EDA with Switching Distributions for Long-Short Portfolio Replication Problems

Shunsuke Shibata*, Yukiko Orito†, Yoshiko Hanada‡ and Hisashi Yamamoto§
*Department of System Design, Tokyo Metropolitan University, Tokyo 191-0065, Japan
Email: shibata-shunsuke@ed.tmu.ac.jp
†Department of Economics, Hiroshima University, Hiroshima 739-8525, Japan
Email: orito@hiroshima-u.ac.jp
‡Faculty of Engineering Science, Kansai University, Osaka 564-8680, Japan
Email: hanada@kansai-u.ac.jp
§Department of System Design, Tokyo Metropolitan University, Tokyo 191-0065, Japan
Email: yamamoto@sd.tmu.ac.jp

***Abstract*— It is desired to replicate the benchmark portfolio when it has delivered good performances. In this paper, our focus is on the portfolio replication problem that the total return of the benchmark portfolio is opened to the public but the proportion-weighted combination is closed to the public. It is difficult to solve this replication problem because we cannot have any techniques to solve the simultaneous equations when the number of unknown valuables is more than the number of equations. In order to solve such a problem, we propose the new Estimation of Distribution Algorithm with the operation switching two distributions in this paper. In the numerical experiments, we show that the portfolios replicated by our proposing algorithm have delivered good performances even in the future periods.**

Keywords- Long-Short Portfolio Replication, Estimation of Distribution Algorithm, Switching Distributions

I. INTRODUCTION

The portfolio optimization problem is to determine the proportion-weighted combination in a portfolio in order to achieve the given investment targets. Many researchers have used various evolutionary methods such as Genetic Algorithms, Simulated Annealing, Tabu Search, Local Search, Memetic Algorithm, Estimation of Distribution Algorithm (EDA) to optimize their portfolios [1],[2],[3],[4]. In all above works, the portfolios consist of the assets with long positions in which they have bought and been held. Such a portfolio is called a long-only portfolio.

In this paper, we are focusing on the long-short portfolio which consists of the assets with not only long positions but also short positions in which they have been borrowed and sold. From a practical viewpoint, an asset management firm (Company A) desires to make the replication portfolio to the portfolio of another firm (Company B) when the portfolio of Company B has delivered better performances than the portfolio of Company A. However, Company B opens only the total performance of portfolio to the public, but other information such as the assets included in the portfolio, the proportion-weighted combination and the rebalancing date is closed to the public. Hence, it is difficult for Company A to replicate the portfolio of Company B.

In this paper, we try to replicate "the replication portfolio" such that its return mimics the return of "the benchmark portfolio" which we want to replicate.

In the portfolio replication problems, however, we need to not only mimic the returns to the benchmark portfolio but also replicate the proportion-weighted combination. When a replication portfolio is an indeterminate solution in a group of solutions which has more than one solution, it is hard to solve the portfolio replication problem. Therefore, we have an important problem that the optimized replication portfolio may differ from the benchmark portfolio although the return of the replication portfolio mimics the return of benchmark portfolio.

In order to avoid this problem, we propose a new EDA with switching the probabilistic distributions. Here, our EDA is an expanded algorithm from the probabilistic model-building genetic algorithm with histograms (Orito et al. [5]). It generates the offspring population by using a probabilistic model given by the parents' population instead of the traditional crossover and mutation operations. In our EDA, the weights of the individual assets of the offspring population are determined by the individual probabilistic distributions from each asset of the parents' population when we assume that the histogram of the individual assets in the parents' population is assigned to the distribution for the individual assets in the offspring population. This means that the weight of each of assets included in the portfolio is determined by each of distributions. Our EDA switches the distributions between two assets with strong correlation of their returns in order to avoid converging to an indeterminate solution.

From the results of the numerical experiments, we show that our EDA can optimize the better replication portfolios than the traditional EDA does. In addition, the returns of the replication portfolio obtained by our EDA mimic the returns of benchmark portfolio in its future period even when our replication portfolio does not consist of the same proportion-weighted combination as the benchmark portfolio.

978-1-4799-0652-9/13 $31.00 © 2013 IEEE
DOI 10.1109/SMC.2013.79

II. LONG-SHORT PORTFOLIO REPLICATION PROBLEM

We describe the long-short portfolio replication problem in this section.

