Department of Industrial and Systems Engineering

Discussion Paper Series

ISE 09-03

On the Role of the Norm Constraint in Portfolio Selection

Jun-ya Gotoh[1], **Akiko Takeda**[2]

February 2009

Chuo University
Bunkyo-ku, Tokyo 112-8551
Japan

[1]Department of Industrial and Systems Engineering, Chuo University
[2]Department of Administration Engineering, Keio University

On the Role of the Norm Constraint in Portfolio Selection

Jun-ya Gotoh 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

February 23, 2009

Abstract

Recently, several optimization approaches for portfolio selection have been proposed in order to alleviate the estimation error in the optimal portfolio. Among such are the norm-constrained variance minimization and the robust portfolio models. In this paper, we examine the role of the norm constraint in the 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. Secondly, the norm constraint is combined with the value-at-risk (VaR) and conditional value-at-risk (CVaR) minimizations. For the combinations, a nonparametric theoretical validation is posed based on the generalization error bound for the ν-support vector machine. Thirdly, the proposed approach is applied to the tracking portfolio problem and computational experiments are conducted. Through the experiments, we see that the norm-constrained minimization of the CVaR-based deviation with a parameter tuning strategy outperforms the traditional models in terms of the out-of-sample tracking error.

Keywords: portfolio optimization, norm constraint, robust portfolio, nonparametric distribution, generalization error, regularization, VaR (value-at-risk), CVaR (conditional value-at-risk), tracking portfolio

1 Introduction

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

Ideally, an optimal portfolio $\boldsymbol{\pi}^*$ is to be achieved as 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 based on the observed historical returns, in place of the ideal function.

Obviously, this framework can be validated by the so-called *law of large numbers*. That is, if the number of observations goes to infinity, the solution $\bar{\boldsymbol{\pi}}$ approaches to $\boldsymbol{\pi}^*$. This validation is, however, 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 where the number of assets is greater than that of historical observations. From a statistical point of view, this may cause the so-called *overfitting*, resulting in a large estimation error of the "optimal" portfolio $\bar{\boldsymbol{\pi}}$.

In fact, many researches have pointed out that the mean-variance model using the sample mean vector and the sample covariance matrix results in a poor out-of-sample performance because of the estimation error in the sample mean and (co)variance (see DeMiguel et al. (2007) and references can be found therein).

Recently, in order to improve the out-of-sample performance of the obtained portfolio, several researches propose to estimate the covariance matrix for the minimal variance model by modifying the sample variance estimate. For example, Ledoit and Wolf (2003, 2004) 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. DeMiguel et al. (2007) additionally impose a norm constraint on the portfolio for the variance minimization criterion by extending the idea of the 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 minimal variance model having the short sale constraint. All of these researches 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 researches agree on that the impact of the estimation error associated with the sample average 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. (2007) 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 for the recent decade. It seeks a good portfolio in the sense that it is feasible even when parameters in the optimization problem takes the least favorable value among a predetermined candidates which are given as the so-called *uncertainty set*. One possible critics to the robust portfolio models is that many of them are proposed without clarifying how to specify the uncertainty set. Exceptionally, some robust approaches take into account the worst case estimation error in a directly 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. Since this factor model-based robust approach adopts the probabilistic confidence region as the uncertainty set, it may not be adequate to call it *robust* for all possible event. Besides, very beginning robust models such as Soyster (1973) and Ben-Tal and Nemirovski (2000) assume nonparametric structure while the factor model approach employs a parametric assumption for constructing an uncertainty set. Except for the factor model-based robust portfolio models, which are motivated by the statistical estimation, it is unclear how to specify the uncertainty set in practice.

Both of the above two approaches – the shrinkage estimation-based minimum variance model and the robust portfolio – seek to alleviate the deterioration of the out-of-sample performance associated with the estimation error of each optimization criterion by considering the estimation of parameters and the selection of a portfolio simultaneously. In this paper, we study the connection between these approaches, in particular, by examining the role of the norm constraint not only in the variance minimization but also in an extended context.

First, we discuss the relation between the norm constraints and the uncertainty sets for the robust portfolio model. Secondly, we consider to expansively apply the norm constraint to two popular downside risk measures: the empirical value-at-risk (VaR) and the empirical conditional value-at-risk (CVaR).

The VaR has been used in capturing a large loss with a small probability in risk management practice. Although it still has much popularity, there have been controversy as to its theoretical property as a risk measure. For example, Cont, Deguest and Scandolo (2007) show that it has a robust property against outlying observations. On the other hands, it has been shown to violates the subadditivity (Artzner et al. 1999), and therefore, has been considered as an undesirable risk measure.

On the other hand, the CVaR is shown to have nice theoretical properties such as the coherence (Artzner et al. 1999) and the consistency with the risk-averse behavior of investors (e.g., Ogryczak and Ruszczýnski 2002), and has obtained a growing popularity also in practice. Moreover, the CVaR is much more attractive than the VaR from an optimization view point because it often leads to a tractable structure of the associated optimization problems (Rockafellar and Uryasev 2002). More interestingly, the authors have pointed out in Gotoh and Takeda (2005) and Takeda (2007) that the ν-support vector machines (ν-SVMs), an optimization-based statistical learning model developed by Schölkopf et al. (2000), has almost the same structure as the CVaR minimization. This fact motivates us to exploit theoretical results developed for the ν-SVMs in the context of the portfolio selection, as in Gotoh and Takeda (2008).

Thirdly, we extend the discussion into the so-called index tracking (mimicking) portfolio problem. Numerical experiments are conducted for demonstrating how the norm-constrained CVaR minimizing tracking portfolio achieves a good out-of-sample performance. In the experiments, we examine how the parameters used for describing the norm constraint and the CVaR objective are tuned, and show some results where the proposed approach involving the parameter tuning can outperform a few standard approaches for the tracking portfolio.

Contribution of this paper can be found in the following points:

- We show that the norm-constrained portfolio optimization can be considered as a robust portfolio optimization formulation with an adequate parameter uncertainty. 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 a nonparametric theory for the 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. Motivated by the theoretical result, 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. (2007). It is worth noting that although the bounds are not tight, the numerical experiments support that this model achieves a good out-of-sample performance. Also, this result also provides a theoretical validation to its robust counterpart.

- In contrast with the traditional models which simply minimize the empirical deviations from a target variable, a novel approach to the tracking portfolio construction is proposed by incorporating the norm constraint. Through some numerical experiments, the norm-constrained CVaR deviation model can achieve a better out-of-sample performance. In particular, tuning parameters of the norm-constrained CVaR deviation model by using the historical observations enhances the tracking performance. Besides, this indicates a possibility that specifying the uncertainty set in a robust portfolio based on the historical observations work effectively.

The structure of the paper is as follows. In the next section, we describe a proposition which relates the norm constraint for the portfolio selection (DeMiguel et al. 2007) and an uncertainty set for the robust portfolio. In Section 3, we consider a norm-constrained CVaR minimization by exploiting the generalization error bound for the ν-SVM (Takeda 2007), and a theoretical validation for the norm-constrained VaR and CVaR minimizations is provided.

In Section 4, we expansively apply the results developed in Section 3 to a tracking portfolio problem. Section 5 is devoted to the numerical experiments, where a norm-constrained tracking portfolio is examined, and it will be shown that adequate parameter tuning leads to improve the out-of-sample tracking performance. Finally, we conclude the paper with some remarks. Proofs for theorems are provided in Appendix.

2 A Robust Optimization Viewpoint for Norm-Constraint

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

As pointed out in Introduction, the 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. (2007) impose the norm constraint to the minimal 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 into each asset.

