行政院國家科學委員會補助專題研究計畫成果報告

※※※※※※※※※※※※※※※※※※※※※※※※※※
※　　　　　　　　　　　　　　　　　　　　　　　　※
※　　建立指數基金的目標規畫模型及其對偶關係　　※
※　　　　　　　　　　　　　　　　　　　　　　　　※
※※※※※※※※※※※※※※※※※※※※※※※※※※

計畫類別：■個別型計畫　　□整合型計畫

計畫編號：NSC89-2213-E-004-011

執行期間：89 年 8 月 1 日至 90 年 7 月 31 日

計畫主持人：劉明郎　副教授

本成果報告包括以下應繳交之附件：

□赴國外出差或研習心得報告一份

□赴大陸地區出差或研習心得報告一份

□出席國際學術會議心得報告及發表之論文各一份

□國際合作研究計畫國外研究報告書一份

執行單位：政治大學應用數學學系

中　華　民　國　九十　年　　　月　　　日

建立指數基金的目標規畫模型及其對偶關係

Goal Programming Models and their Duality Relations Using in the Creating Index Fund

計畫編號：NSC 89-2213-E-004-011
執行期限：自民國 89 年 8 月 1 日起至民國 90 年 7 月 31 日
主持人：劉明郎　國立政治大學應用數學系

中文摘要

愈來愈多投資者或投資機構已接受指數基金為其多項投資策略之一環。建立追隨指數的一籃子股票組合可以建構成混合整數線性目標規畫模型。本文主要係藉由目標規畫模型中對應的對偶變數的研究，將選擇股票進入組合的問題轉為評估被選入股票在組合內的效用。並期望對偶變數的應用能幫助我們發展一個有效率的解法給提出的目標規畫模型。

Abstract

More and more investors or institutes accepted index-tracking stock baskets as one part of their investment strategy. The creating of index-tracking stock can be formulated as a mixed integer linear goal programming model. Goal programming is used to plan a selection of stocks to secure desired basket that will conform 'as close as possible' to a given time period. In this paper the work is changed from selection to evaluation and the dual variables associated with goal programming are brought into play this purpose. The use of dual variables will help us to develop an efficient solution algorithm for the proposed goal programming model.

1. Introduction

The index fund commenced in 1970 when the Wells Fargo Bank introduced the Stagecoach Fund, designed to track the New York Stock Exchange Composite Index. Perhaps due to the overall poor performance of managed funds and the market in general during the late 1960s and early 1970s, and because of a surging awareness of the efficient-market hypothesis and the random-walk theory, this new type of fund has begun to catch the increasing popularity.

Lodge (1993) mentioned that in United Kingdom (UK) over the past three years, only two general UK equity trusts beat the best UK tracker fund. Willis (1996) also described that over the past ten years in the US, only 28% of all stock funds have outpaced the market indexes they were designed to beat. Growing numbers of investors can't think of a good reason to pay mutual-fund managers for that kind of performance. So they are turning to index funds, which aim only to match the market's performance by investing in a portfolio of securities that mirror a broad index, such as the Standard & Poor's 500 (S&P 500).

Kirby (1993) estimates that of the \$1.25 trillion to \$1.5 trillion of the total institutional investment in the stock market in 1992, about 25% to 30% was indexed. As of December 1988, about 55% of the defined benefit assets of the 200 largest pension funds were indexed, with 36% of all index fund investments internally managed.

Pertaining to the method of constructing an index fund, Rudd (1980) proposed an optimization model to construct the index fund. The model is based on quadratic programming. The optimization model is transcribed as follows:

$$\min \; w_p^2$$

subject to:

$$\beta_P = \sum_{i=1}^{n} \beta_i u_i = 1$$

$$\sum_{i=1}^{n} u_i = 1, \;\; u_i \geq 0, \;\; \forall i = 1, 2, \ldots, n.$$

where w_p^2 is portfolio residual variance; β_P is the portfolio beta, defined as the weighted sum of the asset betas, where the weights are the asset holdings and u_i is the proportion of the fund to be invested in the i-th asset.