In this paper, we try to replicate the portfolio of the same proportion-weighted combination as a given benchmark portfolio. We first define the following notations.

N : The number of assets in a replication/benchmark portfolio.

i : Asset i, $i = 1, \cdots, N$.

t : Time basis, $t = 1, \cdots, T$.

$r_i(t)$: Return of Asset i at t.

w_i : Weight of Asset i included in a replication portfolio.

$\mathbf{w}$: Replication portfolio. That is $\mathbf{w} = (w_1, \cdots, w_N)$.

w_i^L : Weight of Asset i with long position included in the replication portfolio $\mathbf{w}$.

w_i^S : Weight of Asset i with short position included in the replication portfolio $\mathbf{w}$.

$\mathbf{w}^L$: Proportion-weighted combination with long position. That is $\mathbf{w}^L = (w_1^L, \cdots, w_N^L)$.

$\mathbf{w}^S$: Proportion-weighted combination with short position. That is $\mathbf{w}^S = (w_1^S, \cdots, w_N^S)$.

$r_{\mathbf{w}}(t)$: Return of a long-short portfolio $\mathbf{w}$ at t.

$\mathbf{w}^B$: Benchmark portfolio. That is $\mathbf{w}^B = (w_1^B, \cdots, w_N^B)$. This portfolio is not given as the usable information in the optimization processes.

$r_{\mathbf{w}^B}(t)$: Return of benchmark portfolio at t. Only this return is given as the usable information in the optimization processes.

α : Financial leverage.

$E_{\mathbf{w}}$: The evaluation value for the replication portfolio.

A. Long-Short Portfolio

In this paper, we focus on a long-short portfolio replication with both of long and short positions. In the long-short portfolio, the proportion-weighted combination consists of the positive real values with long position and the negative real values with short position. Generally, a long-short portfolio consisting of N assets is represented as

$$\mathbf{w} = (w_1, \cdots, w_N), \quad \text{s.t.} \sum_{i=1}^{N} w_i = 1 \qquad (-1 \le w_i \le 1). \tag{1}$$

The equality constraint given by the equation (1) is defined under the condition that we sell the assets with short position and purchase other assets with long position by the funds for short selling. So, a feasible solution must be satisfied the equality constraint given by the equation (1). However, it is hard to make the feasible solution because the positive (negative) weights of feasible solution depend on the sum of negative (positive) weights because of the equality constraint.

In order to avoid this problem, we re-define the long-short portfolio as follows.

$$\begin{aligned} &\mathbf{w} = (w_1^L - \alpha w_1^S, \cdots, w_N^L - \alpha w_1^S), \\ &\mathbf{w}^L = (w_1^L, \cdots, w_N^L), \quad \mathbf{w}^S = (w_1^S, \cdots, w_N^S). \\ &\text{s.t.} \sum_{i=1}^{N} w_i^L = 1, \ \sum_{i=1}^{N} w_i^S = 1, \ 0 \le w_i^L \le 1, \ 0 \le w_i^S \le 1 \end{aligned} \tag{2}$$

where α is a financial leverage between the individual assets with long position and with short position.

In the equation (2), weights of Asset i with long position and with short position can take on the ranges of $0 < w_i^L$ and $0 < w_i^S$, respectively. This equation has an advantage that a weight of asset with long (short) position does not depend on with short (long) position. In this paper, we optimize the long-short portfolio defined by the equation (2).

Note that the portfolio with $\alpha = 1$ means the long-short portfolio in this paper. On the other hand, the portfolios with $\alpha = 0$ and $\alpha = 2$ mean the long-only and the short-only portfolios, respectively. In the numerical experiments, we discuss the results of only the long-short portfolios with $\alpha = 1$.

B. Objective Function

In this paper, we assume that the total return of benchmark portfolio and the assets are given as the usable information for our replication problem. That is, the assets of benchmark portfolio are opened to the public, but the weights of assets are closed to the public.