In DeMiguel et al. (2007), it is shown that when the 2-norm, $\|\boldsymbol{\pi}\|_2 := \sqrt{\boldsymbol{\pi}^\top \boldsymbol{\pi}}$, is employed as the norm $\|\boldsymbol{\pi}\|$ 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 $C = 1$, it is shown to be equivalent to the short sale-constrained minimum variance portfolio. In addition, they show that if the norm term is replaced with the quantity $\|\boldsymbol{\pi}\|_A := \sqrt{\boldsymbol{\pi}^\top \boldsymbol{A} \boldsymbol{\pi}}$ with $\boldsymbol{A}$ the covariance matrix induced from the single factor model, then the resulting portfolio is shown to be equivalent to the shrinkage estimate of the covariance matrix proposed by Ledoit and Wolf (2003).

It is worth noting that the above properties associated with $\|\boldsymbol{\pi}\|_2$ and $\|\boldsymbol{\pi}\|_1$ hold independently of the variance in the objective of (1), but hold based only on the basic constraint of the form $\boldsymbol{e}_n^\top \boldsymbol{\pi} = 1$. Thus, we first present a robust modeling view point on 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:

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

Substituting $-(\boldsymbol{r} - \boldsymbol{r}_0)/s$ as $\boldsymbol{r}$, the desired result is obtained. □

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 where 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 average return $\overline{\boldsymbol{\mu}} := \sum_{t=1}^{T} \boldsymbol{R}_t/T$ with the observed historical return vectors $\boldsymbol{R}_1, ..., \boldsymbol{R}_T$.

Apparently, the nominal return vector $\boldsymbol{r}_0$ and the positive scalar s is introduced so that we could interpret the constraint in a usual 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 put specific values in the parameters. For example, let us consider the case of $\boldsymbol{r}_0 = \boldsymbol{0}$ and $s' = s/C$, in which the equivalence is rewritten as

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

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

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

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

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

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

2.2 Various Type of Norms and Relation to the Uncertainty Sets

As stated in Proposition 1, the 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 by Soyster (1973), which is known to result in a bit too much conservative solution. Besides, since the 1-norm constraint with $C = 1$ is equivalent to the short sale constraint as mentioned in DeMiguel et al. (2007), the short sale constraint, $\boldsymbol{\pi} \geq \boldsymbol{0}$ is equivalent to a Soyster's type robust constraint of the form:

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

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. (2007), the A-norm-constrained variance minimizing portfolio with the covariance matrix of the single-factor model is equivalent to the minimal 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 the 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 the Norms in the Norm-Constraints for Portfolio Selection and the Uncertainty Sets for Robust Portfolio

Norm in Norm Constraint	Norm in Uncertainty Set
$\|\boldsymbol{\pi}\|_1$ (DeMiguel et al. 2007)	$\|\boldsymbol{r}\|_\infty$ (Soyster 1973)
$\|\boldsymbol{\pi}\|_A$ (DeMiguel et al. 2007)	$\|\boldsymbol{r}\|_{A^{-1}}$ (Ben-Tal and Nemirovski 2000)
$\|\boldsymbol{\pi}\|_2$ (DeMiguel et al. 2007)	$\|\boldsymbol{r}\|_2$
$\|\boldsymbol{\pi}\|_\infty$	$\|\boldsymbol{r}\|_1$
$\max\{\|\boldsymbol{\pi}\|_\infty, \|\boldsymbol{\pi}\|_1/p\}$	$\|\!\|\boldsymbol{r}\|\!\|_p$ (Bertsimas, Pachamanova and Sim 2004)

3 Norm-Constrained VaR/CVaR Portfolio and Generalization Error Bound

In this section, we provide a statistical validation to the norm constraint for the VaR and CVaR minimizing portfolio. In connection with the robust optimization view for the norm constraint, the following results indirectly give a statistical foundation to a robust portfolio problem.

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 variable as f if one feels better as it is smaller. For example, the minus return can be employed as a loss, i.e.,

$$f(\boldsymbol{\pi}, \boldsymbol{\mathcal{R}}) = -\mathcal{R}(\boldsymbol{\pi}) = -\boldsymbol{\mathcal{R}}^\top \boldsymbol{\pi}. \tag{2}$$

For $\beta \in (0,1)$, the β-VaR, $\alpha_\beta(\boldsymbol{\pi})$, associated with 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 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, the β-CVaR associated with loss $f(\boldsymbol{\pi}, \boldsymbol{\mathcal{R}})$ 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

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

where $\mathbb{E}[\cdot]$ denotes the operator for the mathematical expectation. According to Rockafellar and Uryasev (2002), the β-CVaR, $\phi_\beta(\boldsymbol{\pi})$, can be approximately regarded as the expected value of loss f greater than the β-VaR, α_β, and therefore, one has $\alpha_\beta(\boldsymbol{\pi}) \leq \phi_\beta(\boldsymbol{\pi})$, as in Figure 1. In practice, similarly to the 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 $\alpha_\beta(\boldsymbol{\pi})$ is nonconvex, in general, even when f is linear in $\boldsymbol{\pi}$. The β-CVaR minimizing portfolio is given by a solution to

$$\min\{\ \phi_\beta(\boldsymbol{\pi})\ :\ \boldsymbol{\pi} \in \Pi\ \} = \min\{\ F_\beta(\boldsymbol{\pi}, \alpha)\ :\ \boldsymbol{\pi} \in \Pi,\ \alpha \in \mathbb{R}\ \}, \tag{3}$$

which can be reformulated as a convex program when loss f is convex in $\boldsymbol{\pi}$ and Π is a convex set. In addition, for an optimal solution $(\boldsymbol{\pi}^*, \alpha^*)$ to (3), α^* gives an approximate value of the 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 not so, α^* is located in a closed interval $[\alpha_\beta(\boldsymbol{\pi}^*), \alpha_\beta^+(\boldsymbol{\pi}^*)]$ where $\alpha_\beta^+(\boldsymbol{\pi}) := \inf\{\alpha : \Phi(\alpha|\boldsymbol{\pi}) > \beta\}$.

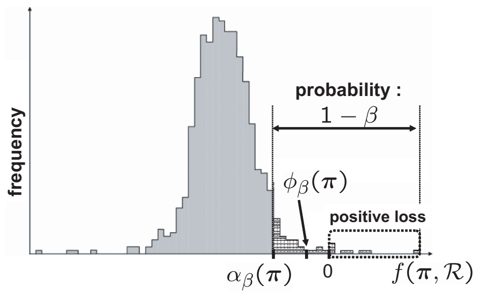

Figure 1: Illustration of the β-VaR, α_β, and the β-CVaR, ϕ_β associated with loss f

As pointed out in Introduction, the minimization of these risk measures cannot be properly implemented in practice because the true distribution is unobservable. In order to approximately obtain an optimal solution $\boldsymbol{\pi}^*$ to these minimizations, the empirical version of the true function is optimized.

