On the Role of Norm Constraints in Portfolio Selection

Jun-ya Gotoh[1] jgoto@indsys.chuo-u.ac.jp
Department of Industrial and Systems Engineering
Chuo University
2-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan

Akiko Takeda takeda@ae.keio.ac.jp
Department of Administration Engineering
Keio University
3-14-1 Hiyoshi, Kohoku, Yokohama, Kangawa 223-8522, Japan

Abstract

Several optimization approaches for portfolio selection have been proposed in order to alleviate the estimation error in the optimal portfolio. Among them are the norm-constrained variance minimization and the robust portfolio models. In this paper, we examine the role of the norm constraint in portfolio optimization from several directions. First, it is shown that the norm constraint can be regarded as a robust constraint associated with the return vector. Second, the reformulations of the robust counterparts of the value-at-risk (VaR) and conditional value-at-risk (CVaR) minimizations contain norm terms and are shown to be highly related to the ν-support vector machine (ν-SVM), a powerful statistical learning method. For the norm-constrained VaR and CVaR minimizations, a nonparametric theoretical validation is posed on the basis of the generalization error bound for the ν-SVM. Third, the norm-constrained approaches are applied to the tracking portfolio problem. Computational experiments reveal that the norm-constrained minimization with a parameter tuning strategy improves on the traditional norm-unconstrained models in terms of the out-of-sample tracking error.

Keywords: portfolio optimization; norm constraint; robust portfolio; tracking portfolio; CVaR (conditional value-at-risk)

1 Introduction

Since the seminal work of Markowitz, portfolio selection has been intensively studied in the fields of operations research and management science. Mathematically, it is a problem of determining a (normalized) weight vector $\boldsymbol{\pi}$ so that the distribution of the resulting random portfolio return $\mathcal{R}(\boldsymbol{\pi}) := \boldsymbol{\mathcal{R}}^\top \boldsymbol{\pi}$, where each component of $\boldsymbol{\mathcal{R}}$ represents the random rate of return of each asset, would have a preferable shape.

Ideally, an optimal portfolio $\boldsymbol{\pi}^*$ is a solution to a constrained optimization whose objective function is represented by a functional on the random return $\mathcal{R}(\boldsymbol{\pi})$. However, since no one knows the true distribution of the asset return $\boldsymbol{\mathcal{R}}$, what we can do in practice is to optimize its empirical counterpart which is estimated on the basis of the observed historical returns, in place of the ideal function.

Obviously, this framework can be validated by the *law of large numbers*. That is, if the number of observations goes to infinity, the solution $\bar{\boldsymbol{\pi}}$ approaches $\boldsymbol{\pi}^*$. This validation is, however, still dubious because in practice, a relatively small number of historical returns are available whereas a relatively large number of portfolio weights are to be estimated. For example, Konno and Yamazaki (1991) apply a mean-risk portfolio model to a practical case in which the number of assets is greater than that of historical observations. From a statistical viewpoint, this may cause *overfitting*, resulting in a large estimation error of the "optimal" portfolio $\bar{\boldsymbol{\pi}}$.

[1]Corresponding author

In fact, many researchers have pointed out that the mean-variance model using the sample mean vector and the sample covariance matrix results in poor out-of-sample performance because of the estimation error in the sample mean and (co)variance (see DeMiguel, Garlappi and Uppal 2009, and references can be found therein).

To improve the out-of-sample performance of the obtained portfolio, several researchers have recently proposed to estimate the covariance matrix for the minimum variance model by modifying the sample variance estimate. For example, Ledoit and Wolf (2003) suggest to use shrinkage estimates of the sample covariance matrix. Jagannathan and Ma (2003) show that imposing the short-sale constraint, $\boldsymbol{\pi} \geq \mathbf{0}$, which is usually imposed in practice, is equivalent to a shrinkage estimation of the covariance matrix. In fact, imposing the short-sale constraint may be why the mean-risk model of Konno and Yamazaki (1991) performs well even when the number of assets is greater than that of the historical data.

DeMiguel et al. (2009) additionally impose a norm constraint on the portfolio for the variance minimization criterion by extending the idea of parameter shrinkage. 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 having the short-sale constraint. Brodie et al. (2009) study 1-norm-constrained minimizations of squared tracking errors. All of the above studies incorporate the shrinkage technique in the sample covariance matrix so as to improve the out-of-sample performance of the minimum variance model.

As for the return estimate, many researchers agree that the impact of the estimation error associated with the sample mean is much worse than that of the (co)variance or other parameters. For example, the following statement in Jagannathan and Ma (2003) declares the uselessness of the sample mean estimate:

> The estimation error in the sample mean is so large nothing much is lost in ignoring the mean altogether when no further information about the population mean is available (pp. 1652-1653).

Indeed, this statement motivates DeMiguel et al. (2009) not to incorporate the return components in their objective function or constraints.

On the other hand, an optimization approach called the *robust portfolio* has been intensively studied in the last decade. It seeks a good portfolio in the sense that it is feasible even when parameters in the optimization problem take on the least favorable values among a set of predetermined candidates (the so-called *uncertainty set*). One possible critique about robust portfolio models is that many of them do not say how to specify the uncertainty set. Certain robust approaches take into account the worst-case estimation error in a direct manner. For example, Goldfarb and Iyengar (2003) nicely combine the multi-factor model and the uncertainty set in the robust portfolio, where the uncertainty set is given as a confidence interval (region) of the parameters of the factor model. Besides, very early robust models such as Soyster (1973) and Ben-Tal and Nemirovski (2000) assume a nonparametric structure, whereas the factor model approach employs a parametric assumption for constructing the uncertainty set. Except for the factor model-based models, it is unclear how robust portfolios should specify the uncertainty set in practice.

The above approaches – the shrinkage estimation-based minimum variance model and the robust portfolio – seek to alleviate the deterioration in out-of-sample performance associated with the estimation error of each optimization criterion by simultaneously considering the estimation of parameters and the selection of a portfolio. In this paper, we study the connection between these approaches by examining the role of the norm constraint not only in variance minimization but also in an extended context. In particular, special attention is paid to the norm-constrained versions of value-at-risk (VaR) and conditional value-at-risk (CVaR) minimizations.

VaR has been used in risk management for capturing large losses that may occur with a small probability. Although it is popular, there has been controversy as to its theoretical properties as a risk measure. For example, Cont, Deguest and Scandolo (2007) show that it is robust against outliers. On the other hand, it has been shown to violate subadditivity (Artzner et al. 1999), and therefore, it is now considered undesirable as a risk measure.

On the other hand, CVaR has nice theoretical properties such as coherence (Artzner et al. 1999) and consistency with the risk-averse behavior of investors (e.g., Ogryczak and Ruszczýnski 2002), and it is increasingly being used in practice. CVaR is also much more attractive than VaR from an optimization viewpoint because it often leads to a tractable associated optimization problem (Rockafellar and Uryasev 2002). Moreover, the authors have pointed out in Gotoh and Takeda (2005) and Takeda (2009) that the ν-support vector machine (ν-SVM), an optimization-based statistical learning model developed by Schölkopf et al. (2000), has almost the same structure as the CVaR minimization. This fact motivated us to exploit theoretical results developed for ν-SVMs in the context of portfolio selection.

We also consider the index tracking (mimicking) portfolio problem. We conducted numerical experiments demonstrating the norm-constrained tracking portfolio's out-of-sample performance. Specifically, we examined how the parameters used for describing the norm constraint and the CVaR objective can be tuned. We present results showing that our approach involving parameter tuning outperforms the absolute-error minimization model, a standard approach for the tracking portfolio, and performs better than the norm-constrained variant.

Our study makes the following contributions:

- We show that the norm-constrained portfolio optimization can be considered to be a robust portfolio optimization formulation with an adequate parameter uncertainty. Although the connection between the norm term and robust formulation has been recently discussed, e.g., in Xu, Caramanis and Mannor (2008), we shall focus on the implications in the financial optimization context. In particular, in combination with VaR or CVaR minimization, the norm-constrained formulation can be naturally interpreted as a robust counterpart of the standard VaR or CVaR minimization. In this sense, the norm-constrained portfolio takes into account the worst-case return in an implicit manner even though it does not explicitly include the return estimate.

- By modifying the generalization theory for ν-SVMs, known as the *generalization error bound* (Schölkopf et al. 2000), we provide a theoretical underpinning to the norm-constrained VaR or CVaR minimization. In light of these theoretical results, we can expect that the norm constraint plays a role in improving the out-of-sample performance, similarly to the norm-constrained minimum variance portfolio in DeMiguel et al. (2009). It is worth noting that although the bounds are not tight, the numerical experiments indicate that this model's out-of-sample performance is good. Also, in connection with the robust optimization for the norm constraint, this result also provides a theoretical validation for its robust counterpart.

- In contrast with the traditional models that simply minimize the empirical deviations from a target variable, we propose a novel approach to tracking portfolio construction by incorporating the norm constraint and CVaR-based deviation. Numerical experiments indicate the norm-constrained CVaR deviation model as well as the norm-constrained absolute-deviation minimization have better out-of-sample performance than the norm-unconstrained counterparts. In particular, using historical observations to tune the parameters of the norm-constrained CVaR deviation model enhances the tracking performance. Moreover, this indicates the possibility that specifying the uncertainty set in a robust portfolio on the basis of historical observations works effectively.

The structure of the paper is as follows. The next section describes a proposition that relates the norm constraint for the portfolio selection (DeMiguel et al. 2009) and an uncertainty set for robust portfolios. In Section 3, we consider norm-constrained VaR and CVaR minimizations, providing a natural connection between the norm-constrained VaR and CVaR minimizations and their robust counterparts. Also, the norm-constrained VaR and CVaR minimizations are theoretically validated by exploiting the generalization error bound for ν-SVM (Takeda 2009). In Section 4, we apply the results developed in Section 3 to a tracking portfolio problem. Section 5 is devoted to the numerical experiments in which a norm-constrained tracking portfolio is examined, and it is shown that adequate parameter tuning leads to better out-of-sample tracking performance. We conclude the paper with some remarks and provide proofs of the theorems in the Appendix.

2 Robust Optimization View on the Norm Constraint

2.1 Relation of Norms in the Norm-Constraint and the Uncertainty Set

As pointed out in the Introduction, portfolio selection shares features with parameter estimation in statistics. Inspired by the *regularization* of the regression parameter as in the ridge regression or the lasso (see, e.g., Hastie, Tibshirani and Friedman 2001), DeMiguel et al. (2009) impose the norm constraint on the minimum variance portfolio optimization which uses the sample covariance matrix $\overline{\boldsymbol{\Sigma}}$ of n assets, as follows:

$$\left| \begin{array}{ll} \min & \boldsymbol{\pi}^\top \overline{\boldsymbol{\Sigma}} \boldsymbol{\pi} \\ \text{s.t.} & \boldsymbol{e}_n^\top \boldsymbol{\pi} = 1 \\ & \|\boldsymbol{\pi}\| \leq C \end{array} \right. \tag{1}$$

where $\|\cdot\|$ is a norm in $\mathbb{R}^n$, $\boldsymbol{e}_n := (1, ..., 1)^\top \in \mathbb{R}^n$, and $C > 0$ is a constant. Here, the first constraint $\boldsymbol{e}_n^\top \boldsymbol{\pi} = 1$ implies that each component of $\boldsymbol{\pi}$ represents the investment ratio for each asset.