In this context, the returns of replication portfolio consisting of N assets between $t = 1$ and $t = T$ are represented as

$$\begin{pmatrix} r_{\mathbf{w}}(1) \\ \vdots \\ r_{\mathbf{w}}(T) \end{pmatrix} = \begin{pmatrix} r_1(1) & \cdots & r_N(1) \\ & \vdots & \\ r_1(T) & \cdots & r_N(T) \end{pmatrix} \begin{pmatrix} w_1^L - \alpha w_1^S \\ \vdots \\ w_N^L - \alpha w_N^S \end{pmatrix}. \tag{3}$$

In the equation (3), $(r_{\mathbf{w}}(1), \cdots, r_{\mathbf{w}}(T))$ is given the data series because the replication portfolio has the same returns as the benchmark portfolio. The returns of individual assets $(r_i(1), \cdots, r_i(T))$ $(i = 1, \cdots, N)$ are also given. Hence, if the number of days data T is equal to the number of assets N, we can obtain the optimal solution $\mathbf{w}$ by solving the simultaneous equations. However, the proportion-weighted combination of benchmark portfolio changes on its rebalancing to keep track of the performances in the future periods. The benchmark portfolio is fixed only in a short periods, so T might be very less than N. In the case studies of $T < N$, it is difficult to find the real optimal solution $\mathbf{w}$.

In order to avoid this problem, we try to optimize the replication portfolio such that its return mimics the return of benchmark portfolio. For evaluating the replication portfolios, we employ the following evaluation value consisting of the errors of returns and the rates of changes of returns between replication and benchmark portfolios as the objective function.

$$E_{\mathbf{w}} = \sum_{t=1}^{T} \left(r_{\mathbf{w}}(t) - r_{\mathbf{w}^B}(t) \right)^2 + \rho \sum_{t=1}^{T-1} \left(1 - \frac{r_{\mathbf{w}}(t+1) - r_{\mathbf{w}}(t)}{r_{\mathbf{w}^B}(t+1) - r_{\mathbf{w}^B}(t)} \right)^2, \tag{4}$$

where ρ is a parameter given in the experiments.

In this paper, therefore, we define the portfolio replication problem as follows.

$$\min_{\mathbf{w}} E_{\mathbf{w}} . \tag{5}$$

$$\text{s.t. } \sum_{i=1}^{N} w_i^L = 1, \sum_{i=1}^{N} w_i^S = 1, 0 \le w_i^L \le 1, 0 \le w_i^S \le 1 .$$

III. EDA WITH SWITCHING DISTRIBUTIONS

Among the current applications of evolutionary algorithms to the optimization problems we find: PBIL (Population-Based Incremental Learning) [6], Compact GA (Compact Genetic Algorithm) [7], the Probabilistic Model-building GA with histograms [5], [8]. These algorithms are useful for the problems that do not depend on the relationship between the variables in a solution.

In this paper, we propose a new EDA with switching the probabilistic distributions. Our EDA is an expanded algorithm from the Probabilistic Model-building GA with histograms [5]. In our EDA, the weights of the individual assets of the offspring population are determined by the individual probabilistic distribution from each asset of the parents' population when we assume that the histogram of the individual assets in the parents' population is assigned to the probabilistic distribution for the individual assets in the offspring population. This means that the weight of each of assets is determined by each of probabilistic distributions.

In the portfolio replication problem, however, we need to not only mimic the returns to the benchmark portfolio but also replicate the proportion-weighted combination. When a replication portfolio is an indeterminate solution in a group of solutions which has more than one solution, it is hard to solve the portfolio replication problem. In this context, our EDA switches the probabilistic distributions between the two assets with strong correlation of their returns in order to avoid converging to an indeterminate solution.

A. Genetic Representation and Evaluation Value

In a long-short portfolio replication problem, a solution means portfolio and a variable of the solution means a weight of each asset included in a portfolio.