Let $\Phi^T(\cdot|\boldsymbol{\pi})$ denote the empirical distribution of the portfolio return $\mathcal{R}(\boldsymbol{\pi})$ 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., $\Phi^T(\alpha|\boldsymbol{\pi}) := |\{t \in \{1, ..., T\} : f(\boldsymbol{\pi}, \boldsymbol{R}_t) \leq \alpha\}|/T$. The β-VaR, $\alpha_\beta(\boldsymbol{\pi})$, is then replaced with the empirical version, $\alpha_\beta^T(\boldsymbol{\pi})$, i.e.,

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

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 then written as follows, respectively:

$$\left|\begin{array}{ll} \min & \alpha_\beta(\boldsymbol{\pi}) \\ \text{s.t.} & \boldsymbol{e}_n^\top \boldsymbol{\pi} = 1,\ \boldsymbol{A}\boldsymbol{\pi} \leq \boldsymbol{b} \\ & \|\boldsymbol{\pi}\| \leq C, \end{array}\right. \qquad \left|\begin{array}{ll} \min & \phi_\beta(\boldsymbol{\pi}) \\ \text{s.t.} & \boldsymbol{e}_n^\top \boldsymbol{\pi} = 1,\ \boldsymbol{A}\boldsymbol{\pi} \leq \boldsymbol{b} \\ & \|\boldsymbol{\pi}\| \leq C, \end{array}\right. \tag{4}$$

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 $\boldsymbol{A}\boldsymbol{\pi} \leq \boldsymbol{b}$, 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 (2), the empirical β-VaR and β-CVaR minimizations with sample returns $\boldsymbol{R}_1, ..., \boldsymbol{R}_T$ can be formulated as follows, respectively:

$$\left|\begin{array}{ll} \min & \alpha \\ \text{s.t.} & \boldsymbol{e}_T^\top \boldsymbol{z} \leq \lfloor(1-\beta)T\rfloor \\ & -\boldsymbol{R}_t^\top \boldsymbol{\pi} - M\boldsymbol{z} \leq \alpha \boldsymbol{e}_T, \quad \boldsymbol{z} \in \{0,1\}^T \\ & \boldsymbol{e}_n^\top \boldsymbol{\pi} = 1,\ \boldsymbol{A}\boldsymbol{\pi} \leq \boldsymbol{b}, \quad \|\boldsymbol{\pi}\| \leq C, \end{array}\right. \tag{5}$$

where M is a sufficiently large number;

$$\left|\begin{array}{ll} \min & \alpha + \frac{1}{(1-\beta)T}\boldsymbol{e}_T^\top \boldsymbol{y} \\ \text{s.t.} & \boldsymbol{y} \geq -\boldsymbol{R}_t^\top \boldsymbol{\pi} - \alpha \boldsymbol{e}_T, \quad \boldsymbol{y} \geq \boldsymbol{0} \\ & \boldsymbol{e}_n^\top \boldsymbol{\pi} = 1, \; \boldsymbol{A}\boldsymbol{\pi} \leq \boldsymbol{b}, \quad \|\boldsymbol{\pi}\| \leq C. \end{array}\right. \tag{6}$$

If the Euclidean norm $\|\boldsymbol{\pi}\|_2$ is adopted as the norm $\|\boldsymbol{\pi}\|$, the VaR minimization (5) is a quadratically constrained 0-1 mixed integer program, which can be solved via a state-of-the-art solver such as ILOG CPLEX11 as long as the size of the problem is not huge and it has so good structure that the sophisticated branch-and-cut algorithm works effectively. However, it is still hard to solve (5) in a practical time even when the number of T or n is more than hundred. Therefore, it is reasonable to apply the linearly representable norms such as $\|\boldsymbol{\pi}\|_1, \|\boldsymbol{\pi}\|_\infty$ and D-norm, in place of $\|\boldsymbol{\pi}\|_2$ when the VaR minimization (5) is applied.

On the other hand, the CVaR minimization (6) is a quadratically constrained linear program, and it can be efficiently solved via an interior point algorithm even when the size of the problem is large. Therefore, the Euclidean norm or A-norm is worth applying when the CVaR minimization is considered.

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

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 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*

$$\mathbb{P}\{f(\boldsymbol{\pi}, \boldsymbol{\mathcal{R}}) > \theta\} \leq (1-\beta) + \sqrt{\frac{2}{T}\left\{\frac{4c^2(C^2+1)(B_R^2+\theta^2)\log_2(2T)}{(\alpha_\beta^T(\boldsymbol{\pi})-\theta)^2} + \ln\frac{2}{\delta \mathrm{e}}\right\}} \tag{7}$$

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

$$\mathbb{P}\{f(\boldsymbol{\pi}, \boldsymbol{\mathcal{R}}) > \theta\} \geq (1-\beta) - \sqrt{\frac{2}{T}\left\{\frac{4c^2(C^2+1)(B_R^2+\theta^2)\log_2(2T)}{(\alpha_\beta^T(\boldsymbol{\pi})-\theta)^2} + \ln\frac{2}{\delta \mathrm{e}}\right\}} \tag{8}$$

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*

$$\mathbb{P}\{f(\boldsymbol{\pi}, \boldsymbol{\mathcal{R}}) > \theta\} \leq (1-\beta) + \sqrt{\frac{2}{T}\left\{\frac{4c^2(C^2+1)(B_R^2+\theta^2)\log_2(2T)}{(\phi_\beta^T(\boldsymbol{\pi})-\theta)^2} + \ln\frac{2}{\delta \mathrm{e}}\right\}} \tag{9}$$

with probability at least $1-\delta$.

See 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 $\phi_\beta^T(\boldsymbol{\pi}) < \theta$ holds.

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, $\phi_\beta^T(\boldsymbol{\pi})$. Here, it is noteworthy that in the above inequalities (7), (8) and (9), 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 normal distribution may wonder if the bounded support assumption is too restrictive. However, the support of the asset return should be bounded because the total amount of money or credit over 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 care about the tail part (edge of the support) of the distribution in a nonparametric manner.

The main concern 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 gives a clue for making the probability smaller.

First of all, we should note that the right-hand sides of (7), (8) and (9) 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. By noting that these bounds hold only when the norm of the portfolio is bounded above by a constant C, solutions to the optimization problems (4) are expected to make the loss probability smaller. Besides, the upper and lower bounds are decreasing in C. However, decreasing C restricts the feasibility of $\boldsymbol{\pi}$, which can lead to increase $\alpha_\beta^T(\boldsymbol{\pi})$ and $\phi_\beta^T(\boldsymbol{\pi})$ and, thus, there is a trade-off between the value of C and the empirical risk measures. In order to harmonize the effects of C and $\boldsymbol{\pi}$ in minimizing the bounds, the parameter C should be tuned when the norm-constrained problems (4) are solved.

Next, let us examine the other parameters. The bounds decrease in T, which is consistent with the law of large numbers. Also, the right-hand sides are decreasing in B_R as well. However, B_R is uncontrollable due to its nature, and fortunately, the knowledge on the accurate value of B_R does not alter the effect of $\boldsymbol{\pi}$ in minimizing the bounds, which is unlike C. Therefore, it is unnecessary to pay attention to its size.

As for β, we cannot figure out the shape of the right-hand sides as a function of β because the second term of them includes 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 observation, we can expect that solving (4) in combination with tuning C and β leads to a lower loss probability. One possible critics to this expectation may arise from the fact that the bounds are not tight, and the minimization may not be effective in decreasing the loss probability. However, the norm-constrained portfolio model motivated by the bound minimization will be shown to achieve a better out-of-sample performance than the other models through the numerical experiments given in Section 5.