In DeMiguel et al. (2009), it is shown that when the 2-norm, $\|\boldsymbol{\pi}\|_2 := \sqrt{\boldsymbol{\pi}^\top \boldsymbol{\pi}}$, is employed as $\|\boldsymbol{\pi}\|$ and $C = 1/\sqrt{n}$, the solution to (1) is equivalent to the equally weighted portfolio, i.e., $\pi_j = 1/n$. On the other hand, when the 1-norm, $\|\boldsymbol{\pi}\|_1 := \sum_{j=1}^n |\pi_j|$, is employed and $C = 1$, the solution is equivalent to the short-sale-constrained minimum variance portfolio. In addition, they show that if the norm term is replaced with $\|\boldsymbol{\pi}\|_A := \sqrt{\boldsymbol{\pi}^\top \boldsymbol{A} \boldsymbol{\pi}}$, with $\boldsymbol{A}$ being the covariance matrix induced from the single factor model, the resulting portfolio is equivalent to the shrinkage estimate of the covariance matrix proposed by Ledoit and Wolf (2003).

Note that the above properties associated with $\|\boldsymbol{\pi}\|_2$ and $\|\boldsymbol{\pi}\|_1$ hold independently of the variance in the objective of (1) and hold on the basis of only on a basic constraint of the form $\boldsymbol{e}_n^\top \boldsymbol{\pi} = 1$. Thus, we first present a robust modeling from the viewpoint of the norm constraint, which is independent of the objective or the other constraint.

Proposition 1 *The norm constraint with a norm $\|\boldsymbol{\pi}\|$ is equivalent to a robust inequality in the following sense:*

$$\|\boldsymbol{\pi}\| \leq C \;\Leftrightarrow\; (\boldsymbol{r} - \boldsymbol{r}_0)^\top \boldsymbol{\pi} \geq -s, \quad \textit{for all } \boldsymbol{r} \in \mathcal{U} := \{\boldsymbol{r} : \|\boldsymbol{r} - \boldsymbol{r}_0\|^* \leq \frac{s}{C}\}$$

where $s > 0$ is a constant, $\boldsymbol{r}_0$ is a nominal vector of $\boldsymbol{r}$, and $\|\cdot\|^*$ represents the dual norm of $\|\cdot\|$, i.e., $\|\boldsymbol{r}\|^* := \sup\{\boldsymbol{r}^\top \boldsymbol{\pi} : \|\boldsymbol{\pi}\| \leq 1\}$.

Proof. By definition, the following relation holds for two mutually dual norms:

$$\|\boldsymbol{\pi}\| \le C \;\Leftrightarrow\; \boldsymbol{r}^\top\boldsymbol{\pi} \le 1, \quad \text{for all } \boldsymbol{r} \in \mathcal{U} := \{\boldsymbol{r} : \|\boldsymbol{r}\|^* \le \frac{1}{C}\}.$$

The desired result is obtained by substituting $-(\boldsymbol{r} - \boldsymbol{r}_0)/s$ for $\boldsymbol{r}$. □

If the vector $\boldsymbol{r}$ is regarded as the return of the investable assets, this proposition indicates that the norm constraint can be interpreted as a robust return constraint such that the portfolio return $\boldsymbol{r}^\top\boldsymbol{\pi}$ is no less than $\boldsymbol{r}_0^\top\boldsymbol{\pi} - s$, where $\boldsymbol{r}_0$ can be considered as the nominal portfolio return, which is possibly the sample mean $\overline{\boldsymbol{\mu}} := \sum_{t=1}^T \boldsymbol{R}_t/T$ of the observed historical return vectors $\boldsymbol{R}_1, ..., \boldsymbol{R}_T$.

The nominal return vector $\boldsymbol{r}_0$ and the positive scalar s are introduced so that we can interpret the constraint in a standard robust representation (Ben-Tal and Nemirovski 2000), and they do not appear in the norm constraint. In order to consider the relation in a more direct manner, we can give the parameters specific values. For example, let us consider the case of $\boldsymbol{r}_0 = \mathbf{0}$ and $s' = s/C$, in which the equivalence is rewritten as

$$\|\boldsymbol{\pi}\| \le C \;\Leftrightarrow\; \boldsymbol{r}^\top\boldsymbol{\pi} \ge -Cs', \quad \text{for all } \boldsymbol{r} \in \mathcal{U} := \{\boldsymbol{r} : \|\boldsymbol{r}\|^* \le s'\}.$$

On the other hand, let us consider imposing the sample return constraint on the norm-constrained feasible region as follows:

$$\boldsymbol{e}_n^\top\boldsymbol{\pi} = 1, \quad \overline{\boldsymbol{\mu}}^\top\boldsymbol{\pi} = \rho, \quad \|\boldsymbol{\pi}\| \le C,$$

where ρ is a constant. From the above observation, this can be rewritten as

$$\boldsymbol{e}_n^\top\boldsymbol{\pi} = 1, \quad \overline{\boldsymbol{\mu}}^\top\boldsymbol{\pi} = \rho, \quad \boldsymbol{r}^\top\boldsymbol{\pi} \ge \rho - Cs' \quad \text{for all } \boldsymbol{r} \in \mathcal{U} := \{\boldsymbol{r} : \|\boldsymbol{r} - \overline{\boldsymbol{\mu}}\|^* \le s'\}.$$

If one employs these as the constraints of (1), the resulting formulation represents the mean-variance model with an additional robust return constraint.

2.2 Various Norms and Their Relation to Uncertainty Sets

As stated in Proposition 1, norm-constrained portfolio optimization can be regarded as a robust portfolio selection with an uncertainty set where the dual norm is employed to describe the uncertainty of the return parameter $\boldsymbol{r}$. Table 1 summarizes the correspondence between the two representations.

It is interesting that the 1-norm for the norm constraint corresponds to the classic robust representation of Soyster (1973), which is known to result in too conservative a solution. Besides, since the 1-norm constraint with $C = 1$ is equivalent to the short sale constraint, as mentioned in DeMiguel et al. (2009), the short-sale constraint, $\boldsymbol{\pi} \ge \mathbf{0}$, is equivalent to a Soyster's type robust constraint of the form

$$(\boldsymbol{r} - \boldsymbol{r}_0)^\top\boldsymbol{\pi} \ge -s, \quad \text{for all } \boldsymbol{r} \in \mathcal{U} := \{\boldsymbol{r} : \|\boldsymbol{r} - \boldsymbol{r}_0\|_\infty := \max_{j=1,\dots,n}\{|r_j - r_{0j}|\} \le s\}.$$

On the other hand, the robust model with ellipsoidal uncertainty corresponds to the A-norm, $\|\boldsymbol{\pi}\|_A$, including the 2-norm as a special case. Interestingly, as pointed out in DeMiguel et al. (2009), the A-norm-constrained variance minimizing portfolio with the covariance matrix of the single-factor model is equivalent to the minimum variance model (Ledoit and Wolf 2003) with a shrinkage estimate using the single-factor covariance matrix for the covariance matrix estimation. In that case, the uncertainty set can be regarded as an ellipsoidal uncertainty derived from an elliptical distribution, which has a density function of the form $p(\boldsymbol{r}) := c'\det[\boldsymbol{A}]^{-1/2}q((\boldsymbol{r} - \boldsymbol{r}_0)^\top\boldsymbol{A}^{-1}(\boldsymbol{r} - \boldsymbol{r}_0))$, where $c' > 0$ is a constant and q is a function on $\mathbb{R}$. Also, the use of the D-norm, $|\!|\!|\boldsymbol{r}|\!|\!|_p$, which is suggested by Bertsimas and Sim (2004) and Bertsimas, Pachamanova and Sim (2004), in the robust portfolio is equivalent to that of its dual norm, $\max\{\|\boldsymbol{\pi}\|_\infty, \|\boldsymbol{\pi}\|_1/p\}$, in the norm-constrained portfolio.

Table 1: Correspondence between Norms in Norm-Constraints for Portfolio Selection and Uncertainty Sets for Robust Portfolios

Norm in Norm Constraint	Norm in Uncertainty Set
$\|\boldsymbol{\pi}\|_1$ (DeMiguel et al. 2009)	$\|\boldsymbol{r}\|_\infty$ (Soyster 1973)
$\|\boldsymbol{\pi}\|_A$ (DeMiguel et al. 2009)	$\|\boldsymbol{r}\|_{A^{-1}}$ (Ben-Tal and Nemirovski 2000)
$\|\boldsymbol{\pi}\|_2$ (DeMiguel et al. 2009)	$\|\boldsymbol{r}\|_2$
$\|\boldsymbol{\pi}\|_\infty$	$\|\boldsymbol{r}\|_1$
$\max\{\|\boldsymbol{\pi}\|_\infty, \|\boldsymbol{\pi}\|_1/p\}$	$\|\!\|\boldsymbol{r}\|\!\|_p$ (Bertsimas, Pachamanova and Sim 2004)

2.3 Relation to Robust Return Maximization

The norm constraint can also be derived from the robust return maximization

$$\max_{\boldsymbol{\pi}} \left\{ \min_{\boldsymbol{\mathcal{R}}_t, t=1,\dots,T} \left\{ \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\mathcal{R}}_t^\top \boldsymbol{\pi} : \boldsymbol{\mathcal{R}}_t \in \mathcal{U}(\boldsymbol{R}_t, \delta), t = 1, \dots, T \right\} : \boldsymbol{e}_n^\top \boldsymbol{\pi} = 1, \boldsymbol{A}\boldsymbol{\pi} \le \boldsymbol{b} \right\}$$

for an uncertainty set of the form

$$\mathcal{U}(\boldsymbol{R}_t, \delta) = \{\boldsymbol{r} \in \mathbb{R}^n : \boldsymbol{r} = \boldsymbol{R}_t - \Delta\boldsymbol{r} \ \text{ for some } \Delta\boldsymbol{r} \text{ satisfying } \|\Delta\boldsymbol{r}\| \le \delta\} \tag{2}$$

with a nominal return vector $\boldsymbol{R}$ and $\delta > 0$. The uncertainty set $\mathcal{U}(\boldsymbol{R}_t, \delta)$ indicates that the observed return vector $\boldsymbol{R}_t$ suffers from a possible perturbation of size δ. Note that, for each $\boldsymbol{\pi}$, the minimization in the objective can be simplified as follows:

$$\min_{\{\|\Delta\boldsymbol{r}_t\| \le \delta\}_{t=1,\dots,T}} \left\{ \frac{1}{T}\sum_{t=1}^{T} (\boldsymbol{R}_t - \Delta\boldsymbol{r}_t)^\top \boldsymbol{\pi} \right\} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{R}_t^\top \boldsymbol{\pi} - \max\{\Delta\boldsymbol{r}^\top \boldsymbol{\pi} : \|\Delta\boldsymbol{r}\| \le \delta\}.$$

The robust counterpart can then be rewritten into the following regularized return maximization problem:

$$\max_{\boldsymbol{\pi}} \left\{ \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{R}_t^\top \boldsymbol{\pi} - \delta\|\boldsymbol{\pi}\|^* : \boldsymbol{e}_n^\top \boldsymbol{\pi} = 1, \boldsymbol{A}\boldsymbol{\pi} \le \boldsymbol{b} \right\}.$$

Considering that the dual norm term in the objective can be equivalently transformed into a constraint with adequate parameter settings, this formulation shows that the norm-based return uncertainty leads to a norm constraint. Thus, the norm constraint or regularization term naturally appears in portfolio optimization when the uncertainty associated with the observed returns is taken into account.

3 Norm-Constrained VaR and CVaR Minimization

In contrast to the previous section, here, we shall focus on empirical VaR and CVaR minimizations in combination with the norm constraints and present a robust interpretation and a nonparametric statistical validation.

