Evolutionary Transfer Optimization for Diversified Investment Using Predictive Distribution of Asset Returns

Kiyoharu Tagawa
Department of Informatics
Faculty of Informatics, Kindai University
Higashi-Osaka, Osaka, Japan
e-mail: tagawa@info.kindai.ac.jp

Yukiko Orito
Department of Management Science
College of Engineering, Tamagawa University
Machida, Tokyo, Japan
e-mail: orito.y@eng.tamagawa.ac.jp

***Abstract*—Bayesian inference is a powerful mathematical tool that enables Artificial Intelligence (AI) to make probabilistic predictions and decisions in environments with incomplete data or high levels of noise. In this paper, the traditional portfolio optimization problem is extended to a data-driven one based on Bayesian inference. Specifically, a probabilistic model of asset returns is given by the predictive distribution and updated whenever a new data set of asset returns becomes available. Thus, each time the predictive distribution is updated, it becomes necessary to search for a new solution to the portfolio optimization problem. In order to solve the data-driven portfolio optimization problem iteratively, a novel evolutionary transfer optimization method is proposed and then its performance is evaluated using actual asset data from the Tokyo Stock Price Index (TOPIX17).**

Keywords-Artificial intelligence; Bayesian inference; portfolio optimization; evolutionary transfer optimization.

I. Introduction

Evolutionary Transfer Optimization (ETO) is a paradigm that integrates Evolutionary Algorithm (EA) with knowledge learning and transfer to achieve higher optimization efficiency and performance [1]. Specific ETO techniques include various EAs for dynamic optimization problems [2][3], EAs for optimization problems in uncertain environments [4], EAs for complex and large-scale problems [5], and EAs based on surrogate models [6][7]. Recently, inspired by the success of neural network transfer learning, attention has shifted toward ETO techniques. These ETO methods leverage information from solutions or parameter values of specific optimization problems to solve similar problems more efficiently.

Portfolio optimization is the process of determining the best portion of investment in different assets according to some objective. Since portfolio optimization is one of the core issues in the field of finance, numerous reports have been published on both its theoretical and practical aspects [8][9]. It should be noted that the problem setting differs between investment and speculation. In speculative investment, asset returns are treated as time-series data. Consequently, Artificial Intelligence (AI) techniques, such as Transformers [10] and Bayesian networks [11], are employed for their prediction. On the other hand, in asset investment, the asset returns are treated as random variables from the Efficient-Market Hypothesis (EMH) [12]. Therefore, in portfolio optimization problems, probabilistic models are usually used to predict asset returns [8][9].

This paper extends the traditional portfolio optimization problem based on the EMH to a data-driven one. Specifically, the predictive distribution in Bayesian inference [13] is used as a probabilistic model for asset returns. In the real world, as new data on asset returns is observed, predictions of them are constantly changing. The predictive distribution allows the probabilistic model for asset returns to be updated according to new data without losing information inherited from historical data. Furthermore, to update the predictive distribution in response to new observed data on asset returns and solve the data-driven portfolio optimization problem iteratively, a novel ETO method using a powerful EA, namely an Adaptive Differential Evolution (ADE) algorithm [14], is proposed. The usefulness of the proposed ETO method is verified through the diversified investment using actual data on asset returns from the Tokyo Stock Price Index (TOPIX17) [15].

The rest of this paper is organized as follows. Section II formulates a portfolio optimization problem based on EMH. Section III extends the portfolio optimization problem to the data-driven one by using the predictive distribution of asset returns. Section IV proposes a novel ETO method for solving the data-driven portfolio optimization problem. Section V evaluates the performance of the proposed ETO method by using actual asset data. Section VI concludes this paper.

II. Portfolio Optimization

Let $x_i = \mathbb{R}$, $i = 1, \cdots, D$ be the investment ratio for i-asset. The portfolio is defined as $\boldsymbol{x} = (x_1, \cdots, x_i, \cdots, x_D) \in \mathbb{R}^D$. One unit investment in i-asset yields a return $\xi_i \in \mathbb{R}$. From the EMH, the return $\xi_i \in \mathbb{R}$ of i-asset is considered to be a random variable following a Gaussian distribution. Thus, the vector $\boldsymbol{\xi} \in \mathbb{R}^D$ of them obeys a multivariate Gaussian distribution:

$$\boldsymbol{\xi} = (\xi_1, \cdots, \xi_i, \cdots, \xi_D) \sim \mathcal{N}(\boldsymbol{\xi} \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{1}$$

where $\boldsymbol{\mu} \in \mathbb{R}^D$ denotes the mean vector, $\boldsymbol{\Sigma} \in \mathbb{R}^{D\times D}$ is the covariance matrix, and $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1}$ is called the precision matrix.

From the asset returns $\boldsymbol{\xi} \in \mathbb{R}^D$ in (1), the return of the portfolio $\boldsymbol{x} \in \mathbb{R}^D$ is evaluated as $r = \boldsymbol{\xi}\,\boldsymbol{x}^T$. The return $r \in \mathbb{R}$ of the portfolio is also a random variable. Let $r_o \in \mathbb{R}$ be the target value for the return $r \in \mathbb{R}$. In the diversified investment considered in this paper, the optimal portfolio $\boldsymbol{x} \in \mathbb{R}^D$ that minimizes the risk $\alpha(\boldsymbol{x})$, or the probability of the return $r \in \mathbb{R}$

falling below the target value $r_o \in \mathbb{R}$, is sought. Therefore, the portfolio optimization problem is formulated as

$$\left[\begin{array}{ll} \min\limits_{\boldsymbol{x}} & \alpha(\boldsymbol{x}) = \Pr(r = \boldsymbol{\xi}\,\boldsymbol{x}^T \le r_o) \\ \text{sub. to} & x_1 + \cdots + x_i + \cdots + x_D = 1 \\ & 0 \le x_i \le 1,\ i = 1, \cdots, D \end{array}\right. \tag{2}$$

where $\Pr(\mathcal{E})$ is the probability that event $\mathcal{E}$ occurs.

