



# Understanding Distributional Ambiguity via Non-robust Chance Constraint

Shumin Ma  
City University of Hong Kong  
shuminma@cityu.edu.hk

Cheuk Hang Leung  
City University of Hong Kong  
chleung87@cityu.edu.hk

Qi Wu\*  
City University of Hong Kong  
qiwu55@cityu.edu.hk

Wei Liu  
Tencent  
wl2223@columbia.edu

Nanbo Peng  
JD Digits  
pengnanbo@jd.com

## ABSTRACT

This paper provides a non-robust interpretation of the distributionally robust optimization (DRO) problem by relating the distributional uncertainties to the chance probabilities. Our analysis allows a decision-maker to interpret the size of the ambiguity set, which is often lack of business meaning, through the chance parameters constraining the objective function. We first show that, for general  $\phi$ -divergences, a DRO problem is asymptotically equivalent to a class of mean-deviation problems. These mean-deviation problems are not subject to uncertain distributions, and the ambiguity radius in the original DRO problem now plays the role of controlling the risk preference of the decision-maker. We then demonstrate that a DRO problem can be cast as a chance-constrained optimization (CCO) problem when a boundedness constraint is added to the decision variables. Without the boundedness constraint, the CCO problem is shown to perform uniformly better than the DRO problem, irrespective of the radius of the ambiguity set, the choice of the divergence measure, or the tail heaviness of the center distribution. Thanks to our high-order expansion result, a notable feature of our analysis is that it applies to divergence measures that accommodate well heavy tail distributions such as the student  $t$ -distribution and the lognormal distribution, besides the widely-used Kullback-Leibler (KL) divergence, which requires the distribution of the objective function to be exponentially bounded. Using the portfolio selection problem as an example, our comprehensive testings on multivariate heavy-tail datasets, both synthetic and real-world, shows that this business-interpretation approach is indeed useful and insightful.

## CCS CONCEPTS

• **Applied computing** → **Multi-criterion optimization and decision-making.**

\*The corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](https://permissions.acm.org).

ICAIF '20, October 15–16, 2020, New York, NY, USA

© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-7584-9/20/10...\$15.00

<https://doi.org/10.1145/3383455.3422522>

## KEYWORDS

Distributionally robust optimization, Chance constraint, KL divergence,  $\phi$ -divergence, Heavy-tail distributions, Portfolio selection

## ACM Reference Format:

Shumin Ma, Cheuk Hang Leung, Qi Wu, Wei Liu, and Nanbo Peng. 2020. Understanding Distributional Ambiguity via Non-robust Chance Constraint. In *ACM International Conference on AI in Finance (ICAIF '20)*, October 15–16, 2020, New York, NY, USA. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3383455.3422522>

## 1 INTRODUCTION

Stochastic optimization is widely used in many machine learning algorithms to optimize the expected performance or loss, e.g., the mean squared error for regressions, or the expected discounted return in the context of reinforcement learning [19]. A sound machine learning model demands reliable estimates of the data-generating distribution. However, uncertainties of the data distribution could arise in many ways: limited observations in the stationary case, time-varying law in the non-stationary case, or the law is subject to policy intervention due to the treatment effect. In robust statistics, formulating a decision-making problem as a DRO problem is a remedy to address the distributional uncertainties in the data [6].

A typical DRO formulation adds an extra layer of optimization over a set of possible distributions, called the *ambiguity set*, and optimizes the decision variables in the worst-case distribution. There are mainly three ways in the literature to define the ambiguity set. The first is the geometric approach, which allows the parameters of the chosen distribution to vary within certain geometric regions [14, 20, 21] such as boxes, ellipsoids, and polyhedrons, etc. The second approach, known as the *moment-based approach*, constructs the ambiguity set by collecting distributions that share the same moment constraints [5, 7, 18, 22]. The last one, the *statistical distance approach*, uses divergence measures or difference functions between two probability distributions to define the ambiguity set as a ball of distributions [1, 6, 8, 16]. The radius of the ball is referred to as the ambiguity radius. Among the three, the moment-based and the statistical distance approaches address law uncertainties. In contrast, the geometric approach only addresses the uncertainties in the parameters of a *a priori* fixed distribution, not in its functional form. It does not help if the correct distribution turns out to be lognormal when you think it is instead normal and fine-tune its mean and variance. However, the cost of advancing from parameter uncertainty to law uncertainty is that you lose the interpretability

of the ambiguity set because the parameters characterizing it are non-business quantities.

This paper provides a solution to address this business-interpretation problem. For business applications, a decision-maker would have a hard time relating, e.g., the radius 0.01 of a KL ball to e.g., product sales, taxi demands, or portfolio returns. The radius 0.01 is not related to any measures of the business objective. An unavoidable headache for her is how she should decide the size of the ambiguity set. Our idea is straightforward. We want to translate the impact of the ambiguity radius, which lacks business meanings, to the impact of the chance parameters constraining the objective function, which now allows a decision-maker to enter her preferences directly related to the business objective. Take asset allocation as an example, our solution can tell a portfolio manager that setting the ambiguity radius to 0.01 would be equivalent to asking the optimization not to let the chances of her portfolio return going below  $-13\%$  be higher than  $2\%$ . In this way, the geometry of the ambiguity set, its radius, is connected directly to her granular preference of the objective, the amount of risks she can tolerate.

This paper makes two primary technical contributions. First, our analysis applies to the heavy-tail distributions (e.g., via the Cressie-Read divergence) [10], besides the usual light-tail cases using the KL divergence. Heavy-tail distributions, e.g., the lognormal distribution and the student  $t$ -distribution, are ubiquitous for many business and finance datasets. A DRO problem with an ambiguity set defined by the KL divergence is solvable, however, only when the distribution of the objective function is exponentially bounded [13], in which case heavy-tail distributions are excluded. Our analysis extends well to the general  $\phi$ -divergence family, including KL divergence, Burg entropy,  $\chi^2$ -distance, Hellinger distance, Cressie-Read divergence, etc. [12, 16]. The second contribution of this paper is that we establish two connections between a DRO problem and a CCO problem. The first one is that when a bounded constraint is added to decision variables, a DRO problem can be cast as a CCO problem without distributional uncertainties. The second connection is that, without the boundedness constraint, the CCO problem is shown to perform uniformly better than the DRO problem, irrespective of the radius of the ambiguity set, the choice of the divergence measure, or the tail heaviness of the center distribution.

The rest of the paper is organized as follows. In Section 2, we provide some background information and the motivation for the proposed optimization problems. Theoretical analysis of the DRO problem and the CCO problem is provided in Section 3. Section 4 establishes the connection between the DRO problem and the CCO problem under an explicit formulation of the portfolio selection problem. Section 5 gives numerical experiments, and Section 6 concludes our findings from both synthetic and empirical data. Due to the page limits, all the proofs are omitted in the main body; however, they can be readily provided once requested.

