Designing Proxies for Stock Market Indices is Computationally Hard

(Abstract)

Ming-Yang Kao*
Department of Computer Science
Yale University
New Haven, CT 06520

Stephen R. Tate†
Department of Computer Science
University of North Texas
Denton, TX 76203

1 Introduction

Market indices are widely used to track the performance of stocks or to design investment portfolios [1]. This paper initiates a rigorous mathematical study of the computational complexity of the art of designing proxies for such indices. While there are several results on selecting such proxies (or portfolios) in an on-line manner (see, for example, [2] and [3]), we look at off-line algorithms for designing proxies based on historical data. In particular, we show that all combinations of three fundamental problems (such as tracking or outperforming a full market index) with four commonly-used indices give NP-complete problems, so are computationally hard.

To formally define market indices, let $\mathcal{B}$ be a set of b stocks in a market. Let $S_{i,t} \geq 0$ be the price of the i-th stock at time t. Let w_i be the number of outstanding shares of the i-th stock. We assume that w_i does not change with time. This paper discusses computational complexity issues regarding four kinds of market indices currently in use [1]. These indices are calculated by the following formulas, which can be multiplied by arbitrary constants to arrive at desired starting index values at time 0.

- The *price-weighted index* of $\mathcal{B}$ at time t is

$$\Phi_1(\mathcal{B}, t) = \frac{\sum_{i=1}^{b} S_{i,t}}{b}.$$

The Dow Jones Industrial Average is calculated in this manner for some $\mathcal{B}$ consisting of thirty stocks.

- The *value-weighted index* of $\mathcal{B}$ at time t is

$$\Phi_2(\mathcal{B}, t) = \frac{\sum_{i=1}^{b} w_i \cdot S_{i,t}}{\sum_{i=1}^{b} w_i \cdot S_{i,0}}.$$

The Standard & Poor's 500 is computed in this way with respect to 500 stocks.

- The *equal-weighted index* of $\mathcal{B}$ at time t is

$$\Phi_3(\mathcal{B}, t) = \sum_{i=1}^{b} \frac{S_{i,t}}{S_{i,0}}.$$

The index published by the Indicator Digest is calculated by this method, involving stocks listed on the New York Stock Exchange.

- The *price-relative index* of $\mathcal{B}$ at time t is

$$\Phi_4(\mathcal{B}, t) = \left(\prod_{i=1}^{b} \frac{S_{i,t}}{S_{i,0}} \right)^{\frac{1}{b}}.$$

The Value Line Index is computed by this formula.

There are numerous reasons why stock investors and money managers would want to invest in a subset of stocks rather than those of a whole market [1]. For instance, small investors certainly do not have sufficient capital to invest in every stock in the market. Logically, such investors would attempt to choose a small subset of stocks which hopefully can perform roughly as well as or even outperform the market as a whole. They then face difficult trade-offs between returns and risks. For these and other reasons of optimization, we formulate three natural computational problems for the design of market indices. Given a market $\mathcal{M}$ consisting of m stocks, we wish to choose a subset $\mathcal{M}_k$ of at most k stocks and calculate an index of $\mathcal{M}_k$, which is called a *k-proxy* of the corresponding index of the whole market $\mathcal{M}$ (we sometimes refer to $\mathcal{M}_k$ as a *portfolio*). Our goal is to choose $\mathcal{M}_k$ so that the resulting k-proxy tracks or outperforms the corresponding index of $\mathcal{M}$. This paper shows that designing proxies for the above four indices based on historical data is computationally hard.

*Supported in part by NSF Grant CCR-9531028.

†Supported in part by NSF Grant CCR-9409945 and Texas Advanced Research Program Grant 1997-003594-019.

2 Problem Formulations

In this section we formally define three basic problems related to selecting k-proxies, or portfolios.

PROBLEM 1. (*tracking an index*)

Input: A market $\mathcal{M}$ of m stocks, their prices $S_{i,t} \geq 0$ for $t = 0, \ldots, f$, their numbers w_i of outstanding shares, a real $\epsilon_1 > 0$, an integer $k > 0$, and some $j \in \{1, 2, 3, 4\}$ to indicate the desired type of index.

Output: A subset $\mathcal{M}_k$ of at most k stocks in $\mathcal{M}$ such that for all $t = 1, \ldots, f$,

$$\left| \frac{\Phi_j(\mathcal{M}_k, t)}{\Phi_j(\mathcal{M}_k, 0)} - \frac{\Phi_j(\mathcal{M}, t)}{\Phi_j(\mathcal{M}, 0)} \right| \leq \epsilon_1 \cdot \frac{\Phi_j(\mathcal{M}, t)}{\Phi_j(\mathcal{M}, 0)}.$$

PROBLEM 2. (*outperforming an index*)

Input: A market $\mathcal{M}$ of m stocks, their prices $S_{i,t} \geq 0$ for $t = 0, \ldots, f$, their numbers w_i of outstanding shares, a real $\epsilon_2 \geq 0$, an integer $k > 0$, and some $j \in \{1, 2, 3, 4\}$ to indicate the desired type of index.