III. Problem Formulation

A. Predictive Distribution

It is assumed that a new data set of asset returns $\boldsymbol{\xi}_n \in \mathbb{R}^D$, $n = 1, \cdots, N$ is observed over a period t as

$$\boldsymbol{\Theta}_t = \{\boldsymbol{\xi}_n \in \mathbb{R}^D \,|\, n = 1, \cdots, N\}. \tag{3}$$

The observed data set in (3) can be obtained periodically. However, the data size N is too small to statistically evaluate the risk $\alpha(\boldsymbol{x})$ in (2). Hence, a probabilistic model, specifically the predictive distribution, is used to obtain sufficient data for evaluating the risk. The observed data on asset returns $\boldsymbol{\xi}_n \in \boldsymbol{\Theta}_t$ in (3) obeys the Gaussian distribution in (1). Thus, the conjugate prior distribution for the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Lambda}$ of the distribution in (1) is a Gaussian-Wishart distribution:

$$(\boldsymbol{\mu},\, \boldsymbol{\Lambda}) \sim \mathcal{NW}(\boldsymbol{\mu},\, \boldsymbol{\Lambda} \,|\, \boldsymbol{m},\, \beta,\, \nu,\, \boldsymbol{W}) \tag{4}$$

where $\boldsymbol{m} \in \mathbb{R}^D$, $\beta \in \mathbb{R}$, $\nu > D - 1$, and $\boldsymbol{W} \in \mathbb{R}^{D \times D}$ [13].

Let $p(\boldsymbol{\xi} \,|\, \boldsymbol{\mu},\, \boldsymbol{\Lambda})$ be the Probability Distribution Function (PDF) of the Gaussian distribution in (1). Let $p(\boldsymbol{\mu},\, \boldsymbol{\Lambda})$ be the PDF of the prior distribution in (4). From Bayes' theorem, the PDF of the posterior distribution $p(\boldsymbol{\mu},\, \boldsymbol{\Lambda} \,|\, \boldsymbol{\Theta}_t)$ is

$$p(\boldsymbol{\mu},\, \boldsymbol{\Lambda} \,|\, \boldsymbol{\Theta}_t) \propto \prod_{n=1}^{N} p(\boldsymbol{\xi}_n \,|\, \boldsymbol{\mu}, \boldsymbol{\Theta})\, p(\boldsymbol{\mu},\, \boldsymbol{\Lambda}). \tag{5}$$

From (5), the posterior distribution becomes a Gaussian-Wishart distribution depending on $\boldsymbol{\xi}_n \in \boldsymbol{\Theta}_t$ in (3) as

$$(\boldsymbol{\mu},\, \boldsymbol{\Lambda}) \sim \mathcal{NW}(\boldsymbol{\mu},\, \boldsymbol{\Lambda} \,|\, \widehat{\boldsymbol{m}},\, \hat{\beta},\, \hat{\nu},\, \widehat{\boldsymbol{W}}) \tag{6}$$

where the parameters also depend on those in (4) as

$$\left(\begin{array}{rcl} \hat{\beta} & = & \beta + N \\ \hat{\boldsymbol{m}} & = & \dfrac{1}{\hat{\beta}} \left(\displaystyle\sum_{n=1}^{N} \boldsymbol{\xi}_n + \beta\, \boldsymbol{m} \right) \\ \widehat{\boldsymbol{W}}^{-1} & = & \displaystyle\sum_{n=1}^{N} \boldsymbol{\xi}_n\, \boldsymbol{\xi}_n^T + \beta\, \boldsymbol{m}\, \boldsymbol{m}^T \\ & & -\hat{\beta}\, \hat{\boldsymbol{m}}\, \hat{\boldsymbol{m}}^T + \boldsymbol{W}^{-1} \\ \hat{\nu} & = & \nu + N. \end{array}\right. \tag{7}$$

By marginalizing out the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Lambda}$ of Gaussian distribution $p(\hat{\boldsymbol{\xi}} \,|\, \boldsymbol{\mu},\, \boldsymbol{\Lambda})$ with the posterior one in (5) as

$$p(\hat{\boldsymbol{\xi}} \,|\, \boldsymbol{\Theta}_t) = \iint p(\hat{\boldsymbol{\xi}} \,|\, \boldsymbol{\mu},\, \boldsymbol{\Lambda})\, p(\boldsymbol{\mu},\, \boldsymbol{\Lambda} \,|\, \boldsymbol{\Theta}_t)\, d\boldsymbol{\mu}\, d\boldsymbol{\Lambda}, \tag{8}$$

the PDF of the predictive distribution $p(\hat{\boldsymbol{\xi}} \,|\, \boldsymbol{\Theta}_t)$ is obtained. From (8), the predictive distribution becomes a multivariate Student's t-distribution depending on $\boldsymbol{\xi}_n \in \boldsymbol{\Theta}_t$ in (3) as

$$\hat{\boldsymbol{\xi}} \sim \mathcal{S}(\hat{\boldsymbol{\xi}} \,|\, \boldsymbol{\mu}_s,\, \boldsymbol{\Lambda}_s,\, \nu_s) \tag{9}$$

where $\hat{\boldsymbol{\xi}} \in \mathbb{R}^D$ is the predicted value for the return $\boldsymbol{\xi} \in \mathbb{R}^D$ in (1). The parameters are obtained from those in (6) as

$$\left(\begin{array}{rcl} \boldsymbol{\mu}_s & = & \hat{\boldsymbol{m}} \\ \boldsymbol{\Lambda}_s & = & \dfrac{(1 - D - \hat{\nu})\, \hat{\beta}}{1 + \hat{\beta}}\, \widehat{\boldsymbol{W}} \\ \nu_s & = & 1 - D + \hat{\nu}. \end{array}\right. \tag{10}$$