3.1 Empirical VaR and CVaR

Let $f(\boldsymbol{\pi}, \boldsymbol{\mathcal{R}})$ denote a random portfolio loss associated with the random vector $\boldsymbol{\mathcal{R}}$. In the following, we assume that $\boldsymbol{\mathcal{R}}$ is independent of $\boldsymbol{\pi}$, as in Rockafellar and Uryasev (2002). In general, we can employ any cost function to be minimized as f. For example, a minus return can be employed as a loss, i.e.,

$$f(\boldsymbol{\pi}, \boldsymbol{\mathcal{R}}) = -\mathcal{R}(\boldsymbol{\pi}) = -\boldsymbol{\mathcal{R}}^\top \boldsymbol{\pi}. \tag{3}$$

For $\beta \in (0, 1)$, the β-VaR, $\alpha_\beta(\boldsymbol{\pi})$, associated with a loss $f(\boldsymbol{\pi}, \boldsymbol{\mathcal{R}})$ is the β-quantile of the distribution of f, i.e.,

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

where $\Phi(\cdot|\boldsymbol{\pi})$ is the distribution function of f. The parameter β is a user-defined parameter for representing a confidence level, and it usually takes a fixed value close to 1, say, 0.95 or 0.99, for capturing a large loss with a small probability.

On the other hand, β-CVaR associated with $f(\boldsymbol{\pi}, \boldsymbol{\mathcal{R}})$ is defined by

$$\phi_\beta(\boldsymbol{\pi}) := \min_\alpha F_\beta(\boldsymbol{\pi}, \alpha),$$

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

$$F_\beta(\boldsymbol{\pi}, \alpha) := \alpha + \frac{1}{1-\beta}\mathbb{E}[f(\boldsymbol{\pi}, \boldsymbol{\mathcal{R}}) - \alpha]^+$$

where $\mathbb{E}[\cdot]$ denotes the operator for the mathematical expectation and $[x]^+ := \max\{x, 0\}$. According to Rockafellar and Uryasev (2002), β-CVaR, $\phi_\beta(\boldsymbol{\pi})$, can be approximately regarded as the expected value of f greater than β-VaR, α_β, and therefore, one gets $\alpha_\beta(\boldsymbol{\pi}) \leq \phi_\beta(\boldsymbol{\pi})$, as in Figure 1. In practice, similarly to VaR, β is usually fixed at a value close to one. $\phi_\beta(\boldsymbol{\pi})$ and $F_\beta(\boldsymbol{\pi}, \alpha)$ are convex functions when f is convex in $\boldsymbol{\pi}$, whereas $\alpha_\beta(\boldsymbol{\pi})$ can be nonconvex even when f is linear in $\boldsymbol{\pi}$. The β-CVaR minimizing portfolio is given by the solution to

$$\min\{\ \phi_\beta(\boldsymbol{\pi})\ :\ \boldsymbol{\pi} \in \Pi\ \} = \min\{\ F_\beta(\boldsymbol{\pi}, \alpha)\ :\ \boldsymbol{\pi} \in \Pi,\ \alpha \in \mathbb{R}\ \}, \tag{4}$$

which can be reformulated as a convex program when f is convex in $\boldsymbol{\pi}$ and Π is a convex set. In addition, for an optimal solution $(\boldsymbol{\pi}^*, \alpha^*)$ to (4), α^* gives an approximate value of β-VaR, $\alpha_\beta(\boldsymbol{\pi}^*)$, as a by-product. More precisely, α^* is equal to $\alpha_\beta(\boldsymbol{\pi}^*)$ if the optimal α^* is unique. Even if it is not so, α^* is located in a closed interval $[\alpha_\beta(\boldsymbol{\pi}^*), \alpha_\beta^+(\boldsymbol{\pi}^*)]$ where $\alpha_\beta^+(\boldsymbol{\pi}) := \inf\{\alpha : \Phi(\alpha|\boldsymbol{\pi}) > \beta\}$.

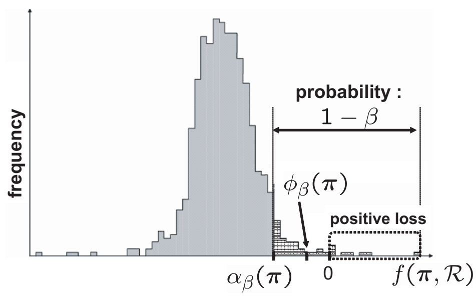

Figure 1: Illustration of β-VaR, α_β, and β-CVaR, ϕ_β associated with loss f

The empirical VaR and CVaR are similarly defined by employing an empirical distribution in place of Φ. Let $\Phi^T(\cdot|\boldsymbol{\pi})$ denote the empirical distribution of the loss based on T observed return data $\boldsymbol{R}_1, ..., \boldsymbol{R}_T$ which are supposed to be independently drawn from the (unknown) distribution Φ, i.e., $\Phi^T(\alpha|\boldsymbol{\pi}) := |\{t \in \{1, ..., T\} : f(\boldsymbol{\pi}, \boldsymbol{R}_t) \le \alpha\}|/T$. β-VaR, $\alpha_\beta(\boldsymbol{\pi})$, is then replaced with the empirical version, $\alpha_\beta^T(\boldsymbol{\pi})$, i.e.,

$$\alpha_\beta^T(\boldsymbol{\pi}) := \min\{\alpha : \Phi^T(\alpha|\boldsymbol{\pi}) \ge \beta\}.$$

On the other hand, the empirical β-CVaR is defined by

$$\phi_\beta^T(\boldsymbol{\pi}) := \min_\alpha F_\beta^T(\boldsymbol{\pi}, \alpha),$$

where

$$F_\beta^T(\boldsymbol{\pi}, \alpha) := \alpha + \frac{1}{(1-\beta)T} \sum_{t=1}^{T} [f(\boldsymbol{\pi}, \boldsymbol{R}_t) - \alpha]^+.$$

The norm-constrained minimizations of the empirical VaR and CVaR are written as follows:

$$\left| \begin{array}{ll} \min & \alpha_\beta^T(\boldsymbol{\pi}) \\ \text{s.t.} & \boldsymbol{e}_n^\top \boldsymbol{\pi} = 1, \ \boldsymbol{A}\boldsymbol{\pi} \le \boldsymbol{b} \\ & \|\boldsymbol{\pi}\| \le C, \end{array} \right. \qquad \left| \begin{array}{ll} \min & \phi_\beta^T(\boldsymbol{\pi}) \\ \text{s.t.} & \boldsymbol{e}_n^\top \boldsymbol{\pi} = 1, \ \boldsymbol{A}\boldsymbol{\pi} \le \boldsymbol{b} \\ & \|\boldsymbol{\pi}\| \le C, \end{array} \right. \tag{5}$$

where the constraints except for $\boldsymbol{e}_n^\top \boldsymbol{\pi} = 1$ and the norm constraint are assumed to be represented by a system of linear inequalities of the form $\boldsymbol{A}\boldsymbol{\pi} \le \boldsymbol{b}$, for simplicity, where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $\boldsymbol{b} \in \mathbb{R}^m$. When the loss function takes the form of (3), the empirical β-VaR and β-CVaR minimizations with sample returns $\boldsymbol{R}_1, ..., \boldsymbol{R}_T$ can be formulated as follows:

$$\left| \begin{array}{ll} \min & \alpha \\ \text{s.t.} & \boldsymbol{e}_T^\top \boldsymbol{z} \le \lfloor (1-\beta)T \rfloor \\ & -\boldsymbol{R}_t^\top \boldsymbol{\pi} - M z_t \le \alpha, \quad z_t \in \{0, 1\}, \ t = 1, ..., T \\ & \boldsymbol{e}_n^\top \boldsymbol{\pi} = 1, \ \boldsymbol{A}\boldsymbol{\pi} \le \boldsymbol{b}, \quad \|\boldsymbol{\pi}\| \le C, \end{array} \right. \tag{6}$$

where M is a sufficiently large number;

$$\left| \begin{array}{ll} \min & \alpha + \frac{1}{(1-\beta)T} \boldsymbol{e}_T^\top \boldsymbol{y} \\ \text{s.t.} & y_t \ge -\boldsymbol{R}_t^\top \boldsymbol{\pi} - \alpha, \ t = 1, ..., T; \quad \boldsymbol{y} \ge \boldsymbol{0} \\ & \boldsymbol{e}_n^\top \boldsymbol{\pi} = 1, \ \boldsymbol{A}\boldsymbol{\pi} \le \boldsymbol{b}, \quad \|\boldsymbol{\pi}\| \le C. \end{array} \right. \tag{7}$$

If the Euclidean norm $\|\boldsymbol{\pi}\|_2$ is used as $\|\boldsymbol{\pi}\|$, the VaR minimization (6) becomes a quadratically constrained 0-1 mixed integer program, which can be solved via a state-of-the-art solver such as ILOG CPLEX12 as long as the size of the problem is not huge and has so good a structure that the sophisticated branch-and-cut algorithm works. However, it is still hard to solve (6) within a practical time period even if T or n is around one hundred. In such case, it is reasonable to use linearly representable norms, such as $\|\boldsymbol{\pi}\|_1, \|\boldsymbol{\pi}\|_\infty$ and D-norm, in place of $\|\boldsymbol{\pi}\|_2$ when the VaR minimization (6) is applied.

On the other hand, the CVaR minimization (7) is a quadratically constrained linear program, and it can be efficiently solved via an interior point algorithm even when the problem is large. Therefore, the Euclidean norm or A-norm can be used when CVaR minimization is applied.

3.2 Interpretation through robust VaR and CVaR minimizations

We have seen that the norm constraint can be considered to be a robust return inequality. Here, we show that the use of norm constraints in the combination with VaR or CVaR minimization can be validated. As in Section 2.3, let us suppose that the observed return suffers from a return uncertainty of the form in (2). The robust counterparts of the VaR and CVaR minimizations can then be respectively formulated as

$$\left|\begin{array}{ll} \min & \alpha \\ \text{s.t.} & \boldsymbol{e}_T^\top \boldsymbol{z} \leq \lfloor (1-\beta)T \rfloor,\ z_t \in \{0,1\},\ t=1,...,T \\ & -(\boldsymbol{R}_t - \Delta \boldsymbol{r}_t)^\top \boldsymbol{\pi} - M z_t \leq \alpha \ \text{ for all } \|\Delta \boldsymbol{r}_t\| \leq \delta,\ t=1,...,T \\ & \boldsymbol{e}_n^\top \boldsymbol{\pi} = 1,\ \boldsymbol{A}\boldsymbol{\pi} \leq \boldsymbol{b} \end{array}\right. \tag{8}$$

and

$$\left|\begin{array}{ll} \min & \alpha + \frac{1}{(1-\beta)T} \boldsymbol{e}_T^\top \boldsymbol{y} \\ \text{s.t.} & y_t \geq -(\boldsymbol{R}_t - \Delta \boldsymbol{r}_t)^\top \boldsymbol{\pi} - \alpha \ \text{ for all } \|\Delta \boldsymbol{r}_t\| \leq \delta,\ t=1,...,T \\ & \boldsymbol{y} \geq \boldsymbol{0},\ \boldsymbol{e}_n^\top \boldsymbol{\pi} = 1,\ \boldsymbol{A}\boldsymbol{\pi} \leq \boldsymbol{b}. \end{array}\right. \tag{9}$$

Proposition 2 *The robust counterparts (8) and (9) of the empirical VaR and CVaR minimization can be reformulated as*

$$\left|\begin{array}{ll} \min & \alpha + \delta \|\boldsymbol{\pi}\|^* \\ s.t. & \boldsymbol{e}_T^\top \boldsymbol{z} \leq \lfloor (1-\beta)T \rfloor \\ & -\boldsymbol{R}_t^\top \boldsymbol{\pi} - M' z_t \leq \alpha,\ z_t \in \{0,1\}, \quad t=1,...,T \\ & \boldsymbol{e}_n^\top \boldsymbol{\pi} = 1,\ \boldsymbol{A}\boldsymbol{\pi} \leq \boldsymbol{b} \end{array}\right. \tag{10}$$

and

$$\left|\begin{array}{ll} \min & \alpha + \frac{1}{(1-\beta)T} \boldsymbol{e}_T^\top \boldsymbol{y} + \delta \|\boldsymbol{\pi}\|^* \\ s.t. & y_t \geq -\boldsymbol{R}_t^\top \boldsymbol{\pi} - \alpha,\ t=1,...,T; \quad \boldsymbol{y} \geq \boldsymbol{0} \\ & \boldsymbol{e}_n^\top \boldsymbol{\pi} = 1,\ \boldsymbol{A}\boldsymbol{\pi} \leq \boldsymbol{b} \end{array}\right. \tag{11}$$

where $\|\boldsymbol{\pi}\|^*$ *denotes the dual norm of* $\|\boldsymbol{\pi}\|$*, and* M' *is a sufficiently large number.*

Proof. Since both of the formulations (10) and (11) can be proved in a similar manner, we shall give only the proof for the VaR minimization. In (8), each inequality of the second constraint can be rewritten as $-\boldsymbol{R}_t^\top \boldsymbol{\pi} + \delta \|\boldsymbol{\pi}\|^* - M z_t \leq \alpha$. We arrive at the regularized version (10) by replacing the term $\alpha - \|\boldsymbol{\pi}\|^*$ with α' and replacing the big constant M by another M' so that the logical condition can be represented by a 0-1 variable. □

This proposition shows that the robust VaR and CVaR minimizations in which the observed returns suffer from some error turn out to be the regularized empirical VaR and CVaR minimizations (10) and (11), respectively, or equivalently, the norm-constrained versions (6) and (7) if the parameters δ and C are set in an adequate manner. Note that this property is based on the positive homogeneity of the two risk measures.

