Minimizing Loss Probability Bounds for Portfolio Selection

Jun-ya Gotoh\textsuperscript{a} \textsuperscript{,} 2 Akiko Takeda\textsuperscript{b}

\textsuperscript{a}Department of Industrial and Systems Engineering, Chuo University
2-13-27 Kasuga, Bunkyo-ka, Tokyo 112-8551, Japan.
E-mail address: jgoto@indsys.chuo-u.ac.jp

\textsuperscript{b}Department of Administration Engineering, Keio University
3-14-1 Hiyoshi, Kohoku, Yokohama, Kangawa 223-8522, Japan
E-mail address: takeda@ae.keio.ac.jp

Abstract

In this paper, we derive a portfolio optimization model by minimizing upper and lower bounds of loss probability. These bounds are obtained under a nonparametric assumption of underlying return distribution by modifying the so-called generalization error bounds for the support vector machine, which has been developed in the field of statistical learning. Based on the bounds, two fractional programs are derived for constructing portfolios, where the numerator of the ratio in the objective includes the value-at-risk (VaR) or conditional value-at-risk (CVaR) while the denominator is any norm of portfolio vector. Depending on the parameter values in the model, the derived formulations can result in a nonconvex constrained optimization, and an algorithm for dealing with such a case is proposed. Some computational experiments are conducted on real stock market data, demonstrating that the CVaR-based fractional programming model outperforms the empirical probability minimization.

Keywords: Finance; Portfolio optimization; CVaR (Conditional Value-at-Risk); SVM (Support Vector Machine); Fractional programming.

1 Introduction

Portfolio Selection Models. The problem of allocating funds into a given set of investable assets is known as portfolio selection. Typical (single period) portfolio selection models determine the distribution of a random return of the form

\[ R(\pi) := \sum_{j=1}^{n} R_j \pi_j, \]

where \( R_j \) represents the random rate of return of asset \( j \), and \( \pi \) represents a portfolio vector, each component representing the investment ratio into each asset. In practice, the criterion for determining a portfolio \( \pi \) is formulated as an optimization problem of the form:

\[ \min \{ F[R(\pi)] : \pi \in \Pi \subset \mathbb{R}^n \}, \] (1)

where the objective is a functional \( F \) of the random vector \( \mathbf{R} := (R_1, ..., R_n)^\top \) on a probability space \((\Omega, \mathcal{F}, \mathbb{P})\) which is independent of \( \pi \), and \( \Pi \subset \mathbb{R}^n \) is a feasible region of the portfolio vector \( \pi \). By definition of \( \pi \), \( \Pi \) is supposed to include a constraint of the form \( e_n^\top \pi = 1 \) where \( e_n := (1, ..., 1)^\top \) is the n-dimensional vector of ones.

For example, the expected utility maximization criterion with some utility function \( U \) can be formulated by adopting \( -E[U(R(\pi))] \) as the objective of (1), where \( E[\cdot] \) denotes the mathematical expectation. More practically, a risk measure such as variance \( \mathbb{V}[R(\pi)] \) or a composite objective considering return as well as risk, e.g., \( \mathbb{V}[R(\pi)] - aE[R(\pi)] \) with a positive constant \( a \), is preferred due to the ease in controlling the characteristics of the distribution of \( R(\pi) \), where

\[ 1 \] This manuscript is the revised version of the authors' paper entitled: "Portfolio Learning via VaR/CVaR Minimization." The title is changed along with the major revision.

\[ 2 \] Corresponding Author. Tel/fax: +81-3-3817-1928. E-mail address: jgoto@indsys.chuo-u.ac.jp (J.Gotoh).

1
\( \mathbb{V}[\cdot] \) denotes the variance operator. Alternative deviation type risk measures such as absolute deviation, \( \mathbb{E}[|R(\pi) - \mathbb{E}[R(\pi)]|] \), have also been used (Konno and Yamazaki, 1991).

Although variance has been the most basic risk measure since the seminal work of Markowitz (1952), drawbacks of such deviation type measures have been pointed out in the literature, and both theoretical and practical attentions have recently cast a spotlight on downside risk measures such as the class of coherent risk measures rather than deviation measures. A classical downside risk measure is Roy (1952)'s safety first criterion which minimizes the probability of portfolio loss being greater than a threshold. More recently, Value-at-Risk (VaR) has gained popularity in 1990s in practice so as to grasp a large loss with a small probability. Although it is intuitive and easy to understand its implication, it faces critiques due to the lack of convexity and, accordingly, coherence (see, e.g., Artzner et al., 1999). In contrast with VaR, Conditional Value-at-Risk (CVaR) has recently been obtaining a growing popularity due to nice properties such as the coherence and the consistency with the second order stochastic dominance (see, e.g., Ogryczak and Ruszczynski, 2002) as well as tractability in its optimization (Rockafellar and Uryasev, 2002).

**Limitation of Existing Approaches.** In a typical situation of portfolio selection, historical return data are given as in Figure 1 (left), where the rate of return of \( n \) assets has been observed in a market for last \( T \) months. For example, in order to adopt the variance as the objective of (1), we first estimate a covariance matrix \( \bar{\Sigma} \) in place of the true covariance matrix \( \Sigma \) by using the data, and next minimize the estimated portfolio variance \( \pi^\top \bar{\Sigma} \pi \) over \( \Pi \) in place of the true objective \( \mathbb{V}[R(\pi)] = \pi^\top \Sigma \pi \), which is never known. This framework can be validated by the law of large numbers, i.e., a large number of samples lead to the true optimal value. In financial practice, however, the number of available samples, \( T \), is often limited due to the limited frequency of the observations, and a significant estimation error may occur in estimating the parameters and, accordingly, the optimal portfolio \( \pi \). Besides, the number of assets, \( n \), is often larger than that of the samples, \( T \). For example, \( T \) is 120 for ten-year monthly observations, while \( n \) can be more than one thousand, and in such a case, estimation error can be much larger.

Numerous papers address estimation error in the estimates of the mean and variance (e.g., Chopra and Ziemba, 1993). According to those, estimation error in the mean estimate is much severer than that in variance. Therefore, recent papers focus on the minimum variance model rather than the mean-variance model (e.g., Jagannathan and Ma, 2003; DeMiguel et al., 2009). Also, there exist papers comparing criteria proposed by various researchers. For example, Simaan (1997) points out that the use of absolute deviation instead of variance may cause larger estimation error when return follows a normal distribution.

Limitation of such arguments, however, is that only estimation error from a certain supposed distribution can be evaluated. More specifically, since we cannot know the true distribution in usual financial practice, we cannot evaluate the true estimation error. In addition, although many researchers assume the normal distribution so as to validate those criteria, it is known that return does not follow a normal distribution, but follows more fat-tailed and/or skewed distribution. Thus, it can be said that the traditional models have little theoretical underpinning for the out-of-sample performance of their optimal portfolio in the real world, or in the absence of a specific parametric distribution assumption.

**Recent Developments.** In contrast to those traditional approaches, several recent researches focus on the out-of-sample performance of the “optimal” portfolio by introducing the idea developed in the area of statistics into the portfolio selection models (1). For example, Jagannathan and Ma (2003) show that imposing the short-sale constraint, \( \pi \geq 0 \), which is usually imposed in practice, on the minimum variance model is equivalent to a shrinkage of the covariance matrix and the out-of-sample performance will be improved. DeMiguel et al. (2009) impose a norm-
constraint on the variance minimization criterion by extending the idea of shrinkage methods in statistics (see, e.g., Hastie et al., 2001). They reveal that the problem formulation with the 2-norm (Euclidean norm) constraint contains the equally weighted portfolio, i.e., \( \pi_j = 1/n \), as a special case while that with the 1-norm constraint contains the minimum variance model with the short-sale constraint. All of these researches incorporate shrinkage techniques in the sample covariance matrix so as to improve the out-of-sample performance of the minimum variance model.