B. Data-Driven Portfolio Optimization

By using the predictive distribution in (9), the predicted data set of asset returns is generated randomly as

$$\widehat{\boldsymbol{\Theta}}_t = \{\hat{\boldsymbol{\xi}}_m \in \mathbb{R}^D \,|\, m = 1, \cdots, M\} \tag{11}$$

where the data size M, $M > N$ is large enough.

From the predicted data set of asset returns $\hat{\boldsymbol{\xi}}_m \in \widehat{\boldsymbol{\Theta}}_t$ in (11), the Empirical Cumulative Distribution Function (ECDF) for the return $r \in \mathbb{R}$ of the portfolio $\boldsymbol{x} \in \mathbb{R}^D$ is defined as

$$F(r \,|\, \boldsymbol{x},\, \widehat{\boldsymbol{\Theta}}_t) = \frac{1}{M} \sum_{m=1}^{M} \mathbb{1}\left(\hat{\boldsymbol{\xi}}_m\, \boldsymbol{x}^T \le r \right) \tag{12}$$

where $\mathbb{1}(\mathcal{E}) = 1$ if $\mathcal{E}$ is satisfied; otherwise $\mathbb{1}(\mathcal{E}) = 0$.

By using the ECDF in (12), the risk $\alpha(\boldsymbol{x})$ in (2) can be estimated from the predicted data set $\widehat{\boldsymbol{\Theta}}_t$ in (11) as

$$\alpha(\boldsymbol{x}) = \Pr(r = \boldsymbol{\xi}\, \boldsymbol{x}^T \le r_o) \simeq F(r_o \,|\, \boldsymbol{x},\, \widehat{\boldsymbol{\Theta}}_t) \tag{13}$$

From (13), the portfolio optimization problem in (2) is formulated as a data-driven optimization problem as

$$\left[\begin{array}{ll} \min\limits_{\boldsymbol{x}} & \alpha(\boldsymbol{x}) = F(r_o \,|\, \boldsymbol{x},\, \widehat{\boldsymbol{\Theta}}_t) \\ \text{sub. to} & x_1 + \cdots + x_i + \cdots + x_D = 1 \\ & 0 \le x_i \le 1,\ i = 1, \cdots, D. \end{array}\right. \tag{14}$$

The ECDF in (12) is composed of the predicted data set $\widehat{\boldsymbol{\Theta}}_t$ in (11) which is generated by using the predictive distribution in (9). As will be discussed later, the predictive distribution $p(\hat{\boldsymbol{\xi}} \,|\, \boldsymbol{\Theta}_t)$ in (8) incorporates information not only from the observed data $\boldsymbol{\xi}_n \in \boldsymbol{\Theta}_t$ but also from all past data. Therefore, the optimal solution of the portfolio optimization problem in (14) depends on all the data observed up to the period t.

IV. Evolutionary Transfer Optimization

A. Adaptive Differential Evolution

A powerful ADE algorithm called jDE [14] is used to solve the portfolio optimization problem in (14), which is defined by the data set $\widehat{\boldsymbol{\Theta}}_t$. ADE holds a set of individuals called the population $\boldsymbol{P}_{t,g}$. In each generation g, the current population $\boldsymbol{P}_{t,g}$ is updated to the next population $\boldsymbol{P}_{t,g+1}$. The termination condition of ADE is given by the maximum number of generations $\overline{g}$. Each individual $\boldsymbol{v}_k = (\boldsymbol{x}_k,\, F_k,\, CR_k) \in \boldsymbol{P}_{t,g}$, $k = 1, \cdots, N_P$ in the population is composed of three elements: $\boldsymbol{x}_k \in \mathbb{R}^D$ is a candidate solution to the portfolio optimization problem in (14); $F_k \in [0,\,1]$ and $CR_k \in [0,\,1]$ are used for the self-adaptive control parameters of ADE.

In order to decide the individuals $\hat{\boldsymbol{v}}_k \in \boldsymbol{P}_{t,g+1}$ in the next generation based on the current generation's ones $\boldsymbol{v}_k \in \boldsymbol{P}_{t,g}$, $k = 1, \cdots, N_P$, the following procedure is applied to each of the individuals $\boldsymbol{v}_k \in \boldsymbol{P}_{t,g}$, $k = 1, \cdots, N_P$ in turn where $\varepsilon_j \in [0,\,1]$ denotes the jth uniform random number.

Step 1: Three individuals $\boldsymbol{v}_a, \boldsymbol{v}_b, \boldsymbol{v}_c \in \boldsymbol{P}_{t,g}$, $a \neq b \neq c \neq k$ are selected at random from the current population $\boldsymbol{P}_{t,g}$ excluding the target one $\boldsymbol{v}_k = (\boldsymbol{x}_k, F_k, CR_k)$.