Also, we should note that the formulation (11) have the same structure as the one-class ν-SVM except for the additional constraints $\boldsymbol{e}_n^\top \boldsymbol{\pi} = 1,\ \boldsymbol{A}\boldsymbol{\pi} \leq \boldsymbol{b}$. Especially, the dual norm term in

(11) corresponds to the regularization term of the ν-SVM. Therefore, this proposition indicates the relation between a statistical learning technique and robust portfolio approach. Since it is empirically known that the regularization term improves the out-of-sample performance of models estimated by simply minimizing the empirical loss function, we expect that the robust portfolio obtained by (11) improves the norm-unconstrained CVaR model in a probabilistic sense as will be shown in the next subsection.

3.3 Generalization Error Bounds with Empirical VaR and CVaR

In this subsection, we provide another validation for the norm-constrained VaR and CVaR minimizations by employing a nonparametric statistical theory known as the generalization error bound.

By modifying the generalization error bound for ν-SVM (Schölkopf et al. 2000) and using the expression of the empirical β-VaR, $\alpha_\beta^T(\boldsymbol{\pi})$, or β-CVaR, $\phi_\beta^T(\boldsymbol{\pi})$, we can obtain upper and lower bounds on the probability that the loss f is greater than a threshold θ under the assumption of a nonparametric distribution.

Theorem 1 *Let $\mathcal{L} := \{\boldsymbol{R} \mapsto \boldsymbol{R}^\top \boldsymbol{\pi} : \|\boldsymbol{\pi}\|_2 \leq C, \|\boldsymbol{R}\|_2 \leq B_R\}$ with constants C and B_R. Let θ be a threshold for the portfolio loss f. T sample return data, $\boldsymbol{R}_1, ..., \boldsymbol{R}_T$, are independently drawn from an unknown probability distribution whose support is contained in $\{\boldsymbol{R} : \|\boldsymbol{R}\|_2 \leq B_R\}$. Then, for any $f(\boldsymbol{\pi}, \cdot) \in \mathcal{L}$ and $\boldsymbol{\pi}$ satisfying $\alpha_\beta^T(\boldsymbol{\pi}) < \theta$, the probability of the loss $f(\boldsymbol{\pi}, \boldsymbol{\mathcal{R}})$ being greater than θ, $\mathbb{P}\{f(\boldsymbol{\pi}, \boldsymbol{\mathcal{R}}) > \theta\}$, is bounded above as*

$$\mathbb{P}\{f(\boldsymbol{\pi}, \boldsymbol{\mathcal{R}}) > \theta\} \leq (1-\beta) + \sqrt{\frac{2}{T}\left\{\frac{4c^2(C^2+1)(B_R^2+\theta^2)\log_2(2T)}{(\alpha_\beta^T(\boldsymbol{\pi})-\theta)^2} + \ln\frac{2}{\delta \mathrm{e}}\right\}} \tag{12}$$

with probability at least $1-\delta$, and $C > 0$ is a constant. On the other hand, for $\boldsymbol{\pi}$ satisfying $\alpha_\beta^T(\boldsymbol{\pi}) > \theta$, the probability is bounded below as

$$\mathbb{P}\{f(\boldsymbol{\pi}, \boldsymbol{\mathcal{R}}) > \theta\} \geq (1-\beta) - \sqrt{\frac{2}{T}\left\{\frac{4c^2(C^2+1)(B_R^2+\theta^2)\log_2(2T)}{(\alpha_\beta^T(\boldsymbol{\pi})-\theta)^2} + \ln\frac{2}{\delta \mathrm{e}}\right\}} \tag{13}$$

with probability at least $1-\delta$.

Corollary 1 *Suppose the same assumption as in Theorem 1. Then, for any $f(\boldsymbol{\pi}, \cdot) \in \mathcal{L}$ and $\boldsymbol{\pi}$ satisfying $\phi_\beta^T(\boldsymbol{\pi}) < \theta$, one has*

$$\mathbb{P}\{f(\boldsymbol{\pi}, \boldsymbol{\mathcal{R}}) > \theta\} \leq (1-\beta) + \sqrt{\frac{2}{T}\left\{\frac{4c^2(C^2+1)(B_R^2+\theta^2)\log_2(2T)}{(\phi_\beta^T(\boldsymbol{\pi})-\theta)^2} + \ln\frac{2}{\delta \mathrm{e}}\right\}} \tag{14}$$

with probability at least $1-\delta$.

See the Appendix for the proof of Theorem 1. Corollary 1 is easily obtained from Theorem 1 since $\alpha_\beta^T(\boldsymbol{\pi}) \leq \phi_\beta^T(\boldsymbol{\pi})$ holds for any $\boldsymbol{\pi}$ and, thus, we have $(\alpha_\beta^T(\boldsymbol{\pi}) - \theta)^2 \geq (\phi_\beta^T(\boldsymbol{\pi}) - \theta)^2$ as long as $\phi_\beta^T(\boldsymbol{\pi}) < \theta$.

These propositions reveal that the unknown loss probability $\mathbb{P}\{f(\boldsymbol{\pi}, \boldsymbol{\mathcal{R}}) > \theta\}$ can be bounded above or below by some quantity involving the empirical β-VaR, $\alpha_\beta^T(\boldsymbol{\pi})$, and β-CVaR, $\phi_\beta^T(\boldsymbol{\pi})$. In the above inequalities (12), (13) and (14), the 2-norm $\|\cdot\|_2$ can be replaced with any norm $\|\cdot\|$ in $\mathbb{R}^n$ by multiplying a constant, due to the equivalence of any two norms in a vector space of finite dimension.

Someone who is used to the assumption of the unbounded support distribution such as in the normal distribution may wonder if the bounded support assumption is too restrictive. However, the support of the asset return is bounded because the total amount of money or credit in the world market is bounded. Needless to say, the boundedness assumption does not exclude the fat-tail property of the return distribution. Instead, the above theorem takes into consideration the tail part (edge of the support) of the distribution in a nonparametric manner.

The main goal of the propositions is not to calculate the tight bound, but to examine what kind of parameters are included in the bound and how they contribute to the unknown loss probability, which will give us a clue about how to make the probability smaller.

First of all, we should note that the right-hand sides of (12), (13) and (14) decrease as $\alpha_\beta^T(\boldsymbol{\pi})$ and $\phi_\beta^T(\boldsymbol{\pi})$ decrease, which implies that minimizing the empirical VaR, $\alpha_\beta^T(\boldsymbol{\pi})$, and CVaR, $\phi_\beta^T(\boldsymbol{\pi})$, for fixed β reduces the bounds of the probability. Since these bounds hold only when the norm of the portfolio is bounded above by a constant C, the solutions to the optimization problems (5) should make the loss probability smaller. In addition, the upper and lower bounds decrease as C decreases. However, decreasing C restricts the feasibility of $\boldsymbol{\pi}$, and this can lead to an increase in $\alpha_\beta^T(\boldsymbol{\pi})$ and $\phi_\beta^T(\boldsymbol{\pi})$. Thus, there is a trade-off between the value of C and the empirical risk measures. In order to reconcile the effects of C and $\boldsymbol{\pi}$ in minimizing the bounds, the parameter C should be tuned when the norm-constrained problems (5) are solved.

Next, let us examine the other parameters. The bounds decrease as T decreases, which is consistent with the law of large numbers. Also, the right-hand sides increase as B_R increases. Although B_R is uncontrollable because of its nature, knowing the accurate value of B_R does not alter the effect of $\boldsymbol{\pi}$ in minimizing the bounds, which is unlike the case of C. Therefore, it is unnecessary to pay attention to the size of B_R. In this sense, we can apply (5) to any set of finite samples independently drawn from a possibly unbounded distribution.

As for β, we cannot figure out the shape of the right-hand sides as functions of β because their second terms include unknown parameters. In keeping with the spirit of the upper and lower bounds minimization, the parameter β should also be tuned as C should be.

From the above observations, we can expect that solving (5) in combination with tuning C and β leads to a lower loss probability. One possible criticism of this expectation is that the bounds are not tight, and hence, the minimization may not be effective in decreasing the loss probability. However, the numerical experiments given in Section 5 show that the norm-constrained portfolio model motivated by bound minimization achieves a better out-of-sample performance than the other models.

Remark 1 *The above argument is similar to the previous work by the authors (Gotoh and Takeda 2008), in which two fractional programming formulations are posed:*

$$\left|\begin{array}{ll} \min & \dfrac{\alpha_\beta^T(\boldsymbol{\pi})-\theta}{\|\boldsymbol{\pi}\|} \\ s.t. & \boldsymbol{e}_n^\top \boldsymbol{\pi}=1,\ \boldsymbol{A}\boldsymbol{\pi}\le \boldsymbol{b}, \end{array}\right. \qquad \left|\begin{array}{ll} \min & \dfrac{\phi_\beta^T(\boldsymbol{\pi})-\theta}{\|\boldsymbol{\pi}\|} \\ s.t. & \boldsymbol{e}_n^\top \boldsymbol{\pi}=1,\ \boldsymbol{A}\boldsymbol{\pi}\le \boldsymbol{b}. \end{array}\right. \tag{15}$$

In Gotoh and Takeda (2008), the authors describe a two-step framework to solving (15) because the fractional problems may be intractable when θ is small. The norm-constrained problems (5) are much easier to solve than they are in the fractional programming formulation, and, thus, it is easier to conduct a cross-validation for the parameter tuning. Of course, there is a possibility that ignoring the difficult case would limit the potential gain in performance. However, the ease of implementing the parameter tuning is a strong motivation for us to introduce the norm constraint into practice.

4 Application to Tracking Portfolio

In this section, we extend the norm-constrained VaR and CVaR minimizations (6) and (7) to tracking (or mimicking) a portfolio.

Let $\mathcal{I}$ be the random return of a target asset (such as a stock price index) to be mimicked, and let us suppose that n assets are available to replicate the target. Then, it is typical that a portfolio $\boldsymbol{\pi} \in \mathbb{R}^n$ is determined so that a proper deviation would be minimized. For example, Gilli and Këllezi (2002) applies the criterion of minimizing the mean absolute deviation $\mathbb{E}[|\mathcal{I} - \mathcal{R}(\boldsymbol{\pi})|]$ by minimizing the sample mean of the empirical absolute errors:

$$\min\{ \frac{1}{T} \sum_{t=1}^{T} |I_t - \boldsymbol{R}_t^\top \boldsymbol{\pi}| \; : \; \boldsymbol{e}_n^\top \boldsymbol{\pi} = 1, \; \boldsymbol{A}\boldsymbol{\pi} \leq \boldsymbol{b}\}, \tag{16}$$

where $\boldsymbol{R}_t := (R_{t,1}, ..., R_{t,n})^\top$, $t = 1, ..., T$, are observed historical return vectors of the n assets, and I_t, $t = 1, ..., T$, are observed index return. Note that (16) can be reformulated as a linear program.