**Similarity between Portfolio Selection and Statistical Learning.** Apart from the portfolio selection context stated above, many statistical models also seek a linear combination of random variables. For example, an outlier (or novelty) detection model called the one-class \( \nu \)-support vector machine (OC-\( \nu \)-SVM), which is known as an unsupervised statistical learning model, determines a linear function of \( d \)-dimensional random vector \( a^\top = (A_1, ..., A_d) \):

\[
A(w) := \sum_{k=1}^{d} w_k A_k
\]

whose factor loadings \( w_i \) are estimated by solving a convex program:

\[
\min \left\{ \frac{1}{2} \| w \|_2^2 - \rho + \frac{1}{\nu m} \sum_{i=1}^{m} [\rho - a_i^\top w^+]^+: w \in \mathbb{R}^d, \rho \in \mathbb{R} \right\},
\]

(2)
where \( \| x \|_2 \) is 2-norm of a vector \( x \), \( [x]^+ := \max\{x, 0\} \), \( \nu \in (0, 1] \) is a constant to be tuned, and \( a_1, ..., a_m \) are observed data of random vector \( a \), given as in Figure 1 (right) where \( a_i := (a_{i1}, ..., a_{id})^\top \), \( i = 1, ..., m \), follow an unknown distribution in \( \mathbb{R}^d \). The outliers are then defined by points \( a_i \) satisfying \( a_i^\top w^* < \rho^* \) where \( (w^*, \rho^*) \) is an optimal solution to (2). The out-of-sample performance of OC-\( \nu \)-SVM as well as the other types of SVMs is underpinned by an error bound which is known as the generalization error bound (see, e.g., Theorem 8.6 of Schölkopf and Smola, 2002). It gives, for example, an upper bound of \( \mathbb{P}\{A(w^*) < \rho^* - \gamma\} \) for \( \gamma > 0 \), which means the probability that the linear function \( (w^*)^\top a \) takes less than \( \rho^* - \gamma \) for a new data sample \( a_{m+1} \).

As easily seen from Figure 1 and the procedures mentioned above, portfolio selection and unsupervised statistical learning have the same structure in the sense that both estimate a linear model \( R(\pi) \) or \( A(w) \) from the batch data as Figure 1 via optimization. In fact, the authors (Gotoh and Takeda, 2005) propose a classification model, which is also a well-studied subject in statistical learning, by incorporating the concept of CVaR, which has been used in the financial context as stated above, and point out that the two-class \( \nu \)-SVM (Schölkopf et al., 2000) and its extension (Perez-Cruz et al., 2003) can be interpreted as a CVaR minimization with a loss associated with the (geometric) margin (see, e.g., Schölkopf and Smola, 2002, for the classification problem and the definition of ‘margin’). Also, one of the authors (Takeda, 2009) shows stronger generalization error bounds for a couple of \( \nu \)-SVMs by employing the notions of...
VaR and CVaR. These facts indicate that VaR and CVaR can play an important role not only in portfolio selection but also in statistical learning.

This paper aims to discuss the interaction in the reverse direction. By employing the results explored for statistical learning, we provide a new sight for the use of VaR and CVaR. More precisely, contributions of this paper are summarized as follows:

▷ Under a nonparametric assumption, upper and lower bounds of the probability of a portfolio loss being greater than a threshold are presented by using (empirical) VaR or CVaR on the basis of the generalization error bound for SVMs. This indicates that empirical VaR or CVaR optimization can work for lowering out-of-sample loss in a probabilistic sense and, thus, the bounds yield a theoretical underpinning for the use of VaR or CVaR.

▷ Fractional programming portfolio models are posed so as to minimize those bounds of loss probability, where the numerator of the objective function to be minimized includes VaR or CVaR whereas the denominator is any norm. Besides, the formulation contains the normal loss probability minimization as a special case, which indicates that the normal probability minimization can achieve a good out-of-sample performance even for nonnormal distribution. Also, tractability of the fractional optimization formulation depends on the sign of the optimal value. To cope with the variable tractability, we develop a two-step algorithm, which yields a local optimizer when the problem is intractable while a global optimizer when tractable.

▷ The generalization bounds also indicate a theoretical underpinning for the ordinary VaR or CVaR minimizing portfolio when the short-sale constraint is imposed. In other words, the short-sale constraint plays a role in improving out-of-sample performance of the VaR and CVaR minimization as in the variance minimization discussed in Jagannathan and Ma (2003).

▷ In addition to the theoretical achievement, we conduct numerical experiments on actual stock market data. The results show that the CVaR-based fractional programming model outperforms the empirical loss probability minimization, and is comparable to the minimum variance model.

The structure of this paper is as follows: In the next section, a few bounds of portfolio loss probability are given after briefly reviewing existing approach for the probability minimizing portfolio and introducing notions of (empirical) VaR and CVaR. Section 3 is devoted to an algorithm for solving the fractional programs. In Section 4, the optimization models will be applied to real stock market data. Finally, Section 5 concludes the paper with some remarks. Proofs of the propositions are given in the end of the paper.

2 Nonparametric Error Bounds of the Loss Probability

2.1 Loss Probability Minimizations

Let \( f(\pi, \mathcal{R}) \) denote a random portfolio loss which is preferable to be as small as possible. In this paper, we adopt

\[
f(\pi, \mathcal{R}) = -\mathcal{R}(\pi) = -\mathcal{R}^\top \pi
\]

for the loss variate in portfolio selection since the return \( \mathcal{R}(= -f) \) is a prominent objective to be as large as possible. In order to determine a portfolio \( \pi \), we consider a criterion of minimizing the probability of the portfolio loss \( f \) being larger than a threshold \( \theta \), i.e., \( P\{f(\pi, \mathcal{R}) > \theta\} \). For example, if one sets \( \theta = 0 \), it indicates the probability of the portfolio return being negative. This is known as Roy’s safety first criterion (Roy, 1952). Noting that minimizing \( P\{f(\pi, \mathcal{R}) > \theta\} \) is
equal to maximizing $\mathbb{P}\{R(\pi, \mathcal{R}) \geq -\theta\}$, we can consider this criterion as the return probability maximization.

When one knows that the return $\mathcal{R}$ follows a normal distribution $N(\mu, \Sigma)$, but does not know the true parameters $\mu$ and $\Sigma$, he/she is likely to estimate the parameters and solve the following fractional program in order to obtain the safety first portfolio:

$$\min \left\{ \frac{-\hat{\mu}^\top \pi - \theta}{\sqrt{\pi^\top \Sigma \pi}} : \pi \in \Pi \right\}, \tag{3}$$

where $\hat{\mu}$ and $\hat{\Sigma}$ are the estimated return vector and covariance matrix, respectively. Typically, the parameters are estimated as $\hat{\mu} := \frac{1}{T} \sum_{t=1}^{T} R_t$ and $\hat{\Sigma} := \frac{1}{T} \sum_{t=1}^{T} (R_t - \hat{\mu})(R_t - \hat{\mu})^\top / (T - 1)$ where $R_1, \ldots, R_T$ are $T$ historical observations of $\mathcal{R}$. In particular, when $-\theta$ is set equal to the risk-free rate, (3) is equivalent to the so-called Sharpe ratio maximization.

In contrast with the normal distribution case, we suppose in the remaining part of the paper that the return $\mathcal{R}$ follows an unknown ($n$-dimensional) distribution, and that $R_1, \ldots, R_T$ are independently drawn from it. Note that most of traditional models have been constructed on the i.i.d. assumption.