Andrews, Ford, and Mallison (1986) examined and compared three methods of constructing the index fund. These methods are *full replication*, *stratification*, and *sampling*.

Two criteria have been exploited by Toy and Zurack (1989) for tracking the Europe-Pacific(Euro-Pac) Index, an international index that do not include U.S. securities. The first criterion is to measure the risk of each stock in the Euro-Pac Index in relation to the total index. The second criterion is to measure the risk of each stock in relation to its home country stock index. These two criteria were utilized in building stock baskets to track the Euro-Pac Index.

Meade and Salkin (1989) compared a number of different approaches to index fund selection. These approaches used either estimated coefficients (in this case the size of the holding of a particular share is determined according to some statistical criterion) or capitalization weighting (in this case the proportion of the fund invested in a particular company is the ratio of the company's market value to that of all the companies in the fund).

The objective was the minimization of historical tracking error. In the first case this objective was pursued using quadratic programming; in the second case a heuristic zero-one selection procedure was used. The effects of adding stratification constraints, where the fund had the same proportion invested in different industrial sectors as the market, was also examined. The index fund based on estimated coefficients was shown to perform better than the capitalization weighted fund. Stratification did not lead to appreciable improvements in tracking performance.

Kornbluth, Meade, and Salkin (1993) formulated a goal-programming duration model to create bond portfolios for the sake of tracking the 5-15 year UK gilts index as closely as possible. They showed that duration moments could be useful for constructing index-tracking bond portfolios.

Tabata and Takeda (1995) considered a bicriteria optimization problem of index fund: To simultaneously minimize the tracking error and the number of securities encompassed in the index fund. In attempting to minimize the tracking error under the given number of securities included in the index fund, they formulated it as a quadratic programming problem with zero-one variables. On account of difficulty in solving this combinatorial optimization problem in reality, they submitted an efficient approach to obtain a local optimal solution, which is a compromise answer to this bicriteria optimization problem.

2. The Model Formulations and their Duality Model

Once we have selected the underlying index, our objective remains to minimize the discrepancy in performance between the replicating portfolio and the index.

Definition 1. The *tracking error* $\varepsilon : R^n \rightarrow R$ is a function which measures the performance discrepancy between the portfolio $\boldsymbol{P}$ and the benchmark.

Suppose I_t is the value of the index (benchmark), V_t is the value of the portfolio at t-th period, p_{it} is the price of the issue i at t-th period, and x_i is the amount of the issue i in the portfolio. The following measure has been discussed in the Liu (2000).

Tracking error measure:

$$\varepsilon(\boldsymbol{P}) = \sum_{t=1}^{T} |V_t - I_t| = \sum_{t=1}^{T} |\sum_{i=1}^{n} p_{it} x_i - I_t|$$

The major target is to create an index fund minimizing the tracking error with the stocks of as few different companies as possible. We use the model proposed in the Liu (2000).

Absolute-Deviation Model:

Model *A-D*:

$$\min \sum_{t=1}^{T} \left| \sum_{i=1}^{n} p_{it} x_i - I_t \right| \qquad (2.1)$$

subject to:

$$\sum_{i=1}^{n} y_i \le N_0 \qquad (2.2)$$

$$x_i \le M y_i , \quad \forall i = 1, 2, \ldots, n. \qquad (2.3)$$

x_i are positive integer,

$$y_i \in \{0, 1\}, \quad \forall i = 1, 2, \ldots, n. \qquad (2.4)$$

Objective function (2.1) is equivalent to the tracking error $\varepsilon(\boldsymbol{P})$. Constraints (2.2) ~ (2.4) are to control the number of the issue in the portfolio. We introduce the binary variables defined as:

$$y_i = \begin{cases} 1 & \text{if } x_i > 0 \\ 0 & \text{otherwise} \end{cases} \quad \forall i = 1, 2, \ldots, n.$$

This relation can be converted into another form in the formulations.

$$x_i \le M y_i , \quad \forall i = 1, 2, \ldots, n$$

By imposing the upper bound N_0 on the sum of y_i, the number of issues in the portfolio will be confined.