In our EDA, j -th solution on the l -th generation's population is represented as

$$\mathbf{w}^{(l,j)} = \left(w_1^{L(i,j)} - \alpha\, w_1^{S(i,j)}, \cdots, w_N^{L(i,j)} - \alpha\, w_1^{S(i,j)}\right),$$

$$\mathbf{w}^{L(l,j)} = \left(w_1^{L(i,j)}, \cdots, w_N^{L(i,j)}\right),\ \mathbf{w}^{S(l,j)} = \left(w_1^{S(i,j)}, \cdots, w_N^{S(i,j)}\right). \tag{6}$$

$$\text{s.t. } \sum_{i=1}^{N} w_i^{L(i,j)} = 1, \sum_{i=1}^{N} w_i^{S(i,j)} = 1,\ 0 \le w_i^{L(i,j)} \le 1, 0 \le w_i^{S(i,j)} \le 1 .$$

The evaluation value is as follows.

$$E_{\mathbf{w}^{(l,j)}} = \sum_{t=1}^{T} \left(r_{\mathbf{w}^{(l,j)}}(t) - r_{\mathbf{w}^B}(t)\right)^2 + \rho \sum_{t=1}^{T-1} \left(1 - \frac{r_{\mathbf{w}^{(l,j)}}(t+1) - r_{\mathbf{w}^{(l,j)}}(t)}{r_{\mathbf{w}^B}(t+1) - r_{\mathbf{w}^B}(t)}\right)^2 . \tag{7}$$

B. Procedure of EDA

From the equation (6), the solution $\mathbf{w}^{(l,j)}$ is derived by the proportion-weighted combinations consisting of long weights of $\mathbf{w}^{L(l,j)}$ and short weights of $\mathbf{w}^{S(l,j)}$. In the processes below, our EDA optimizes to each of $\mathbf{w}^{L(l,j)}$ and $\mathbf{w}^{S(l,j)}$ in order to obtain the optimal portfolio $\mathbf{w}^{(l,j)}$.

The procedure of our EDA, as shown in Fig. 1, consists of the processes from Step (1) to Step (6).

1) Initial State

At the initial generation of $l = 0$, the EDA randomly generates M_{pop} solutions in the initial population $\left\{\mathbf{w}^{(0,j)} \middle| j = 1, \cdots, M_{pop}\right\}$. Note that the weight of the individual assets in the initial population is randomly given under the constraints $\sum_{i=1}^{N} w_i^L = 1$ or $\sum_{i=1}^{N} w_i^S = 1$ with long or short position, respectively.

2) Histograms of Parents' Population

For the long position, the proportion-weighted combination $\mathbf{w}^{L(l,j)}$ on the l -th generation, the EDA divides the interval of the weight of an asset on the range of $[0,1]$ into H discrete intervals. A histogram consists of M_{pop} solutions in the parents' population. Thereby, Bin h $(h = 1, \cdots, H)$ is a bin of the discrete interval on the range of $\left[\frac{h-1}{H}, \frac{h}{H}\right]$. The histogram with frequencies $v_i^{L(l)}[h]$ to h is defined as

$$v_i^{L(l)}[h] = \#\left\{ j \middle| \frac{h-1}{H} \le w_i^{L(i,j)} \le \frac{h}{H} \right\}. \tag{8}$$

$$\left(j \in 1, \cdots, M_{pop},\ i = 1, \cdots, N,\ h = 1, \cdots, H\right)$$

As similar to equation (8), for the weights of short position $\mathbf{w}^{S(l,j)}$, the histogram with frequencies $v_i^{S(l)}[h]$ to h is defined as

$$v_i^{S(l)}[h] = \#\left\{ j \middle| \frac{h-1}{H} \le w_i^{S(i,j)} \le \frac{h}{H} \right\}. \tag{9}$$

$$\left(j \in 1, \cdots, M_{pop},\ i = 1, \cdots, N,\ h = 1, \cdots, H\right)$$

3) Probabilistic Model for Offspring Population

In the histograms defined by equations (8) and (9), the frequency of bin h is 0 if there is no weight to hold the selectable probabilities of bin h. Our EDA gives σ as a minimum for each of all bins.