## 2 PROBLEM SETUP

### 2.1 Notations

Let  $\mathbf{r} \in \mathbb{R}^n$ , an  $n$ -dimensional real-valued random vector, be the vector of asset returns. And suppose the joint probability distribution of  $\mathbf{r}$  is  $\mathbb{P}$ . Let  $\mathbb{P}_0$  be the nominal probability distribution of  $\mathbf{r}$ . Let  $\mathbf{x} \in \mathbb{R}^n$  be the asset allocation strategy, and  $\mathbf{e} \in \mathbb{R}^n$  be a vector with

**Table 1: The two  $\phi$ -divergences used in this paper. The KL divergence applies to light-tail distributions, while the Cressie-Read divergence is compatible with heavy-tail distributions.**

	Kullback-Leibler	Cressie-Read
$\phi(t)$	$t \log(t) - t + 1$	$\frac{1-\theta+t-t^\theta}{\theta(1-\theta)}, \theta \neq 0, 1$
$\phi^*(s)$	$e^s - 1$	$\frac{(1-s(1-\theta))^{\frac{\theta}{\theta-1}}}{\theta} - \frac{1}{\theta}, s < \frac{1}{1-\theta}$

all entries equal to 1. Denote the utility function that is concave in  $\mathbf{x}$  and associated with  $\mathbf{x}$  and  $\mathbf{r}$  by  $f(\mathbf{x}, \mathbf{r})$ . We assume that  $\mathbf{x}$  lies in a convex set  $\mathcal{X}$  and  $\mathbb{P}$  belongs to an ambiguity set  $\mathcal{U}$ . The expectation and variance of a random variable under  $\mathbb{P}$  are represented by  $\mathbb{E}_{\mathbb{P}}[\cdot]$  and  $\mathbb{V}_{\mathbb{P}}[\cdot]$ , respectively.

**Definition 2.1.** ( $\phi$ -divergence) Assume that  $\phi(t)$  is convex for  $t \geq 0$  and that  $\phi(1) = 0$ . Then the  $\phi$ -divergence  $D(\mathbb{Q}||\mathbb{P})$  between distribution  $\mathbb{P}$  and distribution  $\mathbb{Q}$  is defined as:

$$D(\mathbb{Q}||\mathbb{P}) := \int \phi\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right) d\mathbb{P} = \mathbb{E}_{\mathbb{P}}\left[\phi\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)\right] := \mathbb{E}_{\mathbb{P}}[\phi(L)]. \quad (1)$$

The quantity  $L$  in Eq. (1) is called the Radon Nikodym derivative (or likelihood ratio) such that  $L \geq 0$  almost surely and  $\mathbb{E}_{\mathbb{P}}[L] = 1$ . Notice that, for the Radon-Nikodym derivative  $L$  to exist,  $\mathbb{Q}$  must be absolutely continuous w.r.t.  $\mathbb{P}$ . Given the function  $\phi$  for a specific  $\phi$ -divergence, its *conjugate*  $\phi^*$  is defined as  $\phi^*(s) := \sup_{t \geq 0} \{st - \phi(t)\}$ . Table 1 lists the two divergences used in this paper. But it should be mentioned that, our interpretation of the ambiguity radius actually applies to all the  $\phi$ -divergences, including Burg entropy,  $J$ -divergence,  $\chi^2$ -distance, modified  $\chi^2$ -distance, and Hellinger distance. (For more information about the  $\phi$ -divergence family, see [2]).

### 2.2 Motivation

The goal is to maximize the expected utility over a set of admissible allocation strategies  $\mathcal{X}$ , namely,

$$\max_{\mathbf{x} \in \mathcal{X}} \mathbb{E}_{\mathbb{P}}[f(\mathbf{x}, \mathbf{r})]. \quad (2)$$

We introduce the ambiguity set  $\mathcal{U}$  centered at the nominal distribution  $\mathbb{P}_0$  (also called the center distribution in the following context) and controlled by the radius parameter  $\rho > 0$ , that is,  $\mathcal{U} := \{\mathbb{P} : D(\mathbb{P}||\mathbb{P}_0) \leq \rho\}$ . Thus, the distributionally robust counterpart of problem (2) is:

$$\max_{\mathbf{x} \in \mathcal{X}} \min_{\mathbb{P} \in \mathcal{U}} \mathbb{E}_{\mathbb{P}}[f(\mathbf{x}, \mathbf{r})]. \quad (3)$$

For a decision-maker, the ambiguity radius  $\rho$  is critical. One cannot set it too large since the optimal utility decreases in  $\rho$ . However, if it is too small, one loses the robust protection. There is a trade-off in choosing its magnitude in the financial context. In literature, [17] presents the characteristics of the  $\phi$ -divergence between the true distribution  $\mathbb{P}$  and the nominal distribution  $\mathbb{P}_0$ ,  $D(\mathbb{P}||\mathbb{P}_0)$ . Assuming that  $\mathbb{P}$  and  $\mathbb{P}_0$  belong to the same parameterized distribution family with parameter dimension  $d$ , and that  $\phi$  is twice continuously differentiable in a neighborhood of 1 with  $\phi^{(2)}(1) > 0$ , the normalized estimated  $\phi$ -divergence  $\frac{2N}{\phi^{(2)}(1)} D(\mathbb{P}||\mathbb{P}_0)$  asymptotically

(i.e., for the sample size  $N \rightarrow \infty$ ) follows a  $\chi_d^2$ -distribution. This conclusion thus relates the ambiguity radius  $\rho$  to a confidence level at which the true distribution  $\mathbb{P}$  falls within the ambiguity set. [3] provides one methodology, under the Markowitz's mean-variance portfolio selection framework, to select the ambiguity radius  $\rho$  as the smallest radius such that the true asset allocation strategy is included with a given confidence level.

However, in financial practice with real data, the assumption that the true distribution is in the same parameterized family with the center distribution is too strong. A wrong guess of the nominal distribution may lead to a meaningless confidence level interpretation of the ambiguity radius  $\rho$ . Since the DRO approach is believed to provide robust protection against distributional uncertainty, we are motivated to connect the robust protection to protection provided by traditional risk measures. In particular, the heavy-tail nature of distributions that we are concerned with reminds us of the tail probability protection, the optimization based on which is known as CCO problems. Specifically, we define the CCO problem as:

$$\max_{\mathbf{x} \in \mathcal{X}} \mathbb{E}_{\mathbb{P}_0}[f(\mathbf{x}, \mathbf{r})] \quad \text{s.t.} \quad \Pr_{\mathbf{r} \sim \mathbb{P}_0}(\mathbf{x}^T \mathbf{r} \leq -\delta) \leq \epsilon. \quad (4)$$

Here,  $\delta > 0$  characterizes a typical investor's loss threshold and  $\epsilon > 0$  corresponds to the loss probability. The CCO problem in problem (4) shares the same objective function as that of problem (2). The expectation is taken under the nominal distribution  $\mathbb{P}_0$ , not subject to any distributional robustness (the term "non-robust" in the title originates from here). Compared to problem (2), the new component is the chance constraint with parameters  $(\delta, \epsilon)$  characterizing an investor's tolerance to losses.

We would build up a performance-based interpretation of the ambiguity radius  $\rho$  through the parameters of the chance-constrained problem. To be specific, if under some ambiguity radius  $\rho$  and chance constraint parameters  $(\delta, \epsilon)$ , problem (3) and problem (4) achieve the same optimal value, we would say that the robust protection under the ambiguity radius  $\rho$  is similar to that of a tail probability protection. In addition, we would also look into how the choice of the allocation strategy set  $\mathcal{X}$  and the tail heaviness of the nominal distribution  $\mathbb{P}_0$  affect the interpretation of the ambiguity radius  $\rho$ , given that  $\mathcal{X}$  and  $\mathbb{P}_0$  are the shared model settings of the two problems (3) and (4).

### 3 ANALYSIS OF DRO AND CCO PROBLEMS

This section is devoted to the theoretical analysis of problems (3) and (4). We show that, for general  $\phi$ -divergences, problem (3) can be reformulated as a class of mean-deviation problems with the investor's risk preference parameter controlled by the ambiguity radius  $\rho$ . Besides, we provide an approximation framework to solve problem (4).