Charnes, Cooper, and Ferguson (1955) have shown that the optimal values of the variables x_i ($i = 1, 2, \ldots, n$) in objective function (2.1) can be reached through a linear programming formulation. Taking advantage of their result, we can convert Model *A-D* into the consecutive goal programming model. For transforming the Model *A-D*, we introduce new variables d_t^+, d_t^- (d_t^+, $d_t^- \ge 0$).

Goal Programming Model:

Model *GP*:

$$\min \ z = \sum_{t=1}^{T} (d_t^+ + d_t^-) \qquad (2.5)$$

subject to:

$$\sum_{i=1}^{n} y_i \le N_0 \qquad (2.2)$$

$$x_i \le M y_i , \quad \forall i = 1, 2, \ldots, n. \qquad (2.3)$$

$$\sum_{t=1}^{n} p_{it} x_i + d_t^+ - d_t^- = I_t , \ \forall t = 1, 2, \ldots, T.$$

x_i are positive integer,

$$y_i \in \{0, 1\}, \quad \forall i = 1, 2, \ldots, n. \qquad (2.4)$$

$$d_t^+ , \ d_t^- \ge 0 , \ \forall t = 1, 2, \ldots, T. \qquad (2.7)$$

By (2.6) in Model *GP*, d_t^+ and d_t^- can be interpreted as the negative and positive amount by which the portfolio deviates from the current index at *t*-th period.

From constraint (2.4), we can readily perceive the huge complexity of model *GP*. Take the MSCI for instance: 77 issues are included in MSCI, which means there are at lease 2^{77} combinations of all variables y_i, let alone taking the combinations of all variables x_i into real computations. Due to this difficulty, our programs are not solvable within 2 full days even for the simplified case when $n = 40$ and $T = 30$. Such a situation coerced us to do some preprocessing and develop a heuristic to approximate the optimal solution. Thereby, we derive a lemma for generating some naive cuts.

Lemma 1. (Liu, 2000)

Let $U_i = \min_{1 \le t \le T} \{\frac{I_t}{p_{it}}\}, \quad \forall i = 1, 2, \ldots, n,$

then we have naive cuts

$$x_i \le U_i, \quad \forall i = 1, 2, \ldots, n. \quad (2.8)$$

By the investigation of the constraint (2.3), we found that if some of the issues have been selected in the portfolio, *i.e.* $y_i = 1$, the corresponding x_i must be satisfied the constraint

$$x_i \le U_i$$

The importantly of the value of U_i is described by the corresponding dual variables. We can therefore use this idea to develop our heuristic algorithm. First we are taking the relaxation of the model *GP*, by removing the binary variables as followings:

Model *RGP*

$$\min \quad z = \sum_{t=1}^{T} (d_t^+ + d_t^-) \quad (2.5)$$

subject to:

$$x_i \le U_i, \quad \forall i = 1, 2, \ldots, n. \quad (2.8)$$

$$\sum_{t=1}^{n} p_{it} x_i + d_t^+ - d_t^- = I_t, \forall t = 1, 2, \ldots, T. (2.6)$$

$$x_i \ge 0, \quad y_i \in \{0, 1\}, \quad \forall i = 1, 2, \ldots, n. (2.4)$$

$$d_t^+, \ d_t^- \ge 0, \quad \forall t = 1, 2, \ldots, T. \quad (2.7)$$

The corresponding dual model is constructed by introducing two set of variables u_i and v_t as the multiplier of the constraints (2.8) and (2.6), respectively.