Noting that under the above nonparametric assumption, the empirical counterpart of the loss probability is represented by $\frac{1}{T} |\{t : f(\pi, R_t) > \theta, t = 1, \ldots, T\}|$ a straightforward formulation for obtaining the smallest empirical loss probability portfolio is then given by

$$\min \left\{ \frac{1}{T} e_T^\top z : -R_t^\top \pi - \theta \leq M z_t, \ (t = 1, \ldots, T), \ z \in \{0, 1\}^T, \ \pi \in \Pi \right\}, \tag{4}$$

where $M$ is a sufficiently large constant and $e_T$ is the $T$-dimensional column vector of ones. This problem is a 0-1 mixed integer linear program (MILP), which can be efficiently solved by state-of-the-art optimization softwares, e.g., IBM-ILOG CPLEX12, when the size of the problem, $T$ and $n$, is not huge and $\Pi$ is represented by a system of linear or convex quadratic inequalities.

In addition, we assume throughout the paper that random return $\mathcal{R}$ has an implicit bounded support. Although this assumption excludes the normal distribution, it does not weaken its applicability. In fact, the total amount of financial contracts in the world is bounded, and therefore, the return must be bounded.

### 2.2 Loss Probability Bounds and VaR/CVaR Minimization

Under the above nonparametric distribution assumptions, we below derive upper and lower bounds of $\mathbb{P}\{f(\pi, \mathcal{R}) > \theta\}$. In order to describe those bounds, let us first introduce the $\beta$-value-at-risk ($\beta$-VaR) and $\beta$-conditional value-at-risk ($\beta$-CVaR).

Given $T$ observed return vectors $R_1, \ldots, R_T$, the (empirical) $\beta$-VaR associated with loss $f(\pi, \mathcal{R})$ is the $\beta$-quantile defined by

$$\alpha_{\beta}(\pi) := \min \{\alpha : \Phi_T(\alpha | \pi) \geq \beta\}$$

where $\Phi_T(\alpha | \pi) := \frac{1}{T} |\{t \in \{1, \ldots, T\} : f(\pi, R_t) \leq \alpha\}|$ is the empirical distribution function of the loss $f$. $\beta \in (0, 1)$ is a user-defined parameter for representing a confidence level and usually takes a value close to 1, say, 0.95 or 0.99, for capturing a large loss with a small probability. The $\beta$-VaR minimizing portfolio is thus given by a solution to

$$\min \{ \alpha_{\beta}(\pi) : \pi \in \Pi \}, \tag{5}$$

2
which can be reformulated as an MILP when \( \Pi \) is given by a system of linear inequalities.

On the other hand, the (empirical) \( \beta \)-CVaR associated with loss \( f(\pi, \mathcal{R}) \) is defined by

\[
\phi(\pi) := \min_\alpha F(\pi, \alpha),
\]

where \( \beta \in [0, 1) \) and \( F(\pi, \alpha) \) is a convex function on \( \mathbb{R}^n \times \mathbb{R} \), defined by

\[
F(\pi, \alpha) := \alpha + \frac{1}{(1 - \beta)T} \sum_{t=1}^{T} [f(\pi, R_t) - \alpha]^+.
\]

According to Rockafellar and Uryasev (2002), \( \beta \)-CVaR, \( \phi(\pi) \), can be approximately regarded as the expected value of loss \( f \) greater than \( \beta \)-VaR, \( \alpha(\pi) \), and therefore, one has \( \alpha(\pi) \leq \phi(\pi) \), as in Figure 2. In addition, we can see that when \( \beta = 0 \), one has \( \phi(\pi) = -\hat{\mu}^T \pi \) with

\[
\hat{\mu} := \frac{1}{T} \sum_{t=1}^{T} R_t / T,
\]

which implies that the minimization of \( \phi_0(\pi) \) is equal to the maximization of expected return. On the other hand, when \( \beta \in (1 - 1/T, 1) \), one has \( \phi(\pi) = \max\{ -R_t^T \pi : t \in \{1, ..., T\} \} \), which implies that the minimization of \( \phi(\pi) \) is then equal to the maximization of the minimal return (Young, 1998). In practice, similarly to VaR, \( \beta \) is usually fixed at a value close to one. \( \phi(\pi) \) and \( F(\pi, \alpha) \) are piecewise linear convex function, whereas \( \alpha(\pi) \) is nonconvex, in general. This fact leads to a disadvantage of VaR over CVaR in efficient minimization. Also, let us note that both \( \alpha(\pi) \) and \( \phi(\pi) \) as well as \( F(\pi, \alpha) \) are positively homogeneous functions.

The \( \beta \)-CVaR minimizing portfolio is given by a solution to

\[
\min\{ F(\pi, \alpha) : \pi \in \Pi, \alpha \in \mathbb{R} \},
\]

which can be reformulated as a linear program when \( \Pi \) is represented by a system of linear inequalities. In addition, for an optimal solution \((\pi^*, \alpha^*)\) to (6), \( \alpha^* \) gives an approximate value of \( \beta \)-VaR, \( \alpha(\pi^*) \), as a by-product.

![Figure 2: Illustration of \( \beta \)-VaR, \( \alpha \), and \( \beta \)-CVaR, \( \phi \) associated with loss \( f \)](image)

**Remark 1** The authors have pointed out in (Gotoh and Takeda, 2005; Takeda, 2009) that the CVaR minimization with an adequate loss function \( f \) results in the two-class \( \nu \)SVM and \( \nu \)-support vector regression (\( \nu \)SVR). Those results can be extended straightforward into the one-class \( \nu \)-SVM (OC-\( \nu \)SVM) (2). We describe the details in Section B.1 for readers interested in the statistical learning.

By using \( \beta \)-VaR, \( \alpha(\pi) \), and \( \beta \)-CVaR, \( \phi(\pi) \), we have the following theorem:

**Theorem 1** Let \( \theta \) be a threshold for portfolio loss. Suppose that random return vector \( \mathcal{R} \) has a bounded support in the sense that \( \mathcal{R} + \theta e_n \) lies in a ball of radius \( B_R \) centered at the origin,
and that $T$ return vectors, $R_1, ..., R_T$, are independently drawn from $\mathcal{R}$. Then, for any feasible portfolio $\pi$ satisfying $\alpha_\beta(\pi) < \theta$, the probability of the loss $f(\pi, \mathcal{R})$ being greater than $\theta$, $\mathbb{P}\{f(\pi, \mathcal{R}) > \theta\}$, is bounded above as

$$\mathbb{P}\{f(\pi, \mathcal{R}) > \theta\} \leq (1 - \beta) + G\left(\frac{\|\pi\|_2}{\alpha_\beta(\pi) - \theta}\right) \tag{7}$$

with probability at least $1 - \delta$, where $G(x) := \sqrt{\frac{2}{T}} \left( x^2 \cdot 4e^2 B_R^2 \log_2(2T) + \ln\frac{2}{\nu} \right)$ and $c > 0$ is a constant. Besides, for $\pi$ satisfying $\phi_\beta(\pi) < \theta$, one has

$$\mathbb{P}\{f(\pi, \mathcal{R}) > \theta\} \leq (1 - \beta) + G\left(\frac{\|\pi\|_2}{\phi_\beta(\pi) - \theta}\right). \tag{8}$$

**Remark 2** From the viewpoint of the generalization error for OC-$\nu$SVM, the above bounds are different from the known one. We provide a new generalization error bound for OC-$\nu$SVM in Section B.2 for readers interested in statistical learning.

Similarly to the upper bounds, for a portfolio $\pi$ satisfying $\alpha_\beta(\pi) > \theta$, the loss probability is also bounded below by a function associated with $\beta$-VaR.

**Theorem 2** Suppose the same assumption as in Theorem 1. For $\pi$ satisfying $\alpha_\beta(\pi) > \theta$, $\mathbb{P}\{f(\pi, \mathcal{R}) > \theta\}$ is bounded below as

$$\mathbb{P}\{f(\pi, \mathcal{R}) > \theta\} \geq (1 - \beta) - G\left(\frac{\|\pi\|_2}{\alpha_\beta(\pi) - \theta}\right) \tag{9}$$

with probability at least $1 - \delta$.