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:*

$$\left| \begin{array}{ll} \min & \dfrac{\alpha_\beta^T(\boldsymbol{\pi}) - \theta}{\|\boldsymbol{\pi}\|} \\ s.t. & \boldsymbol{e}_n^\top \boldsymbol{\pi} = 1, \;\; \boldsymbol{A}\boldsymbol{\pi} \leq \boldsymbol{b}, \end{array} \right. \qquad \left| \begin{array}{ll} \min & \dfrac{\phi_\beta^T(\boldsymbol{\pi}) - \theta}{\|\boldsymbol{\pi}\|} \\ s.t. & \boldsymbol{e}_n^\top \boldsymbol{\pi} = 1, \;\; \boldsymbol{A}\boldsymbol{\pi} \leq \boldsymbol{b}. \end{array} \right. \tag{10}$$

In Gotoh and Takeda (2008), the authors describe a two-step framework for approaching a solution to (10) because the fractional problems may have an intractable structure when θ is small enough. By comparing to the fractional programming formulation, the norm-constrained problems (4) are much easier to solve and, thus, easier to conduct a cross-validation for the parameter tuning. Of course, there is a possibility that ignoring the difficult case leads to passing *over the good performance. However, easy implementation of the parameter tuning motivates 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 (5) and (6) to a tracking (or mimicking) 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 $\boldsymbol{\pi} \in \mathrm{IR}^n$ is determined so that the squared error $\sigma^2(\boldsymbol{\pi}) := \mathbb{E}[(\mathcal{R}(\boldsymbol{\pi}) - \mathcal{I})^2]$ would be minimized. As already mentioned, it is, however, impossible to evaluate the expectation $\mathbb{E}[\cdot]$ in an exact manner because no one knows the true distribution of $(\mathcal{I}, \boldsymbol{\mathcal{R}})$. Therefore, a tracking portfolio is obtained by solving the empirical version of the criterion:

$$\min\{\frac{1}{T}\sum_{t=1}^{T}|\boldsymbol{R}_t^\top\boldsymbol{\pi} - I_t|^2 \, : \, \boldsymbol{e}_n^\top\boldsymbol{\pi} = 1, \; \boldsymbol{A}\boldsymbol{\pi} \le \boldsymbol{b}\} \tag{11}$$

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.

Some researches have proposed to use the other deviations for measuring error (see, e.g., Prigent 2007). For example, a criterion of minimizing the mean absolute deviation $\mathbb{E}[|\mathcal{R}(\boldsymbol{\pi})-\mathcal{I}|]$ has been applied in Gilli and Këllezi (2002) where the mean absolute deviation minimization is implemented by minimizing the sample average of the empirical absolute errors:

$$\min\{\frac{1}{T}\sum_{t=1}^{T}|\boldsymbol{R}_t^\top\boldsymbol{\pi} - I_t| \, : \, \boldsymbol{e}_n^\top\boldsymbol{\pi} = 1, \; \boldsymbol{A}\boldsymbol{\pi} \le \boldsymbol{b}\}, \tag{12}$$

which 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*

$$y = a_1x_1 + \cdots + a_dx_d \tag{13}$$

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:* $\sum_{i=1}^{m}\{y_i - (a_1x_{i1} + \cdots + a_dx_{id})\}^2$*. 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). Also, the ridge regression and the lasso minimize the regularization terms $\|\boldsymbol{\pi}\|_2$ *and* $\|\boldsymbol{\pi}\|_1$*, respectively, in addition to the sum of squared errors. As already mentioned, this idea is applied to the portfolio context by DeMiguel et al. (2007). Although they do not mention the tracking portfolio, Ledoit and Wolf (2004) apply the shrinkage technique to the tracking portfolio.*

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 the overfitting (see, e.g., Hastie, Tibshirani and Friedman 2001).

In the following, we describe 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

$$f(\boldsymbol{\pi}, \mathcal{I}, \boldsymbol{\mathcal{R}}) = |\mathcal{I} - \mathcal{R}(\boldsymbol{\pi})|. \tag{14}$$

	asset					attribute			
	idx.	1	$\cdots$	n		obj.	1	$\cdots$	d
1	I_1	R_{11}	$\cdots$	R_{1n}	1	y_1	x_{11}	$\cdots$	x_{1d}
date $\vdots$	$\vdots$	$\vdots$		$\vdots$	case $\vdots$	$\vdots$	$\vdots$		$\vdots$
T	I_T	R_{T1}	$\cdots$	R_{Tn}	m	y_m	x_{m1}	$\cdots$	x_{md}
portfolio		π_1	$\cdots$	π_n	coefficients		a_1	$\cdots$	a_d

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

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 $\Phi^T(\alpha) := |\{t : |I_t - \boldsymbol{R}_t^\top \boldsymbol{\pi}| \leq \alpha\}|/T$, while the empirical β-CVaR, $\phi_\beta^T(\boldsymbol{\pi})$, is defined similarly:

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

The tracking portfolio model using the norm-constrained VaR deviation is formulated as follows:

$$\left| \begin{array}{ll} \displaystyle\min_{\alpha,\boldsymbol{\pi},\boldsymbol{w},\boldsymbol{x},\boldsymbol{z}} & \alpha \\ \text{s.t.} & \boldsymbol{e}_T^\top \boldsymbol{z} \leq \lfloor (1-\beta)T \rfloor \\ & \boldsymbol{w} + \boldsymbol{x} - M\boldsymbol{z} \leq \alpha \boldsymbol{e}_T \\ & w_t - x_t = I_t - \boldsymbol{R}_t^\top \boldsymbol{\pi}, \quad t = 1, ..., T \\ & \boldsymbol{w} \geq \boldsymbol{0}, \ \boldsymbol{x} \geq \boldsymbol{0}, \ \boldsymbol{z} \in \{0,1\}^T \\ & \boldsymbol{e}_n^\top \boldsymbol{\pi} = 1, \ \boldsymbol{A}\boldsymbol{\pi} \leq \boldsymbol{b}, \ \|\boldsymbol{\pi}\| \leq C, \end{array} \right. \tag{15}$$

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

$$\left| \begin{array}{ll} \displaystyle\min_{\alpha,\boldsymbol{\pi},\boldsymbol{w},\boldsymbol{x},\boldsymbol{y}} & \alpha + \frac{1}{(1-\beta)T} \boldsymbol{e}_T^\top \boldsymbol{y} \\ \text{s.t.} & \boldsymbol{y} \geq \boldsymbol{w} + \boldsymbol{x} - \alpha \boldsymbol{e}_T \\ & \boldsymbol{y} \geq \boldsymbol{0}, \ \boldsymbol{w} \geq \boldsymbol{0}, \ \boldsymbol{x} \geq \boldsymbol{0} \\ & w_t - x_t = I_t - \boldsymbol{R}_t^\top \boldsymbol{\pi}, \quad t = 1, ..., T \\ & \boldsymbol{e}_n^\top \boldsymbol{\pi} = 1, \ \boldsymbol{A}\boldsymbol{\pi} \leq \boldsymbol{b}, \ \|\boldsymbol{\pi}\| \leq C. \end{array} \right. \tag{16}$$

As mentioned before, the use of the 2-norm $\|\boldsymbol{\pi}\|_2$ in (15) is not recommendable, while that in (16) is worth applying because (16) is still a convex optimization problem. Taking into account the computational tractability and the coherence as a risk measure, we only examine the norm-constrained CVaR minimization (16) with $\|\boldsymbol{\pi}\|_2$ in the computational experiment in the next section.

Remark 3 *As described in Gotoh and Takeda (2005) and Takeda (2007), the β-CVaR minimization is equivalent to the ν-SVMs. From this viewpoint, the above model (16) can be considered as an application of the ν-support vector regression (ν-SVR) to the index tracking. One difference between these two frameworks can be found in the additional constraints imposed on* *the associated optimizations. Actually, 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 (13).*

Remark 4 *Seemingly, the formulation (16) is just a norm-constrained version of the CVaR deviation model in Rockafellar and Uryasev (2002). In addition to the norm constraint, the parameter* β *in (16) plays a different role from that in the standard CVaR model. While* β *in the standard model is fixed by user before the optimization so that he/she can capture the predetermined level risk, the* β *in (16) is to be tuned at the optimization. In the numerical experiment in the next section, we provide a way of tuning* β *as well as* C *based on the historical data.*

Similarly to the previous section, nonparametric bounds of the tracking error probability are also obtained with the empirical β-VaR and β-CVaR of the tracking error (14) by modifying Theorem 1 and Corollary 1.

