UNDERSTANDING DISTRIBUTIONAL AMBIGUITY VIA NON-ROBUST CHANCE CONSTRAINT

Anonymous authors
Paper under double-blind review

ABSTRACT

We propose a non-robust interpretation of the distributionally robust optimization (DRO) problem by relating the impact of uncertainties around the distribution on the impact of constraining the objective through tail probabilities. Our interpretation allows utility maximizers to figure out the size of the ambiguity set through parameters that are directly linked to the chance parameters. We first show that for general ϕ-divergences, a DRO problem is asymptotically equivalent to a class of mean-deviation problems, where the ambiguity radius controls investor's risk preference. Based on this non-robust reformulation, we then show that when a boundedness constraint is added to the investment strategy. The DRO problem can be cast as a chance-constrained optimization (CCO) problem without distributional uncertainties. 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. Besides the widely-used Kullback-Leibler (KL) divergence which requires the distribution of the objective function to be exponentially bounded, our results apply to divergence measures that accommodate well heavy tail distribution such as the student t-distribution and the lognormal distribution. Comprehensive testings on synthetic data and real data are provided.

1 INTRODUCTION

Optimization models with an expectation as the objective function are always used when studying decision making problems. These problems require the knowledge of the "true" distribution, which is uncertain, so its estimation is required. Many studies have been focused on different estimation methods, see for example Merton (1980), Bai et al. (2007), El Karoui et al. (2008), and Qiu et al. (2015). However, it is often not accurate since the number of observations is inadequate. For example, as stated in Chen & Yuan (2016), around $3,000$ months of historical data are needed to give accurate approximations for a portfolio of 25 assets. In reality, it is impossible to have such long time series data sets. An inaccurate estimation of the distribution can cause erroneous judgments for a practitioner based on corresponding solutions. It is necessary to seek an approach that can incorporate uncertainty under a scarcity of data.

Formulating a distributionally robust optimization (DRO) model is recognized as a solution to deal with the issue. In a DRO model, we add an extra layer of inner optimization over an *ambiguity set* which contains distribution as uncertainty. Generally, there are 3 ways to describe the ambiguity set. The first way is the *geometric approach*, in which the ambiguity set is described as desired geometric shapes such as box, ellipsoid and polyhedral, etc. (e.g., Kim et al. (2014), Zhu et al. (2009), and Zhu & Fukushima (2009)). The second way is the *moment-based approach*, and was studied in Delage & Ye (2010), Scarf (1957), Chen et al. (2011), and Zymler et al. (2013) for instance. The ambiguity set is usually defined so that all distribution in it have the same *n-th* moments. The last way is the *statistical distance approach* so that the ambiguity set can be thought as a ball with ambiguity radius ρ such that it contains all the distribution of distance/divergence measures from a center distribution less than ρ.

We adopt the statistical distance approach to define the ambiguity set. This is because under the geometric approach and moment-based approach, it is not assured that the "true" distribution lies in the ambiguity set. As a result, optimal solutions become meaningless although it is easier than using the statistical distance approach. Only a small portion of the DRO problems under the statistical distance approach are known to be solved analytically. For example, Hu & Hong (2013) solves the DRO problem if the ambiguity set is constructed using KL divergence. Instead of seeking an analytical solution, we can apply an expansion approach (see e.g., Lam (2016) and Gotoh et al. (2018)) to solve the DRO problem. The advantage of using the expansion approach is that we can consider solving DRO under general divergence measures.

In this paper, we would like to understand the ambiguity set in the financial context and further find a non-robust interpretation of the ambiguity set, which is characterized by the radius parameter ρ. This is because the value has a great impact on the optimal solution. Small values of ρ exclude the true distribution being examined, while large values of ρ make the DRO problem conservative. Both are not very useful in practice. Pardo (2005) gives a statistical meaning of ρ asymptotically and Ben-Tal et al. (2013) applies it to interpret the ambiguity set. The interpretation of the ambiguity radius is based on the assumption that the true distribution is within the same parameterized distribution family as the center distribution. Such assumption limits the universal understanding of the ambiguity radius. In this paper, we relate the DRO problem to a CCO problem. We find that a DRO problem can be reformulated as a class of optimization problems with mean-deviation as the objective function after using the expansion method to remove randomness. We then reformulate a chance-constrained optimization CCO problem and derive equivalence between the DRO problem and the CCO problem. This equivalence opens a door for financial practitioners to interpret the radius parameter through parameters that are directly linked to investment performance.

The rest of this paper is organized as following. In Section 2, we provide background information and a motivation on the optimization problems being considered, including notations and ϕ-divergences in defining the ambiguity set. In Section 3, we reformulate the DRO problem and CCO problem, respectively, and solve the corresponding optimal decisions and optimal values. Section 4 gives numerical experiments, while Section 5 concludes our findings. All detailed proofs are postponed to the Appendix.

2 PROBLEM SETUP

In this section, we introduce the notations used in this paper and give a brief introduction to the ϕ-divergence for defining the ambiguity set. Furthermore, we present the motivation to explain why we focus on the ambiguity radius in the DRO problem.

Notations. Let $\mathbf{r} \in \mathbb{R}^n$, an n-dimensional real-valued random vector, be the vector of asset returns and suppose that the joint probability distribution of $\mathbf{r}$ is P. Let P_0 be the nominal probability distribution of $\mathbf{r}$. Let $\boldsymbol{x} \in \mathbb{R}^n$ be the asset allocation strategy and $\mathbf{e} \in \mathbb{R}^n$ be a vector with all entries equal to 1, respectively. We assume that $\boldsymbol{x}$ lies in a convex set $\mathbb{X}$ and P lies in an ambiguity set $\mathbb{U}$.

ϕ-divergence. ϕ-divergence is a commonly used statistical distance to describe the ambiguity set $\mathbb{U}$. It quantifies how one probability distribution diverges from another. It's defined by a convex function $\phi(t)$ which satisfies $\phi(1) = 0$, $0\phi\left(\frac{0}{0}\right) := 0$, and $\phi\left(\frac{a}{0}\right) := a \lim_{t\to\infty} \frac{\phi(t)}{t}$ when $a > 0$. Given the function $\phi(t)$, the ϕ-divergence $D(Q||P)$ between distribution Q and distribution P is:

$$D(Q||P) := \int_{\Omega} \phi\left(\frac{dQ}{dP}\right) P(dt) = \mathbb{E}_P\left[\phi\left(\frac{dQ}{dP}\right)\right] := \mathbb{E}_P\left[\phi\left(L\right)\right],$$

The quantity L is called the Radon Nikodym derivative (or likelihood ratio) such that $L \geq 0$ almost surely and $\mathbb{E}_P\left[L\right] = 1$. Associated with the function $\phi(t)$ is its *conjugate function* $\phi^*(s)$, defined as

$$\phi^*(s) := \sup_{t\geq 0}\{st - \phi(t)\}. \tag{1}$$

Table 1 lists the two divergences used in this paper. The KL divergence is commonly used because of the appealing properties its conjugate function has. In order to use the KL divergence, the distribution of the objective function needs to be exponentially bounded, which excludes important heavy tail distribution used ubiquitously for financial asset returns, especially the lognormal distribution and the student t-distribution. The Cressie-Read divergence, also called the α-divergence in Glasserman & Xu (2014), overcomes this limitation and can well accommodate heavy tail distribution. However, our interpretation of the ambiguity radius ρ as a chance constraint applies to all

Table 1: The two ϕ-divergences used in this paper. The KL divergence applies to light-tail distribution, while the Cressie-Read divergence is compatible with heavy tail distribution.

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

the ϕ-divergences, including Burg entropy, J-divergence, χ^2-distance, modified χ^2-distance, and Hellinger distance.

Motivation. The goal is to maximize the expectation of investors' utility function $f(\boldsymbol{x}, \mathbf{r})$ over a set of admissible allocation strategies $\mathbb{X}$, namely, $\max_{\boldsymbol{x}\in\mathbb{X}} \mathbb{E}_{\mathbf{r}\sim P}[f(\boldsymbol{x}, \mathbf{r})]$. Utility function is ubiquitous in economics and it measures one's preference for alternatives. In general, it is either in the form of *exponential utility function* $f(\boldsymbol{x}, \mathbf{r}) = \frac{1-e^{-a\boldsymbol{x}^T\mathbf{r}}}{a}$ $(a \neq 0)$ or in the form of *power utility function* (a.k.a. CRRA utility function) $f(\boldsymbol{x}, \mathbf{r}) = \frac{(\boldsymbol{x}^T\mathbf{r})^{1-\eta}-1}{1-\eta}$ when $\eta \neq 1$ & $\eta > 0$ and $f(\boldsymbol{x}, \mathbf{r}) = \ln(\boldsymbol{x}^T\mathbf{r})$ when $\eta = 1$ (Pratt (1978)). Whatever the concrete utility function is, the stochastic optimization problem itself depends strongly on assumptions and properties of the asset return $\mathbf{r}$, the true distribution of which is usually unknown. To address the distribution uncertainty, a common approach is to formulate a robust counterpart of the original problem, called as a distributionally robust optimization problem. A DRO formulation addresses the uncertainty by constructing an ambiguity set of distributions centered at the nominal distribution P_0 and then optimizing the worst-case utility among all distributions within the ambiguity set. In this paper, we focus on an ambiguity set $\mathbb{U}$ defined by the ϕ-divergence and controlled by a radius parameter $\rho > 0$, i.e., $\mathbb{U} := \{P : D(P||P_0) \leq \rho\}$. Thus, the distributionally robust counterpart of the original stochastic optimization problem is:

$$\max_{\boldsymbol{x}\in\mathbb{X}} \min_{P\in\mathbb{U}} \mathbb{E}_{\mathbf{r}\sim P}[f(\boldsymbol{x}, \mathbf{r})]. \tag{2}$$

The choice of ambiguity radius ρ is critical. One cannot set it too large since the maximal utility decreases in ρ. However, if it is too small, one loses the robust protection. There is a trade-off in choosing its magnitude. This situation requires an alternative interpretation of the ambiguity radius ρ. Pardo (2005) proves that, assuming that the true distribution P and the nominal distribution P_0 belong to the same parameterized distribution family with parameter dimension d, then the normalized estimated ϕ-divergence $\frac{2N}{\phi^{(2)}(1)} D(P||P_0)$ asymptotically follows a χ^2_d-distribution. It thus relates the ambiguity radius ρ to a confidence level at which the true distribution P falls within this ambiguity set. However, in financial practice with real data, the assumption that the true distribution is in the same parameterized family with the center distribution often fails to hold. A wrong guess of the nominal distribution might lead to a meaningless confidence level interpretation of the ambiguity radius ρ. Focusing on the impact of the ambiguity radius ρ on the maximal utility, we realize that, through a CCO problem, the ambiguity radius ρ can be explained by the parameters of a chance-constrained problem, using only information of the nominal distribution P_0.

The CCO problem considered in this paper is:

$$\max_{\boldsymbol{x}\in\mathbb{X}} \quad \mathbb{E}_{\mathbf{r}\sim P_0}[f(\boldsymbol{x}, \mathbf{r})] \qquad s.t. \qquad P_0(f(\boldsymbol{x}, \mathbf{r}) \leq -\delta) \leq \epsilon. \tag{3}$$

It shares the same objective function as that of problem (2). The expectation is taken under the nominal distribution P_0, not subject to distributional robustness. The new component is the probability constraint with the parameters δ and ϵ characterizing *downside risk*, which incorporates the tail of $\mathbf{r}$ into consideration. Similar CCO formulations appear often, from the basic Kelly problem to the newsvendor problem (see for example, Busseti et al. (2016) and Özler et al. (2009)). Our work aims to interpret $\mathbb{U}$ incorporating the tail information of $\mathbf{r}$. To be specific, we try to find the relation between ρ in (2) and (ϵ, δ) in (3). In the coming section, we first describe the expansion approach in solving DRO problem (2). For CCO problem (3) with general chance constraints, there is no consensus on explicit solutions. So as a special case, we focus on the utility function $f(\boldsymbol{x}, \mathbf{r}) := \boldsymbol{x}^T\mathbf{r}$ in the CCO problem. By transforming constraints in (3) to VaR, we obtain analytically the relation between (2) and (3) under an unbounded feasibility set of allocation strategies. For bounded feasibility set, we look into the equivalence on (2) and (3) through numerical experiments. Although our methodology is investigated in the financial context, it can also be applied to other contexts.

3 INTERPRET ρ AS A CHANCE CONSTRAINT

In this section, we discuss how to reformulate (2) and (3) so that we can compare the optimal values of the two problems.

3.1 REFORMULATION OF (2)

The Lagrangian dual to the inner minimization problem in (2) is:

$$\begin{aligned} &\max_{\eta_1 \in \mathbb{R}, \eta_2 \geq 0} \left\{ -\frac{1}{\eta_2} \max_L \left\{ \mathbb{E}_{\mathbf{r} \sim P_0} [-\eta_2 (f(\boldsymbol{x}, \mathbf{r}) + \eta_1) L - \phi(L)] \right\} - \eta_1 - \frac{\rho}{\eta_2} \right\} \\ = &\max_{\eta_1 \in \mathbb{R}, \eta_2 \geq 0} \left\{ -\frac{1}{\eta_2} \mathbb{E}_{\mathbf{r} \sim P_0} [\phi^*(-\eta_2 (f(\boldsymbol{x}, \mathbf{r}) + \eta_1))] - \eta_1 - \frac{\rho}{\eta_2} \right\}. \end{aligned} \tag{4}$$

The last equality is from the definition of conjugate function $\phi^*(s)$ in Eq. (1). Difficulty in solving the dual problem lies in the term $\mathbb{E}_{\mathbf{r} \sim P_0}[\phi^*(-\eta_2(\boldsymbol{x}^T\mathbf{r} + \eta_1))]$. For general ϕ-divergence, $\phi^*(s)$ may not have explicitly formula (see Ben-Tal et al. (2013)). We hereby follow the idea in Gotoh et al. (2018) to express (4) in terms of *Regular Measure of Deviation*. The results are summarized in Theorem 3.1. For notation simplicity, we would shorten the term $\mathbb{E}_{\mathbf{r} \sim P_0}[\cdot]$ to $\mathbb{E}[\cdot]$ without further explanation.