Let us emphasize that in the bounds (7), (8) and (9) the (unobservable) true probability $\mathbb{P}\{f(\pi, \mathcal{R}) > \theta\}$ for a portfolio $\pi$ can be related to the (observable) empirical $\beta$-VaR $\alpha_\beta(\pi)$ and $\beta$-CVaR $\phi_\beta(\pi)$. Although these bounds may not be tight for specific distributions, we expect that minimizing the right-hand side of (7) or (8) helps in lowering the loss probability because 1) on the basis of such a loose error bound, SVMs have shown their success in many applications; 2) asset return distribution cannot be specified in reality, and the above theorems consider the underlying distribution in a robust manner so that they hold true for all possible bounded distributions which include fat-tailed or skewed distributions.

In order to minimize the bounds, we separately conduct the minimization in $\pi$ and that in $\beta$, as in the applications of SVMs. In other words, we minimize the ratio $(\alpha_\beta(\pi) - \theta)/\|\pi\|_2$ or $(\phi_\beta(\pi) - \theta)/\|\pi\|_2$ for various $\beta$s in $(0, 1)$ and choose the best $\beta$ along some criterion by using only historical data. One of advantages of the separate minimization approach is that we need not worry how to estimate the parameters $c$ and $B_R$, or whether the bounds is convex in $\beta$ or not. No need for knowing $B_R$ in optimization is a good news in the sense that the optimization can be applied for any distribution as long as independent samples are available.

Also, we should note that the bounds in Theorems 1 and 2 hold even if $2$-norm is replaced with the other norms since any two norms in a finite dimensional vector space are equivalent, i.e., $\exists C_1, C_2 > 0 : \forall x \in \mathbb{R}^n, C_1\|x\| \leq \|x\|_2 \leq C_2\|x\|$. Therefore, the bounds (7), (8) and (9) can be modified by replacing $2$-norm $\|\pi\|_2$ with any other norm $\|\pi\|$. Therefore, we below consider the following fractional programs for obtaining a portfolio:

$$\min\{\frac{\alpha_\beta(\pi) - \theta}{\|\pi\|} : \pi \in \Pi\}; \tag{10}$$

$$\min\{\frac{\phi_\beta(\pi) - \theta}{\|\pi\|} : \pi \in \Pi\} = \min\{\frac{F_\beta(\pi, \alpha) - \theta}{\|\pi\|} : \pi \in \Pi, \alpha \in \mathbb{R}\}. \tag{11}$$
2.3 On the Role of Short-Sale Constraint in Ordinary VaR/CVaR Minimization

The fact that any norm can be applied in place of 2-norm in Theorems 1 and 2 provides the standard VaR and CVaR minimizations (5) and (6) with a theoretical underpinning when the so-called short-sale constraint, i.e., \( \pi \geq 0 \), is imposed on. Let us consider that 1-norm \( \| \pi \|_1 = 1 \) is applied in place of 2-norm \( \| \pi \|_2 \). By noting the equivalence \( \{ \pi : e_n^\top \pi = 1, \pi \geq 0 \} = \{ \pi : \| \pi \|_1 = 1, \pi \geq 0 \} \), the functions \( \{ \alpha (\pi) - \theta \| \pi \|_1 \} \) in (10) and \( \{ \phi (\pi) - \theta \| \pi \|_1 \} \) in (11) can be replaced with \( \alpha (\pi) - \theta \) and \( \phi (\pi) - \theta \), respectively, when the short-sale constraint is imposed. From this, we have the following propositions.

**Proposition 1** When 1-norm is employed in place of the norm, the fractional program (10) results in the ordinary VaR minimization (5) with the short-sale constraint.

\[
\min\{ \alpha : -R_i^\top \pi - M z_t \leq \alpha, (t = 1, \ldots, T), e_n^\top z \leq [(1 - \beta)T], z \in \{0, 1\}^T, A \pi \leq b, e_n^\top \pi = 1, \pi \geq 0 \}.
\]

Similarly, when 1-norm is employed, (11) results in the ordinary CVaR minimization (6) with the short-sale constraint.

\[
\min\{ \alpha + \frac{1}{(1 - \beta)T}e_n^\top y : y_t \geq -R_i^\top \pi - \alpha, y_t \geq 0, (t = 1, \ldots, T), A \pi \leq b, e_n^\top \pi = 1, \pi \geq 0 \}.
\]

Combined with Theorems 1 and 2, this proposition indicates that imposing the short-sale constraint, \( \pi \geq 0 \), plays a role not only in preventing short-position, but also in improving the out-of-sample performance of the ordinary VaR and CVaR minimizations.

2.4 On the Similarity to the Normal Probability Minimization

It is interesting to see the similarity between the fractional programs (10) and (11) and the normal probability minimization (3). The latter is derived under a parametric distribution whereas the former is derived under nonparametric assumption.

Instead, it is contrasting that (10) and (11) look far different from the empirical loss probability minimization (4) although both are derived under nonparametric assumption.

In particular, noting that \( \phi_0 (\pi) \) is equal to \( -\hat{\mu}^\top \pi \), we observe the normal probability minimization (3) is a special case of (11). In this sense, we can expect that an optimal portfolio via (3) achieves a good out-of-sample performance not only for normally distributed returns, but also for wider class of return distributions.

3 Minimization of the Fractional Functions

3.1 Reformulation of the Fractional Programs into Norm-Constrained Conic Programs

In this section, we pose an algorithm for solving the fractional programs (10) and (11). In the remainder of the paper, we assume that \( \Pi \) is given by a polytope of the form \( \Pi = \{ \pi \in \mathbb{R}^n : e_n^\top \pi = 1, A \pi \leq b \} \) where \( A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m \). This assumption is usually supposed in practical portfolio selection.

By introducing a variable \( \eta > 0 \) satisfying \( \eta = 1/\| \pi \| \), and replacing as \( (\pi, \alpha) \leftarrow \eta \cdot (\pi, \alpha) \), (10) and (11) can be transformed into the following norm constrained optimizations:

\[
\min_{\pi, \eta}\{ \alpha (\pi) - \theta \eta : \| \pi \| = 1, e_n^\top \pi = \eta, A \pi \leq b \eta, \eta \geq 0 \}.
\]
\[
\min_{\pi, \alpha, \eta} \{ F_\beta(\pi, \alpha) - \theta \eta : \| \pi \| = 1, \ e_n^\top \pi = \eta, \ A\pi \leq b\eta, \ \eta \geq 0 \},
\]
respectively, where \( \eta \) can be deleted as follows:

\[
\min_{\pi, \alpha, \eta} \{ \alpha_\beta(\pi) - \theta e_n^\top \pi : \| \pi \| = 1, (A - be_n^\top)\pi \leq 0, \ e_n^\top \pi \geq 0 \};
\]

\[
\min_{\pi, \alpha, \eta} \{ F_\beta(\pi, \alpha) - \theta e_n^\top \pi : \| \pi \| = 1, (A - be_n^\top)\pi \leq 0, \ e_n^\top \pi \geq 0 \}.
\]

The objective of Problem (12) is a nonconvex function while that of (13) is a convex function. However, Problems (13) also results in a nonconvex program because of the norm constraint.

We next show the equivalence of the fractional programs and the norm-constrained programs. First, let us note the following lemma.

**Lemma 1** When \( \Pi \) is bounded, optimal solutions \( \pi \) to (12) and (13) satisfy \( e_n^\top \pi > 0 \).

The norm-constrained optimization problems (12) and (13) are then equivalent to the fractional programs (10) and (11), respectively, in the following sense:

**Proposition 2** Suppose that \( \Pi \) is bounded. One then has

(i) If (12) has an optimal solution \( \pi^* \), then \( \pi^*/\| \pi^* \| \) is an optimal solution to (10), and the two optimal values meet. If (10) has an optimal solution \( \pi^\star \), then \( \pi^\star/\| \pi^\star \| \) is an optimal solution to (12), and the two optimal values meet.