Remark 2 *The framework of the tracking portfolio reminds us of the well-known linear regression analysis, where a linear model*

$$y = a_1 x_1 + \cdots + a_d x_d \tag{17}$$

is estimated from a given set of observed data $\{(y_1, x_{11}, ..., x_{1d}), ..., (y_m, x_{m1}, ..., x_{md})\}$ *by minimizing the sum of the in-sample squared error:* $\sum_{i=1}^{m}\{y_i - (a_1 x_{i1} + \cdots + a_d x_{id})\}^2$*. In fact, the data formats required in those two models are in the same style as shown in Figure 2.*

The mean absolute deviation minimizing regression has been also analyzed in, e.g., Arthanari and Dodge (1981). Moreover, the ridge regression and the lasso respectively minimize the regularization terms $\|\boldsymbol{\pi}\|_2$ *and* $\|\boldsymbol{\pi}\|_1$ *in addition to the sum of squared errors. As already mentioned, this idea has been applied to portfolios by DeMiguel et al. (2009) and Brodie et al. (2009).*

Interestingly, the use of the regularization term (or, equivalently, the norm constraint) in the regression has been shown to select the important variables (assets) and to enable the obtained model to avoid overfitting (see, e.g., Hastie, Tibshirani and Friedman 2001).

		asset			
		idx.	1	$\cdots$	n
date	1	I_1	R_{11}	$\cdots$	R_{1n}
	$\vdots$	$\vdots$	$\vdots$		$\vdots$
	T	I_T	R_{T1}	$\cdots$	R_{Tn}
portfolio			π_1	$\cdots$	π_n

		attribute			
		obj.	1	$\cdots$	d
case	1	y_1	x_{11}	$\cdots$	x_{1d}
	$\vdots$	$\vdots$	$\vdots$		$\vdots$
	m	y_m	x_{m1}	$\cdots$	x_{md}
coefficients			a_1	$\cdots$	a_d

Figure 2: Typical data formats for index tracking (left) and supervised statistical model (right)

In the following, we devise new norm-constrained tracking portfolio models by modifying the empirical β-VaR, $\alpha_\beta^T(\boldsymbol{\pi})$, and β-CVaR, $\phi_\beta^T(\boldsymbol{\pi})$. The tracking error is a random loss defined by

$$f(\boldsymbol{\pi}, \mathcal{I}, \boldsymbol{\mathcal{R}}) = |\mathcal{I} - \mathcal{R}(\boldsymbol{\pi})|. \tag{18}$$

The empirical β-VaR, $\alpha_\beta^T(\boldsymbol{\pi})$, is then defined as the β-quantile of the empirical distribution Φ^T of the tracking error where $\Phi^T(\alpha) := |\{t : |I_t - \boldsymbol{R}_t^\top \boldsymbol{\pi}| \leq \alpha\}|/T$ and the empirical β-CVaR, $\phi_\beta^T(\boldsymbol{\pi})$, is similarly defined:

$$\phi_\beta^T(\boldsymbol{\pi}) = \min_\alpha\{\alpha + \frac{1}{(1-\beta)T}\sum_{t=1}^{T}[|I_t - \boldsymbol{R}_t^\top \boldsymbol{\pi}| - \alpha]^+\}.$$

It is noteworthy that β-CVaR contains the absolute deviation as a special case, i.e., the absolute-deviation is equal to ϕ_0^T. The tracking portfolio model using the norm-constrained VaR deviation is formulated as follows:

$$\left|\begin{array}{ll}\min\limits_{\alpha,\boldsymbol{\pi},\boldsymbol{w},\boldsymbol{x},\boldsymbol{z}} & \alpha \\ \text{s.t.} & \boldsymbol{e}_T^\top \boldsymbol{z} \le \lfloor (1-\beta)T \rfloor \\ & \boldsymbol{w} + \boldsymbol{x} - M\boldsymbol{z} \le \alpha \boldsymbol{e}_T \\ & w_t - x_t = I_t - \boldsymbol{R}_t^\top \boldsymbol{\pi}, \quad t = 1, ..., T \\ & \boldsymbol{w} \ge \boldsymbol{0}, \ \boldsymbol{x} \ge \boldsymbol{0}, \ \boldsymbol{z} \in \{0,1\}^T \\ & \boldsymbol{e}_n^\top \boldsymbol{\pi} = 1, \ \boldsymbol{A}\boldsymbol{\pi} \le \boldsymbol{b}, \ \|\boldsymbol{\pi}\| \le C, \end{array}\right. \tag{19}$$

where M is a sufficiently large number. Similarly, the tracking portfolio using the norm-constrained CVaR deviation is represented by the following convex optimization:

$$\left|\begin{array}{ll}\min\limits_{\alpha,\boldsymbol{\pi},\boldsymbol{w},\boldsymbol{x},\boldsymbol{y}} & \alpha + \frac{1}{(1-\beta)T}\boldsymbol{e}_T^\top \boldsymbol{y} \\ \text{s.t.} & \boldsymbol{y} \ge \boldsymbol{w} + \boldsymbol{x} - \alpha \boldsymbol{e}_T \\ & \boldsymbol{y} \ge \boldsymbol{0}, \ \boldsymbol{w} \ge \boldsymbol{0}, \ \boldsymbol{x} \ge \boldsymbol{0} \\ & w_t - x_t = I_t - \boldsymbol{R}_t^\top \boldsymbol{\pi}, \quad t = 1, ..., T \\ & \boldsymbol{e}_n^\top \boldsymbol{\pi} = 1, \ \boldsymbol{A}\boldsymbol{\pi} \le \boldsymbol{b}, \ \|\boldsymbol{\pi}\| \le C. \end{array}\right. \tag{20}$$

As mentioned before, the use of the 2-norm $\|\boldsymbol{\pi}\|_2$ in (19) is not recommended, while that in (20) is worth applying because (20) is still a convex optimization problem. Taking into account the computational tractability and the coherence as a risk measure, we only examined the norm-constrained CVaR minimization (20) with $\|\boldsymbol{\pi}\|_2$ in the computational experiment described in the next section.

Remark 3 *As described in Gotoh and Takeda (2005) and Takeda (2009), the β-CVaR minimization is equivalent to ν-SVMs. From this viewpoint, the above model (20) can be considered to be an application of ν-support vector regression (ν-SVR) to index tracking. One difference between these two frameworks can be found in the additional constraints imposed on the associated optimizations. In fact, the portfolio problem always has a constraint of the form $\boldsymbol{e}_n^\top \boldsymbol{\pi} = 1$ by definition, while the regression problem has usually no constraint on the parameters $a_1, ..., a_d$ of (17).*

Remark 4 *The formulation (20) is seemingly just a norm-constrained version of the CVaR minimization model in which the absolute deviation is employed as f. As in the case of the norm constraint, the parameter β in (20) plays a different role from that in the standard CVaR model. While β in the standard model is fixed by the user before the optimization so that he/she can look at a predetermined level of risk, the β in (20) is tuned at the time of optimization. In the numerical experiment in the next section, we provide a way of tuning β as well as C based on historical data.*

Similarly to what was done in the previous section, nonparametric bounds of the tracking error probability can be obtained from the empirical β-VaR and β-CVaR of the tracking error (18) by modifying Theorem 1 and Corollary 1.

Theorem 2 *Let θ be the threshold for the portfolio loss. Suppose that the random return vector $(\mathcal{I}, \boldsymbol{\mathcal{R}})$ has a bounded support in the sense that $(\mathcal{I}, -\boldsymbol{\mathcal{R}})$ lie in a ball of radius B_R centered at the origin, and that T return data, $(I_1, \boldsymbol{R}_1), ..., (I_T, \boldsymbol{R}_T)$, are independently drawn from $(\mathcal{I}, \boldsymbol{\mathcal{R}})$. Then, for any feasible portfolio $\boldsymbol{\pi}$ satisfying $\alpha_\beta^T(\boldsymbol{\pi}) < \theta$ and $\|\boldsymbol{\pi}\|_2 \le C$, the probability of the tracking error being greater than θ, $\mathbb{P}\{|\mathcal{I} - \mathcal{R}(\boldsymbol{\pi})| > \theta\}$, is bounded above as*

$$\mathbb{P}\{|\mathcal{I} - \mathcal{R}(\boldsymbol{\pi})| > \theta\} \le (1-\beta) + 2\sqrt{\frac{2}{T}\left\{\frac{4c^2(C^2+1)(B_R+\theta)^2\log_2(2T)}{(\alpha_\beta^T(\boldsymbol{\pi})-\theta)^2} + \ln\frac{2}{\delta \mathrm{e}}\right\}} \tag{21}$$

with probability at least $1-\delta$, where $C > 0$ is a constant. On the other hand, for $\boldsymbol{\pi}$ satisfying $\alpha_\beta^T(\boldsymbol{\pi}) > \theta$ and $\|\boldsymbol{\pi}\|_2 \le C$, the probability is bounded below as

$$\mathbb{P}\{|\mathcal{I} - \mathcal{R}(\boldsymbol{\pi})| > \theta\} \ge (1-\beta) - 2\sqrt{\frac{2}{T}\left\{\frac{4c^2(C^2+1)(B_R+\theta)^2\log_2(2T)}{(\alpha_\beta^T(\boldsymbol{\pi})-\theta)^2} + \ln\frac{2}{\delta \mathrm{e}}\right\}} \tag{22}$$

with probability at least $1-\delta$.

Corollary 2 *Suppose the same assumption as in Theorem 2. Then, for $\boldsymbol{\pi}$ satisfying $\phi_\beta^T(\boldsymbol{\pi}) < \theta$ and $\|\boldsymbol{\pi}\|_2 \le C$, one has*

$$\mathbb{P}\{|\mathcal{I} - \mathcal{R}(\boldsymbol{\pi})| > \theta\} \le (1-\beta) + 2\sqrt{\frac{2}{T}\left\{\frac{4c^2(C^2+1)(B_R+\theta)^2\log_2(2T)}{(\phi_\beta^T(\boldsymbol{\pi})-\theta)^2} + \ln\frac{2}{\delta \mathrm{e}}\right\}}. \tag{23}$$

with probability at least $1-\delta$.

See the Appendix for the proofs.

5 Numerical Experiments

We conducted numerical experiments to examine how our tracking portfolio approach works.

In consideration of the computational efficiency, we only solved the norm-constrained CVaR model (20) with 2-norm. For comparison, the norm-constrained versions of absolute error minimization (16) and the norm-unconstrained versions of the two criteria were also solved.

5.1 Results with Randomly Generated Normal Distributed Data

An experiment was conducted on a randomly generated data set in order to examine how well the norm-constrained CVaR deviation model with parameter tuning decreased the loss probability, $\mathbb{P}\{|\mathcal{I} - \mathcal{R}(\boldsymbol{\pi})| > \theta\}$, in comparison with the standard models. We randomly generated 200-dimensional vectors $\boldsymbol{R}_t$, $t = 1, \ldots, 1000$, which followed a normal distribution $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where μ_j and σ_j were independently drawn from uniform distributions on $(0, 10)$, and the correlation between R_i and R_j for any $i \ne j$ was 0.5. Associated with the generated variables, we defined the target as $I_t = \boldsymbol{e}_{200}^\top \boldsymbol{R}_t / 200$, $t = 1, ..., 1000$. The first 500 returns I_t and $\boldsymbol{R}_t$, $t = 1, \ldots, 500$, were used for constructing tracking portfolios of sizes up to 200, while the latter 500 were used for evaluating the out-of-sample loss. Although a bounded distribution is supposed in Theorem 2 and Corollary 2, our method is still applicable to this situation because 1) B_R can be regarded so large that the generated data are inside the ball of radius B_R; 2) B_R does not appear in our computation.