Theorem 2 *Let* θ *be a threshold for portfolio loss. Suppose that 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* $\|\boldsymbol{\pi}\|_2 \leq C$*, the probability of the tracking error being greater than* θ*,* $\mathbb{P}\{|\mathcal{I} - \mathcal{R}(\boldsymbol{\pi})| > \theta\}$*, is bounded above as*

$$\mathbb{P}\{|\mathcal{I} - \mathcal{R}(\boldsymbol{\pi})| > \theta\} \leq (1-\beta) + 2\sqrt{\frac{2}{T}\left\{\frac{4c^2(C^2+1)(B_R+\theta)^2\log_2(2T)}{(\alpha_\beta^T(\boldsymbol{\pi})-\theta)^2} + \ln\frac{2}{\delta \mathrm{e}}\right\}} \tag{17}$$

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* $\|\boldsymbol{\pi}\|_2 \leq C$*, the probability is bounded below as*

$$\mathbb{P}\{|\mathcal{I} - \mathcal{R}(\boldsymbol{\pi})| > \theta\} \geq (1-\beta) - 2\sqrt{\frac{2}{T}\left\{\frac{4c^2(C^2+1)(B_R+\theta)^2\log_2(2T)}{(\alpha_\beta^T(\boldsymbol{\pi})-\theta)^2} + \ln\frac{2}{\delta \mathrm{e}}\right\}} \tag{18}$$

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* $\|\boldsymbol{\pi}\|_2 \leq C$*, one has*

$$\mathbb{P}\{|\mathcal{I} - \mathcal{R}(\boldsymbol{\pi})| > \theta\} \leq (1-\beta) + 2\sqrt{\frac{2}{T}\left\{\frac{4c^2(C^2+1)(B_R+\theta)^2\log_2(2T)}{(\phi_\beta^T(\boldsymbol{\pi})-\theta)^2} + \ln\frac{2}{\delta \mathrm{e}}\right\}}. \tag{19}$$

with probability at least $1-\delta$*.*

See Appendix for the proofs.

5 Numerical Experiments

In this section, we conduct some numerical experiments for examining how the proposed tracking portfolio approach works.

Considering the computational efficiency, we only solve the norm-constrained CVaR model (16) and compare with the squared error minimization (11), the absolute error minimization (12) and the CVaR deviation model without the norm constraint.

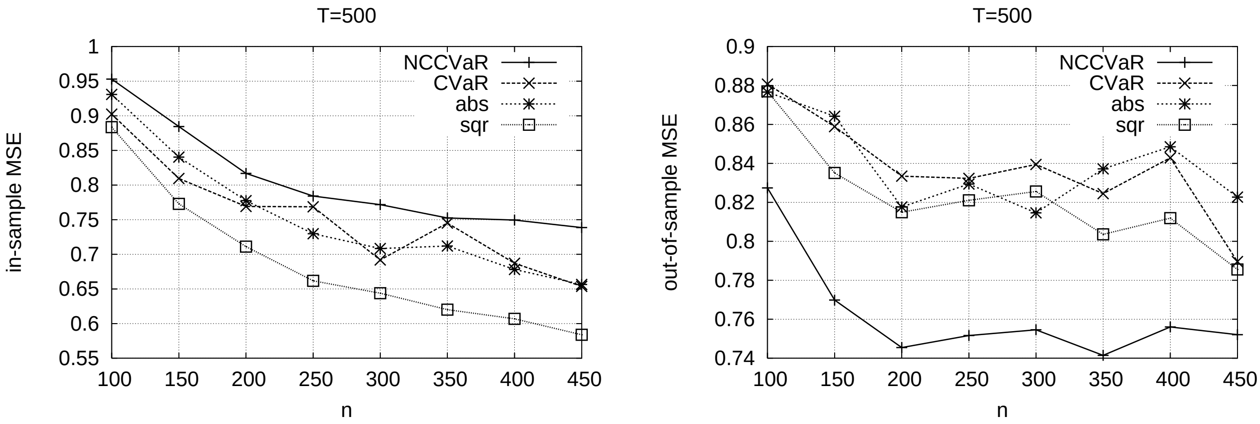

Figure 3: In-sample MSE and out-of-sample MSE for randomly generated dataset
"NCCVaR" stands for the proposed model, "CVaR" stands for the CVaR minimization without the norm constraint, "abs" stands for the absolute deviation minimization (12), and "sqr" stands for the squared deviation minimization (11).

5.1 Results with Randomly Generated Normal Distributed Data

First of all, an experiment is conducted on a randomly generated data set in order to examine how the norm-constrained CVaR deviation model with a parameter tuning works. We randomly generated $\bar{n}$-dimensional vectors $\bar{\boldsymbol{R}}_t$, $t = 1, \ldots, 500$, which follow an $\bar{n}$-dimensional normal distribution $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where μ_j and σ_j are independently drawn from uniform distributions on $(0, 10)$, and the correlation between R_i and R_j for any $i \neq j$ is 0.5. Associated with the generated variables, we define the target as $I_t = \boldsymbol{e}_{\bar{n}}^\top \bar{\boldsymbol{R}}_t / \bar{n}$, $t = 1, ..., 500$. For each $n < \bar{n}$, a submatrix $[\boldsymbol{R}_1, ..., \boldsymbol{R}_{500}] \in \mathrm{I\!R}^{n \times 500}$ is prepared for constructing the tracking portfolio of size up to n, by randomly extracting n rows out of the matrix $[\bar{\boldsymbol{R}}_1, ..., \bar{\boldsymbol{R}}_{500}] \in \mathrm{I\!R}^{\bar{n} \times 500}$. Based on the submatrix, a portfolio is computed via each model. The tracking error is expected to decrease as the number n grows because, by construction, the value $I_t - \boldsymbol{e}_n^\top \boldsymbol{R}_t / n$ plays a role of residual, and it is expected to vanish as n approaches to $\bar{n}$.

As for the norm-constrained CVaR model (16), the parameters β and C should be tuned by using $\boldsymbol{R}_t$, $t = 1, ..., 500$. In this experiment, they are determined so that the in-sample mean squared error (MSE) is minimized over 25 pairs of (β, C)s with $\beta = 0.5, 0.6, ..., 0.9$ and $C = 1/\sqrt{n} + 0.02, 1/\sqrt{n} + 0.04, ..., 1/\sqrt{n} + 0.10$. On the other hand, the out-of-sample performance of the resulting portfolio is evaluated using new 500 vectors $\boldsymbol{R}'_{t'} \in \mathrm{I\!R}^n$, $t' = 1, ..., 500$, which follow the same distribution as $\boldsymbol{R}_t \in \mathrm{I\!R}^n$, $t = 1, ..., 500$. We also tune β for the CVaR minimization model without the norm constraint similarly to the norm-constrained one. That is, β is determined so that the in-sample MSE is minimized over five βs through 0.5 to 0.9.

The line graphs in Figure 3 show the in-sample MSE (left) and the out-of-sample MSE (right) for $n = 100, 150 \ldots, 450$ and $\bar{n} = 500$. As expected, both of the in-sample and out-of-sample MSEs of every model roughly decrease as the number n grows, and the squared error minimization achieves the best in-sample MSE by definition. The most interesting fact is that for any n, the norm-constrained CVaR approach achieves the smallest out-of-sample MSE, while it takes the largest in-sample MSE. Considering that the CVaR minimization without the norm constraint results in large out-of-sample MSEs and is dominated by the square error minimization in any n, we see that the norm constraint with some parameter tuning can improve the out-of-sample performance by avoiding the overfitting.

