\ and\ p \in [0,1]$$

Financial interpretation is that investor gains a lower utility from a venture with some expected payoff and a prospect with a certain payoff, equal to the expected payoff of the venture

Clearly, not all functions satisfy both properties. For example, quadratic functions with positive leading coefficient satisfies non-satiable property but do not satisfy risk-averse property.

The type of utility functions that satisfy both non-satiable and risk-averse property are the following

$$Logarithmic: u(x) = \log(x), r_A(x) = \frac{1}{x}$$

$$Exponential: u(x) = -e^{-ax}, r_A(x) = a$$

$$Power: u(x) = \frac{-x^{-a}}{a}, r_A(x) = \frac{a}{x}$$

In order to measure risk-averseness of a utility function, a term called Arrow-Pratt coefficient of absolute risk aversion (ARA) is introduced. It is defined as,

$$r_A(x) = -\frac{u''(x)}{u'(x)}$$

Studies have been conducted to test whether above utility functions lead to good Weighted Average Value at Risk. The central findings from Dowd and Cotter [11] is that exponential functions leads to intuitive and nicely behaved properties while power utility function does not. For example, for some range of ARA, increase in risk aversion lead to decrease in Weighted Average Value at Risk derived from using power utility functions, which is clearly not how risk-averse function should behave.

For logarithmic utility function, the risk-averseness depends on value of the quantile, whereas the risk-averseness depends on a constant for exponential utility function. In other words, by choosing exponential utility function, the decision maker can adjust the value of ARA to express his/her level of risk-averseness.

Appendix D. Tracking error definition

D.1. Defining Tracking Error

Rudolf, M., Wolter, H., and Zimmerman, H. [28] introduces many ways to define tracking error between the portfolio and Index fund.

1. Mean absolute deviations (MAD): sum of the absolute deviations between the benchmark returns and the portfolio returns.
2. MinMax: maximum deviation between portfolio and benchmark returns is minimized.
3. Downside risk: the return on the portfolio is below the return on the benchmark portfolio. Tracking error can be restricted to the negative deviations between portfolio and

benchmark returns. For example, in case of MAD, the sum of absolute deviation is minimized subject to the restriction that the portfolio return is below the benchmark return (MADD) In case of MinMax, maximum negative deviation is minimized.

In summary,

$$TE_{MAD} = \min I'(|X\beta - Y|),\ \ where\ I=(1,\ldots,1)$$

$$TE_{MADD} = \min I'(|\bar{X}\beta - \bar{Y}|),\ \ where\ \overline{X_t}\beta < \overline{Y_t},\ I=(1,\ldots,1)$$

$$TE_{MinMax} = \min I'(|X\beta - Y|),\ \ where\ I=(1,\ldots,1)$$

$$TE_{MinMaxD} = \min I'(|\bar{X}\beta - \bar{Y}|),\ \ where\ \overline{X_t}\beta < \overline{Y_t},\ I=(1,\ldots,1)$$

$Y = vector\ of\ continously\ compounded\ index\ returns$
$X = matrix\ of\ continously\ compounded\ asset\ returns$
$\beta = portfolio\ weights\ to\ be\ determined\ (decision\ variable)$
$The\ matrix\ \ \bar{X}\ and\ the\ vector\ \bar{Y}\ contain\ only\ those\ lines\ where\ the\ benchmark\ return$
$is\ below\ the\ portfolio\ return$

There are also many variations of above tracking error definitions. Charupat and Miu [5] discuss following definitions.

1. Average absolute difference between the return on the fund and that of the underlying benchmark index

$$TE_1 = \frac{1}{T}\sum_{t=1}^{T}|r_t^{ETF} - r_t^{Index}|$$

2. Root-mean-square deviation of the return on the fund from that of the benchmark.

$$TE_2 = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}\left(r_t^{ETF} - r_t^{Index}\right)^2}$$

3. The standard deviation of the difference between the return on the fund and that of the benchmark

$$TE_3 = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}[\left(r_t^{ETF} - r_t^{Index}\right) - (\bar{r}_t^{ETF} - \bar{r}_t^{Index})]}$$

4. The standard error of the regression of the returns on the funds on that of the benchmark.

$$r_t^{ETF} = \alpha + \beta r_t^{Index} + \varepsilon_t$$

$r_t^{ETF} = return\ of\ ETF\ at\ time\ t$
$r_t^{Index} = return\ of\ Index\ (S\&P500)\ at\ time\ t$

All above definitions involve deviation between portfolio return and Index return. It aims to find a portfolio return distribution which is closest to the return distribution of Index.

Suppose decision maker have found the portfolio tracking Index most closely with respect to tracking error defined as deviation measure. Can he/she be sure that the risk of the portfolio is close to the risk of Index? The answer is affirmative only if he/she uses the deviation measure as a risk measure. If decision maker uses risk models that are not deviation measures then he/she needs to find appropriate probability metric. [25]

D.2. Probability Metric

Rachev, S., Stoyanov, S., and Fabozzi, F. [25] discuss probability metric in detail. In order to guarantee that small distance between return distributions correspond to similar risks, A suitable probability metric needs to be found. For a given risk measure we need to find a probability metric with respect to which the risk measure is a continuous functional

$$|\rho(X) - \rho(Y)| \leq \mu(X, Y)$$

$$\rho = the\ risk\ measure$$

$$\mu = probability\ metric$$

$$|\rho(Index) - \rho(Portfolio)| \leq \int_0^1 |F_X^{-1}(p) - F_Y^{-1}(p)|\, \emptyset(p) dp = \kappa_\phi(Index, Portfolio)$$

$$\kappa_\phi(Index, Portfolio) = weighted\ Kantorovich\ metric$$

If the distance between Index return distribution and Portfolio return distribution is small, as measured by the metric $\kappa_\phi(Index, Portfolio)$ then the risk of Index is close to the risk of Portfolio as measured by the spectral risk measure $wAVaR_\phi$

$$\kappa_\phi(X, Y) = \int_{\mathbb{R}} |F_X(x) - F_Y(y)|\, \emptyset(x) dx$$

$$where\ EX < \infty\ and\ EY < \infty$$

$$F_X(X)\ and\ F_Y(Y)\ are\ the\ probabilities\ that\ X\ and\ Y\ lose\ more\ than\ the\ level\ x.$$

The Kantorovich metric κ_ϕ sums the absolute deviation between the two probabilities for all possible values of the loss level x