#### 3.1 Reformulation of the DRO problem (3)

Consider the inner optimization problem in problem (3):

$$\min_{\mathbb{P} \in \mathcal{U}} \mathbb{E}_{\mathbb{P}}[f(\mathbf{x}, \mathbf{r})]. \quad (5)$$

The Lagrangian dual to problem (5) is:

$$\begin{aligned} & \sup_{\eta_1 \in \mathbb{R}, \eta_2 \geq 0} \left\{ -\frac{1}{\eta_2} \sup_L \left\{ \mathbb{E}_{\mathbb{P}_0}[-\eta_2(f(\mathbf{x}, \mathbf{r}) + \eta_1)L - \phi(L)] \right\} - \eta_1 - \frac{\rho}{\eta_2} \right\} \\ &= \sup_{\eta_1 \in \mathbb{R}, \eta_2 \geq 0} \left\{ -\frac{1}{\eta_2} \mathbb{E}_{\mathbb{P}_0}[\phi^*(-\eta_2(f(\mathbf{x}, \mathbf{r}) + \eta_1))] - \eta_1 - \frac{\rho}{\eta_2} \right\}. \end{aligned}$$

The last equality is derived directly from the definition of the conjugate function of  $\phi$ -divergence. Difficulty in solving the dual problem lies in the term  $\mathbb{E}_{\mathbb{P}_0}[\phi^*(-\eta_2(f(\mathbf{x}, \mathbf{r}) + \eta_1))]$ . We hereby follow the idea in [11] to express optimization (5) in terms of *Regular Measure of Deviation*, with results summarized in Theorem 3.1.

**THEOREM 3.1.** *Let  $\phi$  be a closed proper convex function and  $\phi^*$  be its corresponding conjugate function, respectively. Suppose that under mild conditions, the strong duality holds. Define the regular measure of deviation*

$$\begin{aligned} & \mathcal{D}_{\eta_2, \phi, \mathbb{P}_0}(f(\mathbf{x}, \mathbf{r}) | \mathbb{E}_{\mathbb{P}_0}[f(\mathbf{x}, \mathbf{r})]) \\ &:= \inf_{\eta_1} \left\{ \eta_1 + \frac{1}{\eta_2} \mathbb{E}_{\mathbb{P}_0}[\phi^*(\eta_2(\mathbb{E}_{\mathbb{P}_0}[f(\mathbf{x}, \mathbf{r})] - f(\mathbf{x}, \mathbf{r}) - \eta_1))] \right\}. \end{aligned}$$

Then, optimization (5) is equivalent to :

$$\mathbb{E}_{\mathbb{P}_0}[f(\mathbf{x}, \mathbf{r})] - \inf_{\eta_2 \geq 0} \left\{ \frac{\rho}{\eta_2} + \mathcal{D}_{\eta_2, \phi, \mathbb{P}_0}(f(\mathbf{x}, \mathbf{r}) | \mathbb{E}_{\mathbb{P}_0}[f(\mathbf{x}, \mathbf{r})]) \right\}.$$

Furthermore, the quantity  $\mathcal{D}_{\eta_2, \phi, \mathbb{P}_0}(f(\mathbf{x}, \mathbf{r}) | \mathbb{E}_{\mathbb{P}_0}[f(\mathbf{x}, \mathbf{r})])$  can be expanded as a series of terms, the coefficients of which can be computed under the nominal distribution  $\mathbb{P}_0$ . By doing so, we can reformulate the DRO problem (3) as a single-layer maximization problem.

**LEMMA 3.2.** *Suppose that  $K$  is an even number,  $\phi \in C^{K+1}$  is a convex function which satisfies  $\phi(1) = \phi^{(1)}(1) = 0$  and  $\phi^{(2)}(1) > 0$ . Assume that  $\mathbb{E}_{\mathbb{P}_0}[X^k] < \infty$  for  $k \leq K$  and  $X$  is defined as  $X := f(\mathbf{x}, \mathbf{r}) - \mathbb{E}_{\mathbb{P}_0}[f(\mathbf{x}, \mathbf{r})]$ . Then*

$$\begin{aligned} & \mathcal{D}_{\eta_2, \phi, \mathbb{P}_0}(f(\mathbf{x}, \mathbf{r}) | \mathbb{E}_{\mathbb{P}_0}[f(\mathbf{x}, \mathbf{r})]) \\ &= \sum_{k=1}^{K-1} b_k \mathbb{E}_{\mathbb{P}_0}[(X + \eta_1^*)^{k+1}] \eta_2^k + o(\eta_2^{K-1}), \end{aligned} \quad (6)$$

where  $b_k = \frac{(-1)^{k+1} z^{(k)}(0)}{(k+1)!}$ , and  $\eta_1^*$  is the optimal solution to

$$\min_{\eta_1} \sum_{k=1}^{K-1} b_k \mathbb{E}_{\mathbb{P}_0}[(X + \eta_1)^{k+1}] \eta_2^k.$$

Specifically,  $z(\cdot)$  is a function satisfying  $z(0) = 1$ ,  $z^{(1)}(\cdot) = \frac{1}{\phi^{(2)}(z(\cdot))}$  and  $z^{(k)}(\cdot)$  can be obtained recursively for  $k \geq 2$ .

Note that the above expansion applies to general utility functions  $f(\mathbf{x}, \mathbf{r})$  that are concave in  $\mathbf{x}$ . More importantly, most of the  $\phi$ -divergences (KL divergence, Cressie-Read divergence, Burg entropy,  $J$ -divergence,  $\chi^2$ -distance, modified  $\chi^2$ -distance, and Hellinger distance) satisfy the smoothness conditions. Taking KL and Cressie-Read divergence as example, for  $K = 4$ , we can explicitly solve the terms in Eq. (6), as are shown in the following corollary.

COROLLARY 3.3. Consider  $K = 4$ . We have the 4<sup>th</sup> order expansion of  $\mathcal{D}_{\eta_2, \phi, \mathbb{P}_0}(f(\mathbf{x}, \mathbf{r}) | \mathbb{E}_{\mathbb{P}_0}[f(\mathbf{x}, \mathbf{r})])$ :

$$\begin{aligned} & \mathcal{D}_{\eta_2, \phi, \mathbb{P}_0}(f(\mathbf{x}, \mathbf{r}) | \mathbb{E}_{\mathbb{P}_0}[f(\mathbf{x}, \mathbf{r})]) \\ &= \sum_{k=1}^3 b_k \mathbb{E}_{\mathbb{P}_0} \left[ (X + \eta_1^*)^{k+1} \right] \eta_2^k + o(\eta_2^3), \end{aligned} \quad (7)$$

with  $\eta_1^*$  being the real root to the 3<sup>rd</sup> order equation

$$\begin{aligned} & \sum_{k=1}^3 (k+1) b_k \eta_2^k \cdot \eta_1^k + 12 b_3 \eta_2^3 \mathbb{E}_{\mathbb{P}_0}[X^2] \cdot \eta_1 \\ &+ 4 b_3 \eta_2^3 \mathbb{E}_{\mathbb{P}_0}[X^3] + 3 b_2 \eta_2^2 \mathbb{E}_{\mathbb{P}_0}[X^2] = 0. \end{aligned}$$

For KL divergence, the coefficients are  $b_1 = 1/2$ ,  $b_2 = -1/6$ ,  $b_3 = 1/24$ ; for Cressie-Read divergence with  $\theta > 2$ , the coefficients are  $b_1 = 1/2$ ,  $b_2 = (\theta - 2)/6$ ,  $b_3 = (\theta - 2)(2\theta - 3)/24$ .