Figure 4 shows the percentage of the realized losses f being greater than θ, i.e., $|\{t \in \{1, ..., 500\} : f(\boldsymbol{\pi}, \boldsymbol{R}'_t) > \theta\}|/500$, in the case of $n = 200$. It is worth noting that this percentage

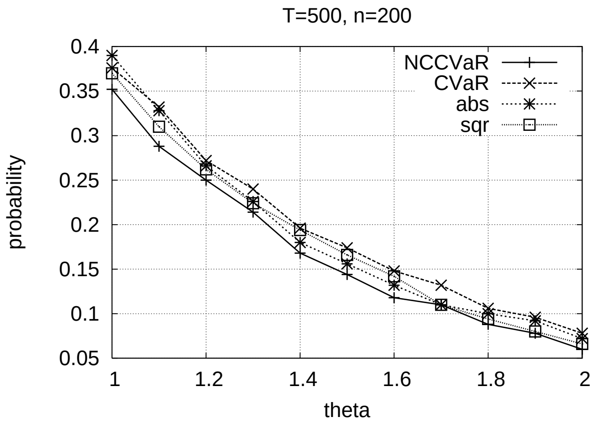

Figure 4: The probability of loss being greater than θ for randomly generated dataset ($n = 200$, $T = 500$).

The best parameters chosen in NCCVaR are $\beta = 0.9$ and $C = 1/\sqrt{n} + 0.02 = 0.09$.

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 three models in this sense. Although Theorems 1 and 2 and Corollaries 1 and 2 seem to provide rough 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 with Actual Market Data

Next we conduct an experiment on real financial market data by using monthly and weekly return data of stocks listed in the Nikkei 225 index at the end of August 2008. The monthly data set that we used in this experiment consists of returns of $n = 186$ companies in 256 consecutive months (21 years) starting from May 1987 to August 2008, whereas the weekly data set consists of returns of $n = 185$ companies in 1,117 consecutive weeks (almost 21 years) starting from April 12, 1987 to August 31, 2008. The targets to be mimicked are the monthly and weekly returns of the Nikkei 225 index, respectively. As in the random data case, n assets consisting of the universe are randomly chosen from $\bar{n} = 186$ 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 obtain 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, $\overline{T} = 136$ times and $\overline{T} = 967$ times rolling horizon evaluations are performed for the monthly and weekly data sets, respectively. As a measure to evaluate the out-of-sample performance, the MSE is employed:

$$\frac{1}{\overline{T}} \sum_{t=1}^{\overline{T}} (I_{t+T} - \boldsymbol{R}_{t+T}^\top \bar{\boldsymbol{\pi}}_t)^2.$$

In the norm-constrained CVaR deviation model, the parameters β and C were tuned systematically as follows. Using the first $\frac{5}{6}T$-period data set, we found a pair of β and C that give the best prediction for the remaining $\frac{1}{6}T$-period data set. For the monthly data set, the best β was found from 0.5 to 0.9 in steps of 0.1 while the best C was found from $1/\sqrt{n} + 0.02$ to $1/\sqrt{n} + 0.10$ in steps of 0.02. For the weekly data set, the best β was found similarly while the best C was found from $1/\sqrt{n} + 0.01$ to $1/\sqrt{n} + 0.05$ in steps of 0.01. Similarly to the random data case, only β is tuned for the norm-unconstrained CVaR model in the same manner.

Figure 5 reports the in-sample and out-of-sample MSEs and the out-of-sample correlation coefficient between the obtained portfolio and the index for the monthly data when the short sale constraint is imposed on the portfolio. Similarly to the random data case, for any n, the out-of-sample MSE is minimized by the norm-constrained CVaR model while the in-sample MSE is minimized by the squared error minimization (11) and the norm-constrained CVaR model results in the largest MSE. This implies that the norm-constrained CVaR model still has advantage over the traditional model having no norm constraint.

Figure 6 reports the MSEs and the correlation coefficient when the short sale constraint is not imposed. The difference between the norm-constrained model and the traditional models is now more emphasized. In fact, the three norm-unconstrained models achieve zero in-sample MSE for n more than 120, while the corresponding out-of-sample MSEs diverge, although we omit these cases in the figures because the optimization problem then has infinite number of optimal solutions and the solution is obtained just by chance. Even for the case of $n < 120$, the difference is very clear because the norm-unconstrained models suffer from the overfitting. In addition, comparing with Figure 5, the performance of every model deteriorates, and we see that the short sale constraint, $\boldsymbol{\pi} \geq \mathbf{0}$, plays a significant role in avoiding the overfitting and improving the out-of-sample performance, not only in the CVaR minimization, but also in the other traditional models. This seems to be because the short sale constraint has something to do with the norm constraint of the form $\|\boldsymbol{\pi}\|_1 \leq C$, as indicated in Section 2.

Corresponding to the monthly data case, Figures 7 and 8 show results for the weekly data. Similarly to the monthly data case, the norm-constrained CVaR minimization remains to show a dominant advantage. Although the advantage found in the figure seems to be smaller than the monthly one, the difference is more significantly supported through a statistical test.

Table 2 shows the p-values of the paired Wilcoxon signed-rank test, which is a non-parametric statistical hypothesis test for the case of two related samples. The null hypothesis of the test is that the out-of-sample MSE of the CVaR model without the norm constraint, the absolute deviation model or the squared deviation model is no more than that of the norm-constrained CVaR model. Small p-value indicates that the out-of-sample MSE of each of traditional models is significantly larger than that of the norm-constrained CVaR model. From this table, except for several cases where the short sale constraint is imposed and the monthly data is used, we see that the norm-constrained CVaR model is significantly superior to the other models.

6 Concluding Remarks

As pointed out in DeMiguel et al. (2007), the norm constraint has something to do with basic notions used in the traditional portfolio optimization such as the short sale constraint, the $1/n$ portfolio, and the shrinkage techniques for the covariance matrix.

In the paper, we explore the role of the norm constraint in portfolio selection from various perspectives. For one thing, the norm constraint can be rewritten in a robust inequality form by using the dual norm in describing the uncertainty set. What is more, the norm constraint in the CVaR or VaR minimization has much to do with the regularization term of the ν-SVMs, and a theoretical validation to the ν-SVMs can be employed in the context of the portfolio optimization.

According to the numerical experiments, we see that the norm-constrained CVaR minimization with some parameter tuning can alleviate the overfitting and overcomes the norm-unconstrained approaches. Even for the norm-unconstrained risk minimizations, we see that the out-of-sample performance is improved by imposing the short sale constraint. This observation indicates that the result of Jagannathan and Ma (2003) holds true not only for the variance minimization, but also for the other portfolio optimization. In fact, imposing the short sale constraint may be the reason why the mean-risk model of Konno and Yamazaki (1991) achieves a good performance even when the number of assets is greater than that of the historical data. Through the experiments we conducted and the existing researches, we can expect that the norm constraint improves the out-of-sample performance for a wide class of risk minimizations.

In addition, it is noteworthy that the parameter tuning based on the historical data works for improving the out-of-sample performance. Considering the relation between the norm constraint and the robust constraint, such a tuning technique can be applied to the tuning of the uncertainty set for nonparametric robust portfolio models.