Step 2: Scale factor $F \in [0,\,1]$ is determined as

$$F = \begin{cases} 0.1 + 0.9\,\varepsilon_1 & \text{if } \varepsilon_2 < 0.1 \\ F_k & \text{otherwise.} \end{cases} \tag{15}$$

Step 3: Crossover rate $CR \in [0,\,1]$ is determined as

$$CR = \begin{cases} \varepsilon_3 & \text{if } \varepsilon_4 < 0.1 \\ CR_k & \text{otherwise.} \end{cases} \tag{16}$$

Step 4: Elements $z_{k,i} \in \mathbb{R}$, $i = 1, \cdots, D$ of a new solution $\boldsymbol{z}_k \in \mathbb{R}^D$ are generated from the four solutions as

$$z_{k,i} = \begin{cases} x_{a,i} + F\,(x_{b,i} - x_{c,i}) & \text{if } \varepsilon_{4+i} < CR \\ x_{k,i} & \text{otherwise.} \end{cases} \tag{17}$$

Furthermore, the new solution $\boldsymbol{z}_k = (z_{k,1}, \cdots, z_{k,D})$ is normalized for the equality constraint in (14) as

$$z_{k,i} = \frac{z_{k,i}}{z_{k,1} + \cdots + z_{k,D}},\ i = 1, \cdots, D. \tag{18}$$

Step 5: Let $\boldsymbol{u}_k = (\boldsymbol{z}_k, F, CR)$ be a new individual.

Step 6: The next individual $\hat{\boldsymbol{v}}_k \in \boldsymbol{P}_{t,g+1}$ is selected as

$$\hat{\boldsymbol{v}}_k = \begin{cases} \boldsymbol{u}_k & \text{if } \alpha(\boldsymbol{z}_k) \leq \alpha(\boldsymbol{x}_k) \\ \boldsymbol{v}_k & \text{otherwise} \end{cases} \tag{19}$$

where $\boldsymbol{u}_k = (\boldsymbol{z}_k, F, CR)$ and $\boldsymbol{v}_k = (\boldsymbol{x}_k, F_k, CR_k)$.

After repeating the above procedure until all individuals $\hat{\boldsymbol{v}}_k \in \boldsymbol{P}_{t,g+1}$ have been determined, the population is updated as $\boldsymbol{P}_{t,g} = \boldsymbol{P}_{t,g+1}$, $g = g+1$. The best solution in the final population $\boldsymbol{P}_{t,\overline{g}}$ is adopted as the solution to the portfolio optimization problem in (14), or the portfolio $\boldsymbol{x}_t^\star \in \mathbb{R}^D$.

B. Proposed ETO Method

In the proposed ETO method, ADE is applied iteratively to the portfolio optimization problem in (14) that is defined by the predicted data set $\widehat{\boldsymbol{\Theta}}_t$ in (11) for $t = 1, 2, \cdots$. The data set $\widehat{\boldsymbol{\Theta}}_t$ is generated by the predictive distribution $p(\hat{\boldsymbol{\xi}} \,|\, \boldsymbol{\Theta}_t)$ in (8). Each time a new data set $\boldsymbol{\Theta}_t$ in (3) is observed in period t, the predictive distribution $p(\hat{\boldsymbol{\xi}} \,|\, \boldsymbol{\Theta}_t)$ is updated via the prior and posterior distributions. If newly observed asset returns $\boldsymbol{\xi}_n \in \boldsymbol{\Theta}_{t+1}$, $n = 1, \cdots, N$ are considered to follow the current predictive distribution $p(\hat{\boldsymbol{\xi}} \,|\, \boldsymbol{\Theta}_t)$, some of superior individuals from the final population $\boldsymbol{P}_{t,\overline{g}}$ are transferred to the next initial one $\boldsymbol{P}_{t+1,0}$, and then ADE is applied to the portfolio optimization problem with the next data set $\widehat{\boldsymbol{\Theta}}_{t+1}$.

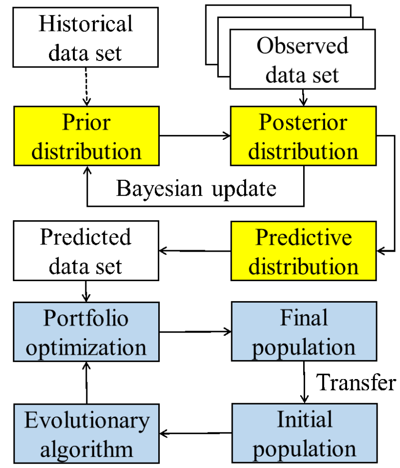


Figure 1. Flowchart of proposed ETO method.

Figure 1 shows the flowchart of the proposed ETO method to solve the data-driven portfolio optimization problem. The procedure of the ETO method is described as follows:

Step 1: Let $t = 1$. The individuals $\boldsymbol{v}_k \in \boldsymbol{P}_{t,0}$, $k = 1, \cdots, N_P$ are generated randomly as the initial population.

Step 2: The first prior distribution $p(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ is obtained by the likelihood estimation [16] from the historical data.

Step 3: From the observed data set $\boldsymbol{\Theta}_t$ in (3) and $p(\boldsymbol{\mu}, \boldsymbol{\Lambda})$, the posterior distribution $p(\boldsymbol{\mu}, \boldsymbol{\Lambda} \,|\, \boldsymbol{\Theta}_t)$ in (5) is obtained.

Step 4: From the posterior distribution in (5), the predictive one $p(\hat{\boldsymbol{\xi}} \,|\, \boldsymbol{\Theta}_t)$ for $\hat{\boldsymbol{\xi}} \in \mathbb{R}^D$ is obtained as shown in (8).