Model *D-RGP*

$$\max \quad w = -\sum_{i=1}^{n} U_i u_i + \sum_{t=1}^{T} I_t v_t \quad (2.9)$$

subject to:

$$\sum_{t=1}^{T} p_{it} v_t - u_i \le 0, \quad \forall t = 1, 2, \ldots, T. (2.10)$$

$$| v_t | \le 1, \quad \forall t = 1, 2, \ldots, T. \quad (2.11)$$

$u_i \geq 0, \ \forall i = 1, 2, \ldots, n.$ (2.12)

v_t unsigned, $\forall t = 1, 2, \ldots, T.$ (2.13)

From the theory of the duality theorem, we have

$$\sum_{t=1}^{T}(d_t^{+*} + d_t^{-*}) = -\sum_{i=1}^{n} U_i u_i^* + \sum_{t=1}^{T} I_t v_i^*$$

for the optimal objective by solving *RGP* and *DRGP*, as indicated by '*', which we now use to provide the interpretations that will help us the determine the impact of the selection certain asset in the portfolio.

u_i^* and v_t^* provide evaluators for the deviations generated from the selected assets and the tracking dates. $U_i u_i^*$ represents the part of the deviational value associated with asset x_i. A value of $u_i^* \geq 0$ reflects the deviation in the objective of model *RGP* of the selected x_i^*, rounding up the fractional value x_i^* to the nearest integer will improve the objective value. An algorithm hence is designed to rounding up and down the fractional x_i^* to keep the tracking error as small as possible while still maintenance the feasibility of the resulting solution.

It is not to solve the dual model *DRGP* since its optimal solution is automatically obtained in the printout from most linear programming computer codes. The algorithm is utilized the impact of the value of u_i^* outlined in the next section.

3. Heuristic Algorithm

The algorithm first solves the model *GP* by relaxing the variable x_i not to be integer, get a solution called x_i^R and y_i^*. The results of y_i^* indicate the selected issues in the portfolio. Now the solution x_i^R obtained from solving the relaxed *GP* is a fractional value which can be represented as $x_i^R = \lfloor x_i^R \rfloor + f_i$ where $\lfloor x_i^R \rfloor$ represent the largest integer less than or equal to x_i^R and $0 \leq f_i \leq 1$. Clearly, the value $\lfloor x_i^R \rfloor$ will be one of the feasible solutions for the model *GP*. If we can adjust the value of x_i^R by rounding up or down to an integer in a smart way while maintenance the resulting solution still feasible for the model *GP* then we have a heuristic solution for the model *GP*. The algorithm is listed as follows:

Algorithm:

Step 1

Preprocessing: Set $S = \left\lfloor \frac{n}{2} \right\rfloor$, and select S issues which have more capital than other stocks. Reset $n = S$. ($\lfloor m \rfloor$ is the largest integer smaller than m)

Step 1.1

Generate cuts by (2.8): $x_i \le \min\{U_i, U\}$, $\forall i = 1, 2, \ldots, S$.

Step 2

Relaxation: Relax each x_i as a positive real number. Solve model *GP* to get the relaxed optimal solution x_i^R and y_i^*, $\forall i = 1, 2, \ldots, S$.

Step 3

Solve Dual Model: Fixing each x_i as zero when $y_i^* = 0$. Solve model *RGP* obtain the dual prices u_i^* to the issues i for whose $y_i^* = 1$.

Step 4

Heuristic Procedures:

Set $CE = 0$ and let $x_i^R = \lfloor x_i^R \rfloor + f_i$, $\forall i = 1, 2, \ldots, S$. $SE = \sum_{i=1}^{S} \bar{p}_i \cdot f_i$.

Step 4.1

Rank u_i^* in an increasing order. Without loss of generality, we assume that: $u_i^* \le u_{i+1}^*$, $\forall i = 1, 2, \ldots, (S-1)$.

Step 4.2

For $i = 1, 2, \ldots, S$: If $(CE + \bar{p}_i) \le SE$ and $x_i^R > 0$, set $x_i^H = \lfloor x_i^R \rfloor + 1$, and $CE = CE + \bar{p}_i$.

Else $x_i^H = \lfloor x_i^R \rfloor$.

Step 4.3

Stopping Criterion:

Set $RE = SE - CE$. If $RE \ge \bar{p}_1$, go to Step 4.2. Else go to Step 5.