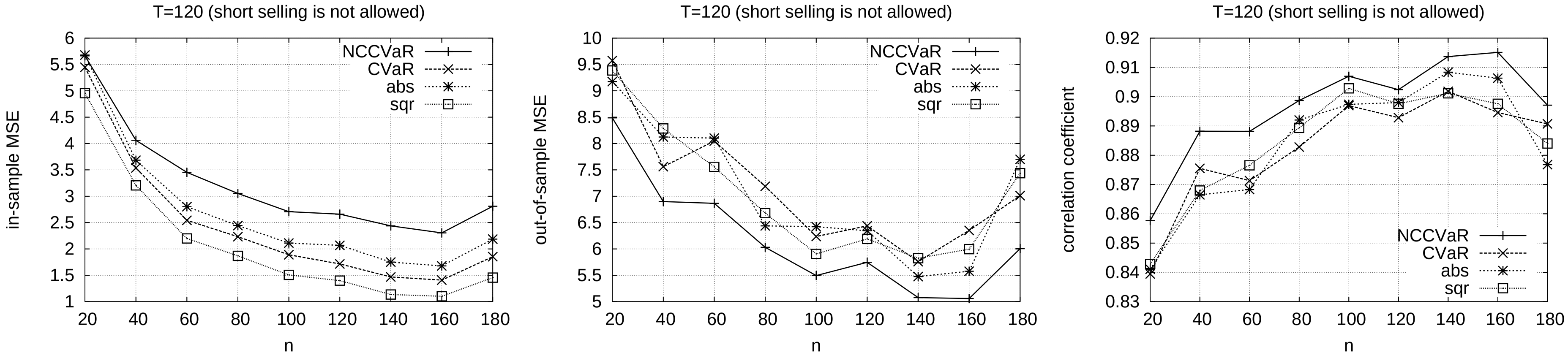

Figure 5: In-sample MSE (left), out-of-sample MSE (middle), out-of-sample correlation coefficient (right) for the monthly data set when the short sale constraint, $\boldsymbol{\pi} \geq \mathbf{0}$, is imposed

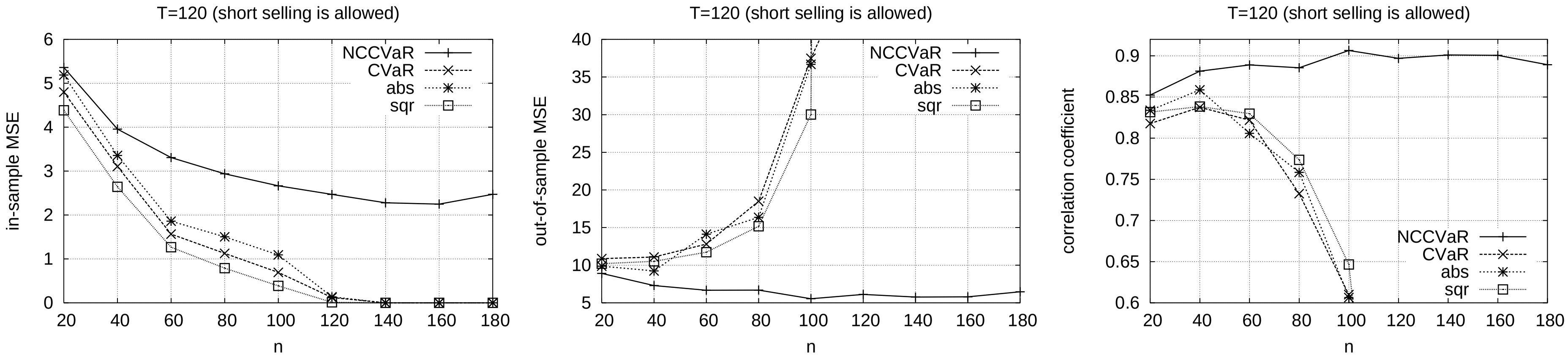

Figure 6: In-sample MSE (left), out-of-sample MSE (middle), out-of-sample correlation coefficient (right) for the monthly data set when short sale constraint is not imposed

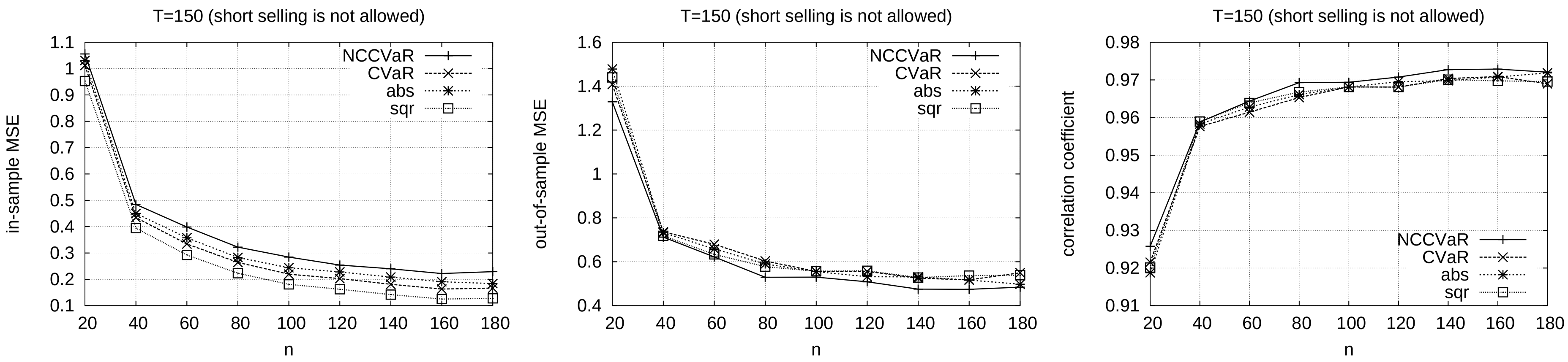

Figure 7: In-sample MSE (left), out-of-sample MSE (middle), out-of-sample correlation coefficient (right) for the weekly data set when short sale constraint is imposed

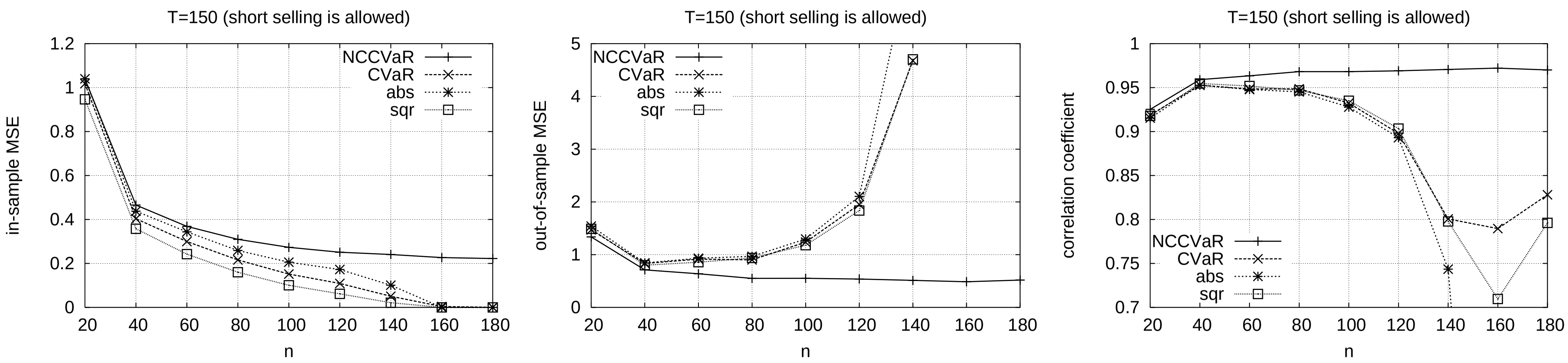

Figure 8: In-sample MSE (left), out-of-sample MSE (middle), out-of-sample correlation coefficient (right) for the weekly data set when it is not imposed

A Proofs of Theorems

The generalization error bounds have been investigated for the two-class classification problem, where given data set with binary labels are used for constructing a function h which 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).