Output: A subset $\mathcal{M}_k$ of at most k stocks in $\mathcal{M}$ such that for all $t = 1, \ldots, f$,

$$\frac{\Phi_j(\mathcal{M}_k, t)}{\Phi_j(\mathcal{M}_k, 0)} \geq (1 + \epsilon_2) \cdot \frac{\Phi_j(\mathcal{M}, t)}{\Phi_j(\mathcal{M}, 0)}.$$

For the final problem, we need a few extra definitions in order to analyze the *volatility* of a set of stocks. Let $\mathcal{B}$ be a set of stocks as defined in §1.

- The *one-period return* of Φ_j for $\mathcal{B}$ at time $t \geq 1$ is

$$R_j(\mathcal{B}, t) = \ln \frac{\Phi_j(\mathcal{B}, t)}{\Phi_j(\mathcal{B}, t-1)}.$$

- The *average return* of Φ_j for $\mathcal{B}$ up to time $t \geq 1$ is

$$\overline{R}_j(\mathcal{B}, t) = \frac{\sum_{i=1}^{t} R_j(\mathcal{B}, i)}{t}.$$

- The *volatility* of Φ_j for $\mathcal{B}$ up to time $t \geq 2$ is

$$\Delta_j(\mathcal{B}, t) = \sqrt{\frac{\sum_{i=1}^{t} \left(R_j(\mathcal{B}, i) - \overline{R}_j(\mathcal{B}, t) \right)^2}{t-1}}.$$

PROBLEM 3. (*sacrificing return for less volatility*)

Input: A market $\mathcal{M}$ of m stocks, their prices $S_{i,t} \geq 0$ for $t = 0, \ldots, f$, their numbers w_i of outstanding shares, two reals $\alpha, \beta > 0$, an integer $k > 0$, and some $j \in \{1, 2, 3, 4\}$ to indicate the desired type of index.

Output: A subset $\mathcal{M}_k$ of at most k stocks in $\mathcal{M}$ such that

$$\frac{\Phi_j(\mathcal{M}_k, t)}{\Phi_j(\mathcal{M}_k, 0)} \geq \alpha \cdot \frac{\Phi_j(\mathcal{M}, t)}{\Phi_j(\mathcal{M}, 0)} \text{ for all } t = 1, \ldots, f;$$

$$\Delta_j(\mathcal{M}_k, s) \leq \beta \cdot \Delta_j(\mathcal{M}, s) \text{ for all } s = 2, \ldots, f.$$

3 Results

In this abstract, we simply quote the main results — all the proofs can be found in the full paper.

THEOREM 3.1. *Let ϵ_1 be any error bound satisfying $0 < \epsilon_1 < 1$ and specified using $n^{O(1)}$ bits in fixed point notation. Then the tracking problem with error bound ϵ_1 is NP-hard for the price-weighted index, value-weighted index, and equal-weighted index.*

THEOREM 3.2. *Let ϵ_2 be any value satisfying $0 < \epsilon_2 < n^c$ for some constant c. Then the problem of outperforming the market average with bound ϵ_2 is NP-hard for the price-weighted index, value-weighted index, equal-weighted index, and price-relative index.*

THEOREM 3.3. *Let α and β be values expressed using $n^{O(1)}$ bits in fixed-point binary notation, and satisfying $0 < \alpha \leq n^{O(1)}$ and $\beta = \Omega\left(\frac{\log k}{\log n}\right)$. Then the problem of sacrificing return for less volatility is NP-complete for the price-weighted index, value-weighted index, equal-weighted index, and price-relative index.*

The one result that must be separated from the above is the tracking problem for the price-relative index. In order for our reduction to work in this case, we were required to reduce the range of possible values for ϵ_1.

THEOREM 3.4. *Let ϵ_1 be any error bound satisfying $0 < \epsilon_1 < 1$ and specified using $O(\log n)$ bits in fixed point notation. Then the tracking problem for a price-relative index with error bound ϵ_1 is NP-hard.*

References

[1] G. J. ALEXANDER, W. F. SHARPE, AND J. V. BAILEY, *Fundamentals of Investments*, Prentice-Hall, Upper Saddle River, NJ, 2nd ed., 1993.

[2] T. M. COVER, *Universal portfolios*, Mathematical Finance, 1 (1991), pp. 1–29.

[3] T. M. COVER AND E. ORDENTLICH, *Universal portfolios with side information*, IEEE Transactions on Information Theory, 42 (1996), pp. 348–363.