[11] gives a similar expansion of  $\mathcal{D}_{\eta_2, \phi, \mathbb{P}_0}(f(\mathbf{x}, \mathbf{r}) | \mathbb{E}_{\mathbb{P}_0}[f(\mathbf{x}, \mathbf{r})])$  in Proposition 3.5. The main difference between our expansion in Eq. (7) and their expansion lies in the calculation of  $\eta_1^*$ . In Eq. (7),  $\eta_1^*$  is directly solved through the polynomial equation, while in [11],  $\eta_1^*$  is an approximated function of  $\eta_2$ .

In the sequel, we take  $K = 2$ , consider the 2<sup>nd</sup> order expansion of  $\mathcal{D}_{\eta_2, \phi, \mathbb{P}_0}(\mathbf{x}^T \mathbf{r} | \mathbf{x}^T \mu)$  and ignore the higher order terms, which gives

$$\begin{aligned} & \min_{\mathbb{P} \in \mathcal{U}} \mathbb{E}_{\mathbb{P}}[f(\mathbf{x}, \mathbf{r})] \\ & \approx \mathbb{E}_{\mathbb{P}_0}[f(\mathbf{x}, \mathbf{r})] - \inf_{\eta_2 \geq 0} \left\{ \frac{\rho}{\eta_2} + \frac{\eta_2}{2\phi^{(2)}(1)} \mathbb{V}_{\mathbb{P}_0}[f(\mathbf{x}, \mathbf{r})] \right\} \\ &= \mathbb{E}_{\mathbb{P}_0}[f(\mathbf{x}, \mathbf{r})] - \sqrt{\frac{2\rho}{\phi^{(2)}(1)} \mathbb{V}_{\mathbb{P}_0}[f(\mathbf{x}, \mathbf{r})]}. \end{aligned}$$

The last equality comes as a result of

$$\inf_{\eta_2 \geq 0} \left\{ \frac{\rho}{\eta_2} + \frac{\eta_2}{2\phi^{(2)}(1)} \mathbb{V}_{\mathbb{P}_0}[f(\mathbf{x}, \mathbf{r})] \right\} = \sqrt{\frac{2\rho \mathbb{V}_{\mathbb{P}_0}[f(\mathbf{x}, \mathbf{r})]}{\phi^{(2)}(1)}},$$

and the minimum is achieved at  $\eta_2 = \sqrt{\frac{2\rho \phi^{(2)}(1)}{\mathbb{V}_{\mathbb{P}_0}[f(\mathbf{x}, \mathbf{r})]}}$ . This suggests, when  $\rho$  is small, the optimal Lagrangian multiplier  $\eta_2$  is also small and the expansion in (6) is accurate. By taking  $\max_{\mathbf{x} \in \mathcal{X}}$  on both sides, we finally achieve the 2<sup>nd</sup> order reformulation of problem (3) in Theorem 3.4.

THEOREM 3.4. Suppose that  $\phi$  is convex, twice continuously differentiable, and that  $\phi(1) = \phi^{(1)}(1) = 0$  and  $\phi^{(2)}(1) > 0$ . The DRO problem in problem (3) is asymptotically equivalent to a mean-deviation problem:

$$\max_{\mathbf{x} \in \mathcal{X}} \left\{ \mathbb{E}_{\mathbb{P}_0}[f(\mathbf{x}, \mathbf{r})] - \sqrt{\frac{2\rho \mathbb{V}_{\mathbb{P}_0}[f(\mathbf{x}, \mathbf{r})]}{\phi^{(2)}(1)}} \right\}. \quad (8)$$

Theorem 3.4 tells that the ambiguity radius  $\rho$  actually controls the investor's risk preference.

### 3.2 Reformulation of the CCO problem (4)

Notice that, the chance constraint in problem (4) is in the same form as the definition of Value-at-Risk (VaR), a risk measure that focuses on the probability of losses. This motivates us to reorganize the tail

chance constraint in problem (4) with VaR. The VaR is defined as the minimal level  $\gamma$  such that the probability that the portfolio loss  $-\mathbf{x}^T \mathbf{r}$  exceeds  $\gamma$  is below  $\epsilon$ :

$$V_\epsilon(\mathbf{x}) := \inf\{\gamma \in \mathbb{R} : \Pr_{\sim \mathbb{P}_0}\{-\mathbf{x}^T \mathbf{r} \geq \gamma\} \leq \epsilon\}.$$

The equivalent form of the chance constraint in problem (4):  $\Pr_{\sim \mathbb{P}_0}\{-\mathbf{x}^T \mathbf{r} \geq \delta\} \leq \epsilon$  implies that,  $\delta$  is included in the set  $\{\gamma \in \mathbb{R} : \Pr_{\sim \mathbb{P}_0}\{-\mathbf{x}^T \mathbf{r} \geq \gamma\} \leq \epsilon\}$ . That is to say, the chance constraint can be reorganized with  $V_\epsilon(\mathbf{x})$ , namely,

$$\Pr_{\sim \mathbb{P}_0}\{-\mathbf{x}^T \mathbf{r} \geq \delta\} \leq \epsilon \Leftrightarrow V_\epsilon(\mathbf{x}) \leq \delta.$$

Hence, given  $\mathbb{E}_{\mathbb{P}_0}[\mathbf{x}^T \mathbf{r}] = \mathbf{x}^T \mu$ , problem (4) can be reformulated as

$$\max_{\mathbf{x} \in \mathcal{X}} \mathbf{x}^T \mu \quad \text{s.t.} \quad V_\epsilon(\mathbf{x}) \leq \delta.$$

If  $\mathbb{P}_0$  is normal, then the VaR can be expressed as

$$V_\epsilon(\mathbf{x}) = \kappa(\epsilon) \sqrt{\mathbf{x}^T \Sigma \mathbf{x}} - \mathbf{x}^T \mu,$$

where  $\kappa(\epsilon) = -\Phi^{-1}(\epsilon)$  and  $\Phi^{-1}(\cdot)$  is the inverse of the cumulative distribution function of the standard normal distribution. If  $\mathbb{P}_0$  is a member of general elliptical distribution family, [15] gives an asymptotic expansion of  $V_\epsilon(\mathbf{x})$ , which takes the form  $\kappa(\epsilon) \sqrt{\mathbf{x}^T \Sigma \mathbf{x}} - \mathbf{x}^T \mu$  asymptotically when  $\epsilon \rightarrow 0$ . For example, if  $\mathbb{P}_0$  is a student  $t$ -distribution with degree of freedom parameter  $\nu$ , then

$$\kappa(\epsilon) = D\epsilon^{-\frac{1}{\nu}}, \quad \text{where } D = \left( \frac{c_n \pi^{\frac{n-1}{2}} \Gamma(\frac{\nu+1}{2})}{\nu \Gamma(\frac{\nu+n}{2})} \right)^{\frac{1}{\nu}}, \quad c_n = \frac{\Gamma(\frac{\nu+n}{2})}{\Gamma(\frac{\nu}{2})} \nu^{\frac{\nu}{2}} \pi^{-\frac{n}{2}},$$

and  $\Gamma(\cdot)$  refers to the gamma function. For distributions other than elliptical distributions,  $\sqrt{\frac{1-\epsilon}{\epsilon}} \sqrt{\mathbf{x}^T \Sigma \mathbf{x}} - \mathbf{x}^T \mu$  is proved to be a valid approximation of  $V_\epsilon(\mathbf{x})$  [4, 9]. These in all provide the approximation of problem (4) reformulated as

$$\max_{\mathbf{x} \in \mathcal{X}} \mathbb{E}_{\mathbb{P}_0}[f(\mathbf{x}, \mathbf{r})] \quad \text{s.t.} \quad \kappa(\epsilon) \sqrt{\mathbf{x}^T \Sigma \mathbf{x}} - \mathbf{x}^T \mu \leq \delta. \quad (9)$$

With the following lemma, we can verify that problem (9) is a convex optimization when  $\kappa(\epsilon) > 0$ . For general feasibility set  $\mathcal{X}$ , problem (9) can always be efficiently solved with second-order cone programming (SOCP).