Hence, for the weights of long position $\mathbf{w}^{L(l,j)}$, the probability of Asset i to h on the l -th generation is defined as

$$p_i^{L(l)}[h] = \frac{1}{M_{pop} + H \cdot \sigma}\left(\sigma + v_i^{L(l)}[h]\right). \tag{10}$$

$$\left(i = 1, \cdots, N,\ h = 1, \cdots, H\right)$$

On the other hand, for the weights of short position $\mathbf{w}^{S(l,j)}$, the probability of Asset i to h on the l -th

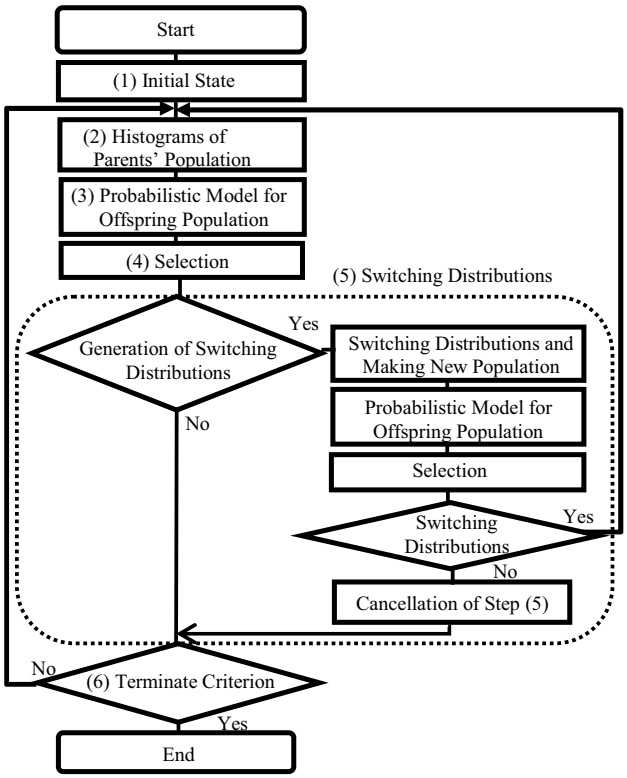

Fig. 1. Algorithm of EDA

generation is defined as

$$p_i^{S(l)}[h] = \frac{1}{M_{pop} + H \cdot \sigma}\left(\sigma + v_i^{S(l)}[h]\right). \tag{11}$$

$$\left(i = 1, \cdots, N,\ h = 1, \cdots, H\right)$$

The individual weights with long or short position are randomly selected according to the distributions given by the equation (10) or (11). Based on both these weights, the total weights of solutions are calculated by equation (6). These new solutions are offspring.

4) Selection

M_{pop} solutions are selected by the elitism selection and the roulette wheel selection from the current parents' and offspring populations for the next generation.

5) Switching Distributions and Making New Population

When a replication portfolio is an indeterminate solution, it is hard to solve the portfolio replication problem. Our EDA switches the distributions between two assets with strong correlation of their returns in order to avoid converging to an indeterminate solution.

Since two assets with a strong correlation have a similar path of their returns, it is expected to improve the current solutions by switching their distributions without breaking the structure of the solutions.

The way of switching the distributions is described as follows.

(a) We calculate the correlation coefficients of the returns between the pairs of all the two assets. We renumber the pairs in a descending order of correlation coefficient, $\{i_{1,1}, i_{1,2}\}$, $\{i_{2,1}, i_{2,2}\}, \cdots$.

(b) Let L_{change} be an initial generation number of switching distributions and δ be the interval of generations for the switching operation. When the generation of the EDA reaches to $l = L_{change} + k \cdot \delta$, the distributions between two assets of $\{i_{k,1}, i_{k,2}\}$ are switched. And the EDA adds 1 to the repetition number of switching operation k.

(c) The EDA re-calculates the evaluation values of all solutions of the renewed parents' population, makes offspring again using the operations of Step (3), and selects new population. If the evaluation value of the renewed best solution is lower than the value of the current best solution, the renewed population is accepted as the new parents' population for the next generations.

6) Terminate Criterion

The EDA repeats from the Step (2) to (5) until the maximal number of the repetitions, $l = L_{\max}$, is satisfied. From the last population, we select one solution that has the lowest evaluation value of all. This solution is the optimum or quasi-optimum long-short replication portfolio obtained by our EDA.

IV. Numerical Experiments

In the numerical experiments, we employ N assets with high turnover on the Tokyo Stock Exchange as assets included in a replication and benchmark portfolios. Each data phase consists of 10 days data for the past period $(t = 1, \cdots, 10)$ and 100 days data in the future period $(t = 11, \cdots, 110)$ from 2005 to 2010. We call them Phase 1 through Phase 13 respectively.