As for the norm-constrained CVaR model (20), the parameters β and C should be tuned by using the first 500 data. In this experiment, they were determined so that the in-sample

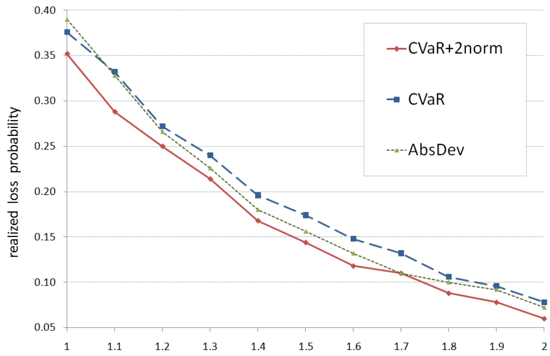

Figure 3: Probability of loss being greater than θ for randomly generated data set ($n = 200$, $T = 500$).

"CVaR+2norm" denotes the 2-norm-constrained CVaR model while "CVaR," and "AbsDev" denote the CVaR and the absolute deviation models without norm constraint, respectively. The best parameters chosen in "CVaR+2norm" are $\beta = 0.9$ and $C = 1/\sqrt{200} + 0.02 = 0.09$.

mean squared error (MSE) was minimized over 25 pairs of (β, C)s with $\beta = 0.5, 0.6, ..., 0.9$ and $C = 1/\sqrt{n} + 0.02, 1/\sqrt{n} + 0.04, ..., 1/\sqrt{n} + 0.10$ for $n = 200$. We also tuned β for the CVaR minimization model without the norm constraint similarly to the norm-constrained one.

Figure 3 shows the percentage of the realized losses f being greater than θ, i.e., $|\{t \in \{501, ..., 1000\} : f(\boldsymbol{\pi}, \boldsymbol{R}_t) > \theta\}|/500$. It is worth noting that this percentage can be considered as an estimate of the probability $\mathbb{P}\{f(\boldsymbol{\pi}, \boldsymbol{\mathcal{R}}) > \theta\}$, which is bounded above or below in Theorem 2 and Corollary 2. From this figure, we see that the norm-constrained CVaR minimization dominates the other two models in this sense. Although Theorems 1 and 2 and Corollaries 1 and 2 seem to provide loose bounds for $\mathbb{P}\{f(\boldsymbol{\pi}, \boldsymbol{\mathcal{R}}) > \theta\}$, we see that the principle of minimizing such bounds works effectively for the randomly generated data.

5.2 Results for Actual Market Data

To examine the practical effect of the norm constraint, we conducted an experiment on monthly and weekly return data of stocks listed in the Nikkei 225 index at the end of October 2009. The monthly data set consisted of returns of 182 companies over 270 consecutive months (22.5 years) starting from May 1987 to October 2009, whereas the weekly data set consisted of returns of 182 companies over 1178 consecutive weeks (about 22.5 years) starting from April 12, 1987 to November 1, 2009. The targets to be mimicked were the monthly and weekly returns of the Nikkei 225 index, respectively. The universe consisting of n assets was randomly chosen from 182 assets.

Using the data for the last $T = 120$ (10 years) consecutive periods from the monthly data set or those for the last $T = 150$ (almost 3 years) consecutive periods from the weekly data set, we obtained a portfolio $\bar{\boldsymbol{\pi}}_t$ by using the historical data $\boldsymbol{R}_t, ..., \boldsymbol{R}_{t+T-1}$ and evaluated the test error, $I_{t+T} - \boldsymbol{R}_{t+T}^\top \bar{\boldsymbol{\pi}}_t$, for a new data point $(I_{t+T}, \boldsymbol{R}_{t+T})$. By iteratively repeating this evaluation, $\overline{T} = 150$ and $\overline{T} = 1028$ rolling horizon evaluations were performed for the monthly and weekly data sets, respectively. The mean squared error (MSE) of the form $\frac{1}{\overline{T}} \sum_{t=1}^{\overline{T}} (I_{t+T} - \boldsymbol{R}_{t+T}^\top \bar{\boldsymbol{\pi}}_t)^2$ was employed as a measure of out-of-sample performance.

In the norm-constrained CVaR deviation model (20), the parameters β and C were systematically tuned as follows. Using the first $\frac{5}{6}T$-period data set, we found a pair of β and C that gave the best prediction in terms of MSE for the remaining $\frac{1}{6}T$-period data set. For each data set, the best β was found from among 0.1 to 0.9 in steps of 0.2 while the best C was found from

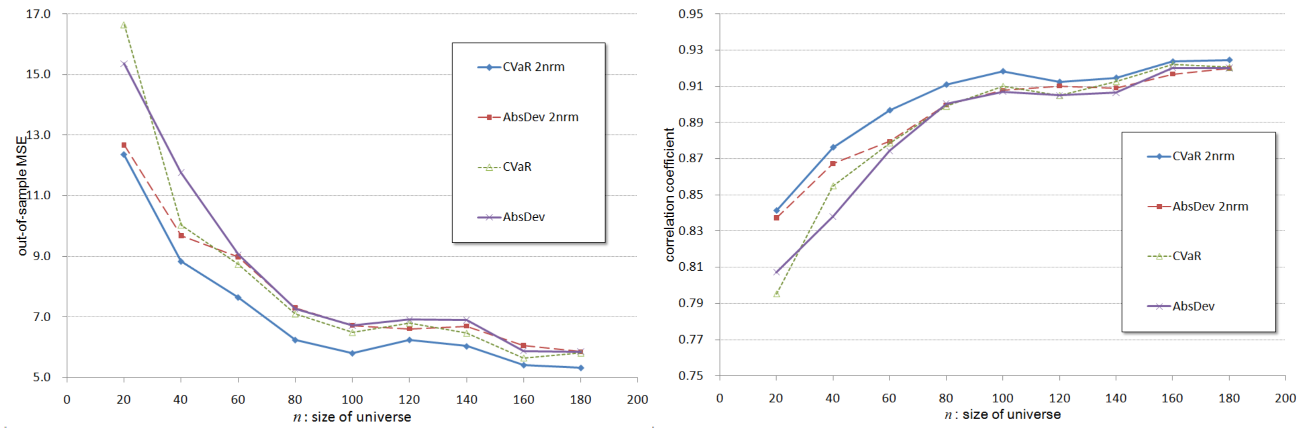

Figure 4: Out-of-sample MSE (left), out-of-sample correlation coefficient (right) of the monthly data set for when the short-sale constraint, $\boldsymbol{\pi} \geq \mathbf{0}$, is imposed

among five candidates $1/\sqrt{n} + k \cdot (1 - 1/\sqrt{n})/10\sqrt{n}$, $k = 1, ..., 5$, where the range of parameter C was determined from a preliminary experiment so that the selection of C would not be meaningless. C and β for the norm-constrained deviation models and the norm-unconstrained CVaR model were tuned in the same manner.

Figures 4 to 7 show the out-of-sample MSE and the out-of-sample correlation coefficient between the obtained portfolio return and the index return. In addition to Figures 4 and 6 showing a situation where short-sales are not allowed, Figures 5 and 7 show the situation where short-sales are allowed.

From these figures, it is easy to see that when short-sales are allowed and n is greater than T, the norm-unconstrained models give poor results in the out-of-sample evaluation where they achieve zero in-sample MSEs. Comparing the cases with and without the short-sale constraint, we see that the short-sale constraint prevents the optimization models from overfitting, which is consistent with the existing research and experience of practitioners.

What is interesting is that even when the short-sale constraint is imposed, two of the norm-constrained models outperform the norm-unconstrained ones in the out-of-sample evaluation. In order to see the significance of this result, Table 2 lists the p-values of a Wilcoxon signed rank sum test in which the advantage of the norm-constrained models over the norm-unconstrained ones is tested under the null hypothesis that the out-of-sample absolute deviation of the former is no less than that of the latter. For the monthly data, except the absolute deviation criterion for four ns, the p-values are less than 0.1; this indicates the norm-constrained version has a smaller absolute deviation than the norm-unconstrained at more than 90% significance; for the weekly data, except for $n = 20$, the advantage is supported at more than 99% significance. This result encourages us to use the norm constraint (or its variation) and to tune the involved parameters using historical data for improving the out-of-sample performance even when the short sale constraint is already imposed. We shall omit the table for the case in which short-sales are allowed since the advantage of the norm-constraint is much more evident.

Also, we see from Figures 4 to 7 that the CVaR-based norm-constrained criterion achieves smaller out-of-sample MSEs than the absolute deviation model especially for relatively large n. Considering that the absolute deviation is a special case of CVaR, this indicates that the tuning of β works.

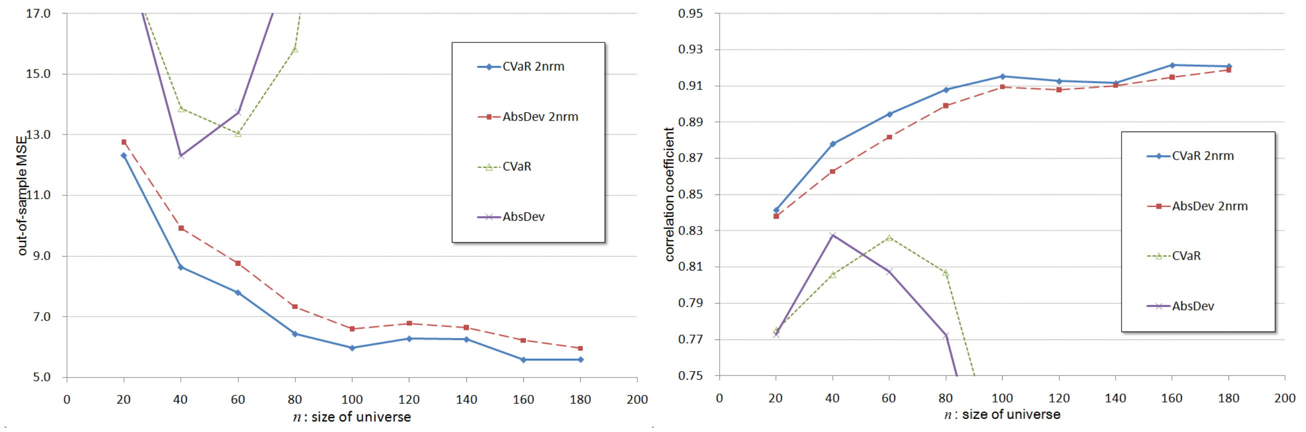

Figure 5: Out-of-sample MSE (left), out-of-sample correlation coefficient (right) of the monthly data set for when the short-sale constraint is not imposed