LEMMA 3.5. Suppose  $a > 0$ . Then the function  $a\sqrt{\mathbf{x}^T \Sigma \mathbf{x}} - \mathbf{x}^T \mu$  is a convex function of  $\mathbf{x}$ .

## 4 EXPLICIT FORMULATIONS OF PORTFOLIO SELECTION

In this section, we propose the explicit formulations for portfolio selection problem with  $f(\mathbf{x}, \mathbf{r}) = \mathbf{x}^T \mathbf{r}$ . It only remains to explicitly specify the set  $\mathcal{X}$ . We begin with the most simple but basic unbounded set  $\mathcal{X} := \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x}^T \mathbf{e} = 1\}$ . We would denote the optimal solution and optimal value to optimization (8) by  $\mathbf{x}^*$  and  $v^*$ , respectively. The corresponding optimal solution and optimal value to optimization (9) are denoted by  $\tilde{\mathbf{x}}^*$  and  $\tilde{v}^*$ , respectively. Throughout the rest of the paper, we would denote  $\mathbb{E}_{\mathbb{P}_0}[\mathbf{r}]$  and covariance matrix of  $\mathbf{r}$  under  $\mathbb{P}_0$  by  $\mu$  and  $\Sigma$ , respectively. Then naturally,  $\mathbb{E}_{\mathbb{P}_0}[\mathbf{x}^T \mathbf{r}] = \mathbf{x}^T \mu$ ,  $\mathbb{V}_{\mathbb{P}_0}[\mathbf{x}^T \mathbf{r}] = \mathbf{x}^T \Sigma \mathbf{x}$ ,  $v^* = \mathbf{x}^{*T} \mu - \sqrt{\frac{2\rho \mathbf{x}^{*T} \Sigma \mathbf{x}^*}{\phi^{(2)}(1)}}$

and  $\tilde{v}^* = \tilde{\mathbf{x}}^{*T} \mu$ .

Recall that in a convex optimization, any local optimum is also a global optimum. This motivates us to study the optimal solution

to problem (8),  $\mathbf{x}^*$ , and the optimal solution to problem (9),  $\tilde{\mathbf{x}}^*$ , through KKT conditions. The results for  $(\mathbf{x}^*, v^*)$  and  $(\tilde{\mathbf{x}}^*, \tilde{v}^*)$  are summarized in Theorem 4.1 and Theorem 4.2, respectively.

**THEOREM 4.1.** *Suppose  $\phi^{(2)}(1) > 0$ . Define  $A := \mathbf{e}^T \Sigma^{-1} \mathbf{e}$ ,  $B := \mu^T \Sigma^{-1} \mathbf{e}$ , and  $C := \mu^T \Sigma^{-1} \mu$ . Then problem (8) with  $f(\mathbf{x}, \mathbf{r}) = \mathbf{x}^T \mathbf{r}$  and the feasibility set  $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x}^T \mathbf{e} = 1\}$  has an optimal solution when  $\rho > \phi^{(2)}(1)(C - B^2/A)/2$ . And the optimal solution  $\mathbf{x}^*$  and optimal value  $v^*$  are:*

$$\mathbf{x}^* = \frac{\Sigma^{-1}(\mu - \lambda^* \mathbf{e})}{B - \lambda^* A}, \quad v^* = \lambda^*,$$

$$\text{where } \lambda^* = \frac{B}{A} - \frac{\sqrt{B^2 - A(C - 2\rho/\phi^{(2)}(1))}}{A}.$$

**THEOREM 4.2.** *Suppose  $\kappa(\epsilon) > 0$ . Problem (9) with  $f(\mathbf{x}, \mathbf{r}) = \mathbf{x}^T \mathbf{r}$  and the feasibility set  $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x}^T \mathbf{e} = 1\}$  has an optimal solution when  $(\epsilon, \delta)$  satisfies  $C - B^2/A < (\kappa(\epsilon))^2 < \delta^2 A + 2\delta B + C$  and  $B + \delta A > 0$ . ( $A$ ,  $B$ , and  $C$  defined in Theorem 4.1.) And the optimal solution  $\tilde{\mathbf{x}}^*$  and optimal value  $\tilde{v}^*$  are:*

$$\tilde{\mathbf{x}}^* = \frac{\Sigma^{-1}[(1 + \tilde{\lambda})\mu - \tilde{\theta}\mathbf{e}]}{(1 + \tilde{\lambda})B - \tilde{\theta}A}, \quad \tilde{v}^* = \tilde{\lambda}\delta + \tilde{\theta},$$

$$\text{where } \tilde{\lambda} = \frac{\sqrt{AC - B^2}}{A\kappa(\epsilon)^2 - AC + B^2} \left( \frac{\kappa(\epsilon)(B + A\delta)}{\sqrt{A\delta^2 + 2\delta B + C - \kappa(\epsilon)^2}} + \sqrt{AC - B^2} \right), \text{ and } \tilde{\theta} = \frac{(C + \delta B)(\tilde{\lambda} + 1) - \tilde{\lambda}\kappa(\epsilon)^2}{B + \delta A}.$$

Furthermore,  $\tilde{v}^* \geq v^*$ , i.e., problem (8) always outperforms problem (9).

In Theorem 4.2, we first identify the sufficient conditions of  $(\epsilon, \delta)$  for the optimization problem (9) to be feasible. The comparison between  $\tilde{v}^*$  and  $v^*$  shows that the CCO reformulation performs uniformly better than the DRO reformulation. Here it should be mentioned that, the outperformance of problem (8) over problem (9) is not so obvious. At the first glance, it does seem quite straightforward that the objective function in problem (9) is always smaller than that in problem (8). While in fact, rather than

comparing  $\mathbf{x}^T \mu - \sqrt{\frac{2\rho \mathbf{x}^T \Sigma \mathbf{x}}{\phi^{(2)}(1)}}$  and  $\mathbf{x}^T \mu$  based on the same asset allocation strategy  $\mathbf{x}$ , we are comparing the two objective functions based on their respective optimal asset allocation strategies, namely,  $\mathbf{x}^{*T} \mu - \sqrt{\frac{2\rho \mathbf{x}^{*T} \Sigma \mathbf{x}^*}{\phi^{(2)}(1)}}$  vs  $\tilde{\mathbf{x}}^{*T} \mu$ .

For more complex sets  $\mathcal{X}$ , we resort to numerical analysis to investigate interpretation of the ambiguity radius  $\rho$  through chance constraint parameters.

## 5 EXPERIMENTS

Sections 5.1 and 5.2 are based on synthetic data to test the reformulation accuracy of the DRO problem (3) and to see how the tail heaviness of the nominal distribution  $\mathbb{P}_0$  affects the interpretation of  $\rho$ . Section 5.3 is devoted to a more detailed understanding of the ambiguity radius  $\rho$  based on the empirical daily returns of 4 asset classes. And Section 5.4 uses intraday 5-minute stock returns to test the value of robust protection in the real portfolio selection problem.