We start with introducing an analysis for the generalization performance of the classification function. Let us address a classification problem of learning a classifier $h(\boldsymbol{R})$ whose decision function $\mathrm{sign}(h(\boldsymbol{R}))$ maps $\boldsymbol{R}$ to -1 or 1 based on 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 classification 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 based on 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 $R[h]$ below based on 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

$$\mathcal{F} = \{\boldsymbol{R} \mapsto \boldsymbol{R}^\top \boldsymbol{\pi} \,:\, \|\boldsymbol{\pi}\|_2 \leq 1,\ \boldsymbol{R} \in V\}. \tag{20}$$

Then, 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

$$\mathbb{P}[\mathcal{I}h(\boldsymbol{\mathcal{R}}) < 0] \leq \frac{1}{T}|\{\, i \,:\, I_i h(\boldsymbol{R}_i) < \gamma \,\}| + \sqrt{\frac{2}{T}\left(\frac{4c^2\bar{B}^2}{\gamma^2}\log_2(2T) - 1 + \log\left(\frac{2}{\delta}\right)\right)}, \tag{21}$$

for any $\gamma > 0$.

Theorem 1 analyzes 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, a threshold constant θ is introduced, 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 (21) of the generalization error for classification. To make the regression problem to fit to the setting of (21), 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 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 $\bar{B} = \sqrt{B_R^2 + \theta^2}$.

Table 2: p-values of the paired Wilcoxon signed-rank test for the Nikkei data set.

(i) monthly data

(a) short selling is not allowed

n	vs. CVaR	vs. abs	vs. sqr
20	8.93e-04	9.74e-03	0.01897
40	0.1148	0.01225	0.05604
60	0.1610	3.09e-03	0.2194
80	4.30e-04	0.09568	0.01847
100	0.06277	0.1565	0.1737
120	0.03265	0.2432	0.04689
140	6.64e-03	7.79e-03	3.39e-03
160	1.53e-05	0.01521	3.02e-05
180	3.34e-03	6.23e-06	1.31e-04

(b) short selling is allowed

n	vs. CVaR	vs. abs	vs. sqr
20	1.86e-05	5.40e-05	6.86e-04
40	4.48e-06	6.65e-04	9.98e-07
60	6.03e-08	1.30e-08	1.57e-08
80	1.05e-12	2.45e-10	1.31e-10
100	< 2.2e-16	< 2.2e-16	< 2.2e-16
120	< 2.2e-16	< 2.2e-16	< 2.2e-16

(ii) weekly data

(a) short selling is not allowed

n	vs. CVaR	vs. abs	vs. sqr
20	7.37e-03	5.18e-05	9.57e-03
40	2.40e-03	0.03443	0.01634
60	1.93e-04	2.30e-03	0.06016
80	1.15e-08	5.48e-08	2.64e-05
100	2.42e-05	1.94e-05	2.04e-05
120	5.50e-10	5.86e-05	5.38e-08
140	1.57e-06	2.30e-08	1.51e-05
160	3.08e-08	5.12e-08	7.54e-10
180	7.54e-07	1.92e-04	5.18e-07

(b) short selling is allowed

n	vs. CVaR	vs. abs	vs. sqr
20	8.93e-05	2.06e-06	1.18e-04
40	2.16e-09	1.86e-06	2.11e-04
60	< 2.2e-16	< 2.2e-16	2.52e-16
80	< 2.2e-16	< 2.2e-16	< 2.2e-16

"CVaR" stands for the CVaR deviation without the norm constraint, "abs" stands for the mean absolute deviation model (12), and "sqr" stands for the squared deviation model (11). "< 2.2e-16" denotes the p-value is less than 2.2e-16.

When $\alpha_\beta^T(\boldsymbol{\pi}) < \theta$, we regard $(\theta - \alpha_\beta^T(\boldsymbol{\pi}))/\|\tilde{\boldsymbol{\pi}}\|_2$ as γ in (21), *i.e.,* the threshold for training error. Then we get (7), that is,

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

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

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

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

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

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 (21) 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

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

where $(\alpha_\beta^T(\boldsymbol{\pi}) - \theta)/\|\tilde{\boldsymbol{\pi}}\|_2 > 0$ corresponds to γ in (21). Then we get (8) 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

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

□

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, 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 (21) 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 training errors of $\tilde{\boldsymbol{R}}^{(i)\top} \tilde{\boldsymbol{\pi}}/\|\tilde{\boldsymbol{\pi}}\|_2$, $i = 1, 2$. Noticing that $\|\tilde{\boldsymbol{\pi}}\|_2 \leq \sqrt{C^2+1}$ and $\|\tilde{\boldsymbol{R}}^{(i)}\|_2 \leq B_R + \theta$, (17) follows. Moreover, we can prove (18) and (19) similarly to the proofs of Theorem 1 and Corollary 1, respectively. □

Acknowledgment

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

References

Arthanari, T.S., Dodge, Y. 1981. Mathematical Programming in Statistics. John Wiley & Sons, Inc.

Artzner, P., F. Delbaen, J.M. Eber, D. Heath. 1999. Coherent Measures of Risk. *Mathematical Finance* **9** 203–228.

Bartlett, P.L. 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., A. Nemirovski. 2000. Robust solutions of linear programming problems contaminated with uncertain data. *Math. Programming* **88** 411–424.

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

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

Chopra, V.K., W.T. Ziemba. 1993. The Effect of Errors in Means, Variance, and Covariance on Optimal Portfolio Choice. *Journal of Portfolio Management* **19** 6–11.

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

DeMiguel, V., L. Garlappi, F.J. Nogales, R. Uppal. 2007. A Generalized Approach to Portfolio Optimization: Improving Performance By Constraining Portfolio Norms. London Business School.

Gilli, M., E. Këllezi. 2002. The threshold accepting heuristic for index tracking, *Financial Engineering, E-Commerce, and Supply Chain*, eds. P.Pardalos and V.K.Tsitsiringos, Kluwer Applied Optimization Series, 1–18.

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

Gotoh, J., A. Takeda. 2008. Portfolio Learning via VaR/CVaR Minimization, Department of Industrial and Systems Engineering Discussion Paper Series ISE08-04, Chuo University.

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

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

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

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

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

Ledoit, O., M. Wolf. 2004. Honey, I shrunk the sample covariance matrix. *Journal of Portfolio Management* **30** 110–119.

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

Prigent, Jean-Luc. 2007. Portfolio Optimization and Performance Analysis. Chapman & Hall/CRC.

Rockafellar, T.R., S. Uryasev. 2002. Conditional Value-at-Risk for General Loss Distributions. *Journal of Banking and Finance* **26** 1443–1471.

Schölkopf, B., A.J. Smola, R.C. Williamson, P.L. Bartlett. 2000. New support vector algorithms. *Neural Computation* **12** 1207–1245.

Soyster, A. L. 1973. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research* **21** 1154–1157.

Takeda, A. 2007. Generalization Performance of nu-Support Vector Classifier Based on Conditional Value-at-Risk Minimization. to appear in *Neurocomputing*.

Williamson, R.C., A.J. Smola, B. Schölkopf. 2001. Generalization Performance of Regularization Networks and Support Vector Machines via Entropy Numbers of Compact Operators. *IEEE Transactions on Information Theory* **47** 2516–2532.