Theorem 3.1. *Let $\phi(t)$ be a closed proper convex function and the conjugate function $\phi^*(s)$ defined in Eq. (1), respectively. Suppose that under mild conditions, the strong duality holds. Defining*

$$\mathcal{D}_{\eta_2, \phi, P_0}(f(\boldsymbol{x}, \mathbf{r}) | \mathbb{E}[f(\boldsymbol{x}, \mathbf{r})]) := \min_{\eta_1} \left\{ \eta_1 + \frac{1}{\eta_2} \mathbb{E}\left[\phi^* \left(\eta_2 (\mathbb{E}[f(\boldsymbol{x}, \mathbf{r})] - f(\boldsymbol{x}, \mathbf{r}) - \eta_1) \right) \right] \right\}.$$

Then, (2) is equivalent to:

$$\max_{\boldsymbol{x} \in \mathbb{X}} \left\{ \mathbb{E}[f(\boldsymbol{x}, \mathbf{r})] - \min_{\eta_2 \geq 0} \left[\frac{\rho}{\eta_2} + \mathcal{D}_{\eta_2, \phi, P_0}(f(\boldsymbol{x}, \mathbf{r}) | \mathbb{E}[f(\boldsymbol{x}, \mathbf{r})]) \right] \right\}. \tag{5}$$

Through Theorem 3.1, we can eliminate the distribution uncertainty in (2) and reformulate the DRO problem to a deterministic optimization problem as in (5). The only unknown term $\mathcal{D}_{\eta_2, \phi, P_0}(f(\boldsymbol{x}, \mathbf{r}) | \mathbb{E}[f(\boldsymbol{x}, \mathbf{r})])$ is indeed a function of ϕ^*. Therefore, we apply the expansion method to the quantity $\mathcal{D}_{\eta_2, \phi, P_0}(f(\boldsymbol{x}, \mathbf{r}) | \mathbb{E}[f(\boldsymbol{x}, \mathbf{r})])$ and expand it as a series of terms, the coefficients of which can be computed under the nominal distribution P_0. In such a way, we can reformulate (2) as a single-layer maximization problem.

Lemma 3.2. *Suppose that n is an even number, $\phi(t) \in \mathcal{C}^{n+1}$ is a convex function which satisfies $\phi(1) = \phi^{(1)}(1) = 0$ and $\phi^{(2)}(1) > 0$. Assume $\mathbb{E}[X^k] < \infty$ for $k \leq n$ and X defined by $X := f(\boldsymbol{x}, \mathbf{r}) - \mathbb{E}[f(\boldsymbol{x}, \mathbf{r})]$, then*

$$\mathcal{D}_{\eta_2, \phi, P_0}(f(\boldsymbol{x}, \mathbf{r}) | \mathbb{E}[f(\boldsymbol{x}, \mathbf{r})]) = \sum_{k=1}^{n-1} b_k \mathbb{E}\left[(X + \eta_1^*)^{k+1} \right] \eta_2^k + o(\eta_2^{n-1}), \tag{6}$$

where $b_k = \frac{(-1)^{k+1} z^{(k)}(0)}{(k+1)!}$, and η_1^* is the optimal solution to $\min_{\eta_1} \sum_{k=1}^{n-1} b_k \mathbb{E}\left[(X + \eta_1)^{k+1} \right] \eta_2^k$. 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 most of the ϕ-divergences (KL divergence, Cressie-Read divergence, Burg entropy, J-divergence, χ^2-distance, modified χ^2-distance, and Hellinger distance) satisfy the smoothness conditions. Taking KL divergence and Cressie-Read divergence as example, for $n = 4$, we can explicitly solve the terms in Eq. (6), as shown in the following corollary.

Corollary 3.3. *Consider the case $n = 4$. Define $X := f(\boldsymbol{x}, \mathbf{r}) - \mathbb{E}[f(\boldsymbol{x}, \mathbf{r})]$. Then the 4^{th} order expansion of $\mathcal{D}_{\eta_2, \phi, P_0}(f(\boldsymbol{x}, \mathbf{r}) | \mathbb{E}[f(\boldsymbol{x}, \mathbf{r}])$ is:*

$$\mathcal{D}_{\eta_2, \phi, P_0}(f(\boldsymbol{x}, \mathbf{r}) | \mathbb{E}[f(\boldsymbol{x}, \mathbf{r}]) = \sum_{k=1}^{3} b_k \mathbb{E}\left[(X + \eta_1^*)^{k+1} \right] \eta_2^k + o(\eta_2^3), \tag{7}$$

with η_1^* being the real root to the following cubic equation

$$\sum_{k=1}^{3}(k+1)b_k\eta_2^k\cdot\eta_1^k+12b_3\eta_2^3\mathbb{E}[X^2]\cdot\eta_1+(4b_3\eta_2^3\mathbb{E}[X^3]+3b_2\eta_2^2\mathbb{E}\left[X^2\right])=0. \quad (8)$$

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 $b_1 = 1/2$, $b_2 = (\theta-2)/6$, $b_3 = (\theta-2)(2\theta-3)/24$.

Gotoh et al. gives a similar expansion of $\mathcal{D}_{\eta_2,\phi,P_0}(f(\boldsymbol{x},\mathbf{r})|\mathbb{E}[f(\boldsymbol{x},\mathbf{r})])$ (see Proposition 3.5 of Gotoh et al. (2018). The main difference between our expansion in Eq. (7) and the expansion given by Gotoh *et al.* lies in the calculation of η_1^*. In Eq. (7), η_1^* is directly solved through the polynomial equation, while in Gotoh et al. (2018), η_1^* is approximated as a function of η_2. In the sequel, we consider the 2^{nd} order expansion of $\mathcal{D}_{\eta_2,\phi,P_0}(f(\boldsymbol{x},\mathbf{r})|\mathbb{E}[f(\boldsymbol{x},\mathbf{r})])$, which is $\frac{\eta_2\mathbb{E}[X^2]}{2\phi^{(2)}(1)}$. Substituting this back to (5), we achieve the 2^{nd} order reformulation of (2) in Theorem 3.4.

Theorem 3.4. *Suppose that $\phi(t)$ is convex, twice continuously differentiable, and that $\phi(1) = \phi^{(1)}(1) = 0$ and $\phi^{(2)}(1) > 0$. Define $X := f(\boldsymbol{x},\mathbf{r}) - \mathbb{E}[f(\boldsymbol{x},\mathbf{r})]$. (2) is asymptotically equivalent to a mean-deviation problem:*

$$\max_{\boldsymbol{x}\in\mathbb{X}}\left\{\mathbb{E}[f(\boldsymbol{x},\mathbf{r})]-\sqrt{\frac{2\rho\mathbb{E}[X^2]}{\phi^{(2)}(1)}}\right\}. \quad (9)$$

Theorem 3.4 tells that the ambiguity radius ρ controls the investor's risk preference. Notice that, in the 2^{nd} order reformulation of (2), the optimal Lagrangian multiplier $\eta_2^* = \sqrt{\frac{2\rho\phi^{(2)}(1)}{\mathbb{E}[X^2]}}$, which increases simultaneously with ρ. This suggests, when ρ is small, the optimal Lagrangian multiplier η_2^* is also small and the expansion in equation (6) is accurate. Later theoretical analysis over the DRO problem would be based on the 2^{nd} order reformulation in Eq. (9). As stated before, we consider $f(\boldsymbol{x},\mathbf{r}) = \boldsymbol{x}^T\mathbf{r}$ in the sequel such that $\mathbb{E}[f(\boldsymbol{x},\mathbf{r})]$ is replaced by $\boldsymbol{x}^T\mu$ and $\mathbb{E}[X^2]$ is replaced by $\boldsymbol{x}^T\Sigma\boldsymbol{x}$ in Eq. (9), where we denote $\mathbb{E}[\mathbf{r}] = \mu$ and $\mathrm{Cov}(\mathbf{r}) = \Sigma$.

3.2 Reformulation of (3)

Note that when $f(\boldsymbol{x},\mathbf{r}) = \boldsymbol{x}^T\mathbf{r}$, the chance constraint in (3) is of the same form as the definition of VaR, which focuses on the probability of losses. This motivates us to reorganize the tail chance constraint in (3) with VaR. The VaR is defined as the minimal level γ such that the probability that the portfolio loss $-\boldsymbol{x}^T\mathbf{r}$ exceeds γ is below ϵ, i.e., $\mathrm{V}_\epsilon(\boldsymbol{x}) := \inf\{\gamma\in\mathbb{R} : P_0\{-\boldsymbol{x}^T\mathbf{r}\geq\gamma\}\leq\epsilon\}$. With the definition of $\mathrm{V}_\epsilon(\boldsymbol{x})$, we have the equivalent form of chance constraint $P_0\{-\boldsymbol{x}^T\mathbf{r}\geq\delta\}\leq\epsilon$ as $\mathrm{V}_\epsilon(\boldsymbol{x})\leq\delta$. Hence, (3) can be reformulated as

$$\max_{x\in\mathbb{X}} \quad \boldsymbol{x}^T\mu \qquad s.t. \qquad \mathrm{V}_\epsilon(\boldsymbol{x})\leq\delta.$$

If P_0 is a normal distribution, then $\mathrm{V}_\epsilon(\boldsymbol{x}) = \kappa(\epsilon)\sqrt{\boldsymbol{x}^T\Sigma\boldsymbol{x}}-\boldsymbol{x}^T\mu$, where $\kappa(\epsilon) = -\Phi^{-1}(\epsilon)$. For general elliptical distribution, Lesniewski et al. (2016) gives an asymptotic expansion of $\mathrm{V}_\epsilon(\boldsymbol{x})$, which takes the form $\kappa(\epsilon)\sqrt{\boldsymbol{x}^T\Sigma\boldsymbol{x}}-\boldsymbol{x}^T\mu$ asymptotically when $\epsilon\to 0$. For example, if P_0 is student t distribution with degree of freedom parameter ν, then $\kappa(\epsilon) = D\epsilon^{-\frac{1}{\nu}}$, where $D = \left(\frac{c_d\pi^{\frac{d-1}{2}}\Gamma(\frac{\nu+1}{2})}{\nu\Gamma(\frac{\nu+d}{2})}\right)^{\frac{1}{\nu}}$ and $c_d = \frac{\Gamma(\frac{\nu+d}{2})}{\Gamma(\frac{\nu}{2})}\nu^{\frac{\nu}{2}}\pi^{-\frac{d}{2}}$. For distributions other than elliptical distributions, Ghaoui et al. (2003) suggests an approximation of $\kappa(\epsilon)$: $\kappa(\epsilon) = \sqrt{\frac{1-\epsilon}{\epsilon}}$. This suggests that (3) can be reformulated as a second-order cone optimization problem:

$$\max_{x\in\mathbb{X}} \quad \boldsymbol{x}^T\mu \qquad s.t. \qquad \kappa(\epsilon)\sqrt{\boldsymbol{x}^T\Sigma\boldsymbol{x}}-\boldsymbol{x}^T\mu\leq\delta. \quad (10)$$

3.3 Solutions

It has been shown that (2) and (3) can be reformulated to a deterministic mean-deviation problem in (9) and a linear optimization with a second-order cone constraint in (10), respectively. In this section, we look into the optimal solution and the optimal value to optimizations (9) and (10) to interpret the impact of the ambiguity radius ρ. We would denote the optimal solution and the optimal value to optimization (9) by $\boldsymbol{x}^*$ and v^*, respectively. Correspondingly, the optimal solution and the optimal value to optimization (10) are denoted by $\tilde{\boldsymbol{x}}^*$ and $\tilde{v}^*$. We begin with the unbounded feasibility set $\mathbb{X} := \left\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x}^T\boldsymbol{e} = 1\right\}$. By examining the convexity of the function $a\sqrt{\boldsymbol{x}^T\Sigma\boldsymbol{x}} - \boldsymbol{x}^T\mu$, we know that problem (9) is a convex optimization and problem (10) is a convex optimization only when $\kappa(\epsilon) > 0$. Recall that in a convex optimization, any local optimum is certainly a global optimum. This motivates us to study the optimal solution to problem (9), $\boldsymbol{x}^*$, and the optimal solution to problem (10), $\tilde{\boldsymbol{x}}^*$, through KKT conditions. The results for $(\boldsymbol{x}^*, v^*)$ and $(\tilde{\boldsymbol{x}}^*, \tilde{v}^*)$ are summarized in Theorem 3.5 and Theorem 3.6, respectively.

Theorem 3.5. *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 for optimization (9) with the feasibility set $\mathbb{X} := \left\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x}^T\mathbf{e} = 1\right\}$, we have:*

- *If $\rho > \phi^{(2)}(1)(C - \frac{B^2}{A})/2$, then the optimal solution and optimal value for problem (9) are*
$$\boldsymbol{x}^* = \frac{\Sigma^{-1}(\mu - \lambda^*\mathbf{e})}{\mu^T\Sigma^{-1}\mathbf{e} - \lambda^*\mathbf{e}^T\Sigma^{-1}\mathbf{e}}, \quad v^* = \lambda^* = B/A - \sqrt{B^2 - A\left(C - 2\rho/\phi^{(2)}(1)\right)}/A.$$

- *If $\rho \leq \phi^{(2)}(1)(C - \frac{B^2}{A})/2$, then no local optimal solution for problem (9) and the optimal value $v^* = +\infty$.*

Notice that, for (2) with an unbounded feasibility set, the optimal value is $+\infty$ when there is no distributional uncertainty (equivalently, $\rho = 0$). Thus, Theorem 3.5 establishes a threshold for the ambiguity radius ρ, beyond which the distributional uncertainty is indeed effective.

Theorem 3.6. *Given A, B, and C defined in Theorem 3.5. Suppose $\kappa(\epsilon) > 0$, and then for optimization (10) with the feasibility set $\mathbb{X} := \left\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x}^T\mathbf{e} = 1\right\}$, the following conclusions hold:*

- *If (ϵ, δ) satisfies $C - B^2/A < (\kappa(\epsilon))^2 < \delta^2 A + 2\delta B + C$ and $B + \delta A > 0$, then the optimal solution and optimal value are:*
$$\tilde{\boldsymbol{x}}^* = \frac{\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\mathbf{e}]}{\mathbf{e}^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\mathbf{e}]}, \quad \tilde{v}^* = \tilde{\lambda}\delta + \tilde{\theta},$$
where $\tilde{\lambda} = (AC - B^2)/(A\kappa(\epsilon)^2 - AC + B^2) + (\kappa(\epsilon)(B + A\delta)\sqrt{AC - B^2})/((A\kappa(\epsilon)^2 - AC + B^2)\sqrt{A\delta^2 + 2B\delta + C - \kappa(\epsilon)^2})$ and $\tilde{\theta} = ((C + \delta B)(\tilde{\lambda} + 1) - \tilde{\lambda}\kappa(\epsilon)^2)/(B + \delta A)$. Furthermore, when the optimal value v^* for problem (9) is finite, $\tilde{v}^* \geq v^*$.

- *If $B > 0$ and if (ϵ, δ) satisfies $C - B^2/A < (\kappa(\epsilon))^2 < \delta^2 A + 2\delta B + C$ and $B + \delta A < 0$, then the feasibility set for problem (10) is $\emptyset$.*

- *If (ϵ, δ) satisfies $(\kappa(\epsilon))^2 \leq C - B^2/A$, then there exists no local optimal solution for problem (10) and the optimal value $\tilde{v}^* = +\infty$.*

In Theorem 3.6, we first provide the sufficient and necessary conditions of (ϵ, δ) for the optimization problem (10) to have a finite optimal value. We further compare the finite optimal values to problem (10) and to problem (9), $\tilde{v}^*$ and v^*, and find that the CCO reformulation performs uniformly better than the DRO reformulation. In addition, we identify one sufficient condition under which optimization (10) is infeasible. We also identify one sufficient condition for the chance constraint to be redundant in optimization (10), and the optimal value tends to be positive infinity.