Here, we discuss the results of only the long-short portfolios with $\alpha = 1$. We call our proposing EDA with switching distributions "EDA1" and the traditional EDA without switching distributions "EDA2", respectively.

A. Benchmark Portfolio

In the numerical experiments, the benchmark portfolio $\mathbf{w}^B$ is represented as

$$\mathbf{w}^B = \left(w_1^{B,L} - \alpha\, w_1^{B,S}, \cdots, w_N^{B,L} - \alpha\, w_1^{B,S}\right), \tag{12}$$

$$\mathbf{w}^{B,L} = \left(w_1^{B,L}, \cdots, w_N^{B,L}\right), \quad \mathbf{w}^{B,S} = \left(w_1^{B,S}, \cdots, w_N^{B,S}\right).$$

$$\text{s.t. } \sum_{i=1}^{N} w_i^{B,L} = 1, \sum_{i=1}^{N} w_i^{B,S} = 1, 0 \le w_i^{B,L} \le 1, 0 \le w_i^{B,S} \le 1$$

Of course, this benchmark portfolio is not given as the usable information in the optimization processes. We can use only the total return of this benchmark portfolio as the usable information.

B. Parameters of EDA1 and EDA2

In the numerical experiments, the parameters of the EDA1 and EDA2 are set as follows.

The number of assets in portfolio: $N = 100$

TABLE I. RESULTS OF EDA 1 AND EDA2

Phase	EDA1			EDA2		
	Evaluation value	*Error of returns*	*Rate of changes*	*Evaluation value*	*Error of returns*	*Rate of changes*
1	6.62E-08	6.57E-08	5.09E-02	2.59E-06	2.54E-06	5.31E+00
2	5.80E-07	5.19E-07	6.07E+00	2.15E-06	1.23E-06	9.18E+01
3	4.33E-08	4.31E-08	2.23E-02	1.24E-06	1.23E-06	7.96E-01
4	1.03E-07	1.01E-07	1.44E-01	1.15E-06	1.07E-06	8.31E+00
5	1.43E-08	1.41E-08	2.42E-02	4.53E-07	4.23E-07	2.98E+00
6	7.39E-08	7.02E-08	3.71E-01	1.38E-06	1.14E-06	2.36E+01
7	5.33E-07	4.93E-07	4.03E+00	5.65E-06	3.83E-06	1.82E+02
8	8.59E-08	8.58E-08	1.21E-02	3.38E-06	3.37E-06	9.59E-01
9	4.61E-08	4.49E-08	1.30E-01	1.32E-06	1.22E-06	1.06E+01
10	7.58E-08	7.54E-08	3.45E-02	2.48E-06	2.47E-06	6.11E-01
11	2.54E-08	2.51E-08	2.49E-02	8.57E-07	8.33E-07	2.41E+00
12	8.83E-08	8.82E-08	9.15E-03	1.32E-06	1.31E-06	5.95E-01
13	4.39E-08	4.17E-08	2.18E-01	1.10E-06	1.09E-06	6.77E-01

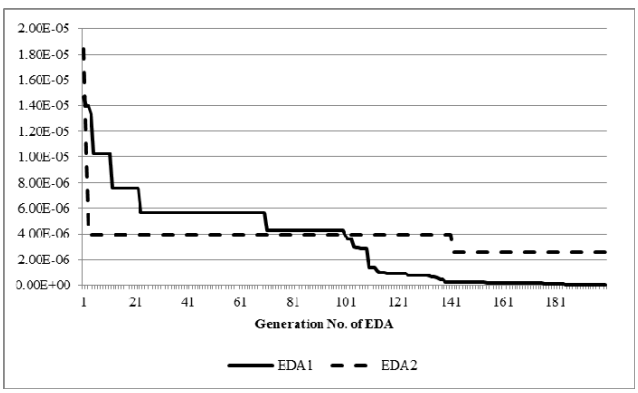

Fig. 2. Evaluation values of replication portfolios obtained by EDAs 1 and 2 as functions of generations