### 5.1 Reformulation accuracy of problem (3)

In this section, we numerically test the accuracies of the 2<sup>nd</sup> order and the 4<sup>th</sup> order reformulations with respect to the original robust problem (3). The  $\phi$ -divergence we take is KL divergence, under which problem (3) can be exactly solved. And we take the exact optimal value as a benchmark to compare the 2<sup>nd</sup> order and the 4<sup>th</sup> order reformulations.

Table 2 records the relative errors (in the 3<sup>rd</sup> & 4<sup>th</sup> columns) w.r.t. the exact optimal value (the 2<sup>nd</sup> column) under KL divergence. It shows that the higher order improvement is particularly notable when data exhibits a heavier tail. In the case of Cressie-Read divergence, which we do not record in the table due to the page limit, we observe a 50 times improvement: when  $\rho$  is set to 0.78, relative error for the 4<sup>th</sup> order reformulation is 1.53%, while it is 56.54% for the 2<sup>nd</sup> order reformulation given that the optimal value is  $-0.2787$ . Here, we assume the ambiguity set under the KL divergence centers at a six-dimensional multivariate exponential distribution with mean=0.2, std=0.2, skewness=2, and kurtosis=6. We set the dimensions to be i.i.d to see a clean impact from the heavy tail. And the center distribution  $\mathbb{P}_0$  under Cressie-Read divergence is multivariate  $t$ . We see that the larger the size of the ambiguity set (i.e., larger  $\rho$ ), the better the improvement of the 4<sup>th</sup> order reformulation. In fact, the error reduction is about 10 folds in this example. However, using the 2<sup>nd</sup> order equivalent formulation is good enough to solve problem (3) when  $\rho$  is small.

**Table 2: Relative errors of the 4<sup>th</sup> order reformulation and 2<sup>nd</sup> order reformulation w.r.t. the optimal value of problem (3). Ambiguity sets are defined by KL divergence centered at a 6-d exponential distribution.**

	Optimal value	Relative errors	
		4 <sup>th</sup> order	2 <sup>nd</sup> order
$\rho = \mathbf{0.01}$	0.1887	0.0002%	0.1172%
$\mathbf{0.02}$	0.1841	0.0038%	0.2397%
$\mathbf{0.03}$	0.1807	0.0128 %	0.3659%
$\mathbf{0.04}$	0.1778	0.0274%	0.4951%
$\mathbf{0.05}$	0.1753	0.0479%	0.6270%
$\mathbf{0.06}$	0.1730	0.0748%	0.7613%
$\mathbf{0.07}$	0.1710	0.1082%	0.8979%
$\mathbf{0.08}$	0.1691	0.1483%	1.037%
$\mathbf{0.09}$	0.1673	0.1951%	1.778%

### 5.2 Interpretation of $\rho$ under distributions with different tail heavinesses

This experiment shows that tail heaviness of the nominal distribution  $\mathbb{P}_0$  indeed affects the interpretation of the ambiguity radius  $\rho$ . We focus on three distributions for 5 assets: multivariate normal, lognormal distribution and student  $t_3$ -distribution. The set of allocation strategies is bounded below by -1, and the ambiguity radius  $\rho$  is fixed at 0.27. We plot the results of equivalent  $(\epsilon, \delta)$  in Figure 1. It shows that, first, the ambiguity radius  $\rho$  can be explained by a set of pairs  $(\epsilon, \delta)$  in terms of the impact on the optimal value. Second, tail heaviness affects the interpretation of  $\rho$  and distributions

with heavier tail result in a larger loss threshold for a given loss probability  $\epsilon$ .

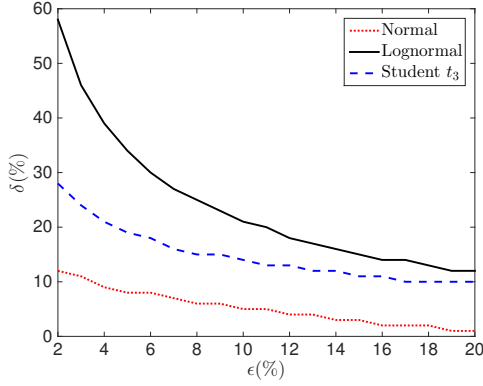


Figure 1: Given  $\rho = 0.27$ , tail heaviness affects the equivalent loss threshold  $\delta$ .

### 5.3 Empirical studies with daily asset returns

To see more clearly the financial interpretation of the ambiguity radius  $\rho$ , we undergo experiments based on empirical data. We extract past 40 years' daily simple returns of four major asset classes: Equity indexes (DAX, FTSE, HSI, NASDAQ, NIKKEI250, SP500), US Treasuries (2year, 10year, 30year), Currencies (AUD, CHF, EUR, GBP, JPY) and Commodities (Crude oil, Silver, Gold). For the DRO problem, we use the Cressie-Read divergence instead of KL divergence since all data exhibits quite heavy tail. For the CCO problem, we choose the negative daily return threshold  $-\delta$  to be the 3% empirical quantile of the daily simply return series for each asset class so that they can differ across assets. We choose the chance level  $\epsilon$  to be 2% and 5%, mincing (rounded) event frequencies at quarterly (4 out of 252) and monthly (12 out of 252) so that investors can relate  $\epsilon$  to the degree of event rareness. The portfolio weights are constrained to be bounded below by -1. Both multivariate  $t$ - and normal distributions are tested as the center  $\mathbb{P}_0$  of the ambiguity set  $\mathcal{U}$  when fitting data. Also, we test both the  $4^{th}$  order and  $2^{nd}$  order reformulations of the DRO problem.

Table 3 (a)-(d) report the equivalent ambiguity radius  $\rho$  of the DRO problem, together with the corresponding optimal portfolio return (annualized), at a given pair of CCO parameters  $(\epsilon, \delta)$  for the four asset classes, respectively. Take Table 3(a) as an example. There are 2 rows, 4 columns and 8 entries in total. Each row corresponds to the choice of the parameter  $\epsilon$ , and each column corresponds to the choice of the reformulation framework of the DRO problem and the choice of the center distribution  $\mathbb{P}_0$ . The upper number in one entry records the equivalent ambiguity radius  $\rho$ , while the lower number in the round brackets records the corresponding optimal portfolio annualized return. With other parameter fixed, we compare the optimal portfolio returns between the multivariate student  $t$ -distribution and the normal distribution, and label the entry numbers with a larger portfolio return in bold black.

We read from Table 3 that, by relating the size parameter  $\rho$  of the ambiguity set in the DRO problem to the CCO chance parameters, it then becomes tangible, without which even the appropriate

Table 3: The equivalent ambiguity radius  $\rho$  of the DRO problem for the four asset classes: (a) Equity, (b) US Treasury, (c) Currency, and (d) Commodity. The loss threshold  $\delta$  is taken as the negative value of the 3% empirical quantile of the daily simple return series for each asset class, thus is different across assets. We compare the portfolio performance within each asset class based on the choice of the center distribution (either multivariate student  $t$ - or normal distributions) under both the  $4^{th}$  order and  $2^{nd}$  order reformulations of the DRO problem. The percentage number in the round brackets under the equivalent ambiguity radius  $\rho$  records the corresponding optimal portfolio annualized return. Bold numbers emphasize the better portfolio return performance at a given pair of  $(\epsilon, \delta)$  under a given solution framework of the DRO problem.

(a) Equity:  $\delta = 3.35\%$ .

	$4^{th}$ order		$2^{nd}$ order	
	Student $t$	Normal	Student $t$	Normal
$\epsilon = 2\%$	<b>3.5e-4</b> (30.7%)	1.2e-4 (15.3%)	<b>6.1e-4</b> (30.7%)	1.2e-4 (15.3%)
$\epsilon = 5\%$	<b>3.4e-4</b> (39.2%)	1.2e-4 (19.8%)	<b>6.1e-4</b> (39.2%)	1.2e-4 (19.8%)

(b) US Treasury:  $\delta = 6.58\%$ .