4 Experiments

In Section 3, we proved that the CCO reformulation performs better than the DRO reformulation when the optimal values are both finite for an unbounded strategy set. For a bounded strategy set, SOCP problem with boundedness is generally explicitly unsolvable. So we resort to numerical analysis. In Section 4.1, we numerically test the reformulation accuracy of optimization (9) to (2). In Section 4.2, we conduct experiments to understand the ambiguity radius ρ via the chance constraints and find that the tail heaviness indeed affects the interpretation of ρ. In Section 4.3, we test the interpretation of the ambiguity radius based on empirical data.

4.1 REFORMULATION ACCURACY OF (2)

In this section, we numerically test the reformulation accuracies of the 2^{nd} order and 4^{th} order reformulations with respect to (2). The ϕ-divergence we take includes both KL divergence and Cressie-Read divergence. It is known that under KL divergence, (2) can be exactly solved. Under Cressie-Read divergence, we take the optimal value solved by the Robust Monte Carlo method introduced in Glasserman & Xu (2014) as the benchmark to compare the results.

Table 2 records under KL divergence, the relative errors (in the 3^{rd} & 4^{th} rows) w.r.t. the exact optimal value (the 2^{nd} row). It is demonstrated 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 ρ is set to 0.78, relative error for the 4^{th} order reformulation is 1.53%, while it is 56.54% for the 2^{nd} order reformulation given that the optimal value is -0.2787. Here, we assume that 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 clear impact from the heavy tail. And the center distribution P_0 under Cressie-Read divergence is multivariate t. We see that the larger the size of the ambiguity set, i.e., larger ρ, the better the improvement of the 4^{th} order reformulation. In fact, the error reduction is about 10 folds in this example. However, using the 2^{nd} order equivalent formulation is good enough to study relations between (2) and (3) when ρ is small.

Table 2: Relative errors of the 4^{th} order reformulation and 2^{nd} order reformulation w.r.t. the optimal value of (2). Ambiguity sets are defined by KL divergence centered at 6-d exponential distribution.

ρ	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
v_{opt1}	0.1887	0.1841	0.1807	0.1778	0.1753	0.1730	0.1710	0.1691	0.1673
4^{th} order	0.0002%	0.0038%	0.0128 %	0.0274%	0.0479%	0.0748%	0.1082%	0.1483%	0.1951%
2^{nd} order	0.1172%	0.2397%	0.3659%	0.4951%	0.6270%	0.7613%	0.8979%	1.037%	1.778%

4.2 INTERPRETATION OF ρ UNDER DISTRIBUTION WITH DIFFERENT TAIL HEAVINESSES

This experiment shows that the tail heaviness of the nominal distribution P_0 indeed affects the interpretation of the ambiguity radius ρ. We focus on three distribution of 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 ρ is fixed at 0.27. We say that the ambiguity radius ρ can be explained by a chance constraint with parameters (ϵ, δ) if the optimal value of problem (9) is equal to that of problem (10) under the same distribution P_0. We plot the results of equivalent (ϵ, δ) in Figure 1. It shows that, first, the ambiguity radius ρ can be explained by a set of pairs (ϵ, δ) in terms of the impact on the optimal value. Second, tail heaviness affects the interpretation of ρ and distribution with heavier tail result in a larger loss threshold for a given loss probability ϵ.

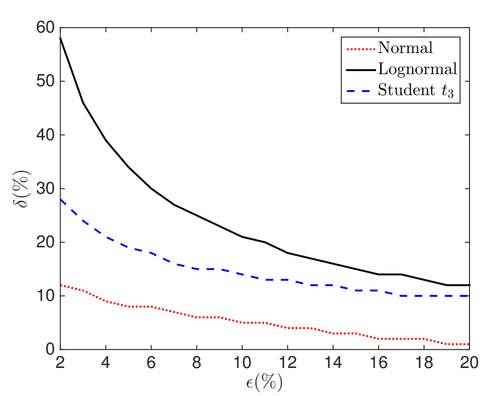

Figure 1: Given $\rho = 0.27$, tail heaviness affects the equivalent loss threshold δ.

4.3 EMPIRICAL STUDIES

To see more clearly the financial interpretation of the ambiguity radius ρ, 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/USD, CHF/USD, EUR/USD, GBP/USD, JPY/USD), 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 1%, 3% and 5% empirical quantile of the daily simply return series for each asset class so that they can differ across assets. We choose the chance level ϵ ranging from 1% to 20%. The results are given in the section A.1 such that Table 4, Table 5, Table 6 and Table 7 corresponds to asset class equity indexes, US Treasuries, currencies and commodities accordingly.

In order to study the relation between ρ in (2) as well as δ and ϵ in (3) on the four assets classes at the same time, we report the equivalent ambiguity radius ρ of the DRO problem, together with the corresponding optimal portfolio return (annualized), at a given pair of the CCO parameters (ϵ, δ) for the four asset classes simultaneously in Table 3. In the table, we extract the results when δ corresponds to 3% and ϵ to be 2% and 5%, which mincing (rounded) event frequencies at quarterly (4 out of 252) and monthly (12 out of 252) so that investors can relate ϵ to the degree of event rareness. The portfolio weights are limited to be bounded below by -1. Both multivariate t- and normal distribution are tested as the center P_0 of the ambiguity set $\mathbb{U}$ when fitting data. Also, we test the 4^{th} order and 2^{nd} order reformations of the DRO problem.

We read from Table 3 that, by relating the size of ambiguity set ρ in the DRO problem to the CCO chance parameters, it then becomes tangible, without which even the appropriate order is hard to guess. In our tests, its magnitude can range from 10^{-3} to 10^{-14} depending on asset classes and the investor's tolerance level. What's more, the heavy tail nature of financial data demands the use of divergence measures such as the Cressie-Read divergence, which allows heavy tail distributions 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 distribution used ubiquitously for financial asset returns, e.g., the student t-distribution. Among the 16 tests in Table 3, the larger returns in bold show 12 favor fitting data by assuming that P_0 is multivariate t-distributed.

Table 3: Equivalent ambiguity radius ρ and optimal annualized return R at a given pair of (ϵ, δ).

$\binom{\rho_t, \rho_n}{R_t, R_n}$	Equity: -δ = -3.35%		Bond: -δ = -6.58%		FX: -δ = -1.40%		Commodity: -δ = -4.4%	
	4^{th} order	2^{nd} order	4^{th} order	2^{nd} order	4^{th} order	2^{nd} order	4^{th} order	2^{nd} order
ϵ = 2%	**3.5e-4**, 1.2e-4	**6.1e-4**, 1.2e-4	**2e-6**, 2.8e-14	**9.5e-6**, 2.8e-14	2.6e-4, **6.1e-5**	3.1e-4, **6.1e-5**	**9.6e-5**, 3.7e-9	**1.5e-4**, 3.7e-9
	30.7%, 15.3%	**30.7**%, 15.3%	**-1.1**%, -2.6%	**-1.1**%, -2.6%	2.3%, **3.6**%	2.3%, **3.6**%	**17.3**%, 4.6%	**17.3**%, 4.6%
ϵ = 5%	**3.4e-4**, 1.2e-4	**6.1e-4**, 1.2e-4	**2e-6**, 2.8e-14	**4.8e-6**, 2.8e-14	1.5e-4, **3.1e-5**	3.1e-4, **3.1e-5**	**6.5e-5**, 1.9e-9	**7.6e-4**, 1.9e-9
	39.2%, 19.8%	**39.2**%, 19.8%	**0.7**%, -2.6%	**0.7**%, -2.6%	4.4%, **5.0**%	4.4%, **5.0**%	**22.7**%, 4.6%	**22.6**%, 4.6%

5 CONCLUSIONS

We delved into the ambiguity radius for DRO problems with a distribution ambiguity set regulated by ϕ-divergence. We showed that for general ϕ-divergences, a DRO portfolio optimization problem is asymptotically equivalent to a mean-deviation problem, where the ambiguity radius controls an investor's risk preference parameter. Theoretical analysis over the mean-deviation problem sets a threshold for the ambiguity radius, across which the optimal value suffers from a drastic phase transition. It is only beyond that radius threshold can the distributional uncertainty take effect. We also showed both numerically and theoretically 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.

REFERENCES

ZD Bai, BQ Miao, GM Pan, et al. On asymptotics of eigenvectors of large sample covariance matrix. *The Annals of Probability*, 35(4):1532–1572, 2007.

Aharon Ben-Tal and Marc Teboulle. An old-new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance*, 17(3):449–476, 2007.

Aharon Ben-Tal, Dick Den Hertog, Anja De Waegenaere, Bertrand Melenberg, and Gijs Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357, 2013.

Enzo Busseti, Ernest K Ryu, and Stephen Boyd. Risk-constrained kelly gambling. *The Journal of Investing*, 25(3):118–134, 2016.

Jiaqin Chen and Ming Yuan. Efficient portfolio selection in a large market. *Journal of Financial Econometrics*, 14(3):496–524, 2016.

Li Chen, Simai He, and Shuzhong Zhang. Tight bounds for some risk measures, with applications to robust portfolio selection. *Operations Research*, 59(4):847–865, 2011.

Erick Delage and Yinyu Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations research*, 58(3):595–612, 2010.

Noureddine El Karoui et al. Spectrum estimation for large dimensional covariance matrices using random matrix theory. *The Annals of Statistics*, 36(6):2757–2790, 2008.

Laurent El Ghaoui, Maksim Oks, and Francois Oustry. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations research*, 51(4):543–556, 2003.

Paul Glasserman and Xingbo Xu. Robust risk measurement and model risk. *Quantitative Finance*, 14(1):29–58, 2014.

Jun-ya Gotoh, Michael Jong Kim, and Andrew EB Lim. Robust empirical optimization is almost the same as mean–variance optimization. *Operations Research Letters*, 46(4):448–452, 2018.

Zhaolin Hu and L Jeff Hong. Kullback-leibler divergence constrained distributionally robust optimization. *Available at Optimization Online*, 2013.

Zhaolin Hu, L Jeff Hong, and Anthony Man-Cho So. Ambiguous probabilistic programs. *Available at Optimization Online*, 2013.

Woo Chang Kim, Jang Ho Kim, and Frank J Fabozzi. Deciphering robust portfolios. *Journal of Banking & Finance*, 45:1–8, 2014.

Henry Lam. Robust sensitivity analysis for stochastic systems. *Mathematics of Operations Research*, 41(4):1248–1275, 2016.

Andrew Lesniewski, Heng Sun, and Qi Wu. Asymptotics of portfolio tail risk metrics for elliptically distributed asset returns. *Available at SSRN 2748970*, 2016.

Robert C Merton. On estimating the expected return on the market: An exploratory investigation. *Journal of financial economics*, 8(4):323–361, 1980.

Aysun Özler, Barış Tan, and Fikri Karaesmen. Multi-product newsvendor problem with value-at-risk considerations. *International Journal of Production Economics*, 117(2):244–255, 2009.

Leandro Pardo. *Statistical Inference Based on Divergence Measures*. CRC Press, 2005.

John W Pratt. Risk aversion in the small and in the large. In *Uncertainty in Economics*, pp. 59–79. Elsevier, 1978.

Huitong Qiu, Fang Han, Han Liu, and Brian Caffo. Robust portfolio optimization. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett (eds.), *Advances in Neural Information Processing Systems 28*, pp. 46–54. Curran Associates, Inc., 2015. URL `http://papers.nips.cc/paper/5714-robust-portfolio-optimization.pdf`.

R Tyrrell Rockafellar and Stan Uryasev. The fundamental risk quadrangle in risk management, optimization and statistical estimation. *Surveys in Operations Research and Management Science*, 18(1-2):33–53, 2013.

R Tyrrell Rockafellar and Roger J-B Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.

Herbert E Scarf. A min-max solution of an inventory problem. Technical report, RAND CORP SANTA MONICA CALIF, 1957.

Shushang Zhu and Masao Fukushima. Worst-case conditional value-at-risk with application to robust portfolio management. *Operations research*, 57(5):1155–1168, 2009.

Shushang Zhu, Duan Li, and Shouyang Wang. Robust portfolio selection under downside risk measures. *Quantitative Finance*, 9(7):869–885, 2009.

Steve Zymler, Daniel Kuhn, and Berç Rustem. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 137(1-2):167–198, 2013.

A APPENDIX

A.1 NUMERICAL RESULTS

In this section, we give numerical results on ρ and annualized return R under different δ and ϵ using 4^{th} and 2^{nd} order of formulation of the DRO problem for the four major asset classes. Two types of distribution, multivariate distribution t and multivariate normal distribution are considered.

Table 4: (Equity Indexes) Equivalent ambiguity radius ρ and optimal annualized return R at a given pair of (ϵ, δ). '—' means that there is no equivalent ρ for the given pair of (ϵ, δ).