(ii) If (13) has an optimal solution \( (\pi, \alpha) \), then \( (\pi, \alpha)/(e_n^\top \pi) \) is an optimal solution to (11), and the two optimal values meet. If (11) has an optimal solution \( (\pi, \alpha) \), then \( (\pi, \alpha)/\| \pi \| \) is an optimal solution to (13), and the two optimal values meet.

Proof of this is omitted since it can be shown in a straightforward manner.

### 3.2 Nonconvexity of the Norm Constraint and the Sign of the Optimal Values

We next analyze the relation between the tractability of the norm constraint and the sign of the optimal value of the reformulated problems by extending the results of Gotoh and Takeda (2005). For simplicity, let \( h(\pi, \alpha) \) denote a positively homogeneous function of degree one, and consider

\[
\min_{\pi, \alpha} \{ h(\pi, \alpha) : \pi \in K, \| \pi \| = 1 \},
\]

where \( K \subset \mathbb{R}^n \) is a cone. It is easy to see that both (12) are (13) are in this form.

By slightly modifying the discussion of Gotoh and Takeda (2005), one has the following statements:

**Proposition 3** Let \( (\pi^\star, \alpha^\star) \) be an optimal solution to the problem of the form

\[
\min_{\pi, \alpha} \{ h(\pi, \alpha) : \pi \in K, \| \pi \| \leq 1 \}.
\]

Then, the followings hold.

1. If the optimal value of (14) is negative, one has \( \| \pi^\star \| = 1 \).
2. If the optimal value of (14) is positive, one has \( \| \pi^\star \| = 0 \).
Proposition 3 reveals that if the optimal value of (14) is negative, solving Problem (15), which has a convex feasible region, results in a solution to Problem (14), which has a nonconvex region. As a result, the sign of the optimal value of the fractional programs (10) and (11) indicates the difficulty associated with the convexity of the feasible region of (12) and (13), respectively.

The following corollary can be obtained from the contraposition of the statements in the above proposition.

**Corollary 1** Let \((\pi^*, \alpha^*)\) be an optimal solution of Problem (15). Then, the followings hold.

1. If \(\|\pi^*\| = 1\) holds, then \((\pi^*, \alpha^*)\) is an optimal solution of Problem (14). In this case, the optimal value is nonpositive, i.e., \(h(\pi^*, \alpha^*) \leq 0\).
2. If \(0 < \|\pi^*\| < 1\) holds, then \(\frac{1}{\|\pi^*\|}(\pi^*, \alpha^*)\) is an optimal solution of Problem (14), and the optimal values of both (14) and (15) are 0, i.e., \(h(\pi^*, \alpha^*) = h(\pi^*, \alpha^*)/\|\pi^*\| = 0\).
3. If \(\|\pi^*\| = 0\) holds, the optimal value of (14) is nonnegative.

This corollary indicates when the relaxed problem (15) becomes equivalent to the original one (14).

### 3.3 Algorithm for Solving the Norm-Constrained Problems

Based on Proposition 3 and Corollary 1, we below describe a two-step framework for approaching a solution to the norm-constrained conic problem (14).

**Algorithm 1** Two-Step Framework for the Norm-Constrained Conic Program

**Step 1.** Solve (15) and let \((\hat{\pi}, \hat{\alpha})\) be an optimal solution.

- **Case (a):** \(\|\hat{\pi}\| = 1\) Quit. \((\hat{\pi})\) is optimal to (14).
- **Case (b):** \(0 < \|\hat{\pi}\| < 1\) Quit. \((\hat{\pi}/\|\hat{\pi}\|)\) is optimal to (14).
- **Case (c):** \(\|\hat{\pi}\| = 0\) Go to Step 2.

**Step 2.** Solve (14) in an approximate manner.

In this framework, if the optimal value of (14) is nonpositive, an optimal solution to (14) is found at Step 1 by solving (15). Otherwise, a solution would be found at Step 2 by solving the equality norm-constrained problem (14) in a heuristic manner via Algorithm 2. If Step 1 of Algorithm 1 results in a solution \(\hat{\pi}\) satisfying \(\|\hat{\pi}\| = 1\) in solving (12) or (13), we see that \(\alpha_\beta(\hat{\pi}) - \theta e_n^\top \hat{\pi} \leq 0\) or \(\phi_\beta(\hat{\pi}) - \theta e_n^\top \hat{\pi} \leq 0\), respectively. On the other hand, if Step 1 results in \(\hat{\pi} = 0\), we have \(\alpha_\beta(\hat{\pi}) - \theta e_n^\top \hat{\pi} \geq 0\) or \(\phi_\beta(\hat{\pi}) - \theta e_n^\top \hat{\pi} \geq 0\). Thus, if the VaR minimization (12) and the CVaR minimization (13) terminate in Step 1 and the optimal value is nonzero, the resulting solutions \(\hat{\pi}/e_n^\top \hat{\pi}\) are underpinned by Theorem 1. In addition, if the VaR minimization (12) terminates in Step 2, the solution is underpinned by Theorem 2.

In this paper, the following subroutine is applied for Step 2 of the above algorithm. It should be noted that this subroutine may result in a non-global solution, but it will stop at a solution in finite iterations under a mild assumption (Gotoh and Takeda, 2005).

### 3.4 Norm-Constrained VaR Minimization

In the following two subsections, we describe the details for implementing the algorithm for solving (12) and (13) respectively because the resulting optimization problems have different difficulties.
Algorithm 2 Subroutine for Step 2.

Step 2.0. Obtain a nontrivial feasible solution \( \pi \neq 0 \), by, for example, solving a tractable problem:
\[
\min \{ h(\pi, \alpha) : A\pi \leq b, e_n^\top \pi = 1 \}.
\] (16)

Let \((\pi^0, \alpha^0)\) be an optimal solution to (16). Replacing as \( \pi^0 \leftarrow \pi^0 / \|\pi^0\|_2 \) and \( k \leftarrow 1 \).

Step 2.k. With \( \pi^{k-1} \), solve a linearly constrained problem:
\[
\min \{ h(\pi, \alpha) : (A - be_n^\top)\pi \leq 0, e_n^\top \pi \geq 0, (\pi^{k-1})^\top \pi = 1 \}.
\] (17)

Let \( \pi^k \) be an optimal solution to (17). If \( \|\pi^k\|_2 = 1 \), terminate the algorithm with a solution \( \pi^k \). Otherwise, replace as \( \pi^k \leftarrow \pi^k / \|\pi^k\|_2 \) and \( k \leftarrow k + 1 \), and repeat Step 2.k.

Upper Bound Minimization at Step 1. When the norm-constrained VaR minimization (12) is considered, the relaxed problem in Step 1 results in the following mixed integer quadratically constrained program (MIQCP):
\[
\min \{ \alpha - \theta e_n^\top \pi : \begin{align*}
-R_t^\top \pi - Mz_t &\leq \alpha, (t = 1, ..., T), 
-e_T^\top z &\leq [(1 - \beta)T], 
z &\in \{0, 1\}^T, 
(A - be_n^\top)\pi &\leq 0, e_n^\top \pi \geq 0, (\pi^k)^\top \pi \leq 1
\end{align*}\},
\] (18)
where \( \lfloor x \rfloor \) denotes the maximum integer less than or equal to \( x \), and \( M \) is a sufficiently large constant. Not-so-large MIQCPs can be solved in a practical time by employing the state-of-the-art optimization software. For example, it took about 170 seconds on average for solving a problem of size \( n = 20 \) and \( T = 120 \) on a workstation environment used in the numerical experiment in Section 4. However, computation time will exponentially grow as the problem size increases. Besides, when iterative computations are required for parameter tuning, computation time for solving single problem should be smaller. Considering such a heavy computational burden, we only solve the CVaR-based model (11) in the next section.