Step 5: Using the predictive distribution in (9), the predicted data set of asset returns $\hat{\boldsymbol{\xi}}_m \in \widehat{\boldsymbol{\Theta}}_t$ in (11) is generated.

Step 6: Using the ECDF in (12), the portfolio optimization problem in (14) is defined by the data set $\widehat{\boldsymbol{\Theta}}_t$ in (11).

Step 7: ADE is applied to the portfolio optimization problem. As a result, the best portfolio $\boldsymbol{x}_t^\star \in \mathbb{R}^D$ is obtained.

Step 8: Update the prior distribution in (4) by the posterior one in (6) as $\boldsymbol{m} = \hat{\boldsymbol{m}}$, $\beta = \hat{\beta}$, $\nu = \hat{\nu}$, and $\boldsymbol{W} = \widehat{\boldsymbol{W}}$.

Step 9: If observed asset returns $\boldsymbol{\xi}_n \in \boldsymbol{\Theta}_{t+1}$ are considered to follow the current predictive distribution $p(\hat{\boldsymbol{\xi}}, \boldsymbol{\Theta}_t)$, the top $\tau \times\ 100\%$ individuals from the final population $\boldsymbol{P}_{t,\overline{g}}$ are transferred to the next initial one $\boldsymbol{P}_{t+1,0}$. Other individuals in $\boldsymbol{P}_{t+1,0}$ are generated randomly. Otherwise, all individuals in the initial population $\boldsymbol{P}_{t+1,0}$ are generated randomly.

Step 10: Let $t = t+1$. Go back to Step 3.

C. Kolmogorov-Smirnov Test

In order to test whether the newly observed asset returns follow the current predictive distribution in Step 9 of the ETO method, the Kolmogorov-Smirnov test (K-S test) is used.

Let $\boldsymbol{x}_t^\star \in \mathbb{R}^D$ be the best solution obtained by ADE to the portfolio optimization problem in (14), which is defined by the predicted data set $\widehat{\boldsymbol{\Theta}}_t$. Using the same solution $\boldsymbol{x}_t^\star \in \mathbb{R}^D$, two ECDFs of the return $r \in \mathbb{R}$ are formed from the predicted asset returns $\hat{\boldsymbol{\xi}}_m \in \widehat{\boldsymbol{\Theta}}_t$ and the observed ones $\boldsymbol{\xi}_n \in \boldsymbol{\Theta}_{t+1}$. From

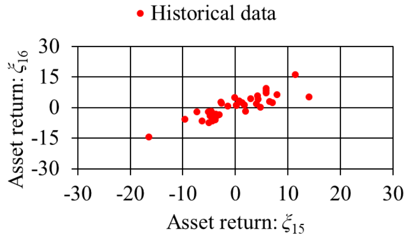
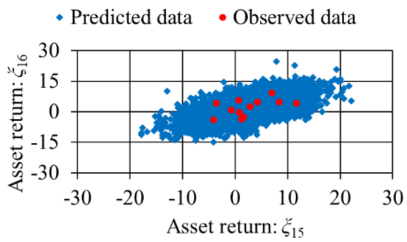
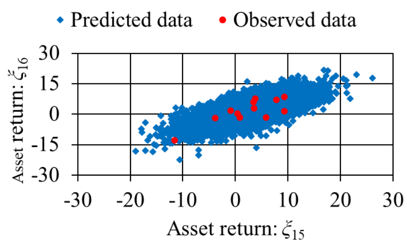


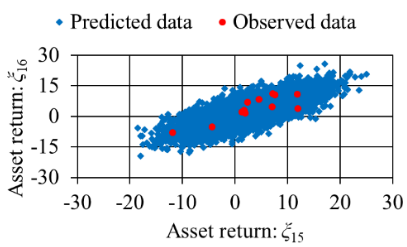
Figure 2. Historical data for prior distribution.



(a) Period: $t = 1$



(b) Period: $t = 2$



(c) Period: $t = 3$

Figure 3. Predicted $\hat{\boldsymbol{\xi}}_m \in \widehat{\boldsymbol{\Theta}}_t$ and observed $\boldsymbol{\xi}_n \in \boldsymbol{\Theta}_{t+1}$.

the two ECDFs of $r \in \mathbb{R}$, the null hypothesis (H_0) and the alternative hypothesis (H_1) of the K-S test are defined as

$$\begin{aligned} H_0 &: F(r_n \,|\, \boldsymbol{x}_t^\star,\, \widehat{\boldsymbol{\Theta}}_t) = F(r_n \,|\, \boldsymbol{x}_t^\star,\, \boldsymbol{\Theta}_{t+1}),\ \ n = 1, \cdots, N \\ H_1 &: F(r_n \,|\, \boldsymbol{x}_t^\star,\, \widehat{\boldsymbol{\Theta}}_t) \neq F(r_n \,|\, \boldsymbol{x}_t^\star,\, \boldsymbol{\Theta}_{t+1}),\ \ n = 1, \cdots, N \end{aligned}$$

where $r_n = \boldsymbol{\xi}_n \, \boldsymbol{x}_t^{\star T} \in \mathbb{R}$ is the return of the portfolio $\boldsymbol{x}_t^\star \in \mathbb{R}^D$ evaluated by the newly observed asset returns $\boldsymbol{\xi}_n \in \boldsymbol{\Theta}_{t+1}$.