$\binom{\rho_t, \rho_n}{R_t, R_n}$	Equity: -δ = -2.68%		Equity: -δ = -3.35%		Equity: -δ = -4.64%	
	4^{th} order	2^{nd} order	4^{th} order	2^{nd} order	4^{th} order	2^{nd} order
ϵ=1%	3.5e-04 ,2.4e-04	1.2e-03 ,2.4e-04	3.5e-04 ,1.2e-04	1.2e-03 ,1.2e-04	3.5e-04 ,1.2e-04	6.1e-04 ,1.2e-04
	18.8% ,9.76%	18.8% ,9.76%	25.2% ,13.2%	25.2% ,13.2%	35% ,19.3%	35% ,19.3%
ϵ=2%	3.5e-04 ,1.2e-04	1.2e-03 ,1.2e-04	3.5e-04 ,1.2e-04	6.1e-04 ,1.2e-04	3.0e-04 ,6.1e-05	6.1e-04 ,6.1e-05
	24.4% ,11.7%	24.4% ,11.7%	30.7% ,15.3%	30.7% ,15.3%	42% ,22.1%	42% ,22.1%
ϵ=3%	3.5e-04 ,1.2e-04	1.2e-03 ,1.2e-04	3.5e-04 ,1.2e-04	6.1e-04 ,1.2e-04	2.3e-04 ,6.1e-05	3.1e-04 ,6.1e-05
	27.5% ,13.1%	27.5% ,13.1%	34.2% ,17%	34.2% ,17%	46.9% ,24.4%	46.9% ,24.4%
ϵ=4%	3.5e-04 ,1.2e-04	6.1e-04 ,1.2e-04	3.5e-04 ,1.2e-04	6.1e-04 ,1.2e-04	1.9e-04 ,—	3.1e-04 ,6.1e-05
	29.8% ,14.3%	29.8% ,14.3%	37% ,18.5%	37% ,18.5%	50.6% ,26.4%	50.6% ,26.4%
ϵ=5%	3.5e-04 ,1.2e-04	6.1e-04 ,1.2e-04	3.4e-04 ,1.2e-04	6.1e-04 ,1.2e-04	1.6e-04 ,—	3.1e-04 ,3.1e-05
	31.7% ,15.4%	31.7% ,15.4%	39.2% ,19.8%	39.2% ,19.8%	53.7% ,28.3%	53.7% ,28.3%
ϵ=6%	3.5e-04 ,1.2e-04	6.1e-04 ,1.2e-04	3.1e-04 ,6.1e-05	6.1e-04 ,6.1e-05	1.3e-04 ,—	1.5e-04 ,3.1e-05
	33.3% ,16.5%	33.3% ,16.5%	41.2% ,21.1%	41.2% ,21.1%	56.4% ,30.1%	56.4% ,30.1%
ϵ=7%	3.5e-04 ,1.2e-04	6.1e-04 ,1.2e-04	2.9e-04 ,6.1e-05	3.1e-04 ,6.1e-05	1.1e-04 ,—	1.5e-04 ,3.1e-05
	34.7% ,17.5%	34.7% ,17.5%	43% ,22.4%	43% ,22.4%	58.7% ,31.8%	58.7% ,31.8%
ϵ=8%	3.5e-04 ,1.2e-04	6.1e-04 ,1.2e-04	2.6e-04 ,6.1e-05	3.1e-04 ,6.1e-05	9.9e-05 ,—	1.5e-04 ,3.1e-05
	36% ,18.5%	36% ,18.5%	44.5% ,23.6%	44.5% ,23.6%	60.7% ,33.3%	60.7% ,33.3%
ϵ=9%	3.5e-04 ,1.2e-04	6.1e-04 ,1.2e-04	2.5e-04 ,6.1e-05	3.1e-04 ,6.1e-05	8.7e-05 ,—	1.5e-04 ,1.5e-05
	37.2% ,19.5%	37.2% ,19.5%	46% ,24.9%	46% ,24.9%	62.4% ,34.9%	62.4% ,34.9%
ϵ=10%	3.5e-04 ,6.1e-05	6.1e-04 ,6.1e-05	2.3e-04 ,—	3.1e-04 ,6.1e-05	7.7e-05 ,—	1.5e-04 ,1.5e-05
	38.2% ,20.6%	38.2% ,20.6%	47.3% ,26.2%	47.3% ,26.2%	64% ,36.4%	64% ,36.4%
ϵ=11%	3.4e-04 ,6.1e-05	6.1e-04 ,6.1e-05	2.1e-04 ,—	3.1e-04 ,6.1e-05	6.8e-05 ,—	7.6e-05 ,1.5e-05
	39.2% ,21.6%	39.2% ,21.6%	48.6% ,27.5%	48.6% ,27.5%	65.5% ,37.9%	65.5% ,37.9%
ϵ=12%	3.3e-04 ,6.1e-05	6.1e-04 ,6.1e-05	2.0e-04 ,—	3.1e-04 ,3.1e-05	6.1e-05 ,—	7.6e-05 ,7.6e-06
	40.2% ,22.7%	40.2% ,22.7%	49.7% ,28.9%	49.7% ,28.9%	66.9% ,39.5%	66.9% ,39.5%
ϵ=13%	3.2e-04 ,6.1e-05	6.1e-04 ,6.1e-05	1.9e-04 ,—	3.1e-04 ,3.1e-05	5.4e-05 ,—	7.6e-05 ,7.6e-06
	41% ,23.8%	41% ,23.8%	50.8% ,30.2%	50.8% ,30.2%	68.1% ,41.1%	68.1% ,41.1%
ϵ=14%	3.0e-04 ,6.1e-05	6.1e-04 ,6.1e-05	1.8e-04 ,—	3.1e-04 ,3.1e-05	4.9e-05 ,—	7.6e-05 ,7.6e-06
	41.9% ,24.9%	41.9% ,24.9%	51.8% ,31.6%	51.8% ,31.6%	69.3% ,42.6%	69.3% ,42.6%
ϵ=15%	2.9e-04 ,—	6.1e-04 ,6.1e-05	1.7e-04 ,—	3.1e-04 ,3.1e-05	4.4e-05 ,—	7.6e-05 ,3.8e-06
	42.7% ,26.1%	42.7% ,26.1%	52.8% ,32.9%	52.8% ,32.9%	70.5% ,44.2%	70.5% ,44.2%
ϵ=16%	2.8e-04 ,—	3.1e-04 ,6.1e-05	1.6e-04 ,—	3.1e-04 ,1.5e-05	4.0e-05 ,—	7.6e-05 ,3.8e-06
	43.4% ,27.4%	43.4% ,27.4%	53.7% ,34.3%	53.7% ,34.3%	71.5% ,45.6%	71.5% ,45.6%
ϵ=17%	2.7e-04 ,—	3.1e-04 ,3.1e-05	1.5e-04 ,—	3.1e-04 ,1.5e-05	3.6e-05 ,—	7.6e-05 ,3.8e-06
	44.1% ,28.7%	44.1% ,28.7%	54.6% ,35.6%	54.6% ,35.6%	72.5% ,46.7%	72.5% ,46.7%
ϵ=18%	2.6e-04 ,—	3.1e-04 ,3.1e-05	1.4e-04 ,—	1.5e-04 ,1.5e-05	3.2e-05 ,—	3.8e-05 ,3.8e-06
	44.8% ,30%	44.8% ,30%	55.5% ,37.1%	55.5% ,37.1%	73.5% ,47.5%	73.5% ,47.5%
ϵ=19%	2.5e-04 ,—	3.1e-04 ,3.1e-05	1.3e-04 ,—	1.5e-04 ,1.5e-05	2.9e-05 ,—	3.8e-05 ,1.9e-06
	45.5% ,31.3%	45.5% ,31.3%	56.3% ,38.6%	56.3% ,38.6%	74.5% ,48.4%	74.5% ,48.4%
ϵ=20%	2.4e-04 ,—	3.1e-04 ,3.1e-05	1.3e-04 ,—	1.5e-04 ,7.6e-06	2.7e-05 ,—	3.8e-05 ,1.9e-06
	46.2% ,32.7%	46.2% ,32.7%	57% ,40.1%	57% ,40.1%	75.3% ,49.1%	75.3% ,49.1%

Table 5: (US Treasuries) Equivalent ambiguity radius ρ and optimal annualized return R at a given pair of (ϵ, δ).

$\left(\begin{smallmatrix}\rho_t, \rho_n\\ R_t, R_n\end{smallmatrix}\right)$	Bond: -δ = -5.18%		Bond: -δ = -6.58%		Bond: -δ = -9.98%	
	4^{th} order	2^{nd} order	4^{th} order	2^{nd} order	4^{th} order	2^{nd} order
ϵ=1%	1.0e-10 ,1.5e-08	Inf ,1.5e-08	3.0e-06 ,5.8e-11	1.9e-05 ,5.8e-11	2.0e-06 ,5.7e-14	9.5e-06 ,5.7e-14
	Inf ,-2.7%	Inf ,-2.7%	-2.09% ,-2.58%	-2.09% ,-2.58%	-1.09% ,-2.57%	-1.09% ,-2.57%
ϵ=2%	3.0e-06 ,1.9e-09	1.9e-05 ,1.9e-09	2.0e-06 ,2.8e-14	9.5e-06 ,2.8e-14	2.0e-06 ,2.8e-14	4.8e-06 ,2.8e-14
	-1.65% ,-2.63%	-1.65% ,-2.63%	-1.07% ,-2.57%	-1.07% ,-2.57%	0.143% ,-2.57%	0.143% ,-2.57%
ϵ=3%	2.0e-06 ,2.9e-11	9.5e-06 ,2.9e-11	2.0e-06 ,2.8e-14	9.5e-06 ,2.8e-14	1.0e-08 ,2.8e-14	2.4e-06 ,2.8e-14
	-1.05% ,-2.58%	-1.05% ,-2.58%	-0.393% ,-2.57%	-0.393% ,-2.57%	1.1% ,-2.57%	1.1% ,-2.57%
ϵ=4%	2.0e-06 ,2.8e-14	9.5e-06 ,2.8e-14	2.0e-06 ,2.8e-14	4.8e-06 ,2.8e-14	1.0e-08 ,2.8e-14	1.2e-06 ,2.8e-14
	-0.58% ,-2.57%	-0.58% ,-2.57%	0.174% ,-2.57%	0.174% ,-2.57%	1.93% ,-2.57%	1.93% ,-2.57%
ϵ=5%	2.0e-06 ,2.8e-14	4.8e-06 ,2.8e-14	2.0e-06 ,2.8e-14	4.8e-06 ,2.8e-14	1.0e-08 ,2.3e-10	1.5e-07 ,2.3e-10
	-0.17% ,-2.57%	-0.17% ,-2.57%	0.677% ,-2.57%	0.677% ,-2.57%	2.68% ,-2.59%	2.68% ,-2.59%
ϵ=6%	2.0e-06 ,5.7e-14	4.8e-06 ,5.7e-14	1.0e-08 ,2.8e-14	2.4e-06 ,2.8e-14	1.0e-08 ,4.7e-10	3.6e-14 ,4.7e-10
	0.201% ,-2.57%	0.201% ,-2.57%	1.14% ,-2.57%	1.14% ,-2.57%	3.17% ,-2.6%	3.17% ,-2.6%
ϵ=7%	2.0e-06 ,2.8e-14	4.8e-06 ,2.8e-14	1.0e-08 ,2.8e-14	1.2e-06 ,2.8e-14	1.0e-08 ,9.3e-10	3.6e-14 ,9.3e-10
	0.545% ,-2.57%	0.545% ,-2.57%	1.57% ,-2.57%	1.57% ,-2.57%	3.17% ,-2.6%	3.17% ,-2.6%
ϵ=8%	1.0e-08 ,2.8e-14	2.4e-06 ,2.8e-14	1.0e-08 ,2.8e-14	6.0e-07 ,2.8e-14	1.0e-08 ,9.3e-10	3.6e-14 ,9.3e-10
	0.868% ,-2.57%	0.868% ,-2.57%	1.98% ,-2.57%	1.98% ,-2.57%	3.17% ,-2.61%	3.17% ,-2.61%
ϵ=9%	1.0e-08 ,1.9e-09	2.4e-06 ,1.9e-09	1.0e-08 ,2.8e-14	3.0e-07 ,2.8e-14	1.0e-08 ,9.3e-10	3.6e-14 ,9.3e-10
	1.17% ,-2.62%	1.17% ,-2.62%	2.36% ,-2.57%	2.36% ,-2.57%	3.17% ,-2.61%	3.17% ,-2.61%
ϵ=10%	1.0e-08 ,4.7e-10	2.4e-06 ,4.7e-10	1.0e-08 ,1.9e-09	1.5e-07 ,1.9e-09	1.0e-08 ,1.9e-09	3.6e-14 ,1.9e-09
	1.47% ,-2.6%	1.47% ,-2.6%	2.74% ,-2.62%	2.74% ,-2.62%	3.17% ,-2.62%	3.17% ,-2.62%
ϵ=11%	1.0e-08 ,2.3e-10	1.2e-06 ,2.3e-10	1.0e-08 ,9.3e-10	2.3e-09 ,9.3e-10	1.0e-08 ,1.9e-09	3.6e-14 ,1.9e-09
	1.75% ,-2.59%	1.75% ,-2.59%	3.09% ,-2.61%	3.09% ,-2.61%	3.17% ,-2.63%	3.17% ,-2.63%
ϵ=12%	1.0e-08 ,2.8e-14	6.0e-07 ,2.8e-14	1.0e-08 ,9.3e-10	3.6e-14 ,9.3e-10	1.0e-08 ,3.7e-09	3.6e-14 ,3.7e-09
	2.02% ,-2.57%	2.02% ,-2.57%	3.17% ,-2.61%	3.17% ,-2.61%	3.17% ,-2.63%	3.17% ,-2.63%
ϵ=13%	1.0e-08 ,2.8e-14	6.0e-07 ,2.8e-14	1.0e-08 ,9.3e-10	7.1e-14 ,9.3e-10	1.0e-08 ,3.7e-09	3.6e-14 ,3.7e-09
	2.28% ,-2.57%	2.28% ,-2.57%	3.17% ,-2.6%	3.17% ,-2.6%	3.17% ,-2.63%	3.17% ,-2.63%
ϵ=14%	1.0e-08 ,2.8e-14	3.0e-07 ,2.8e-14	1.0e-08 ,9.3e-10	3.6e-14 ,9.3e-10	1.0e-08 ,3.7e-09	3.6e-14 ,3.7e-09
	2.54% ,-2.57%	2.54% ,-2.57%	3.17% ,-2.6%	3.17% ,-2.6%	3.17% ,-2.64%	3.17% ,-2.64%
ϵ=15%	1.0e-08 ,2.8e-14	7.5e-08 ,2.8e-14	1.0e-08 ,9.3e-10	3.6e-14 ,9.3e-10	1.0e-08 ,3.7e-09	3.6e-14 ,3.7e-09
	2.78% ,-2.57%	2.78% ,-2.57%	3.17% ,-2.61%	3.17% ,-2.61%	3.17% ,-2.64%	3.17% ,-2.64%
ϵ=16%	1.0e-08 ,2.3e-10	9.3e-09 ,2.3e-10	1.0e-08 ,9.3e-10	3.6e-14 ,9.3e-10	1.0e-08 ,3.7e-09	3.6e-14 ,3.7e-09
	3.03% ,-2.59%	3.03% ,-2.59%	3.17% ,-2.61%	3.17% ,-2.61%	3.17% ,-2.64%	3.17% ,-2.64%
ϵ=17%	1.0e-08 ,2.3e-10	3.6e-14 ,2.3e-10	1.0e-08 ,9.3e-10	3.6e-14 ,9.3e-10	1.0e-08 ,3.7e-09	3.6e-14 ,3.7e-09
	3.17% ,-2.59%	3.17% ,-2.59%	3.17% ,-2.61%	3.17% ,-2.61%	3.17% ,-2.64%	3.17% ,-2.64%
ϵ=18%	1.0e-08 ,4.7e-10	3.6e-14 ,4.7e-10	1.0e-08 ,1.9e-09	3.6e-14 ,1.9e-09	1.0e-08 ,3.7e-09	3.6e-14 ,3.7e-09
	3.17% ,-2.6%	3.17% ,-2.6%	3.17% ,-2.62%	3.17% ,-2.62%	3.17% ,-2.64%	3.17% ,-2.64%
ϵ=19%	1.0e-08 ,4.7e-10	7.1e-14 ,4.7e-10	1.0e-08 ,1.9e-09	3.6e-14 ,1.9e-09	1.0e-08 ,3.7e-09	3.6e-14 ,3.7e-09
	3.17% ,-2.6%	3.17% ,-2.6%	3.17% ,-2.62%	3.17% ,-2.62%	3.17% ,-2.64%	3.17% ,-2.64%
ϵ=20%	1.0e-08 ,9.3e-10	3.6e-14 ,9.3e-10	1.0e-08 ,1.9e-09	3.6e-14 ,1.9e-09	1.0e-08 ,3.7e-09	3.6e-14 ,3.7e-09
	3.17% ,-2.61%	3.17% ,-2.61%	3.17% ,-2.63%	3.17% ,-2.63%	3.17% ,-2.64%	3.17% ,-2.64%

Table 6: (Currencies) Equivalent ambiguity radius ρ and optimal annualized return R at a given pair of (ϵ, δ).