Lower Bound Minimization at Step 2. When Step 1 of Algorithm 1 results in a meaningless solution satisfying \( \bar{\pi} = 0 \), we have to solve MILPs iteratively in Step 2. This can still be a hard task, but it is much better than directly solving the equality-constrained problem of the form (14). In addition, the number of iterations in Step 2 is expected to be small according to Gotoh and Takeda (2005).

3.5 Norm-Constrained CVaR Minimization

Upper Bound Minimization at Step 1. When the norm-constrained CVaR minimization (13) is considered, the relaxed problem (15) with \( h(\pi, \alpha) = F_\beta(\pi, \alpha) \) in Step 1 results in the following quadratically constrained program (QCP):
\[
\min \{ \alpha + \frac{1}{(1-\beta)T} e_T^\top y - \theta e_n^\top \pi : \begin{align*}
y_t &\geq -R_t^\top \pi - \alpha, (t = 1, ..., T), 
y_T &\geq 0, (A - be_n^\top)\pi \leq 0, e_n^\top \pi \geq 0, (\pi^k)^\top \pi \leq 1, \end{align*}\},
\] (19)
which can be solved in an efficient manner by employing a standard optimization software.

CVaR Minimization at Step 2. Although no theoretical support has been obtained from Theorem 1 or 2 for the CVaR minimization in the case of \( \phi_\beta(\pi) \geq \theta \), we expect to attain a lower loss probability even in that case because 1) \( \beta \)-CVaR is a tight upper bound of \( \beta \)-VaR especially \( \forall \beta \in (0, 1) \).
when $\beta$ is large; 2) the nonconvex case which is to be solved in Step 2 sometimes ended up with better out-of-sample results in the classification problem (Gotoh and Takeda, 2005). The CVaR minimizations which are to be repeatedly solved in Step 2 are linear programs, and are more tractable than QCP which appears in Step 1.

4 Numerical Example

In this section, we apply the methods we discussed to real stock market data which consist of 256 monthly returns of stocks listed on the NIKKEI225 index at the end of August 2008. The investable stock set of size $n$ is randomly chosen from 225 stocks listed on the index, $n = 20, 40, 60, 80, 100, 120$.

In order to evaluate the out-of-sample performance of each portfolio optimization model along the time series, we adopt a rolling horizon scheme as follows. At the $\tau$-th time window, $\tau = T + 1, ..., 256$, a portfolio $\pi_\tau^*$ is determined by using $T = 120$ return observations $R_{\tau-T}, ..., R_{\tau-1}$ and is applied to the next month return $R_\tau$. We repeat this $136 (= 256 - T)$ times for the whole given data by sliding the time window.

We solved the CVaR-based fractional program (11) with 2-norm via the algorithm described in Section 3. The parameter $\beta$ was dynamically chosen as follows. Each time a portfolio is determined by using historical data of size $T$, the first $\frac{5}{6}T$ is used for computing a portfolio with some $\beta$ and the obtained portfolio is applied to the remaining $\frac{1}{6}T$ for testing its performance. This is repeated for $\beta = 0.5, 0.6, ..., 0.9$, and one $\beta$ is chosen among the five so that the portfolio achieves the highest (modified) Sharpe ratio (see, e.g., DeMiguel et al., 2009) for the test data. With the chosen $\beta$, we afresh compute a portfolio $\pi_\tau^*$ by using the data of size $T$ and apply it to the next return $R_\tau$.

Loss Probability. First of all, we compare the empirical probability minimization (4) and the CVaR-based bound minimization (11) in terms of loss probability $P\{f(\pi, R) > \theta\}$. As stated in Section 2, the former minimizes the empirical version of the probability whereas the latter does a (seemingly loose) bound of the probability. The thresholds $\theta$ in (4) and (11) are set equal to the 95-percentiles of the realized distributions $R_{\tau-T}^\top \pi_{\tau-1}^*, ..., R_{\tau-1}^\top \pi_{\tau-1}^*$ with previous portfolios $\pi_{\tau-1}^*$ of (4) and (11), respectively. This is expected to reduce the frequency of large loss which exceeds the 95-percentile.

Figures 3 (a) to (c) show the out-of-sample loss distributions $|\{\tau = 121, ..., 256 : f(\pi_\tau^*, R_\tau) > \theta\}|/136$ for $n = 20, 60, 100$. Note that these can be regarded as a realization of the loss probability $P\{f(\pi^*, R) > \theta\}$ with an optimal portfolio $\pi^*$. We see from these figures that more or less, the CVaR-based bound minimization (11) achieves smaller loss probability than the empirical probability minimization (4). Especially, the former dominates the latter in Figure 3 (b). Although there is no longer dominance relation in Figure 3 (c), the dispersion becomes larger except for middle $\theta$. Besides, for every $n$, the CVaR minimization (11) achieves smaller loss probability at big $\theta$s.

Mean and Standard Deviation. Next we present the mean-risk property of the CVaR minimization. Returns of the $1/n$-portfolio (DeMiguel, Garlappi and Uppal, 2009), i.e., the equally weighted portfolio, and solutions to the empirical probability minimization (4) and minimum variance model are also computed for comparison.

Figure 4 shows the mean and standard deviation of out-of-sample returns for various $n$s. We see that the CVaR minimization (11) achieved smaller standard deviation than the $1/n$-portfolio and the empirical probability minimization (4). In particular, the CVaR minimization dominates the empirical probability minimization in the mean-risk sense, i.e., smaller deviation.
and larger mean. Although it has no dominating or dominated relation to the minimum variance model, its higher return behavior (under relatively low risk) is promising since in the literature the minimum variance model has been shown to perform as well as the mean-variance model.

**Tuning Strategy of** $\beta$. Next we check the effectiveness of the dynamic tuning strategy of $\beta$ in the CVaR minimization (11). To this end, we compare the dynamic strategy with static ones in which $\beta$ is fixed to one of the values 0.5, 0.7, 0.9. Figures 5 (a) to (c) show cumulative returns of the CVaR minimizations with the different strategies for $n = 20, 60, 100$, respectively. The best $\beta$ for the static strategies varies depending on $n$. In fact, the static strategy with $\beta = 0.9$ outperforms the other two for $n = 20$, whereas the difference becomes small for $n = 100$. On the other hand, the dynamic tuning constantly shows good performances for all $n$, and the advantage over the static strategies becomes clearer as the number of assets grows. This suggests that the Sharpe ratio criterion for the dynamic tuning of $\beta$ works to a certain extent.
Behavior of the Two-step Algorithm. Finally let us mention the behavior of the two-step algorithm. Table 1 reports how many times the algorithm terminated in Step 1 or Step 2 among the 136 trials of obtaining portfolios for the out-of-sample evaluations. The rows “total wrt. $\beta$” show the numbers of $\beta$s chosen by the dynamic tuning in the rolling horizon scheme. The number of problems whose solutions were found at Step 2 indicates how often the problem resulted in the nonconvex optimization discussed in Section 3. We see from the table that the dynamic tuning often attained large $\beta$s with which we have to deal with the nonconvex structure. Considering the good performance of the dynamic tuning strategy relies in part on large $\beta$s, the nonconvex cases are also worth solving as in the classification problem (Gotoh and Takeda, 2005).

5 Concluding Remarks

In this paper, we have posed general portfolio optimization models, which are formulated as fractional programs whose objective is a ratio where the numerator includes VaR or CVaR while the denominator is any norm of portfolio vector. The fractional programming formulation is considered as a generalization of traditional approaches in the following sense: 1) the CVaR-based formulation includes the classical safety first criterion under normal distribution assumption as a special case ($\beta = 0, \|\pi\| = \sqrt{\pi^\top \Sigma \pi}$); 2) It includes the ordinary VaR and CVaR minimizations when the short-sale constraint is imposed and $\|\pi\| = \|\pi\|_1$. The formulation is derived so as to reduce upper and lower bounds of loss probability, which are obtained by modifying the generalization error bound for an SVM. In this sense, those bounds bring a theoretical underpinning to the criterion and the use of VaR and CVaR in portfolio selection.