If the null hypothesis is not rejected, the newly observed asset returns $\boldsymbol{\xi}_n \in \boldsymbol{\Theta}_{t+1}$ are considered to follow the predictive distribution $p(\hat{\boldsymbol{\xi}} \,|\, \boldsymbol{\Theta}_t)$ as well as the predicted asset returns $\hat{\boldsymbol{\xi}}_m \in \widehat{\boldsymbol{\Theta}}_t$ in (11). On the other hand, if the null hypothesis is rejected, $\boldsymbol{\xi}_n \in \boldsymbol{\Theta}_{t+1}$ are not considered to obey $p(\hat{\boldsymbol{\xi}} \,|\, \boldsymbol{\Theta}_t)$.

V. Numerical Experiment

A. Experimental Setup

The monthly data from the Tokyo Stock Price Index (TOPIX17) over the past six years ($2019 - 2024$) are used for asset returns [15]. Therefore, the number of assets is $D = 17$. The first prior distribution $p(\boldsymbol{\mu},\, \boldsymbol{\Lambda})$ in Step 2 of the proposed ETO method is obtained from the monthly data on the first three years. The monthly data from the remaining three years are used for the observed data $\boldsymbol{\Theta}_t$, $t = 1,\, 2, 3$ in (3). Since monthly data is used, the data size of $\boldsymbol{\Theta}_t$ is $N = 12$.

Figure 2 shows two asset returns $\xi_{n,15}$ and $\xi_{n,16} \in \mathbb{R}$ from the historical data on asset returns $\boldsymbol{\xi}_n \in \mathbb{R}^{17}$, $n = 1, \cdots, 36$. A positive correlation is observed between them. In the same way, Figure 3 shows the observed data $\boldsymbol{\xi}_n \in \boldsymbol{\Theta}_t$, $N = 12$ in (3) and the predicted data $\hat{\boldsymbol{\xi}}_m \in \widehat{\boldsymbol{\Theta}}_t$, $M = 10^4$ in (11) for the periods $t = 1,\, 2,\, 3$. From the data size of $\widehat{\boldsymbol{\Theta}}_t$, the margin of error in the probability estimate is around $1/\sqrt{M} = 0.01$.

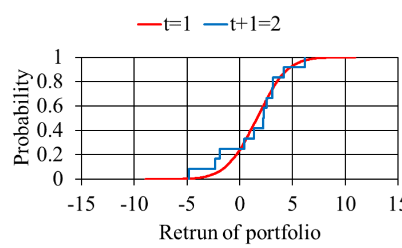
B. Performance of ETO method

The proposed ETO method is applied to the portfolio optimization problems in (14) defined by the three predicted data sets $\widehat{\boldsymbol{\Theta}}_t$, $t = 1,\, 2,\, 3$ and the target value $r_o = 0.01$.

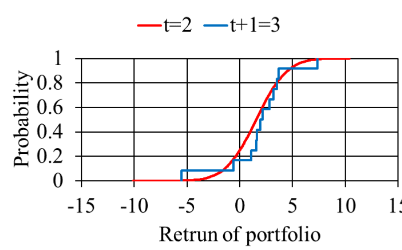
The population size of ADE is $N_P = 100$. The maximum number of generations is chosen as $\overline{g} = 100$. For the transfer rate τ, five values are selected from the range 0 to 1. If $\tau = 0$ then no individual is transferred from $\boldsymbol{P}_{t,\overline{g}}$ to $\boldsymbol{P}_{t+1,0}$. If $\tau = 1$ then all individuals are transferred from $\boldsymbol{P}_{t,\overline{g}}$ to $\boldsymbol{P}_{t+1,0}$.

Figure 4 compares the two ECDFs of returns $r \in \mathbb{R}$ used for the K-S test: $F(r \,|\, \boldsymbol{x}_t^\star,\, \widehat{\boldsymbol{\Theta}}_t)$ and $F(r \,|\, \boldsymbol{x}_t^\star,\, \boldsymbol{\Theta}_{t+1})$, $t = 1,\, 2$. The best solution $\boldsymbol{x}_t^\star \in \mathbb{R}^D$ is obtained with $\tau = 0.25$.

The null hypothesis of the K-S test has not been rejected for every value of τ at a significance level: 0.05. Therefore,

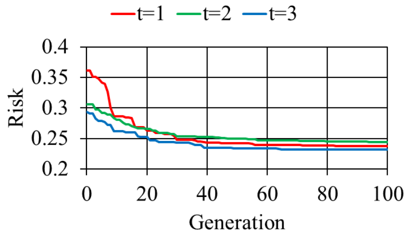


(a) $F(r \,|\, \boldsymbol{x}_1^\star,\, \widehat{\boldsymbol{\Theta}}_1)$ and $F(r \,|\, \boldsymbol{x}_1^\star,\, \boldsymbol{\Theta}_2)$

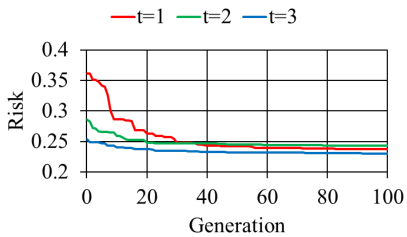


(b) $F(r \,|\, \boldsymbol{x}_2^\star,\, \widehat{\boldsymbol{\Theta}}_2)$ and $F(r \,|\, \boldsymbol{x}_2^\star,\, \boldsymbol{\Theta}_3)$

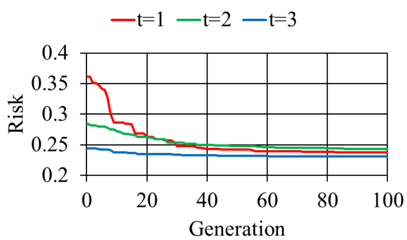
Figure 4. Two ECDFs of $r \in \mathbb{R}$ for K-S test.