$\binom{\rho_t, \rho_n}{R_t, R_n}$	FX: -δ = -1.16%		FX: -δ = -1.40%		FX: -δ = -1.89%	
	4^{th} order	2^{nd} order	4^{th} order	2^{nd} order	4^{th} order	2^{nd} order
ϵ=1%	Inf ,1.2e-04	Inf ,1.2e-04	3.5e-04 ,6.1e-05	6.1e-04 ,6.1e-05	2.1e-04 ,3.1e-05	3.1e-04 ,3.1e-05
	-Inf ,1.76%	-Inf ,1.76%	0.17% ,2.89%	0.17% ,2.89%	3.21% ,4.68%	3.21% ,4.68%
ϵ=2%	3.5e-04 ,6.1e-05	6.1e-04 ,1.2e-04	2.6e-04 ,6.1e-05	3.1e-04 ,6.1e-05	1.4e-04 ,3.1e-05	1.5e-04 ,3.1e-05
	0.21% ,2.53%	0.21% ,2.53%	2.28% ,3.61%	2.28% ,3.61%	4.85% ,5.47%	4.85% ,5.47%
ϵ=3%	3.0e-04 ,6.1e-05	6.1e-04 ,6.1e-05	2.1e-04 ,6.1e-05	3.1e-04 ,6.1e-05	1.0e-04 ,3.1e-05	1.5e-04 ,3.1e-05
	1.59% ,3.04%	1.59% ,3.04%	3.23% ,4.14%	3.23% ,4.14%	5.87% ,6.04%	5.87% ,6.04%
ϵ=4%	2.6e-04 ,6.1e-05	3.1e-04 ,6.1e-05	1.8e-04 ,3.1e-05	3.1e-04 ,6.1e-05	8.3e-05 ,1.5e-05	1.5e-04 ,1.5e-05
	2.31% ,3.46%	2.31% ,3.46%	3.9% ,4.59%	3.9% ,4.59%	6.66% ,6.52%	6.66% ,6.52%
ϵ=5%	2.3e-04 ,6.1e-05	3.1e-04 ,6.1e-05	1.5e-04 ,3.1e-05	3.1e-04 ,3.1e-05	6.8e-05 ,1.5e-05	7.6e-05 ,1.5e-05
	2.84% ,3.83%	2.84% ,3.83%	4.43% ,4.98%	4.43% ,4.98%	7.3% ,6.96%	7.3% ,6.96%
ϵ=6%	2.0e-04 ,6.1e-05	3.1e-04 ,6.1e-05	1.4e-04 ,3.1e-05	1.5e-04 ,3.1e-05	5.7e-05 ,7.6e-06	7.6e-05 ,7.6e-06
	3.26% ,4.17%	3.26% ,4.17%	4.87% ,5.34%	4.87% ,5.34%	7.85% ,7.37%	7.85% ,7.37%
ϵ=7%	1.9e-04 ,6.1e-05	3.1e-04 ,6.1e-05	1.2e-04 ,3.1e-05	1.5e-04 ,3.1e-05	5.0e-05 ,—	7.6e-05 ,7.6e-06
	3.62% ,4.5%	3.62% ,4.5%	5.26% ,5.68%	5.26% ,5.68%	8.33% ,7.77%	8.33% ,7.77%
ϵ=8%	1.7e-04 ,3.1e-05	3.1e-04 ,3.1e-05	1.1e-04 ,3.1e-05	1.5e-04 ,3.1e-05	4.4e-05 ,—	7.6e-05 ,3.8e-06
	3.93% ,4.81%	3.93% ,4.81%	5.6% ,6%	5.6% ,6%	8.77% ,8.15%	8.77% ,8.15%
ϵ=9%	1.6e-04 ,3.1e-05	3.1e-04 ,3.1e-05	1.0e-04 ,1.5e-05	1.5e-04 ,1.5e-05	3.9e-05 ,—	7.6e-05 ,3.8e-06
	4.21% ,5.1%	4.21% ,5.1%	5.91% ,6.31%	5.91% ,6.31%	9.16% ,8.49%	9.16% ,8.49%
ϵ=10%	1.5e-04 ,3.1e-05	3.1e-04 ,3.1e-05	9.5e-05 ,1.5e-05	1.5e-04 ,1.5e-05	3.6e-05 ,—	7.6e-05 ,1.9e-06
	4.47% ,5.39%	4.47% ,5.39%	6.2% ,6.62%	6.2% ,6.62%	9.48% ,8.79%	9.48% ,8.79%
ϵ=11%	1.4e-04 ,3.1e-05	3.1e-04 ,3.1e-05	8.8e-05 ,1.5e-05	1.5e-04 ,1.5e-05	1.0e-08 ,—	3.8e-05 ,1.9e-06
	4.7% ,5.68%	4.7% ,5.68%	6.46% ,6.94%	6.46% ,6.94%	9.76% ,9.08%	9.76% ,9.08%
ϵ=12%	1.4e-04 ,3.1e-05	1.5e-04 ,3.1e-05	8.1e-05 ,1.5e-05	1.5e-04 ,1.5e-05	3.1e-05 ,—	3.8e-05 ,9.5e-07
	4.92% ,5.96%	4.92% ,5.96%	6.7% ,7.25%	6.7% ,7.25%	10% ,9.36%	10% ,9.36%
ϵ=13%	1.3e-04 ,1.5e-05	1.5e-04 ,1.5e-05	7.6e-05 ,—	1.5e-04 ,7.6e-06	2.9e-05 ,—	3.8e-05 ,9.5e-07
	5.12% ,6.24%	5.12% ,6.24%	6.93% ,7.57%	6.93% ,7.57%	10.2% ,9.64%	10.2% ,9.64%
ϵ=14%	1.2e-04 ,1.5e-05	1.5e-04 ,1.5e-05	7.1e-05 ,—	1.5e-04 ,7.6e-06	2.7e-05 ,—	3.8e-05 ,4.8e-07
	5.3% ,6.53%	5.3% ,6.53%	7.14% ,7.9%	7.14% ,7.9%	10.4% ,9.91%	10.4% ,9.91%
ϵ=15%	1.2e-04 ,1.5e-05	1.5e-04 ,1.5e-05	6.7e-05 ,—	7.6e-05 ,3.8e-06	2.6e-05 ,—	3.8e-05 ,2.4e-07
	5.48% ,6.82%	5.48% ,6.82%	7.35% ,8.21%	7.35% ,8.21%	10.6% ,10.1%	10.6% ,10.1%
ϵ=16%	1.1e-04 ,1.5e-05	1.5e-04 ,1.5e-05	6.3e-05 ,—	7.6e-05 ,3.8e-06	2.4e-05 ,—	3.8e-05 ,1.2e-07
	5.65% ,7.12%	5.65% ,7.12%	7.54% ,8.51%	7.54% ,8.51%	10.8% ,10.2%	10.8% ,10.2%
ϵ=17%	1.1e-04 ,7.6e-06	1.5e-04 ,7.6e-06	5.9e-05 ,—	7.6e-05 ,1.9e-06	2.3e-05 ,—	3.8e-05 ,6.0e-08
	5.81% ,7.43%	5.81% ,7.43%	7.73% ,8.79%	7.73% ,8.79%	11% ,10.3%	11% ,10.3%
ϵ=18%	1.0e-04 ,—	1.5e-04 ,7.6e-06	5.6e-05 ,—	7.6e-05 ,1.9e-06	2.2e-05 ,—	3.8e-05 ,3.0e-08
	5.96% ,7.75%	5.96% ,7.75%	7.9% ,9.06%	7.9% ,9.06%	11.1% ,10.4%	11.1% ,10.4%
ϵ=19%	9.7e-05 ,—	1.5e-04 ,7.6e-06	5.4e-05 ,—	7.6e-05 ,9.5e-07	2.0e-05 ,—	3.8e-05 ,1.5e-08
	6.11% ,8.08%	6.11% ,8.08%	8.07% ,9.34%	8.07% ,9.34%	11.3% ,10.4%	11.3% ,10.4%
ϵ=20%	9.3e-05 ,—	1.5e-04 ,3.8e-06	5.1e-05 ,—	7.6e-05 ,9.5e-07	1.9e-05 ,—	3.8e-05 ,1.5e-08
	6.25% ,8.39%	6.25% ,8.39%	8.24% ,9.61%	8.24% ,9.61%	11.4% ,10.5%	11.4% ,10.5%

Table 7: (Commodities) Equivalent ambiguity radius ρ and optimal annualized return R at a given pair of (ϵ, δ).

$\binom{\rho_t, \rho_n}{R_t, R_n}$	Commodity: -δ = -3.60%		Commodity: -δ = -4.40%		Commodity: -δ = -6.64%	
	4^{th} order	2^{nd} order	4^{th} order	2^{nd} order	4^{th} order	2^{nd} order
ϵ=1%	1.8e-04 ,3.7e-09	3.1e-04 ,3.7e-09	1.3e-04 ,3.7e-09	1.5e-04 ,3.7e-09	7.1e-05 ,1.9e-09	7.6e-05 ,1.9e-09
	11.1% ,4.52%	11.1% ,4.52%	14% ,4.54%	14% ,4.54%	21.3% ,4.58%	21.3% ,4.58%
ϵ=2%	1.3e-04 ,3.7e-09	1.5e-04 ,3.7e-09	9.6e-05 ,3.7e-09	1.5e-04 ,3.7e-09	4.8e-05 ,1.9e-09	7.6e-05 ,1.9e-09
	14.1% ,4.53%	14.1% ,4.53%	17.3% ,4.55%	17.3% ,4.55%	26% ,4.59%	26% ,4.59%
ϵ=3%	1.1e-04 ,3.7e-09	1.5e-04 ,3.7e-09	8.3e-05 ,1.9e-09	1.5e-04 ,1.9e-09	3.7e-05 ,1.9e-09	7.6e-05 ,1.9e-09
	16% ,4.54%	16% ,4.54%	19.5% ,4.56%	19.5% ,4.56%	28.7% ,4.6%	28.7% ,4.6%
ϵ=4%	9.6e-05 ,3.7e-09	1.5e-04 ,3.7e-09	7.2e-05 ,1.9e-09	7.6e-05 ,1.9e-09	3.1e-05 ,9.3e-10	3.8e-05 ,9.3e-10
	17.4% ,4.54%	17.4% ,4.54%	21.2% ,4.56%	21.2% ,4.56%	30.5% ,4.61%	30.5% ,4.61%
ϵ=5%	8.7e-05 ,3.7e-09	1.5e-04 ,3.7e-09	6.5e-05 ,1.9e-09	7.6e-05 ,1.9e-09	2.6e-05 ,9.3e-10	3.8e-05 ,9.3e-10
	18.6% ,4.55%	18.6% ,4.55%	22.6% ,4.57%	22.6% ,4.57%	31.8% ,4.62%	31.8% ,4.62%
ϵ=6%	8.0e-05 ,1.9e-09	1.5e-04 ,1.9e-09	5.8e-05 ,1.9e-09	7.6e-05 ,1.9e-09	2.2e-05 ,9.3e-10	3.8e-05 ,9.3e-10
	19.6% ,4.55%	19.6% ,4.55%	23.9% ,4.57%	23.9% ,4.57%	32.9% ,4.62%	32.9% ,4.62%
ϵ=7%	7.4e-05 ,1.9e-09	1.5e-04 ,1.9e-09	5.3e-05 ,1.9e-09	7.6e-05 ,1.9e-09	2.0e-05 ,9.3e-10	3.8e-05 ,9.3e-10
	20.6% ,4.56%	20.6% ,4.56%	25% ,4.58%	25% ,4.58%	33.9% ,4.63%	33.9% ,4.63%
ϵ=8%	7.1e-05 ,1.9e-09	7.6e-05 ,1.9e-09	4.8e-05 ,1.9e-09	7.6e-05 ,1.9e-09	2.0e-05 ,4.7e-10	1.9e-05 ,4.7e-10
	21.4% ,4.56%	21.4% ,4.56%	26% ,4.59%	26% ,4.59%	34.7% ,4.64%	34.7% ,4.64%
ϵ=9%	6.5e-05 ,1.9e-09	7.6e-05 ,1.9e-09	4.5e-05 ,1.9e-09	7.6e-05 ,1.9e-09	1.6e-05 ,4.7e-10	1.9e-05 ,4.7e-10
	22.1% ,4.57%	22.1% ,4.57%	26.8% ,4.59%	26.8% ,4.59%	35.5% ,4.65%	35.5% ,4.65%
ϵ=10%	6.5e-05 ,1.9e-09	7.6e-05 ,1.9e-09	4.1e-05 ,1.9e-09	7.6e-05 ,1.9e-09	1.4e-05 ,4.7e-10	1.9e-05 ,4.7e-10
	22.8% ,4.57%	22.8% ,4.57%	27.6% ,4.6%	27.6% ,4.6%	36.2% ,4.65%	36.2% ,4.65%
ϵ=11%	5.8e-05 ,1.9e-09	7.6e-05 ,1.9e-09	3.7e-05 ,9.3e-10	7.6e-05 ,9.3e-10	1.4e-05 ,2.3e-10	1.9e-05 ,2.3e-10
	23.4% ,4.58%	23.4% ,4.58%	28.2% ,4.6%	28.2% ,4.6%	36.8% ,4.66%	36.8% ,4.66%
ϵ=12%	5.7e-05 ,1.9e-09	7.6e-05 ,1.9e-09	3.7e-05 ,9.3e-10	7.6e-05 ,9.3e-10	1.1e-05 ,2.3e-10	1.9e-05 ,2.3e-10
	24% ,4.58%	24% ,4.58%	28.7% ,4.61%	28.7% ,4.61%	37.4% ,4.67%	37.4% ,4.67%
ϵ=13%	5.3e-05 ,1.9e-09	7.6e-05 ,1.9e-09	3.6e-05 ,9.3e-10	3.8e-05 ,9.3e-10	1.1e-05 ,1.2e-10	1.9e-05 ,1.2e-10
	24.6% ,4.59%	24.6% ,4.59%	29.2% ,4.61%	29.2% ,4.61%	38% ,4.68%	38% ,4.68%
ϵ=14%	5.2e-05 ,1.9e-09	7.6e-05 ,1.9e-09	3.3e-05 ,9.3e-10	3.8e-05 ,9.3e-10	9.0e-06 ,5.8e-11	1.9e-05 ,5.8e-11
	25.2% ,4.59%	25.2% ,4.59%	29.7% ,4.62%	29.7% ,4.62%	38.5% ,4.69%	38.5% ,4.69%
ϵ=15%	5.0e-05 ,1.9e-09	7.6e-05 ,1.9e-09	3.2e-05 ,9.3e-10	3.8e-05 ,9.3e-10	9.0e-06 ,2.9e-11	9.5e-06 ,2.9e-11
	25.7% ,4.6%	25.7% ,4.6%	30.1% ,4.62%	30.1% ,4.62%	39% ,4.69%	39% ,4.69%
ϵ=16%	4.8e-05 ,9.3e-10	7.6e-05 ,9.3e-10	3.1e-05 ,9.3e-10	3.8e-05 ,9.3e-10	9.0e-06 ,1.5e-11	9.5e-06 ,1.5e-11
	26.2% ,4.6%	26.2% ,4.6%	30.5% ,4.63%	30.5% ,4.63%	39.5% ,4.7%	39.5% ,4.7%
ϵ=17%	4.5e-05 ,9.3e-10	7.6e-05 ,9.3e-10	2.9e-05 ,4.7e-10	3.8e-05 ,4.7e-10	7.0e-06 ,2.3e-13	9.5e-06 ,2.3e-13
	26.6% ,4.61%	26.6% ,4.61%	30.8% ,4.64%	30.8% ,4.64%	39.9% ,4.71%	39.9% ,4.71%
ϵ=18%	4.3e-05 ,9.3e-10	7.6e-05 ,9.3e-10	2.8e-05 ,4.7e-10	3.8e-05 ,4.7e-10	7.0e-06 ,1.4e-14	9.5e-06 ,1.4e-14
	27% ,4.61%	27% ,4.61%	31.2% ,4.64%	31.2% ,4.64%	40.4% ,4.71%	40.4% ,4.71%
ϵ=19%	4.2e-05 ,9.3e-10	7.6e-05 ,9.3e-10	2.8e-05 ,4.7e-10	3.8e-05 ,4.7e-10	6.0e-06 ,1.4e-14	9.5e-06 ,1.4e-14
	27.4% ,4.62%	27.4% ,4.62%	31.5% ,4.65%	31.5% ,4.65%	40.8% ,4.71%	40.8% ,4.71%
ϵ=20%	4.1e-05 ,9.3e-10	7.6e-05 ,9.3e-10	2.6e-05 ,4.7e-10	3.8e-05 ,4.7e-10	6.0e-06 ,1.4e-14	9.5e-06 ,1.4e-14
	27.7% ,4.62%	27.7% ,4.62%	31.8% ,4.66%	31.8% ,4.66%	41.2% ,4.71%	41.2% ,4.71%