	$4^{th}$ order		$2^{nd}$ order	
	Student $t$	Normal	Student $t$	Normal
$\epsilon = 2\%$	<b>2e-6</b> (-1.1%)	2.8e-14 (-2.6%)	<b>9.5e-6</b> (-1.1%)	2.8e-14 (-2.6%)
$\epsilon = 5\%$	<b>2e-6</b> (0.7%)	2.8e-14 (-2.6%)	<b>4.8e-6</b> (0.7%)	2.8e-14 (-2.6%)

(c) Currency:  $\delta = 1.40\%$ .

	$4^{th}$ order		$2^{nd}$ order	
	Student $t$	Normal	Student $t$	Normal
$\epsilon = 2\%$	2.6e-4 (2.3%)	<b>6.1e-5</b> (3.6%)	3.1e-4 (2.3%)	<b>6.1e-5</b> (3.6%)
$\epsilon = 5\%$	1.5e-4 (4.4%)	<b>3.1e-5</b> (5.0%)	3.1e-4 (4.4%)	<b>3.1e-5</b> (5.0%)

(d) Commodity:  $\delta = 4.4\%$ .

	$4^{th}$ order		$2^{nd}$ order	
	Student $t$	Normal	Student $t$	Normal
$\epsilon = 2\%$	<b>9.6e-5</b> (17.3%)	3.7e-9 (4.6%)	<b>1.5e-4</b> (17.3%)	3.7e-9 (4.6%)
$\epsilon = 5\%$	<b>6.5e-5</b> (22.7%)	1.9e-9 (4.6%)	<b>7.6e-4</b> (22.6%)	1.9e-9 (4.6%)

order is hard to guess. In our tests, its magnitude can range from  $10^{-4}$  to  $10^{-14}$  depending on asset classes and on the investor's tolerance level. What's more, the heavy-tail nature of financial data demands the usage of divergence measures (e.g., the Cressie-Read divergence) that allow heavy-tail distribution if one takes the robust

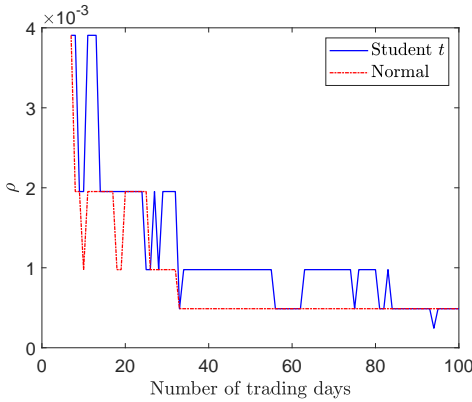


approach for portfolio optimization. Ambiguity sets constructed by the KL divergence, however, require the objective function to be exponentially bounded, which exclude important heavy-tail distributions used ubiquitously for financial asset returns, e.g., the student  $t$ -distribution. Among the 16 tests in Table 3, the larger return in bold shows 12 favor fitting data with  $\mathbb{P}_0$  as multivariate  $t$ -distributed.

#### 5.4 High frequency empirical setting

We collect the intraday 5-minute asset returns of 15 stocks<sup>1</sup> that are selected from the 50 Hang Seng Index constituent stocks based on the market cap and daily turnover. The data spans from Dec 1st, 2014 to Dec 1st, 2017, and consists of roughly 39,390 observations with information of the first and the last half hours in each trading day excluded.

The first experiment illustrates the trend of equivalent ambiguity radius  $\rho$  as more empirical data is available. As in the last experiment, we use the Cressie-Read divergence and set the loss probability  $\epsilon = 3\%$  and  $\delta = 0.28\%$  (the 3% empirical quantile of the return series over 100 trading days). The asset allocation strategy is bounded from below by  $-1$ . We apply the 4<sup>th</sup> order reformulation to solve the DRO problem and test both multivariate  $t$ - and normal distribution as the nominal distribution  $\mathbb{P}_0$ . To begin with, we compute the equivalent ambiguity radius  $\rho$  based on the first 6 consecutive trading days of 5-minute return series. Then we move forward to include one more trading day's sample data and obtain the next equivalent  $\rho$ . Figure 2 plots the series of equivalent  $\rho$  with each  $\rho$  stamped with how many trading days' data the computation is based on.



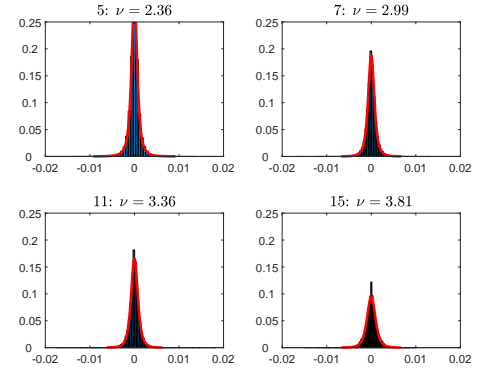
**Figure 2: With  $(\epsilon, \delta) = (3\%, 0.28\%)$  fixed, the equivalent ambiguity radius  $\rho$  goes down and converges as data of more trading days is available.**

Figure 2 shows that, to achieve the same level of tail probability protection, the equivalent ambiguity radius  $\rho$  goes down and converges as more data is available. Such a conclusion is within expectation because the more available data, the more information and thus fewer uncertainties are over the underlying distribution. The second observation accords with the conclusion in Figure 1,

<sup>1</sup>The ticker codes for the selected 15 stocks are: 00001, 00005, 00016, 00027, 00388, 00688, 00700, 00883, 00939, 00941, 01299, 01398, 01928, 02318, 03988.

that is, even with the same empirical data set, the tail heaviness assumption of the center distribution affects the interpretation of the ambiguity radius  $\rho$ . Robust portfolio optimization centered with heavy-tail distributions requires a larger range of robust protections to achieve the same tail probability level.

Then, we fit the returns of each single stock to a univariate student  $t$ -distribution to verify that the distribution of high frequency financial data indeed exhibits heavy tail. The degree of freedom parameter, which quantifies the tail heaviness, is shown to range from 2.36 to 3.81 among the 15 stocks. Figure 3 shows the fitting results of 4 stocks accompanied with the degree of freedom parameter  $\nu$  in the title position. As it suggests, assuming the nominal distribution of the returns as a student  $t$ -distribution is rather reasonable.



**Figure 3: Fitting performance in a student  $t$ -distribution for stock 5, 7, 11 and 15. The degree of freedom parameter  $\nu$  is shown to range from 2.36 to 3.81, which verifies that the distributions of intraday 5-minute returns are indeed heavy-tailed.**

The last experiment focuses on the value of robust protection in portfolio optimization. In real practice, portfolio optimization under a distributional robust framework is needed to protect investors from uncertainties arising from both the limited historical data and future distributional changes. It is necessary for a trader to frequently rebalance the portfolio to accommodate fluctuations in distributions. As we would demonstrate, the robust protection actually helps improve the portfolio performance, especially when compared with portfolios that are selected either based on the nominal distribution (namely, problem (2)) or under the classical Mean Variance framework. The Mean Variance model we take is:

$$\min_{x \in \mathcal{X}} x^T \Sigma x \quad \text{s.t.} \quad x^T \mu \geq r_{\text{target}}.$$

We divide the whole 3-year datasets into two consecutive parts. With the first 2-year data, we fit it to a 15-d student  $t$ -distribution and establish the equivalent ambiguity radius  $\rho = 2.4\text{e-}4$  and optimal return  $0.35\text{e-}4$ , given chance constraint parameters  $(\epsilon, \delta) = (3\%, 39\text{e-}4)$ . Then with the last-year data as a test set, we backtest the portfolio performance with three asset allocation strategies solved respectively by the DRO problem, the nominal optimization problem, and the Mean Variance problem. For the DRO problem, we fix  $\rho = 2.4\text{e-}4$ , and for the Mean Variance problem, we fix  $r_{\text{target}}$

$= 0.35e-4$ . Under each optimization framework, the asset allocation strategy is not constant throughout the whole testing period. We rebalance the portfolio in the frequency of every 5 minutes/hour/half day/day. For each rebalancing, we always use its past 4 months of trading data to solve the optimal allocation strategy and then apply the strategy to next incoming 5 minutes/hour/half day/day. Table 4 summarizes the statistics of the return series based on different strategies and rebalancing frequencies.