Step 5

Reassign x_1^H an integer closest to $(x_1^H + \frac{RE}{\bar{p}_1})$ and we have the heuristic solution: x_i^H, $\forall i = 1, 2, \ldots, S$.

The basic idea behind the heuristic is to get the relaxed optimal solution x_i^R (step 2 of our algorithm), set the initial x_i^H to be the value from rounding x_i^R down, and enlarge the values of some x_i^H to rectify the sum of average error per day (*SE*) incurred by rounding. For retaining the original number of issues in relaxed solution, only those x_i with $x_i^R > 0$ can be enlarged. For correcting *SE*, we first rank u_i^* in an increasing order (step 4.1), and then enlarge the value x_i (with $x_i^R > 0$) by 1 unit per time according to this order (step 4.2). Whenever x_i increase by one unit, *SE* is corrected by the amount $\bar{p}_i$. So we set a number *CE* to represent the cumulative corrected error and use residual error $RE = SE - CE$ to check if the heuristic procedures should stop.

4. Conclusion

This report proposed a method for constructing an index fund. We have built a goal programming model which is concise, flexible, and realistic. Unlike other optimization models in the literature, the linear objective function and linear constraints in model *GP* are substantially easier to solve. This merit greatly improves the efficiency of our optimization model. On the other hand, model *GP* provided us a ground for ensuring the solution is optimal in view of the objective function adopted here, i.e. tracking error function. This feature eliminates the drawbacks of other simplified methods.

Model *GP* presented here is a mixed integer linear programming. Its enormous complexity pushes us to develop a solution algorithm for actually dealing the real cases of this problem. The solution algorithm consists primarily of three parts: Preprocessing, generating cuts, and the relaxation-based heuristic. Empirical results express that the algorithm can produce the solution efficiently while maintaining the good tracking ability of its solution.

From our computational results (omitted in this report), we can see that each portfolio tracks the index precisely during the first month after its construction. On account of this phenomenon, when rebalancing is allowed, we suggest that once the index fund is set up, exploit the newly arriving data together with model *GP* to engender a revised portfolio every month, and subsequently rebalance the index fund conforming to this revised portfolio. Such a rebalancing strategy is easily implemented. We surmise this rebalancing strategy may render the tracking performance of the index fund become better.

Reference

Andrews, C., D. Ford and K. Mallinson, "The design of index funds and alternative methods of replication," *The Investment Analyst* **82** (October), 16-23 (1986).

Charnes, A., W. W. Cooper, and R. O. Ferguson, "Optimal estimation executive compensation by linear programming," *Management Science* **1**, 138-151 (1955).

Givant, M., "Index fund returns not exact match," *Pensions & Investment Age* (May), 17-19 (1988).

Kirby, R. G., "The key word is 'passive'," *Journal of Investing* (Spring), 21 (1993).

Kornbluth, J. S. H., N. Meade and G. R. Salkin, "Bond index funds-a duration approach," *Journal of Operational Research Society* **44**, 945-953 (1993).

Liu, M. L., "Construct Index Fund vio Integer Goal Programming", Technical report, NSC89-2416-H-004-034, 2000.

Lodge, S., "Tracking the market," *Investors Chronicle* **104**, 12-13 (1993).

Meade, N. and G. R. Salkin, "Index funds-construction and performance measurement," *Journal of Operational Research Society* **40**, 871-879 (1989).

Roll, R., "A mean/variance analysis of tracking error," *Journal of Portfolio Management* (Summer), 13-23 (1992).

Rudd, A., "Optimal selection of passive portfolios," *Financial Management* (Spring), 57-66 (1980).

Tabata, Y. and E. Takeda, "Bicriteria

optimization problem of designing an index fund," *Journal of Operational Research Society* **46**, 1023-1032 (1995).

Toy, W. W. and M. A. Zurack, "Tracking the Euro-Pac Index," *Journal of Portfolio Management* (Winter), 55-58 (1989).

Willis, C., "Why every investor should own an index fund," *Working Woman* **21**, 34 (1996).