A.2 PROOFS

We are going to present proofs about the theorems stated in the main paper in details.

We first introduce the definition of *Regular Measure of Deviation*, which was introduced in Rockafellar & Uryasev (2013) (also see Gotoh et al. (2018)) and is useful in the proof of Theorem 3.1.

Definition A.1. *Given any probability space* $(\Omega, \mathcal{F}, P)$*, let* $\mathcal{L}^2(\Omega)$ *denote the space of square-integrable random variables, i.e.,* $\mathbb{E}_{\mathrm{r}\sim P}[X^2] < \infty$*. A functional* $\mathcal{D} : \mathcal{L}^2(\Omega) \to [0, \infty]$ *is said to be a regular measure of deviation if it is closed convex and satisfies:*

1. $\mathcal{D}(c) = 0$ *for any constant* $c \in \mathbb{R}$*;*
2. $\mathcal{D}(Z) > 0$ *for any (non-constant) random variable* $Z \in \mathcal{L}^2(\Omega)$.

With the definition, we are able to complete the proof of Theorem 3.1.

Proof of Theorem 3.1. We need to check that the quantity $\mathcal{D}_{\eta_2,\phi,P_0}(f(\boldsymbol{x},\mathbf{r})|\mathbb{E}[f(\boldsymbol{x},\mathbf{r})])$ defined in Theorem 3.1 satisfies the conditions stated in definition A.1. It is easy and is thus omitted.

Write $\min_{P\in\mathbb{U}}\mathbb{E}_{\mathrm{r}\sim P}\left[f(\boldsymbol{x},\mathbf{r})\right]$ as

$$\min_{P}\mathbb{E}_{\mathrm{r}\sim P}\left[f(\boldsymbol{x},\mathbf{r})\right] \quad s.t. \quad D(P||P_0)\leq\rho.$$

Applying changing of measure, it is equivalent to solving

$$\min_{L}\mathbb{E}\left[f(\boldsymbol{x},\mathbf{r})L\right] \quad s.t. \quad \mathbb{E}\left[\phi(L)\right]\leq\rho \quad and \quad \mathbb{E}\left[L\right]=1. \tag{A.1}$$

Here, L is the likelihood ratio. The Lagrangian function is

$$l(L,\eta_1,\eta_2)=\mathbb{E}\left[f(\boldsymbol{x},\mathbf{r})L\right]+\eta_2(\mathbb{E}\left[\phi(L)\right]-\rho)+\eta_1(\mathbb{E}\left[L\right]-1).$$

Under mild conditions such that strong duality holds, Eq. (A.1) is equivalent to solving its dual $\max_{\eta_1\in\mathbb{R},\eta_2\geq 0}\min_L l(L,\eta_1,\eta_2)$, which is

$$\max_{\eta_1\in\mathbb{R},\eta_2\geq 0}\min_{L}\left\{\mathbb{E}\left[f(\boldsymbol{x},\mathbf{r})L\right]+\eta_2(\mathbb{E}\left[\phi(L)\right]-\rho)+\eta_1(\mathbb{E}\left[L\right]-1)\right\}. \tag{A.2}$$

Eq. (A.2) allows us to eliminate L. Now,

$$\begin{aligned}
&\max_{\eta_1\in\mathbb{R},\eta_2\geq 0}\min_{L}\left\{\mathbb{E}\left[f(\boldsymbol{x},\mathbf{r})L\right]+\eta_2(\mathbb{E}\left[\phi(L)\right]-\rho)+\eta_1(\mathbb{E}\left[L\right]-1)\right\}\\
=&\max_{\eta_1\in\mathbb{R},\eta_2\geq 0}-\eta_2\rho-\eta_1-\eta_2\max_{L}\left\{\mathbb{E}\left[-\frac{(f(\boldsymbol{x},\mathbf{r})+\eta_1)}{\eta_2}L\right]-\mathbb{E}\left[\phi(L)\right]\right\}.
\end{aligned}$$

Suppose that the maximization sign '$\max$' and the expectation sign $\mathbb{E}$ can be interchanged. Making use of the definition of conjugate $\phi^*(\cdot)$, we have

$$\begin{aligned}
&\max_{\eta_1\in\mathbb{R},\eta_2\geq 0}-\eta_2\rho-\eta_1-\eta_2\left\{\mathbb{E}\left[\max_{L}\left(-\frac{(f(\boldsymbol{x},\mathbf{r})+\eta_1)}{\eta_2}L-\phi(L)\right)\right]\right\}\\
=&\max_{\eta_1\in\mathbb{R},\eta_2\geq 0}-\eta_2\rho-\eta_1-\eta_2\left\{\mathbb{E}\left[\phi^*\left(-\frac{(f(\boldsymbol{x},\mathbf{r})+\eta_1)}{\eta_2}\right)\right]\right\}\\
=&-\min_{\eta_2\geq 0}\frac{\rho}{\eta_2}+\min_{\eta_1\in\mathbb{R}}\left\{\eta_1+\frac{1}{\eta_2}\mathbb{E}\left[\phi^*\left(-\eta_2(f(\boldsymbol{x},\mathbf{r})+\eta_1)\right)\right]\right\}\\
=&-\min_{\eta_2\geq 0}\left\{\frac{\rho}{\eta_2}+\mathcal{D}_{\eta_2,\phi,P_0}(f(\boldsymbol{x},\mathbf{r})|\mathbb{E}\left[f(\boldsymbol{x},\mathbf{r})\right])-\mathbb{E}\left[f(\boldsymbol{x},\mathbf{r})\right]\right\}\\
=&\mathbb{E}\left[f(\boldsymbol{x},\mathbf{r})\right]-\min_{\eta_2\geq 0}\left\{\frac{\rho}{\eta_2}+\mathcal{D}_{\eta_2,\phi,P_0}(f(\boldsymbol{x},\mathbf{r})|\mathbb{E}\left[f(\boldsymbol{x},\mathbf{r})\right])\right\}.
\end{aligned}$$

The second last equality holds, since if we let $\bar{\eta}_1=\eta_1-\mathbb{E}\left[f(\boldsymbol{x},\mathbf{r})\right]$, we have

$$\begin{aligned}
\mathcal{D}_{\eta_2,\phi,P_0}(f(\boldsymbol{x},\mathbf{r})|\mathbb{E}[f(\boldsymbol{x},\mathbf{r})]) &:= \min_{\eta_1}\left\{\eta_1+\frac{1}{\eta_2}\mathbb{E}\left[\phi^*\left(\eta_2(\mathbb{E}[f(\boldsymbol{x},\mathbf{r})]-f(\boldsymbol{x},\mathbf{r})-\eta_1)\right)\right]\right\}\\
&=\min_{\bar{\eta}_1}\left\{\bar{\eta}_1+\frac{1}{\eta_2}\mathbb{E}\left[\phi^*\left(\eta_2(-f(\boldsymbol{x},\mathbf{r})-\bar{\eta}_1)\right)\right]\right\}+\mathbb{E}\left[f(\boldsymbol{x},\mathbf{r})\right].
\end{aligned}$$

The second last equality holds by letting $\bar{\eta_1}=\eta_1-\mathbb{E}\left[f(\boldsymbol{x},\mathbf{r})\right]$. Lastly, we apply Result A.2 which guarantees that the maximization sign '$\max$' and the expectation sign $\mathbb{E}$ can be interchanged (see for example, Ben-Tal & Teboulle (2007) and Hu et al. (2013)). The proof is completed. ☐

Result A.2. *Let Ω be a σ-finite measure space, and let $\mathcal{Y}:=\mathcal{L}^p(\Omega,\mathcal{F},P)$, $p\in[1,+\infty]$. Let $g:\mathbb{R}\times\Omega\to(-\infty,+\infty]$ be a normal integrand, and define on $\mathcal{Y}$ the integral functional $I_g(y):=\int_\Omega g(y(\omega),\omega)P(d\omega)$. Then,*

$$\inf_{x\in\mathcal{Y}}\int_\Omega g(y(\omega),\omega)P(d\omega)=\int_\Omega\inf_s g(s,\omega)P(d\omega),$$

provided that the left-hand side is finite. Moreover,

$$\bar{y}\in\arg\min_{y\in\mathcal{Y}}I_g(y)=\bar{y}(\omega)\in\arg\min_{s\in\mathbb{R}}g(s,\omega), a.e.\ on\ \omega\in\Omega.$$

Proof of Result A.2. The proof can be found in Theorem 14.60 of "*Variational analysis* (Rockafellar & Wets, 2009)". $\square$

Followed by Theorem 3.1, we prove Lemma 3.2. What we are going to do is to express the quantity $\mathcal{D}_{\eta_2,\phi,P}(f(\boldsymbol{x},\mathbf{r})|\mathbb{E}\left[f(\boldsymbol{x},\mathbf{r})\right])$ as a series.

Proof of Lemma 3.2. Denoting $X = f(\boldsymbol{x},\mathbf{r}) - \mathbb{E}\left[f(\boldsymbol{x},\mathbf{r})\right]$, we recall that

$$\begin{aligned}\mathcal{D}_{\eta_2,\phi,P_0}(f(\boldsymbol{x},\mathbf{r})|\mathbb{E}\left[f(\boldsymbol{x},\mathbf{r})\right]) &= \min_{\eta_1}\left\{\eta_1 + \frac{1}{\eta_2}\mathbb{E}\left[\phi^*\left(\eta_2(\mathbb{E}\left[f(\boldsymbol{x},\mathbf{r})\right] - f(\boldsymbol{x},\mathbf{r}) - \eta_1)\right)\right]\right\}\\ &= \min_{\eta_1}\left\{\eta_1 + \frac{1}{\eta_2}\mathbb{E}\left[\phi^*\left(\eta_2(-X-\eta_1)\right)\right]\right\}.\end{aligned}$$

Applying Taylor expansion on $\phi^*(\cdot)$ around 0 under the assumption that $\mathbb{E}\left[X^k\right]$ exists for all $k \leq n$, the expansion of $\phi^*\left(\eta_2(-X-\eta_1)\right)$ up to order n is:

$$\begin{aligned}&\sum_{k=2}^{n}\frac{\eta_2^{k-1}(-1)^k\left(\phi^*\right)^{(k)}(0)}{k!}\mathbb{E}\left[\left(X+\eta_1\right)^k\right]\\ =&\sum_{k=1}^{n-1}\frac{\eta_2^{k}(-1)^{k+1}\left(\phi^*\right)^{(k+1)}(0)}{(k+1)!}\mathbb{E}\left[\left(X+\eta_1\right)^{k+1}\right].\end{aligned}$$

It remains to compute $\left(\phi^*\right)^{(k+1)}(\xi)$. We start from the definition of $\phi^*(\xi) = \sup_{z\geq 0}\{z\xi - \phi(z)\}$. Let $z(\xi)$ be the optimizer, then we obtain $\phi^*(\xi) = z(\xi)\xi - \phi(z(\xi))$. Differentiating it gives $(\phi^*)^{(1)}(\xi) = z^{(1)}(\xi)\xi + z(\xi) - \phi^{(1)}(z(\xi))z^{(1)}(\xi) = z(\xi)$. The last equality follows from the first order optimality condition, which gives that $\xi - \phi^{(1)}(z(\xi)) = 0$. As a result, $(\phi^*)^{(k+1)}(\xi) = z^{(k)}(\xi)$ and it suffices to find $z^{(k)}(\xi)$.

The term $z^{(k)}(\xi)$ can be found from the first order optimality condition. If we differentiate it on both sides w.r.t. ξ, we obtain $\phi^{(2)}(z(\xi))z^{(1)}(\xi) = 1$. This means that $z^{(1)}(\xi) = \frac{1}{\phi^{(2)}(z(\xi))}$ and hence $z^{(k)}(\xi) = \left[\frac{1}{\phi^{(2)}(z(\xi))}\right]^{(k-1)}$ for $k \geq 2$. Denoting η_1^* be the optimal solution to $\min_{\eta_1}\sum_{k=1}^{n-1} b_k\mathbb{E}\left[\left(X+\eta_1\right)^{k+1}\right]\eta_2^k$.