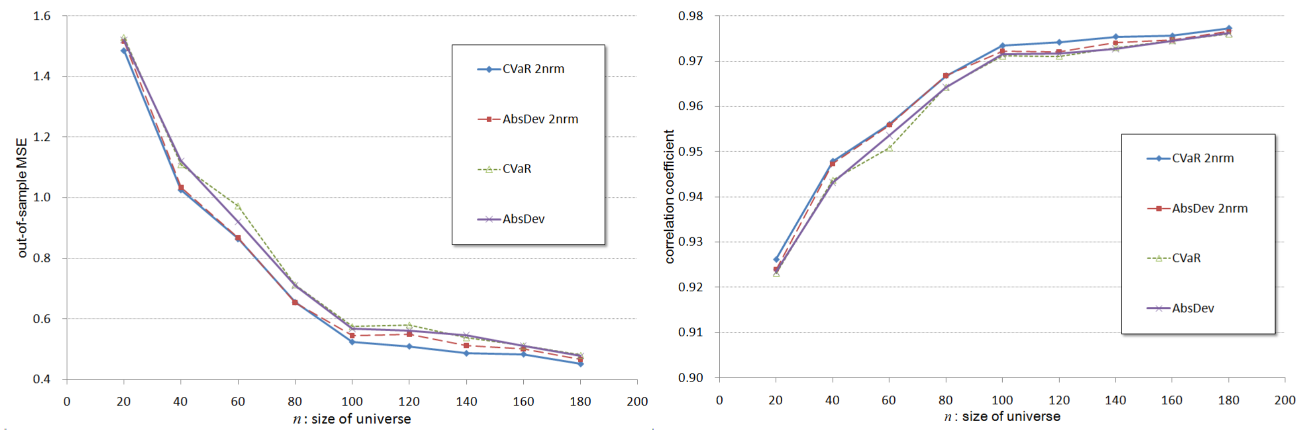

Figure 6: Out-of-sample MSE (left), out-of-sample correlation coefficient (right) of the weekly data set for when the short-sale constraint is imposed

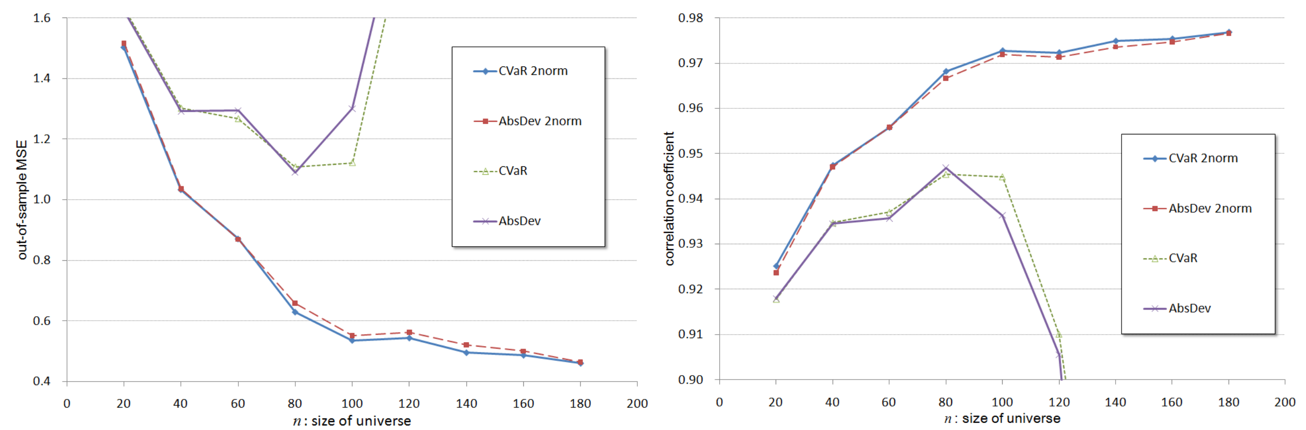

Figure 7: Out-of-sample MSE (left), out-of-sample correlation coefficient (right) of the weekly data set for when the short-sale constraint is not imposed

Table 2: p-values of the Wilcoxon signed rank sum test on the role of norm constraint in the two models

		(a) monthly data		(b) weekly data	
		CVaR	AbsDev	CVaR	AbsDev
	20	0.05158	0.15660	0.04301	0.03125
	40	0.00084	0.00030	0.00023	0.00115
	60	0.00012	0.00783	0.00001	0.00001
	80	0.00209	0.02042	0.00000	0.00025
n	100	0.02986	0.20290	0.00002	0.00022
	120	0.07997	0.03749	0.00000	0.00005
	140	0.01780	0.21530	0.00000	0.00000
	160	0.00142	0.13210	0.00000	0.00000
	180	0.00115	0.00833	0.00000	0.00002

Each column shows p-values of the null hypothesis that each criterion with the norm-constraint has a larger deviation than without norm-constraint. "CVaR" indicates results on the CVaR models; "AbsDev" does on the absolute deviation models.

6 Concluding Remarks

As pointed out in DeMiguel et al. (2009), the norm constraint has something to do with basic notions used in the traditional portfolio optimizations such as the short-sale constraint, the $1/n$ portfolio, and shrinkage techniques for the covariance matrix.

In this paper, we explored the role of the norm constraint in portfolio selection from various perspectives. For one thing, the norm constraint can be rewritten as a robust inequality by using the dual norm in describing the uncertainty set. In addition to such interpretations along the line of robust optimization, the norm constraint in the CVaR or VaR minimization has much to do with the regularization term of the ν-SVMs, and the theory of ν-SVMs can be employed in the context of the portfolio optimization. From these viewpoints, the norm constraint fits in with VaR and CVaR.

The numerical experiments showed that the norm-constrained models with some parameter tuning can alleviate the overfitting and improve on the norm-unconstrained approaches even when the short-sale constraint is already imposed. Especially, compared with the use of the standard deviations, accurately estimating CVaR requires a large number, T, of scenarios, and thus, incorporating a norm constraint can be a quick remedy to the fragility of the CVaR minimization. We expect that the norm constraint will improve the out-of-sample performance for a wide class of risk minimizations.

In addition, the parameter tuning based on historical data improves the out-of-sample performance. Although we chose in the experiments the parameters that achieved the best MSEs, the MSE criterion can be replaced with another one. For example, we can employ the empirical loss probability in place of MSE in order to seek smaller loss probability. The reason why we did not employ the criterion is that the threshold parameter θ should be determined in an arbitrary manner and that comparing the probabilities often results in situations where some tie-breaking rule is needed. We can expect that another criterion results in better out-of-sample performance. Besides, considering the relation between the norm constraint and the robust constraint, such a tuning technique can be used to tune the uncertainty set for nonparametric robust portfolio models.

We compared the CVaR models with the absolute deviation models (16) because the CVaR was defined with the loss function of the form (18). If the squared loss $(\mathcal{I} - \mathcal{R}(\boldsymbol{\pi}))^2$ is employed as the loss function, the modified CVaR model can be compared with the squared deviation model. We expect from the numerical results that the modified CVaR model also shows better results than the squared deviation model.

A Proof of Theorem

On the Generalization Error Bounds The generalization error bounds have been investigated for the two-class classification problem in which a given data set with binary labels is used for constructing a function h that indicates which class an input data point belongs to. In this section, we will provide an analysis for the generalization performance of a regression function by exploiting the bounds already obtained for the classification case (Bartlett 1998, Williamson, Smola and Schölkopf 2001).

Consider the classification problem of learning a classifier $h(\boldsymbol{R})$ whose decision function $\mathrm{sign}(h(\boldsymbol{R}))$ maps $\boldsymbol{R}$ to -1 or 1 on the basis of training samples $\{(\boldsymbol{R}_1, s_1), \ldots, (\boldsymbol{R}_T, s_T)\}$ where $\boldsymbol{R}_t \in \mathbb{R}^n$, $s_t \in \{\pm 1\}$ for $t \in 1, \ldots, T$. A typical example of such a problem is medical diagnosis. Suppose that $\boldsymbol{R}$ denotes a vector of characteristics of a tumor and I denotes whether the tumor $\boldsymbol{R}$ is "benign"($s_t = 1$) or "malignant"($s_t = -1$). What we want to do is to infer if a new patient's tumor is benign or not on the basis of the past data of patients, $\{(\boldsymbol{R}_1, s_1), \ldots, (\boldsymbol{R}_T, s_T)\}$ which will be called the training samples.

We assume that the training samples are independently drawn from an unknown probability distribution on $\mathbb{R}^n \times \{\pm 1\}$. The goal of the classification task is to obtain a classifier h that minimizes the generalization error (or the risk):

$$R[h] := \mathbb{P}\{\mathcal{I} h(\boldsymbol{\mathcal{R}}) < 0\} = \mathbb{P}\{\mathrm{sign}(h(\boldsymbol{\mathcal{R}})) \neq \mathcal{I}\}$$

which corresponds to the misclassification rate for unseen test samples.

We show an upper bound for $R[h]$ by referring to Bartlett (1998) and Williamson, Smola and Schölkopf (2001). Let V denote the ball of radius $\bar{B}$ in $\mathbb{R}^n$, *i.e.,* $V = \{\boldsymbol{R} \in \mathbb{R}^n \,:\, \|\boldsymbol{R}\|_2 \leq \bar{B}\}$, let $\mathcal{F}$ be a class of real-valued functions on V defined by

$$\mathcal{F} = \{\boldsymbol{R} \mapsto \boldsymbol{R}^\top \boldsymbol{\pi} \,:\, \|\boldsymbol{\pi}\|_2 \leq 1,\ \boldsymbol{R} \in V\}. \tag{24}$$

It is shown in Bartlett (1998) that there is a constant c such that with probability at least $1 - \delta$, a classifier $h(\boldsymbol{R}) = \boldsymbol{R}^\top \boldsymbol{\pi} \in \mathcal{F}$ has a test error $R[h]$ such that

$$\mathbb{P}[\mathcal{I} h(\boldsymbol{\mathcal{R}}) < 0] \leq \frac{1}{T}|\{\ i \ :\ I_i h(\boldsymbol{R}_i) < \gamma\ \}| + \sqrt{\frac{2}{T}\left(\frac{4c^2\bar{B}^2}{\gamma^2}\log_2(2T) - 1 + \log\left(\frac{2}{\delta}\right)\right)}, \tag{25}$$

for any $\gamma > 0$.

Theorem 1 gives a generalization performance of a regression function $f(\boldsymbol{\pi}, \boldsymbol{R})$ that belongs to $\mathcal{L} := \{\boldsymbol{R} \mapsto \boldsymbol{R}^\top \boldsymbol{\pi} : \|\boldsymbol{\pi}\|_2 \leq C, \|\boldsymbol{R}\|_2 \leq B_R\}$. In order to fit the loss of the regression to that of the classification, we introduce a threshold constant θ beyond which the loss is considered to be misclassified. We therefore aim to provide a bound on the probability that a randomly drawn test sample $\boldsymbol{R}$ will have the generalization error greater than θ, *i.e.,* $\mathbb{P}\{f(\boldsymbol{\pi}, \boldsymbol{\mathcal{R}}) > \theta\}$.

Proof of Theorem 1. The statement is proved by modifying the inequality (25) of the generalization error for classification. To make a regression problem to fit to the setting of (25), we consider a bound on the probability $\mathbb{P}\{f(\boldsymbol{\pi}, \boldsymbol{\mathcal{R}}) > \theta\}$ for test samples $\boldsymbol{R}$ generated from the same distribution as the training samples. We rewrite $\theta - f(\boldsymbol{\pi}, \boldsymbol{R})$ as $\tilde{\boldsymbol{R}}^\top \tilde{\boldsymbol{\pi}}$, where

$$\tilde{\boldsymbol{\pi}} = \begin{pmatrix} 1 \\ \boldsymbol{\pi} \end{pmatrix}, \quad \tilde{\boldsymbol{R}} = \begin{pmatrix} \theta \\ -\boldsymbol{R} \end{pmatrix}.$$

Note that the function $\tilde{\boldsymbol{R}}^\top \tilde{\boldsymbol{\pi}}/\|\tilde{\boldsymbol{\pi}}\|_2$ is a member of $\mathcal{F}$ with $\bar{B} = \sqrt{B_R^2 + \theta^2}$.