**Table 4: Statistics of the 3 return series constructed by 5-minute/hourly/half-day/daily rebalanced allocation strategies solved by the DRO problem, the nominal problem, and the Mean Variance problem, respectively.**

	DRO	Nominal	Mean Variance
<b>5-minute rebalancing</b>			
Mean (e-4)	<b>3.68</b>	3.24	0.77
Variance(e-6)	56.5	199	3.48
Skewness	<b>0.66</b>	0.17	0.61
<b>Hourly rebalancing</b>			
Mean (e-4)	<b>3.19</b>	2.67	0.59
Variance(e-6)	56.3	200	3.52
Skewness	<b>0.64</b>	0.13	0.53
<b>Half-day rebalancing</b>			
Mean (e-4)	<b>2.7</b>	2.3	0.48
Variance(e-6)	55.9	198	3.47
Skewness	<b>0.62</b>	0.09	0.43
<b>Daily rebalancing</b>			
Mean (e-4)	<b>1.69</b>	1.0	0.24
Variance(e-6)	55.6	190	3.43
Skewness	<b>0.66</b>	0.054	0.52

Table 4 shows that, the dynamic allocation strategy under a robust framework always outperforms that without a robust protection and the classical Mean Variance strategy. The outperformance can be at most 7 times, depending on the rebalancing frequency. And the DRO strategy keeps a medium level of volatility, neither too aggressive nor too conservative to gain low returns. What's more, the highest skewness for the DRO strategy also highlights its inclination to more gains than losses. Last but not the least, although the outperformance of a DRO strategy is consistent between different rebalancing frequencies, an investor benefits from more frequent rebalancing with returns far more than doubled under whatever portfolio selection framework.

## 6 CONCLUSIONS

We delved into the ambiguity radius for DRO problems with a distributional ambiguity set defined by  $\phi$ -divergence. We showed that for general  $\phi$ -divergences, a DRO optimization problem is asymptotically equivalent to a mean-deviation problem, where the risk preference parameter is controlled by the ambiguity radius. We used a portfolio selection example to demonstrate that, when the investment strategy is bounded, the ambiguity radius can be cast as a chance constraint in a deterministic optimization with

the same objective. Otherwise, within the set of unbounded investment strategies, a chance-constrained deterministic optimization consistently performs better than the DRO problem. Through extensive experiments with both synthetic and empirical data, we concluded that, to achieve the same level of tail probability protection, a DRO problem centered at heavy-tail distributions requires a larger ambiguity set.

## ACKNOWLEDGMENTS

Qi WU acknowledges the GRF support from the Hong Kong Research Grants Council under 14211316 and 14206117.

## REFERENCES

- [1] Soroosh Shafieezadeh Abadeh, Peyman Mohajerin Esfahani, and Daniel Kuhn. 2015. Distributionally robust logistic regression. In *Advances in Neural Information Processing Systems*. 1576–1584.
- [2] Aharon Ben-Tal, Dick Den Hertog, Anja De Waegenare, Bertrand Melenberg, and Gijs Rennen. 2013. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science* 59, 2 (2013), 341–357.
- [3] Jose Blanchet, Lin Chen, and Xun Yu Zhou. 2018. Distributionally robust mean-variance portfolio selection with Wasserstein distances. *arXiv preprint arXiv:1802.04885* (2018).
- [4] Pierre Bonami and Miguel A Lejeune. 2009. An exact solution approach for portfolio optimization problems under stochastic and integer constraints. *Operations research* 57, 3 (2009), 650–670.
- [5] Li Chen, Simai He, and Shuzhong Zhang. 2011. Tight bounds for some risk measures, with applications to robust portfolio selection. *Operations Research* 59, 4 (2011), 847–865.
- [6] Ruidi Chen and Ioannis Ch Paschalidis. 2018. A robust learning approach for regression models based on distributionally robust optimization. *The Journal of Machine Learning Research* 19, 1 (2018), 517–564.
- [7] Erick Delage and Yinyu Ye. 2010. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations research* 58, 3 (2010), 595–612.
- [8] Peyman Mohajerin Esfahani and Daniel Kuhn. 2018. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming* 171, 1-2 (2018), 115–166.
- [9] Laurent El Ghaoui, Maksim Oks, and Francois Oustry. 2003. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations research* 51, 4 (2003), 543–556.
- [10] Paul Glasserman and Xingbo Xu. 2014. Robust risk measurement and model risk. *Quantitative Finance* 14, 1 (2014), 29–58.
- [11] Jun-ya Gotoh, Michael Jong Kim, and Andrew EB Lim. 2018. Robust empirical optimization is almost the same as mean-variance optimization. *Operations Research Letters* 46, 4 (2018), 448–452.
- [12] Tatsunori Hashimoto, Megha Srivastava, Hongseok Namkoong, and Percy Liang. 2018. Fairness Without Demographics in Repeated Loss Minimization. In *International Conference on Machine Learning*. 1934–1943.
- [13] Zhaolin Hu and L Jeff Hong. 2013. Kullback-Leibler divergence constrained distributionally robust optimization. *Available at Optimization Online* (2013).
- [14] Woo Chang Kim, Jang Ho Kim, and Frank J Fabozzi. 2014. Deciphering robust portfolios. *Journal of Banking & Finance* 45 (2014), 1–8.
- [15] Andrew Lesniewski, Heng Sun, and Qi Wu. 2016. Asymptotics of portfolio tail risk metrics for elliptically distributed asset returns. *Available at SSRN 2748970* (2016).
- [16] Hongseok Namkoong and John C Duchi. 2017. Variance-based regularization with convex objectives. In *Advances in neural information processing systems*. 2971–2980.
- [17] Leandro Pardo. 2005. *Statistical Inference Based on Divergence Measures*. CRC Press.
- [18] Herbert E Scarf. 1957. *A min-max solution of an inventory problem*. Technical Report. RAND CORP SANTA MONICA CALIF.
- [19] Philip Thomas and Erik Learned-Miller. 2019. Concentration inequalities for conditional value at risk. In *International Conference on Machine Learning*. 6225–6233.
- [20] Shushang Zhu and Masao Fukushima. 2009. Worst-case conditional value-at-risk with application to robust portfolio management. *Operations research* 57, 5 (2009), 1155–1168.
- [21] Shushang Zhu, Duan Li, and Shouyang Wang. 2009. Robust portfolio selection under downside risk measures. *Quantitative Finance* 9, 7 (2009), 869–885.
- [22] Steve Zymmler, Daniel Kuhn, and Berç Rustem. 2013. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming* 137, 1-2 (2013), 167–198.