The numerical results show that the CVaR-based probability bound minimization outperforms the empirical probability minimization in terms of realized loss probability and mean-risk performance. Also, the CVaR-based minimization achieves a promising portfolio performance, yielding relatively small risk and higher return than the benchmarks. This may be in part due to employing the Sharpe ratio criterion for the dynamic tuning of $\beta$. In fact, the dynamic strategy achieved better performance than the static strategies.

On the other hand, we did not conduct experiments on the VaR-based formulation. This is because prohibitively large computation are required for solving hundreds of problems of practical size. Therefore, efficient solution to the VaR-based model is also to be developed.
Table 1: Classification of Problems: Termination Step and Chosen Value of $\beta$

\begin{tabular}{l|ccccc|c}
\hline
 & 0.5 & 0.6 & 0.7 & 0.8 & 0.9 & total \\
\hline
Termination Step & & & & & & \\
\hline
$n = 20$ & Step 1. (convex) & 38 & 11 & 17 & 15 & 5 & 86 \\
 & Step 2. (nonconvex) & 0 & 0 & 0 & 4 & 46 & 50 \\
 & total wrt. $\beta$ & 38 & 11 & 17 & 19 & 51 & 136 \\
$n = 60$ & Step 1. (convex) & 41 & 7 & 26 & 15 & 10 & 99 \\
 & Step 2. (nonconvex) & 0 & 0 & 0 & 3 & 34 & 37 \\
 & total wrt. $\beta$ & 41 & 7 & 26 & 19 & 44 & 136 \\
$n = 100$ & Step 1. (convex) & 22 & 27 & 15 & 31 & 10 & 105 \\
 & Step 2. (nonconvex) & 0 & 0 & 0 & 2 & 29 & 31 \\
 & total wrt. $\beta$ & 22 & 27 & 15 & 33 & 39 & 136 \\
\hline
\end{tabular}

Acknowledgment

Research of the first author is supported by MEXT Grant-in-Aid for Young Scientists (B) 20710120. Research of the second author is supported by MEXT Grant-in-Aid for Young Scientists (B) 19710124.

A Proof of the Propositions

A.1 Proof of Theorem 1

First of all, let us show the case of $\theta = 0$, i.e., the upper bound of $P\{f(\pi, R) > 0\}$.

Let $V$ denote a ball of radius $B_R$ in $\mathbb{R}^n$, i.e., $V = \{R \in \mathbb{R}^n : \|R\|_2 \leq B_R\}$, and $F$ a class of real-valued functions on $V$, defined by $F := \{R \mapsto \pi^\top R : R \in V, \|\pi\|_2 \leq 1\}$. Let $\gamma$ be the margin of class separation. For $\gamma > 0$ and $g \in F$, define the error estimate

$$\text{Err}^\gamma[g] := \frac{1}{T}|\{t \in \{1, \ldots, T\} : g(R_t) < \gamma\}|.$$

Then, the discussion on the generalization error of (Bartlett, 1998; Schölkopf et al., 2000) leads to the following statement for a homogeneous function $g$ in $F$: With probability at least $1 - \delta$, every $g$ in $F$ has error $\text{Err}[g] := P\{g(R) < 0\}$ such as

$$\text{Err}[g] \leq \text{Err}^\gamma[g] + G(1/\gamma).$$

Then, the discussion on the generalization error of (Bartlett, 1998; Schölkopf et al., 2000) leads to the following statement for a homogeneous function $g$ in $F$: With probability at least $1 - \delta$, every $g$ in $F$ has error $\text{Err}[g] := P\{g(R) < 0\}$ such as

$$\text{Err}[g] \leq \text{Err}^\gamma[g] + G(1/\gamma).$$

In order to derive the generalization performance for a portfolio $\pi$, we substitute $\alpha'_\beta := -\alpha_\beta(\pi)/\|\pi\|_2 > 0$ as the margin $\gamma$. Let $q(\pi) := -f(\pi, R)/\|\pi\|_2 = R^\top \pi/\|\pi\|_2$. Noting $q(\pi) \in F$, we have the error estimate

$$\text{Err}^\gamma[q(\pi)] = \frac{1}{T}|\{t \in \{1, \ldots, T\} : f(\pi, R_t) > \alpha_\beta(\pi)\}|.$$
which is bounded above by $(1-\beta)$ by definition of $\alpha_\beta(\pi)$. Therefore, $\text{Err}[q(\pi)] = \mathbb{P}\{f(\pi, R) > 0\}$ is bounded above by $(1-\beta) + G(\|\pi\|^2/\alpha_\beta(\pi))$. If $\phi_\beta(\pi) < 0$, the bound is, furthermore, bounded above by $(1-\beta) + G(\|\pi\|^2/\phi_\beta(\pi))$ because of $\alpha_\beta(\pi) \leq \phi_\beta(\pi)$.

As for the case of $\theta \neq 0$, by noting $e_n^\top \pi = 1$, one has
\[
\mathbb{P}\{-R^\top \pi > \theta\} = \mathbb{P}\{- (R + \theta e_n)^\top \pi > 0\}.
\]
Replacing $R$ with $R + \theta e_n$, the upper bounds are obtained. \hfill \Box

### A.2 Proof of Theorem 2

As in the proof of Theorem 1, we first consider the case of $\theta = 0$. In this case we regard $\alpha_\beta(\pi)/\|\pi\|$ as the margin $\gamma$, and obtain $\text{Err}[-q(\pi)] = 1 - \text{Err}[q(\pi)] \leq \text{Err}[-q(\pi)] + G(\|\pi\|^2/\alpha_\beta(\pi))$. Then, $1 - \text{Err}[-q(\pi)] = \frac{1}{T}\{|t : f(\pi, R_t) > \alpha_\beta(\pi)|\}$ is bounded below by $(1-\beta)$ by definition. By considering $e_n^\top \pi = 1$, we establish the theorem for the case of $\theta \neq 0$ as in the proof of Theorem 1. \hfill \Box

### A.3 Proof of Lemma 1

Suppose that $\Pi$ is bounded. Then, adding a set of box constraints of the form $-Me_n \leq \pi \leq Me_n$ to Problems (10) and (11) does not alter their feasible regions, where $M > 0$ is a large constant. The feasible regions of Problems (12) and (13) then satisfy $-M(e_n^\top \pi)e_n \leq \pi \leq M(e_n^\top \pi)e_n$. From this, if one has $e_n^\top \pi = 0$ at optimality, any optimal solution $\pi$ should satisfy $\pi = 0$, which contradicts the norm constraint $\|\pi\|_2 = 1$ of (12) and (13). \hfill \Box

### A.4 Proof of Proposition 3

Let $(\tilde{\pi}^*, \tilde{\alpha}^*)$ be an optimal solution to (14). Then for any feasible solution $(\tilde{\pi}, \tilde{\alpha})$ of (14), one has $h(\tilde{\pi}^*, \tilde{\alpha}^*) \leq h(\tilde{\pi}, \tilde{\alpha})$. Noting that any feasible solution to (15) is expressed as $\eta \cdot (\tilde{\pi}, \tilde{\alpha})$ for $\eta \in [0, 1]$, the followings are easily obtained: when $h(\tilde{\pi}^*, \tilde{\alpha}^*) < 0$, $(\tilde{\pi}^*, \tilde{\alpha}^*)$ achieves the minimum value in (15), i.e., $\pi^* = \tilde{\pi}^*$, and when $h(\tilde{\pi}^*, \tilde{\alpha}^*) > 0$, $0 \cdot (\tilde{\pi}^*, \tilde{\alpha}^*)$ is the unique minimum solution in (15), i.e., $\pi^* = 0$. \hfill \Box

### B Connection of the $\beta$-CVaR Minimization to OC-$\nu$SVM

In this part, we describe a connection of the $\beta$-CVaR Minimization to OC-$\nu$SVM and generalization bound for OC-$\nu$SVM which is different from that in Schölkopf and Smola (2002).