We then show that $z(0) = 1$. From the definition of $\phi^*(\xi)$, we know that $\phi^*(0) = 0$ so that $\phi(z(0)) = 0$. If $z(0) := a \neq 1$, then $\phi(x) = 0$ for any x lying between 1 and a according to the definition of ϕ. This means that $\phi^{(n)}(x) = 0$ for any $n \in \mathbb{N}$ and x is a point lying between 1 and a. Note that $\phi^{(2)}(1) \neq 0$ and ϕ is an ∞-differentiable function. Ee can find a point $\bar{x}$ lying in between 1 and a (indeed, close to 1) such that $\phi^{(2)}(\bar{x}) \neq 0$, so an contradiction arises. The proof of Lemma is completed. $\square$

We move to the proof of Corollary 3.3:

Proof of Corollary 3.3. We prove for the case when ϕ-divergence is chosen as Cressie Read divergence with $\theta > 2$. The case when ϕ-divergence is chosen as KL divergence is similar and is omitted. First, we compute the first four derivatives of ϕ-divergence when it is a Cressie Read divergence, which gives

$$\phi^{(1)}(t) = \frac{1-t^{\theta-1}}{1-\theta},\ \phi^{(2)}(t) = t^{\theta-2},\ \phi^{(3)}(t) = (\theta-2)t^{\theta-3},\ \phi^{(4)}(t) = (\theta-2)(\theta-3)t^{\theta-4}.$$

In the proof of Lemma 3.2, we know that $(\phi^*)^{(k)}(\cdot)$ can be obtained from $z^{(1)}(\cdot)$, $z^{(2)}(\cdot)$ and $z^{(3)}(\cdot)$. We omit detailed deviations, and we can obtain $z^{(1)}(0) = 1$, $z^{(2)}(0) = -(\theta-2)$ and $z^{(3)}(0) =$

$(\theta-2)(2\theta-3)$. Now, assuming $X = f(\boldsymbol{x}, \mathbf{r}) - \mathbb{E}[f(\boldsymbol{x}, \mathbf{r})]$, this means that

$$\begin{aligned}
&\mathcal{D}_{\eta_2,\phi,P_0}(f(\boldsymbol{x},\mathbf{r})|\mathbb{E}[f(\boldsymbol{x},\mathbf{r})]) \\
&= \min_{\eta_1} \left\{ \frac{\eta_2 \mathbb{E}\left[(X+\eta_1)^2\right]}{2!} \right. \\
&\quad \left. + \frac{\eta_2^2(\theta-2)\mathbb{E}\left[(X+\eta_1)^3\right]}{3!} + \frac{\eta_2^3(\theta-2)(2\theta-3)\mathbb{E}\left[(X+\eta_1)^4\right]}{4!} \right\}.
\end{aligned} \tag{A.3}$$

We need to show that the local optimizer η_1 of Eq. (A.3) is a root of a cubic equation. Indeed, the objective function is a polynomial of η_1 with degree 4, which can be rewritten as $a_0 + a_1\eta_1 + a_2\eta_1^2 + a_3\eta_1^3 + a_4\eta_1^4$, where

$$a_0 = \frac{\eta_2\mathbb{E}\left[X^2\right]}{2} + \frac{(\theta-2)\eta_2^2\mathbb{E}\left[X^3\right]}{6} + \frac{(\theta-2)(2\theta-3)\eta_2^3\mathbb{E}\left[X^4\right]}{24}, \tag{A.4}$$

$$a_1 = \frac{(\theta-2)\eta_2^2\mathbb{E}\left[3X^2\right]}{6} + \frac{(\theta-2)(2\theta-3)\eta_2^3\mathbb{E}\left[4X^3\right]}{24}, \tag{A.5}$$

$$a_2 = \frac{\eta_2}{2} + \frac{(\theta-2)(2\theta-3)\eta_2^3\mathbb{E}\left[6X^2\right]}{24}, \tag{A.6}$$

$$a_3 = \frac{(\theta-2)\eta_2^2}{6}, \tag{A.7}$$

$$a_4 = \frac{(\theta-2)(2\theta-3)\eta_2^3}{24}. \tag{A.8}$$

Since $\theta > 2$, we have $(\theta-2)(2\theta-3) > 0$ and so $\frac{(\theta-2)(2\theta-3)\eta_2^3}{4!} > 0$. We can therefore find a local minimum (global minimum). The local optimal points should satisfy $a_1 + 2a_2\eta_1 + 3a_3\eta_1^2 + 4a_4\eta_1^3 = 0$. It is not difficult to obtain Eq. (8) from Eqs. (6), (A.5), (A.6), (A.7) and (A.8).

It remains to show that the local optimal obtained is indeed a local minimum. This can be done by examining if $2a_2 + 6a_3\eta_1 + 12a_4\eta_1^2 \geq 0$ for local optimal η_1. Indeed,

$$\begin{aligned}
&2a_2 + 6a_3\eta_1 + 12a_4\eta_1^2 \\
=&2\eta_2\left(\frac{(\theta-2)(2\theta-3)}{4}\eta_1^2\eta_2^2 + \frac{\theta-2}{2}\eta_1\eta_2 + \frac{1}{2} + \frac{(\theta-2)(2\theta-3)\eta_2^2\mathbb{E}\left[X^2\right]}{4}\right) \\
=&\frac{2\eta_2}{4}\left((\theta-2)(2\theta-3)\left[\eta_1\eta_2 + \frac{1}{2\theta-3}\right]^2 + \frac{3\theta-4}{2\theta-3} + (\theta-2)(2\theta-3)\eta_2^2\mathbb{E}\left[X^2\right]\right) \geq 0.
\end{aligned}$$

Therefore, we can further simplify $\mathcal{D}_{\eta_2,\phi,P_0}(f(\boldsymbol{x},\mathbf{r})|\mathbb{E}_{\mathbf{r}\sim P_0}[f(\boldsymbol{x},\mathbf{r})])$ if η_1 exists for a real solution of $a_1 + 2a_2\eta_1 + 3a_3\eta_1^2 + 4a_4\eta_1^3 = 0$ and can be expressed explicitly.

The remaining in the proof shows how to find the explicit formula of the local optimizer η_1. We examine $\triangle$ for a triple polynomial equation. When $\triangle > 0$, we know that the local optimal η_1 can be expressed as $u - \frac{p}{3u} - \frac{a_3}{4a_4}$ such that

$$p = \frac{a_2}{2a_4} - \frac{3a_3^2}{16a_4^2}, \quad q = \frac{a_3^3}{32a_4^3} - \frac{a_2a_3}{8a_4^2} + \frac{a_1}{4a_4}, \quad \triangle = q^2 + \frac{4p^3}{27}, \quad u = \sqrt[3]{\frac{-q+\sqrt{\triangle}}{2}},$$

which follows from Cardano formula for a cubic equation. This follows from the fact that $p \geq 0$, since

$$\begin{aligned}
p &= \frac{a_2}{2a_4} - \frac{3a_3^2}{16a_4^2} \\
&= \frac{1}{16a_4^2}\left[8\left(\frac{\eta_2}{2} + \frac{(\theta-2)(2\theta-3)\eta_2^3\mathbb{E}_{\mathbf{r}\sim P}\left[6X^2\right]}{24}\right)\left(\frac{(\theta-2)(2\theta-3)\eta_2^3}{24}\right) - 3\left(\frac{(\theta-2)\eta_2^2}{6}\right)^2\right] \\
&= \frac{\eta_2^4}{192a_4^2}\left[(\theta-2)(3\theta-4) + (\theta-2)^2(2\theta-3)^2\eta_2^2\mathbb{E}\left[X^2\right]\right] \geq 0.
\end{aligned}$$

This completes the proof. $\square$

Before presenting proofs of Theorem 3.5 and 3.6, it is necessary to study any properties of the function $f(\boldsymbol{x}) = a\sqrt{\boldsymbol{x}^T\Sigma\boldsymbol{x}} - \boldsymbol{x}^T b$ with $a > 0$ and a vector b. Indeed, it is a convex function. The proof is tedious and can be completed using Cauchy Schwartz inequality. Now, we are ready to present the proofs of Theorem 3.5 and Theorem 3.6. The proofs are presented sequentially.

Proof of Theorem 3.5. According to the symbols defined and let $b = \sqrt{\frac{2\rho}{\phi^{(2)}(1)}}$ for notational simplification, Eq. (9) is equivalent to

$$-\min_{\boldsymbol{x}\in\mathbb{X}}\left\{-\boldsymbol{x}^T\mu + b\sqrt{\boldsymbol{x}^T\Sigma\boldsymbol{x}}\right\} \Leftrightarrow \begin{cases} & -\min_{\boldsymbol{x}}\left\{-\boldsymbol{x}^T\mu + b\sqrt{\boldsymbol{x}^T\Sigma\boldsymbol{x}}\right\} \\ s.t. & \boldsymbol{x}^T\boldsymbol{e} = 1 \end{cases}.$$

Clearly, it is a convex optimization problem. Applying Lagrangian multiplier method, let us examine

$$-\mu + \frac{b\Sigma\boldsymbol{x}}{\sqrt{\boldsymbol{x}^T\Sigma\boldsymbol{x}}} + \lambda\boldsymbol{e} = 0, \tag{A.9}$$

$$\boldsymbol{x}^T\boldsymbol{e} = 1. \tag{A.10}$$

Let $\boldsymbol{x}(\lambda) = \frac{\Sigma^{-1}(\mu-\lambda\boldsymbol{e})}{\mu^T\Sigma^{-1}\boldsymbol{e}-\lambda\boldsymbol{e}^T\Sigma^{-1}\boldsymbol{e}}$. From the assumption of λ, λ satisfies

$$\boldsymbol{e}^T\Sigma^{-1}\boldsymbol{e}\cdot\lambda^2 - 2\lambda\mu^T\Sigma^{-1}\boldsymbol{e} + (\mu^T\Sigma^{-1}\mu - b^2) = 0$$

for two real roots, where we denote them as

$$\lambda_+ = \frac{\mu^T\Sigma^{-1}\boldsymbol{e} + \sqrt{(\mu^T\Sigma^{-1}\boldsymbol{e})^2 - (\boldsymbol{e}^T\Sigma^{-1}\boldsymbol{e})(\mu^T\Sigma^{-1}\mu - b^2)}}{\boldsymbol{e}^T\Sigma^{-1}\boldsymbol{e}},$$
$$\lambda_- = \frac{\mu^T\Sigma^{-1}\boldsymbol{e} - \sqrt{(\mu^T\Sigma^{-1}\boldsymbol{e})^2 - (\boldsymbol{e}^T\Sigma^{-1}\boldsymbol{e})(\mu^T\Sigma^{-1}\mu - b^2)}}{\boldsymbol{e}^T\Sigma^{-1}\boldsymbol{e}}.$$

If $(\mu^T\Sigma^{-1}\boldsymbol{e})^2 - (\boldsymbol{e}^T\Sigma^{-1}\boldsymbol{e})(\mu^T\Sigma^{-1}\mu - b^2) > 0$, then we have

$$\begin{aligned} & -\mu + \frac{b\Sigma\boldsymbol{x}(\lambda)}{\sqrt{\boldsymbol{x}(\lambda)^T\Sigma\boldsymbol{x}(\lambda)}} + \lambda\boldsymbol{e} \\ = & -\mu + \frac{b(\mu-\lambda\boldsymbol{e})}{\sqrt{(\mu-\lambda\boldsymbol{e})^T\Sigma^{-1}(\mu-\lambda\boldsymbol{e})}}\frac{\sqrt{(\mu^T\Sigma^{-1}\boldsymbol{e}-\lambda\boldsymbol{e}^T\Sigma^{-1}\boldsymbol{e})^2}}{\mu^T\Sigma^{-1}\boldsymbol{e}-\lambda\boldsymbol{e}^T\Sigma^{-1}\boldsymbol{e}} + \lambda\boldsymbol{e} = 0, \end{aligned}$$

iff $\frac{\sqrt{(\mu^T\Sigma^{-1}\boldsymbol{e}-\lambda\boldsymbol{e}^T\Sigma^{-1}\boldsymbol{e})^2}}{\mu^T\Sigma^{-1}\boldsymbol{e}-\lambda\boldsymbol{e}^T\Sigma^{-1}\boldsymbol{e}} = 1$. Note that $\mu^T\Sigma^{-1}\boldsymbol{e} - \lambda_+\boldsymbol{e}^T\Sigma^{-1}\boldsymbol{e} < 0 < \mu^T\Sigma^{-1}\boldsymbol{e} - \lambda_-\boldsymbol{e}^T\Sigma^{-1}\boldsymbol{e}$, so the point $\mu^T\Sigma^{-1}\boldsymbol{e} - \lambda_+\boldsymbol{e}^T\Sigma^{-1}\boldsymbol{e}$ is rejected. Therefore, there is one local optimal solution and it is at $\boldsymbol{x}(\lambda_-)$ with optimal values $\frac{\mu^T\Sigma^{-1}(\mu-\lambda_-\boldsymbol{e})}{\mu^T\Sigma^{-1}\boldsymbol{e}-\lambda_-\boldsymbol{e}^T\Sigma^{-1}\boldsymbol{e}} - b\sqrt{\frac{\mu^T\Sigma^{-1}\mu-2\lambda_-\mu^T\Sigma^{-1}\boldsymbol{e}+\lambda_-^2\boldsymbol{e}^T\Sigma^{-1}\boldsymbol{e}}{(\mu^T\Sigma^{-1}\boldsymbol{e}-\lambda_-\boldsymbol{e}^T\Sigma^{-1}\boldsymbol{e})^2}} = \lambda_-$, where the last equality follows from Eq. (A.9).

If $(\mu^T\Sigma^{-1}\boldsymbol{e})^2 - (\boldsymbol{e}^T\Sigma^{-1}\boldsymbol{e})(\mu^T\Sigma^{-1}\mu - b^2) < 0$, no real roots for λ_+ and λ_- exist. This means that no local optimal solutions exist.

We can draw the same conclusion as in Theorem 3.5 using the same notations in Theorem 3.5. $\square$

Finally, we give the proof of Theorem 3.6.