(a) Transfer rate: $\tau = 0$



(b) Transfer rate: $\tau = 0.25$



(c) Transfer rate: $\tau = 1$

Figure 5. Convergence graphs of ADE.

TABLE I. RISK OF BEST SOLUTION

τ	$t = 2$	$t = 3$
0	0.244 (6.72×10^{-7})	0.231 (1.09×10^{-7})
0.25	0.243 (1.71×10^{-7})	0.230 (1.69×10^{-7})
0.5	0.243 (4.26×10^{-7})	0.231 (1.99×10^{-7})
0.75	0.243 (3.39×10^{-7})	0.230 (1.24×10^{-7})
1	0.244 (1.64×10^{-6})	0.230 (3.67×10^{-8})

TABLE II. STEEL-DWASS TEST ABOUT RISK

τ	τ	$t = 2$	$t = 3$
0	0.25	33.5 (0.001)	37.975 (0.001)
	0.5	19.475 (0.214)	23.9 (0.048)
	0.75	37.475 (0.001)	25.675 (0.004)
	1	3.925 (0.998)	52.825 (0.000)
0.25	0.5	-14.025 (0.594)	-14.075 (0.466)
	0.75	3.975 (0.938)	-12.3 (0.598)
	1	-29.575 (0.008)	14.85 (0.447)
0.5	0.75	18 (0.263)	1.775 (0.999)
	1	-15.55 (0.455)	28.925 (0.017)
0.75	1	-33.55 (0.006)	27.15 (0.002)

the top $\tau \times$ 100% individuals from the final population $\boldsymbol{P}_{t,\overline{g}}$ are transferred to the next initial one $\boldsymbol{P}_{t+1,0}$ for $t = 1, 2$.

Figure 5 shows convergence graphs of ADE using three transfer rates. The horizontal axis is the generation g of ADE. The vertical axis is the risk $\alpha(\boldsymbol{x}_k^\star)$ of the best solution $\boldsymbol{x}_k^\star \in \mathbb{R}^D$ obtained by ADE in each generation. From Figure 5, we can confirm that the best solutions $\boldsymbol{x}_k^\star \in \mathbb{R}^D$ in $\boldsymbol{P}_{t,g}$ have been converged sufficiently at the final generation $\overline{g} = 100$.

From Figure 5, the effect of the proposed ETO method is confirmed. The shapes of the convergence graphs for the period $t = 1$ are almost identical. On the other hand, the shapes for the period $t \geq 2$ vary depending on the transfer rate τ. This dependency arises because the solutions $\boldsymbol{x}_k \in \mathbb{R}^D$ in the initial population $\boldsymbol{P}_{t,0}$ fluctuate according to the value of τ.

Table I shows the risks $\alpha(\boldsymbol{x}_t^\star)$, $t = 2, 3$ of the best solutions $\boldsymbol{x}_t^\star \in \mathbb{R}^D$ averaged over 20 runs. The variance of the risks is also shown in parenthesis. From Table I, the risk $\alpha(\boldsymbol{x}_t^\star)$ of the solution $\boldsymbol{x}_t^\star \in \mathbb{R}^D$ obtained by ADE appears to depend on the transfer rate τ. Hence, the Kruskal-Wallis test (K-W test) is used to determine whether there is a significant difference between the risks resulting from different τ values. The results of K-W test indicate that the risks in Table I significantly depend on the values of τ in both of $t = 2$ and $t = 3$.

The Steel-Dwass test (S-D test) is designed for comparing multiple groups to identify which specific groups differ from each other. Table II shows the results of S-D test about the risks $\alpha(\boldsymbol{x}_t^\star)$, $t = 2, 3$ in Table I. Table II shows the difference in risk ranking obtained with the left and right τ values. The greater the difference in ranking, the lower the risk obtained with the right τ value. The results of S-D test are also shown by p-value in parenthesis. From Table II, the best solution obtained by the ETO method with $\tau = 0.25$ is clearly better than the best solution obtained with $\tau = 0$. No individual is transferred if $\tau = 0$. However, $\tau = 0.25$ is competitive with $\tau = 0.75$ in terms of the quality of the best solution.

From Table I and Table II, it can be confirmed that the transfer of superior solutions performed by the ETO method enhances the quality of the solution obtained by ADE.

C. Analysis of Portfolio

The best portfolios $\boldsymbol{x}_t^\star \in \mathbb{R}^D$, $t = 1, 2, 3$ for the data-driven portfolio optimization problem in (14) are obtained by the proposed ETO method with the transfer rate $\tau = 0.25$. The investment ratios $x_i \in \mathbb{R}$, $i = 1, \cdots, D$ provided by the three best portfolios $\boldsymbol{x}_t^\star \in \mathbb{R}^D$, $t = 1, 2, 3$ are shown in Figure 6. From Figure 6, the best portfolio $\boldsymbol{x}_t^\star \in \mathbb{R}^D$ varies depending on the period t. However, most of the assets with a zero ratio

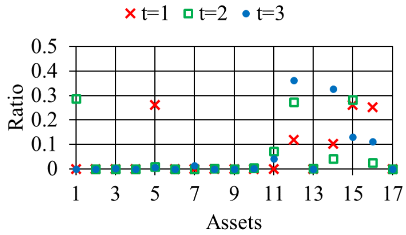


Figure 6. Best portfolio $\boldsymbol{x}_t^\star \in \mathbb{R}^D$, $D = 17$.