#### B.1 Interpretation of OC-$\nu$SVM as a CVaR Minimization

Extending the discussion in (Gotoh and Takeda, 2005; Takeda, 2009), we can provide an interpretation of the formulation (2) for OC-$\nu$SVM in terms of CVaR minimization.

**Proposition 4** The minimization of the CVaR associated with the loss $-\alpha^\top w/\|w\|_2$:
\[
\min\{ \alpha + \frac{1}{(1-\beta)\mu} \sum_{i=1}^{m} [-\alpha_i^\top w/\|w\|_2 - \alpha]_+ : w \in \mathbb{R}^d, \alpha \in \mathbb{R} \} \tag{21}
\]
is equivalent to OC-$\nu$SVM (2) with change of variable and parameter as $\alpha = -\rho/\|w\|_2$ and $1-\beta = \nu$ if the optimal value of (21) is negative.
Proof. In order to prove this statement, we first observe that (21) is equivalent to the following norm-constrained optimization via changing variable and parameter:

\[ \min \{ -\rho + \frac{1}{\nu m} \sum_{i=1}^{m} [\rho - a_i^\top w]^+ : \|w\|_2 = 1, w \in \mathbb{R}^d, \rho \in \mathbb{R} \}. \]  \hspace{1cm} (22)

Thus, what we would like to prove is that as long as the optimal value of (22) is negative, OC-\(\nu\)-SVM (2) provides the same linear function as that of (22) except the scaling. Next, note the following lemma:

Lemma 2 Let \((w^*, \rho^*)\) be an optimal solution to (2). Then, (i) the optimal value of (22) is negative if and only if (ii) \(\|w^*\|_2 > 0\).

Proof of the lemma

(i)\(\Rightarrow\)(ii) Let \((\tilde{w}, \tilde{\rho})\) be an optimal solution to (22). Suppose that \(\|w^*\|_2 = 0\) as well as (i), i.e., \(-\tilde{\rho} + \frac{1}{\nu m} \sum_{i=1}^{m} [\tilde{\rho} - a_i^\top \tilde{w}]^+ < 0\). On the other hand, one observes that the optimal value of (2) is zero since \(\frac{1}{2} - 0 - \rho^* + \frac{1}{\nu m} \sum_{i=1}^{m} [\rho^*]^+ = \frac{1}{2} \max\{(1 - \nu)\rho^*, -\nu \rho^*\} \geq 0\). Note that \((k\tilde{w}, k\tilde{\rho})\) is feasible to (2) for any \(k > 0\), and its objective value is \(\frac{1}{2} k^2 \|\tilde{w}\|_2^2 - k\tilde{\rho} + \frac{1}{\nu m} \sum_{i=1}^{m} [k\tilde{\rho} - (a_i, k\tilde{w})]^+\). However, this can be negative for sufficiently small \(k > 0\), contradicting the optimality of \((w^*, \rho^*)\).

(ii)\(\Rightarrow\)(i) Suppose that the objective value of (22) is nonnegative as well as \(\|w^*\|_2 > 0\). One then has \(0 \leq -\tilde{\rho} + \frac{1}{\nu m} \sum_{i=1}^{m} [\tilde{\rho} - a_i^\top \tilde{w}]^+ < \frac{1}{2} \|w^*\|_2^2 - \rho^* + \frac{1}{\nu m} \sum_{i=1}^{m} [\rho^* - (a_i, w^*)]^+\). \((k\tilde{w}, k\tilde{\rho})\) is feasible to (2) for any \(k > 0\), and the corresponding objective value is \(\frac{1}{2} k^2 \|\tilde{w}\|_2^2 + k(-\tilde{\rho} + \frac{1}{\nu m} \sum_{i=1}^{m} [\tilde{\rho} - (a_i, \tilde{w})]^+)\), which can be smaller than the optimal value of (2) for sufficiently small \(k > 0\), which contradicts the optimality of \((w^*, \rho^*)\).

Based on this lemma, we can see that \((\frac{w^*}{\|w^*\|_2}, \|w^*\|_2^2)\) is an optimal solution to (22) when the optimal value of (22) is negative. On the contrary, when the optimal value of (22) is positive, (2) results in a meaningless solution, \(w^* = 0\), while (22) returns a solution with \(\tilde{w} \neq 0\).

\[\square\]

B.2 A Generalization Error Bound for OC-\(\nu\)-SVM

From the view point of the generalization error for OC-\(\nu\)-SVM, the bounds (7) and (8) are different from the known one. Indeed, the bound given in Schölkopf and Smola (2002) are held only at the optimal solution of the problem (2). On the other hand, (7) and (8) bound the loss probability at any vector satisfying \(e_\alpha^\top \pi = 1\). Therefore, by slightly modifying the proof of Theorem 1, we can obtain the following bounds for OC-\(\nu\)-SVM.

Proposition 5 (Generalization bounds for OC-\(\nu\)-SVM) Let \(\theta_A\) be a threshold. Let \(\alpha_{1-\nu}(w)\) be the \((1 - \nu)\)-VaR associated with the loss \(-A(w) = -a^\top w\). Suppose that the \(n\)-dimensional random vector \(a\) has a bounded support in the sense that the \((n+1)\)-dimensional vector \(\begin{pmatrix} a \\ \theta_A \end{pmatrix}\) lies in a ball of radius \(B_A\) centered at the origin, and that \(\alpha_1, \ldots, \alpha_m\) are independently observed. Then, for any weight vector \(w\) satisfying \(\alpha_{1-\nu}(w) < \theta_A\), the probability of the function \(A(w)\) being smaller than \(-\theta_A\) is bounded above as

\[ \mathbb{P}\{A(w) < -\theta_A\} \leq \nu + G\left(\frac{\|w\|_2 + 1}{\alpha_{1-\nu}(w) - \theta_A}\right) \hspace{1cm} (23) \]

17
with probability at least $1 - \delta$, where $G(x) := \sqrt{\frac{2}{m} \left\{ x^2 \cdot 4c^2B_A^2 \log_2(2m) + \ln \frac{2}{\delta} \right\}}$ and $c > 0$ is a constant. Besides, let $\phi_{1-\nu}(w)$ be the $(1 - \nu)$-CVaR associated with the loss $-A(w)$. Then, for $w$ satisfying $\phi_{1-\nu}(w) < \theta_A$, one has
\[ P\{A(w) < -\theta_A\} \leq \nu + G\left(\frac{\|w\|_2 + 1}{\phi_{1-\nu}(w) - \theta_A}\right). \] (24)

When one sets $\theta_A = -\rho^*$ and $w = w^*$, the bound (24) is equal to that of Schölkopf et al. (2000).

**Proof.** Using $\hat{w} = \begin{pmatrix} w \\ 1 \end{pmatrix}$ and $\hat{a} = \begin{pmatrix} a \\ \theta_A \end{pmatrix}$, we regard a homogeneous function $\hat{a}^\top \hat{w}/\|\hat{w}\|_2$ as $g$ in (20) and set a margin as $\gamma = (\theta_A - \alpha_{1-\nu}(w))/\|\hat{w}\|_2 > 0$. Noticing that $\frac{1}{m}\{|i \in \{1, ..., m\}: -a_i^\top w > \alpha_{1-\nu}(w)\| \leq \nu$, we have a bound (23) of the generalization error $\text{Err}[g] = P\{g(R) < 0\}$ as shown in the proof of Theorem 1. The bound (24) is also achieved when $\phi_{1-\nu}(w) < \theta_A$. \quad \square

**References**