When $\alpha_\beta^T(\boldsymbol{\pi}) < \theta$, we regard $(\theta - \alpha_\beta^T(\boldsymbol{\pi}))/\|\tilde{\boldsymbol{\pi}}\|_2$ as γ in (25), *i.e.,* the threshold for training error. Then we get (12), that is,

$$\mathbb{P}\{\theta - f(\boldsymbol{\pi}, \boldsymbol{\mathcal{R}}) < 0\} \le (1-\beta) + \sqrt{\frac{2}{T}\left(\frac{4c^2(C^2+1)(B_R^2+\theta^2)}{(\alpha_\beta^T(\boldsymbol{\pi}) - \theta)^2}\log_2(2T) - 1 + \log\left(\frac{2}{\delta}\right)\right)},$$

since $\|\tilde{\boldsymbol{\pi}}\|_2 \le \sqrt{C^2+1}$ and

$$\frac{1}{T}|\{\ i\ :\ \theta - f(\boldsymbol{\pi}, \boldsymbol{R}_i) < \theta - \alpha_\beta^T(\boldsymbol{\pi})\}| = 1 - \beta^+(\boldsymbol{\pi}) \le 1 - \beta.$$

Here $\beta^+(\boldsymbol{\pi})$ is an upper bound of β provided in Rockafellar and Uryasev (2002) as

$$\beta^+(\boldsymbol{\pi}) = \Phi^T(\alpha_\beta^T(\boldsymbol{\pi})\,|\,\boldsymbol{\pi}) = \frac{1}{T}|\{t : f(\boldsymbol{\pi}, \boldsymbol{R}_t) \le \alpha_\beta^T(\boldsymbol{\pi})\}|,$$

where $\Phi^T(\cdot|\boldsymbol{\pi})$ is the empirical distribution of the loss $f(\boldsymbol{\pi}, \boldsymbol{\mathcal{R}})$.

For the case that $\alpha_\beta^T(\boldsymbol{\pi}) > \theta$, $(\theta - \alpha_\beta^T(\boldsymbol{\pi}))/\|\tilde{\boldsymbol{\pi}}\|_2$ cannot be taken for γ because γ should be positive. To resolve this issue, we prepare a classifier $-\theta + f(\boldsymbol{\pi}, \boldsymbol{R})$ whose decisions are opposed to those of $\theta - f(\boldsymbol{\pi}, \boldsymbol{R})$. Applying (25) to the function $(-\theta + f(\boldsymbol{\pi}, \boldsymbol{R}))/\|\tilde{\boldsymbol{\pi}}\|_2 = -\tilde{\boldsymbol{R}}^\top \tilde{\boldsymbol{\pi}}/\|\tilde{\boldsymbol{\pi}}\|_2$, we get

$$\begin{aligned}
&\mathbb{P}\{-\theta + f(\boldsymbol{\pi}, \boldsymbol{\mathcal{R}}) < 0\} = 1 - \mathbb{P}\{\theta - f(\boldsymbol{\pi}, \boldsymbol{\mathcal{R}}) < 0\} \\
&\le \frac{1}{T}|\{\ i\ : -\theta + f(\boldsymbol{\pi}, \boldsymbol{R}_i) < -\theta + \alpha_\beta^T(\boldsymbol{\pi})\ \}| \\
&\qquad + \sqrt{\frac{2}{T}\left(\frac{4c^2(C^2+1)(B_R^2+\theta^2)}{(\alpha_\beta^T(\boldsymbol{\pi}) - \theta)^2}\log_2(2T) - 1 + \log\left(\frac{2}{\delta}\right)\right)},
\end{aligned}$$

where $(\alpha_\beta^T(\boldsymbol{\pi}) - \theta)/\|\tilde{\boldsymbol{\pi}}\|_2 > 0$ corresponds to γ in (25). Then we get (13) by using a lower bound of β proved in Rockafellar and Uryasev (2002):

$$\beta^-(\boldsymbol{\pi}) := \Phi(\alpha_\beta^T(\boldsymbol{\pi})^-\,|\,\boldsymbol{\pi}) = \frac{1}{T}|\{t : f(\boldsymbol{\pi}, \boldsymbol{R}_t) < \alpha_\beta^T(\boldsymbol{\pi})\}|$$

to

$$1 - \frac{1}{T}|\{\ i\ :\ f(\boldsymbol{\pi}, \boldsymbol{R}_i) < \alpha_\beta^T(\boldsymbol{\pi})\ \}| = 1 - \beta^-(\boldsymbol{\pi}) \ge 1 - \beta.$$

□

Proofs of Theorem 2 and Corollary 2. We consider the distribution of $f(\boldsymbol{\pi}, I, \boldsymbol{\mathcal{R}}) = |\mathcal{I} - \boldsymbol{\mathcal{R}}^\top \boldsymbol{\pi}|$ and find a bound on the probability of the loss $f(\boldsymbol{\pi}, \mathcal{I}, \boldsymbol{\mathcal{R}})$ being greater than θ, $\mathbb{P}\{f(\boldsymbol{\pi}, \mathcal{I}, \boldsymbol{\mathcal{R}}) > \theta\}$. Here note that

$$\mathbb{P}\{|\mathcal{I} - \boldsymbol{\mathcal{R}}^\top \boldsymbol{\pi}| > \theta\} = \mathbb{P}\{\theta - \mathcal{I} + \boldsymbol{\mathcal{R}}^\top \boldsymbol{\pi} < 0\} + \mathbb{P}\{\theta + \mathcal{I} - \boldsymbol{\mathcal{R}}^\top \boldsymbol{\pi} < 0\}.$$

The function $\theta - I + \boldsymbol{R}^\top \boldsymbol{\pi}$ and $\theta + I - \boldsymbol{R}^\top \boldsymbol{\pi}$ are described as $\tilde{\boldsymbol{R}}^{(1)\top} \tilde{\boldsymbol{\pi}}$ and $\tilde{\boldsymbol{R}}^{(2)\top} \tilde{\boldsymbol{\pi}}$, respectively, by using

$$\tilde{\boldsymbol{\pi}} = \begin{pmatrix} 1 \\ \boldsymbol{\pi} \end{pmatrix}, \quad \tilde{\boldsymbol{R}}^{(1)} = \begin{pmatrix} \theta - I \\ \boldsymbol{R} \end{pmatrix}, \quad \tilde{\boldsymbol{R}}^{(2)} = \begin{pmatrix} \theta + I \\ -\boldsymbol{R} \end{pmatrix}.$$

Since $\tilde{\boldsymbol{R}}^{(1)\top}\tilde{\boldsymbol{\pi}}/\|\tilde{\boldsymbol{\pi}}\|_2$ and $\tilde{\boldsymbol{R}}^{(2)\top}\tilde{\boldsymbol{\pi}}/\|\tilde{\boldsymbol{\pi}}\|_2$ are functions of $\mathcal{F}$, the generalization analysis of (25) is applicable to $\mathbb{P}\{\theta - \mathcal{I} + \boldsymbol{\mathcal{R}}^\top\boldsymbol{\pi} < 0\}$ and $\mathbb{P}\{\theta + \mathcal{I} - \boldsymbol{\mathcal{R}}^\top\boldsymbol{\pi} < 0\}$.

Let $\alpha_\beta^T(\boldsymbol{\pi})$ be the β-VaR for the distribution $|I_t - \boldsymbol{R}_t^\top\boldsymbol{\pi}|$, $t = 1, \ldots, T$. When $\alpha_\beta^T(\boldsymbol{\pi}) < \theta$, we use the threshold $(\theta - \alpha_\beta^T(\boldsymbol{\pi}))/\|\tilde{\boldsymbol{\pi}}\|_2$ for evaluating the training errors of $\tilde{\boldsymbol{R}}^{(i)\top}\tilde{\boldsymbol{\pi}}/\|\tilde{\boldsymbol{\pi}}\|_2$, $i = 1, 2$. By noticing that $\|\tilde{\boldsymbol{\pi}}\|_2 \le \sqrt{C^2+1}$ and $\|\tilde{\boldsymbol{R}}^{(i)}\|_2 \le B_R + \theta$, (21) follows. Moreover, we can prove (22) and (23) similarly to the proofs of Theorem 1 and Corollary 1. □

Acknowledgment

The first author was supported by a MEXT Grant-in-Aid for Young Scientists (B) 20710120. The second author was supported by a MEXT Grant-in-Aid for Young Scientists (B) 19710124.

References

Arthanari TS, Dodge Y (1981) Mathematical Programming in Statistics. John Wiley & Sons, Inc., NJ.

Artzner P, Delbaen F, Eber JM, Heath D (1999) Coherent Measures of Risk. Mathematical Finance 9: 203–228.

Bartlett PL (1998) The Sample Complexity of Pattern Classification with Neural Networks: The Size of the Weights is More Important than the Size of the Network. IEEE Transactions on Information Theory 44: 525–536.

Ben-Tal A, Nemirovski A (2000) Robust solutions of linear programming problems contaminated with uncertain data. Math. Programming 88: 411–424.

Bertsimas D, Pachamanova D, Sim M (2004) Robust linear optimization under general norms. Operations Research Letters 32: 510–516.

Bertsimas D, Sim M (2004) The Price of Robustness. Operations Research 52: 35–53.

Brodie J, Daubechiesa I, De Mol C, Giannone D, Lorisc I (2009) Sparse and Stable Markowitz Portfolios, PNAS 106: 12267–12272.

Cont R, Deguest R, Scandolo G (2007) Robustness and sensitivity analysis of risk measurement procedures. Financial Engineering Report No.2007-06 Center for Financial Engineering, Columbia University.

DeMiguel V, Garlappi L, Nogales FJ, Uppal R (2009) A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms. Management Science 55: 798–812.

DeMiguel V, Garlappi L, Uppal R (2009) Optimal versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy? Review of Financial Studies 22: 1915–1953.

Gilli M, Këllezi E (2002) The threshold accepting heuristic for index tracking. In: Pardalos P, Tsitsiringos VK (ed) Financial Engineering, E-Commerce, and Supply Chain. Kluwer, Dordrecht, pp.1–18.

Goldfarb D, Iyengar G (2003) Robust portfolio selection problems. Mathematics of Operations Research 28: 1–38.

Gotoh J, Takeda A (2008) Minimizing Loss Probability Bounds for Portfolio Selection, Department of Industrial and Systems Engineering Discussion Paper Series ISE08-04, Chuo University. `http://www.indsys.chuo-u.ac.jp/~jgoto/`

Gotoh J, Takeda A (2005) A Linear Classification Model Based on Conditional Geometric Score. Pacific Journal on Optimization 1: 277–296.

Hastie T, Tibshirani R, Friedman J (2001) The Elements of Statistical Learning –Data Mining, Inference, and Prediction–. Springer, NY.

Jagannathan R, Ma T (2003) Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps. The Journal of Finance 4: 1651–1683.

Konno H, Yamazaki H (1991) Mean-Absolute Deviation Portfolio Optimization Model and Its Applications to Tokyo Stock Market. Management Science 37: 519–531.

Ledoit O, Wolf M (2003) Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. Journal of Empirical Finance 10: 603–621.

Ogryczak W, Ruszczýnski A (2002) Dual stochastic dominance and related mean–risk models. SIAM Journal on Optimization 13: 60–78.

Prigent JL (2007) Portfolio Optimization and Performance Analysis. Chapman & Hall/CRC, Florida.

Rockafellar TR, Uryasev S (2002) Conditional Value-at-Risk for General Loss Distributions. Journal of Banking and Finance 26: 1443–1471.

Schölkopf B, Smola AJ, Williamson RC, Bartlett PL (2000) New support vector algorithms. Neural Computation 12: 1207–1245.

Soyster AL (1973) Convex programming with set-inclusive constraints and applications to inexact linear programming. Operations Research 21: 1154–1157.

Takeda A (2009) Generalization Performance of ν-Support Vector Classifier Based on Conditional Value-at-Risk Minimization. Neurocomputing 72: 2351–2358.

Williamson RC, Smola AJ, Schölkopf B (2001) Generalization Performance of Regularization Networks and Support Vector Machines via Entropy Numbers of Compact Operators. IEEE Transactions on Information Theory 47: 2516–2532.

Xu H, Caramanis C, Mannor S (2008) Robust Regression and Lasso. Proceedings of the Neural Information Processing Systems Conference.