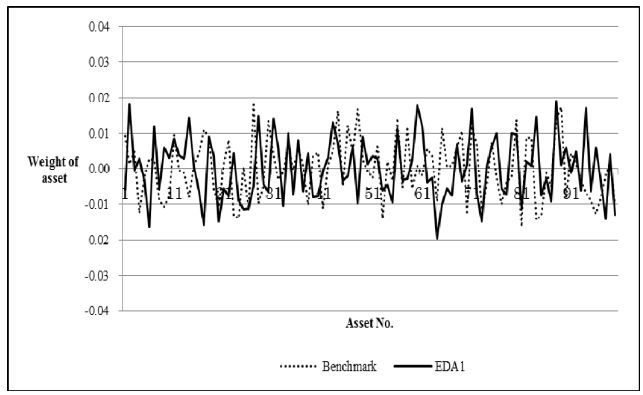

Fig. 3. Weights of benchmark portfolio and replication portfolio obtained by EDA1

The number of days data in the past period: $T = 10$
Financial leverage: $\alpha = 1$ (Long-Short Portfolio)
Weight parameter of equation (7): $\rho = 1.0E - 08$
Population size: $M_{pop} = 100$
Offspring population size: $M_{off} = 2 \cdot M_{pop}$
The number of bins: $H = 500$ (The width of the search space of a bin is $1/500 = 0.002$.)
Minimal height of bin of equations (10) and (11): $\sigma = 5$
Initial generation of switching distributions: $L_{change} = 100$
The interval of generation for switching distributions: $\delta = 1$
Generation size: $L_{\max} = 200$
Algorithm run: 10

C. *Effectiveness of EDA1*

In the numerical experiments, we have adopted the optimization algorithms, EDA1 and EDA2, in order to optimize the replication portfolios to the benchmark portfolio. For each phase, the evaluation value, the error of returns and the rate of changes of returns between replication and benchmark portfolios are shown in Table I. In addition, for Phase 1, the evaluation values of EDA1 and EDA2 as functions of generations of algorithms are shown in Fig. 2.

Table I says that, through all the phases, the evaluation value, the error of returns and the rates of changes of returns obtained by EDA1 is lower than those of EDA2. In other words, EDA1 can optimize the better replication portfolio such that its return mimics the return of benchmark portfolio than EDA2 does. Fig. 2 says that the evaluation value of EDA1 on the last generation is lower than that of EDA2. Especially, the value of EDA1 converges to 0 after on the 100th generation. For other phases, we have obtained similar results to those for Phase 1. Therefore, we can conclude that the operation of switching distributions of EDA1 (Step (5) of EDA1) works well to find the better replication portfolio than that of the traditional EDA, EDA2.

The proportion-weighted combinations of the replication and the benchmark portfolios are shown in Fig. 3. The horizontal axis gives the Asset number i, $i = 1, \cdots, N$, and the vertical axis gives the weight of Asset i. On the other hand,

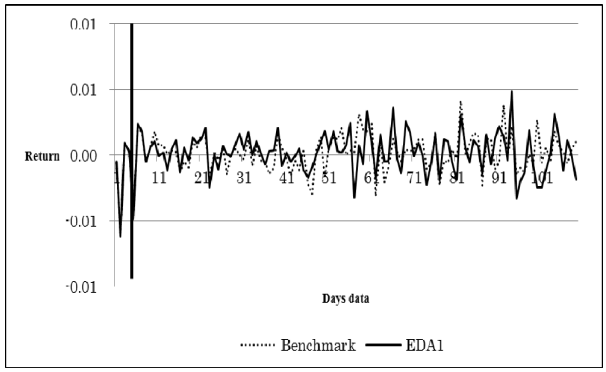

Fig. 4. Returns of benchmark portfolio and replication portfolio obtained by EDA1

for Phase 1, the returns of benchmark portfolio and the replication portfolios obtained by EDA1 are shown in Fig. 4. Here, on the horizontal axis, the days data from $t=1$ to $t=10$ is on the past period for optimizing the replication portfolios and the data from $t=11$ to $t=110$ is on their future period.

The results of Figs. 3 and 4 say that the returns of the replication portfolios are almost the same as the returns of benchmark portfolio in the past period though the EDA1 could not replicate the benchmark portfolio. In the future period, however, the returns of the replication portfolios differ from those of benchmark portfolio. We will discuss this problem in the next section.