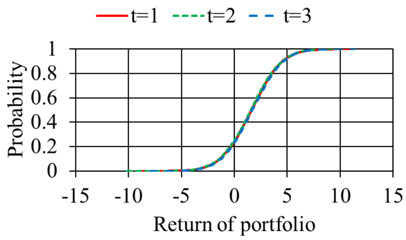


Figure 7. ECDF $F(r \,|\, \boldsymbol{x}_t^\star,\, \widehat{\boldsymbol{\Theta}}_t)$ by best portfolio $\boldsymbol{x}_t^\star \in \mathbb{R}^D$.

($x_i = 0$) are common to all three portfolios. In other words, these portfolios correctly identify non-performing assets.

Figure 7 shows the ECDFs of returns $r \in \mathbb{R}$ by the three best portfolios: $F(r \,|\, \boldsymbol{x}_t^\star,\, \widehat{\boldsymbol{\Theta}}_t)$, $t = 1,\, 2,\, 3$. From Figure 7, there is no significant difference in the shapes of ECDFs. Therefore, the risk $\alpha(\boldsymbol{x}_t^\star)$ achieved by the best portfolio $\boldsymbol{x}_t^\star \in \mathbb{R}^D$ does not differ drastically across the three periods $t = 1,\, 2,\, 3$.

From Figure 6 and Figure 7, the structure of the best solution $\boldsymbol{x}_t^\star \in \mathbb{R}^D$ to the data-driven portfolio optimization problem examined in this paper does not change significantly over the three periods. As a result, the transfer of superior solutions performed by the ETO method has worked effectively.

VI. Conclusion and Future Work

By using the predictive distribution for asset returns, the traditional portfolio optimization problem has been extended to a data-driven one. As shown in Figure 1, the predictive distribution is updated via the prior and posterior distributions whenever a new observed data set becomes available. As a result, information from all past data is aggregated into the probabilistic model of asset returns. This paper presents a novel ETO method utilizing ADE to effectively address the data-driven portfolio optimization problem. In the ETO method, superior solutions from the final population are transferred to the next initial population. Finally, the usefulness of the ETO method has been demonstrated by using actual asset data.

In future work, we will adaptively determine the transfer rate for the proposed ETO method from the state of the population. Furthermore, the proposed ETO method is expected to be applied to a variety of data-driven optimization problems.

Acknowledgment

This work was partially supported by JSPS (Japan Society for Promotion Science) KAKENHI, 22K12187, 24K15113.

References

[1] K. C. Tan, L. Feng, and M. Jiang, “Evolutionary transfer optimization - a new frontier in evolutionary computation research,” *IEEE Computational Intelligence Magazine*, vol. 16, no. 1, pp. 22–33, 2021.
[2] T. T. Nguyen, S. Yang, and J. Branke, “Evolutionary dynamic optimization - a survey of the state of the art,” *Swarm and Evolutionary Computation*, vol. 6, pp. 1–24, 2012.
[3] S. Das, A. Mandal, and R. Mukherjee, “An adaptive differential evolution algorithm for global optimization in dynamic environments,” *IEEE Trans. on Cybernetics*, vol. 44, no. 6, pp. 966–978, 2024.
[4] Y. Jin and J. Branke, “Evolutionary optimization in uncertain environments - a survey,” *IEEE Trans. on Evolutionary Computation*, vol. 9, no. 3, pp. 303–317, 2005.
[5] M. Iqbal, W. N. Browne, and M. Zhang, “Reusing building blocks of extracted knowledge to solve complex, large-scale Boolean problems,” *IEEE Trans. on Evolutionary Computation*, vol. 18, no. 4, pp. 465–480, 2014.
[6] D. Zhan and H. Xing, “A fast kriging-assisted evolutionary algorithm based on incremental learning,” *IEEE Trans. on Evolutionary Computation*, vol. 25, no. 5, pp. 941–955, 2021.
[7] Y. Liu, J. Liu, J. Ding, S. Yang, and Y. Jin, “A surrogate-assisted differential evolution with knowledge transfer for expensive incremental optimization problems,” *IEEE Trans. on Evolutionary Computation*, vol. 28, no. 4, pp. 1039–1053, 2024.
[8] E. J. Elton, M. J. Gruber, S. J. Brown, and W. N. Goetzmann, *Modern Portfolio Theory and Investment Analysis (9th ed.)* John Wiley & Sons, 2014.
[9] M. Mokhtar, A. Shuib, and D. Mohamad, “Mathematical programming models for portfolio optimization problem: A review,” *International Journal of Mathematical and Computational Sciences*, vol. 8, no. 2, pp. 428–435, 2014.
[10] C. Wang, Y. Chen, S. Zhang, and Q. Zhang, “Stock market index prediction using deep transformer model,” *Expert Systems with Applications*, vol. 208, pp. 1–10, 2022.
[11] A. B. Mard, B. Hnich, A. Lahiani, and S. Mefteh-Wali, “Comprehensive stock market insight: Bayesian networks for multi-output forecasting,” *Computational Economics*, vol. 66, no. 5, pp. 4329–4349, 2025.
[12] E. F. Fama, “Efficient capital markets: A review of theory and empirical work,” *The Journal of Finance*, vol. 25, no. 2, pp. 383–417, 1970.
[13] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
[14] J. Brest, S. Greiner, B. Boskovic, M. Mernik, and V. Zumer, “Self-adaptive control parameters in differential evolution: Comparative study on numerical benchmark problems,” *IEEE Trans. on Evolutionary Computation*, vol. 10, no. 6, pp. 646–657, 2006.
[15] Kabutan, https://kabutan.jp, Accessed: 2025/01/20.
[16] R. J. Rossi, *Mathematical Statistics: An Introduction to Likelihood Based Inference*. John Wiley & Sons, 2018.