D. Evaluating Replication Portfolios in the Future Period

The results of Figs. 3 and 4 suggest that EDA1 can make the replication portfolio whose returns are almost the same as the returns of benchmark portfolio, but cannot replicate the benchmark portfolio.

In order to improve the replication portfolios' performances in the future periods, we optimize the replication portfolio by using EDA1 in the past period consisting of longer days data $T=90$ than $T=10$. For each phase, for $T=10$ and $T=90$, the error sum of squares of returns between replication and benchmark portfolios in the future period are shown in Table II, respectively.

Table II says that, through all the phases, the error sum of squares in the future period for $T=90$ is lower than those for $T=10$ except Phase 9. The replication portfolios obtained by EDA1 has a good ability to mimic the return of benchmark portfolio even in the future periods when we can adopt much past data for optimizing the replication portfolios.

V. CONCLUSIONS

The long-short portfolio replication problem is a very difficult problem that it's required to not only mimic the return to the benchmark portfolio but also replicate the proportion-weighted combination.

For solving such a problem, we have proposed the new EDA switching the probabilistic distributions of two assets with a strong correlation of their returns in this paper.

TABLE II. ERROR SUM OF SQUARES OF RETURNS IN THE FUTURE PERIOD

Phase	T=10	T=90
1	2.04E-04	7.26E-05
2	6.24E-04	3.95E-04
3	1.77E-04	1.59E-04
4	2.20E-04	1.18E-04
5	3.95E-04	2.27E-04
6	5.40E-04	2.38E-04
7	5.12E-04	3.88E-04
8	4.97E-04	1.94E-04
9	1.24E-03	1.43E-03
10	6.39E-04	3.97E-04
11	3.71E-04	2.17E-04
12	3.18E-04	1.80E-04
13	1.76E-04	1.25E-04

In the numerical experiments, we showed that our EDA works well to find the better replication portfolios than those of the traditional EDA. In addition, we clarified that our EDA can make the replication portfolio whose returns are the same as those of benchmark portfolio even in the future period.

ACKNOWLEDGMENT

This research was partially supported by Grant-in-Aid for Young Scientists (B) #25730148 and #24700234 from Japan Society for the Promotion of Science (JSPS).

REFERENCES

[1] Y. Xia, B. Liu, S. Wang, and K. K. Lai, "A Model for Portfolio Selection with Order of Expected Returns", Computers & Operations Research, vol. 27, pp. 409–422, 2000.

[2] C. C. Lin and Y. T. Liu, "Genetic Algorithms for Portfolio Selection Problems with Minimum Transaction Lots", European Journal of Operational Research, vol. 185–1, pp. 393–404, 2008.

[3] T. J. Chang, N. Meade, J. E. Beasley and Y. M. Sharaiha, "Heuristics for Cardinality Constrained Portfolio Optimization", Computers & Operations Research, vol. 27, pp. 1271–1302, 2000.

[4] Y. Crama and M. Schyns, "Simulated Annealing for Complex Portfolio Selection Problems", European Journal of Operational Research, vol.150, pp. 546–571, 2003.

[5] Y. Orito, H. Yamamoto, and Y. Tsujimura, "Equality Constrained Long-Short Portfolio Replication by Using Probabilistic Model-building GA", Proceedings of WCCI 2012 IEEE World Congress on Computational Intelligence, IEEE Congress on Evolutionary Computation, pp.513-520, 2012.

[6] S. Baluja, "Population-based Incremental Learning: A Method for Integrating Genetic Search Based Function Optimization and Competitive Learning", Technical Report, No. CMU-CS-94-163, Carnegie Mellon University, 1994.

[7] G. R. Harik, F. G. Lobo, and D. E. Goldberg, "The Compact Genetic Algorithm", Technical Report, No. 97006, IlliGAL Report, 1997.

[8] S. Tsutsui, M. Pelikan, and D. E. Goldberg, "Probabilistic Model-building Genetic Algorithms Using Marginal Histograms in Continuous Domain", Proceedings of the International Conference on Knowledge-Based and Intelligent Information & Engineering Systems 2001, pp. 112-121, 2001