Proof of Theorem 3.6. According to the symbols defined, problem (10) in the main paper is equivalent to

$$\begin{cases} & -\min_{\boldsymbol{x}\in\mathbb{X}}\{-\boldsymbol{x}^T\mu\} \\ s.t. & \kappa(\epsilon)\sqrt{\boldsymbol{x}^T\Sigma\boldsymbol{x}} - \boldsymbol{x}^T\mu \le \delta \end{cases} \Leftrightarrow \begin{cases} & -\min_{\boldsymbol{x}}\{-\boldsymbol{x}^T\mu\} \\ s.t. & \boldsymbol{x}^T\boldsymbol{e} = 1 \\ & \kappa(\epsilon)\sqrt{\boldsymbol{x}^T\Sigma\boldsymbol{x}} - \boldsymbol{x}^T\mu \le \delta \end{cases}.$$

Applying KKT, it is equivalent to examine

$$-(1+\tilde{\lambda})\mu + \frac{\tilde{\lambda} a \Sigma \boldsymbol{x}}{\sqrt{\boldsymbol{x}^T \Sigma \boldsymbol{x}}} + \tilde{\theta}\boldsymbol{e} = 0, \tag{A.11}$$

$$\tilde{\theta}(\boldsymbol{x}^T\boldsymbol{e} - 1) = 0, \tag{A.12}$$

$$\tilde{\lambda}\{\kappa(\epsilon)\sqrt{\boldsymbol{x}^T\Sigma\boldsymbol{x}} - \boldsymbol{x}^T\mu - \delta\} = 0, \tag{A.13}$$

$$\tilde{\lambda} \geq 0. \tag{A.14}$$

From Eq. (A.11), we know that $\boldsymbol{x}$ is linearly dependent of $(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}$. Let $\boldsymbol{x} = \boldsymbol{x}(\tilde{\lambda}, \tilde{\theta}) = \frac{\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]}{\boldsymbol{e}^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]}$. What is left is to find $\tilde{\lambda}$ and $\tilde{\theta}$. From Eq. (A.13), $\tilde{\lambda} \neq 0$, otherwise we obtain $\tilde{\theta}\boldsymbol{e} = \mu$ from Eq. (A.11), which is a contradiction in general. Hence, we only need to consider

$$-(1+\tilde{\lambda})\mu + \frac{\tilde{\lambda} a \Sigma \boldsymbol{x}}{\sqrt{\boldsymbol{x}^T \Sigma \boldsymbol{x}}} + \tilde{\theta}\boldsymbol{e} = 0, \tag{A.15}$$

$$\kappa(\epsilon)\sqrt{\boldsymbol{x}^T\Sigma\boldsymbol{x}} - \boldsymbol{x}^T\mu - \delta = 0. \tag{A.16}$$

Now, the terms $\boldsymbol{x}^T\mu$, $\Sigma\boldsymbol{x}$, and $\sqrt{\boldsymbol{x}^T\Sigma\boldsymbol{x}}$ are:

$$\boldsymbol{x}^T\mu = \frac{\mu^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]}{\boldsymbol{e}^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]}, \quad \Sigma\boldsymbol{x} = \frac{[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]}{\boldsymbol{e}^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]},$$
$$\sqrt{\boldsymbol{x}^T\Sigma\boldsymbol{x}} = \frac{\sqrt{[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]}}{|\boldsymbol{e}^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]|}.$$

This suggests that Eq. (A.15)

$$-(1+\tilde{\lambda})\mu + \frac{\tilde{\lambda}\kappa(\epsilon)[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]}{\boldsymbol{e}^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]} \frac{|\boldsymbol{e}^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]|}{\sqrt{[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]}} + \tilde{\theta}\boldsymbol{e} = 0, \tag{A.17}$$

and Eq. (A.16)

$$\frac{\kappa(\epsilon)\sqrt{[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]}}{|\boldsymbol{e}^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]|} - \frac{\mu^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]}{\boldsymbol{e}^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]} - \delta = 0$$
$$\Rightarrow \frac{\kappa(\epsilon)\sqrt{[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]}}{|\boldsymbol{e}^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]|} = \frac{\mu^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]}{\boldsymbol{e}^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]} + \delta. \tag{A.18}$$

We obtain the first relation between $\tilde{\lambda}$ and $\tilde{\theta}$ by substituting Eq. (A.18) into Eq. (A.17):

$$\tilde{\lambda}(\kappa(\epsilon))^2 = \mu^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}] + \delta\boldsymbol{e}^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]. \tag{A.19}$$

Now, from Eq. (A.17), together with the constraint that $\boldsymbol{x}^T\boldsymbol{e} = 1$, we have the second relation

$$(1+\tilde{\lambda})\left(\mu^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]\right)\left(\boldsymbol{e}^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]\right) - \tilde{\theta}(\boldsymbol{e}^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}])^2 =$$
$$+ \tilde{\lambda}\kappa(\epsilon)\sqrt{[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]}|\boldsymbol{e}^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]|.$$

Squaring the above relation on both sides, we have the reformulations of the left hand side and the right hand side:
L.H.S

$$\left\{(1+\tilde{\lambda})\left(\mu^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]\right)\left(\boldsymbol{e}^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]\right) - \tilde{\theta}(\boldsymbol{e}^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}])^2\right\}^2$$
$$= \left(\boldsymbol{e}^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]\right)^2\left\{(1+\tilde{\lambda})\left(\mu^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}]\right) - \tilde{\theta}(\boldsymbol{e}^T\Sigma^{-1}[(1+\tilde{\lambda})\mu - \tilde{\theta}\boldsymbol{e}])\right\}^2.$$

R.H.S

$$\begin{aligned}&\left\{\tilde{\lambda}\kappa(\epsilon)\sqrt{[(1+\tilde{\lambda})\mu-\tilde{\theta}\boldsymbol{e}]^T\Sigma^{-1}[(1+\tilde{\lambda})\mu-\tilde{\theta}\boldsymbol{e}]}|\boldsymbol{e}^T\Sigma^{-1}[(1+\tilde{\lambda})\mu-\tilde{\theta}\boldsymbol{e}]|\right\}^2\\ =&(\tilde{\lambda}\kappa(\epsilon))^2[(1+\tilde{\lambda})\mu-\tilde{\theta}\boldsymbol{e}]^T\Sigma^{-1}[(1+\tilde{\lambda})\mu-\tilde{\theta}\boldsymbol{e}](\boldsymbol{e}^T\Sigma^{-1}[(1+\tilde{\lambda})\mu-\tilde{\theta}\boldsymbol{e}])^2.\end{aligned}$$

Both sides have the term $(\boldsymbol{e}^T\Sigma^{-1}[(1+\tilde{\lambda})\mu-\tilde{\theta}\boldsymbol{e}])^2$ which can be eliminated. So it reduces to have

$$\begin{aligned}&\left\{(1+\tilde{\lambda})\left(\mu^T\Sigma^{-1}[(1+\tilde{\lambda})\mu-\tilde{\theta}\boldsymbol{e}]\right)-\tilde{\theta}(\boldsymbol{e}^T\Sigma^{-1}[(1+\tilde{\lambda})\mu-\tilde{\theta}\boldsymbol{e}])\right\}^2\\ =&(\tilde{\lambda}\kappa(\epsilon))^2[(1+\tilde{\lambda})\mu-\tilde{\theta}\boldsymbol{e}]^T\Sigma^{-1}[(1+\tilde{\lambda})\mu-\tilde{\theta}\boldsymbol{e}]\\ \Leftrightarrow&[(1+\tilde{\lambda})\mu-\tilde{\theta}\boldsymbol{e}]^T\Sigma^{-1}[(1+\tilde{\lambda})\mu-\tilde{\theta}\boldsymbol{e}]=(\tilde{\lambda}\kappa(\epsilon))^2.\end{aligned}$$

Lastly, from the L.H.S of Eq. (A.17) and the fact that $[(1+\tilde{\lambda})\mu-\tilde{\theta}\boldsymbol{e}]^T\Sigma^{-1}[(1+\tilde{\lambda})\mu-\tilde{\theta}\boldsymbol{e}]=(\tilde{\lambda}\kappa(\epsilon))^2$ proved, it can be seen that

$$\frac{|\boldsymbol{e}^T\Sigma^{-1}[(1+\tilde{\lambda})\mu-\tilde{\theta}\boldsymbol{e}]|}{\boldsymbol{e}^T\Sigma^{-1}[(1+\tilde{\lambda})\mu-\tilde{\theta}\boldsymbol{e}]}=1\Leftrightarrow\boldsymbol{e}^T\Sigma^{-1}[(1+\tilde{\lambda})\mu-\tilde{\theta}\boldsymbol{e}]>0.$$

Moreover, if we can solve $\tilde{\lambda}$ and $\tilde{\theta}$, from Eq. (A.15) and Eq. (A.16), it can be seen that the optimal value, which equals to $\tilde{\lambda}\boldsymbol{x}^T\mu$, can be expressed as $\tilde{\lambda}(-\boldsymbol{x}^T\mu+\kappa(\epsilon)\sqrt{\boldsymbol{x}^T\Sigma\boldsymbol{x}})+\tilde{\theta}=\tilde{\lambda}\delta+\tilde{\theta}$.

Now, applying the given notations that $A=\boldsymbol{e}^T\Sigma^{-1}\boldsymbol{e}$, $B=\mu^T\Sigma^{-1}\boldsymbol{e}$, and $C=\mu^T\Sigma^{-1}\mu$, we can summarize the following conditions:

$$\tilde{\lambda}\kappa(\epsilon)^2=[(1+\tilde{\lambda})C-\tilde{\theta}B]+\delta[(1+\tilde{\lambda})B-\tilde{\theta}A], \tag{A.20}$$

$$(\tilde{\lambda}\kappa(\epsilon))^2=[(1+\tilde{\lambda})^2C-2\tilde{\theta}(1+\tilde{\lambda})B+\tilde{\theta}^2A], \tag{A.21}$$

$$\tilde{\lambda}\geq 0, \tag{A.22}$$

$$(1+\tilde{\lambda})B-\tilde{\theta}A>0. \tag{A.23}$$

We then solve $\tilde{\lambda}$ explicitly.

From Eq. (A.19), we have

$$\tilde{\theta}=\frac{-\tilde{\lambda}\kappa(\epsilon)^2+(C+\delta B)+\tilde{\lambda}(C+\delta B)}{B+\delta A}. \tag{A.24}$$

Substituting the expression of $\tilde{\theta}$ into Eq. (A.21), we acquire a quadratic equation as follows

$$\begin{aligned}(\tilde{\lambda}\kappa(\epsilon))^2=&(1+\tilde{\lambda})^2C-2\left[\frac{-\tilde{\lambda}\kappa(\epsilon)^2+(C+\delta B)+\tilde{\lambda}(C+\delta B)}{B+\delta A}\right](1+\tilde{\lambda})B\\ &+\left(\frac{-\tilde{\lambda}\kappa(\epsilon)^2+(C+\delta B)+\tilde{\lambda}(C+\delta B)}{B+\delta A}\right)^2A\\ \Rightarrow(\tilde{\lambda}\kappa(\epsilon))^2(B+\delta A)^2=&(1+\tilde{\lambda})^2(B+\delta A)^2C\\ &-2(B+\delta A)[-\tilde{\lambda}\kappa(\epsilon)^2+(C+\delta B)(1+\tilde{\lambda})](1+\tilde{\lambda})B\\ &+\left(-\tilde{\lambda}\kappa(\epsilon)^2+(C+\delta B)(1+\tilde{\lambda})\right)^2A\\ \Rightarrow M\tilde{\lambda}^2-2G\tilde{\lambda}+H=0,&\end{aligned}$$

where the terms M, G, and H are

$$\begin{aligned}M=&[\kappa(\epsilon)^2-(A\delta^2+2B\delta+C)][AC-B^2-A\kappa(\epsilon)^2],\\ G=&(B^2-AC)[\kappa(\epsilon)^2-(A\delta^2+2B\delta+C)],\\ H=&(B^2-AC)(A\delta^2+2B\delta+C)\end{aligned}$$

after simplifications. Now, the roots of $\tilde{\lambda}$ can be computed as

$$\tilde{\lambda}_+ = \frac{G + \sqrt{G^2 - MH}}{M}, \tag{A.25}$$

$$\tilde{\lambda}_- = \frac{G - \sqrt{G^2 - MH}}{M}. \tag{A.26}$$

Simplifying the term $G^2 - MH$ using the definition that $K = (B^2 - AC)$ and $L = A\delta^2 + 2B\delta + C$, we obtain

$$\begin{aligned} &G^2 - MH = K^2[\kappa(\epsilon)^2 - L]^2 - KL\{[AL + AC - B^2]\kappa(\epsilon)^2 + KL - A\kappa(\epsilon)^4\} \\ =&K\kappa(\epsilon)^2[(K + AL)(\kappa(\epsilon)^2 - L) - (K + AC - B^2)L] = K\kappa(\epsilon)^2[(B + \delta A)^2(\kappa(\epsilon)^2 - L)] = \kappa(\epsilon)^2(B + \delta A)^2 G. \end{aligned}$$

The second last equality follows since $K + AL = B^2 - AC + \delta^2 A^2 + 2\delta AB + AC = (B + \delta A)^2$ and $(K + AC - B^2) = 0$. We can get $\tilde{\theta}_+$ and $\tilde{\theta}_-$ by substituting $\tilde{\lambda}_+$ and $\tilde{\lambda}_-$ in Eq. (A.24), respectively.

$\tilde{\lambda}_+$ and $\tilde{\lambda}_-$ are given in Eq. (A.25) and Eq. (A.26). Note that $G^2 - MH = \kappa(\epsilon)^2(B + \delta A)^2[(B^2 - AC)(\kappa(\epsilon)^2 - (A\delta^2 + 2B\delta + C)]$. Hence $G^2 - MH > 0 \Leftrightarrow (B^2 - AC)(\kappa(\epsilon)^2 - (A\delta^2 + 2B\delta + C) > 0 \Leftrightarrow 0 \leq \kappa(\epsilon)^2 < A\delta^2 + 2B\delta + C$ since $B^2 - AC \leq 0$ by Cauchy inequality.

The signs of G and H are important in completing the remaining of the proofs, from which we find that $G \geq 0$ and $H \leq 0$. Now, we are able to prove the statement case by case.

Case 1: $B + \delta A > 0$ *and* $\frac{AC - B^2}{A} < \kappa(\epsilon)^2 < A\delta^2 + 2B\delta + C$.

In this case, $M = [\kappa(\epsilon)^2 - (A\delta^2 + 2B\delta + C)][AC - B^2 - A\kappa(\epsilon)^2] > 0$. Now, we can consider the signs of $\tilde{\lambda}_+$ and $\tilde{\lambda}_-$. Clearly, $G^2 - MH \geq G^2 \Rightarrow \sqrt{G^2 - MH} \geq G$. This gives that $\tilde{\lambda}_+ \geq 0$ and $\tilde{\lambda}_- \leq 0$ so that the only choice for $\tilde{\lambda}$ is $\tilde{\lambda}_+$. Substituting $\tilde{\lambda}_+$ obtained into Eq. (A.24), we obtain

$$\tilde{\theta}_+ = \frac{-\tilde{\lambda}_+\kappa(\epsilon)^2 + (1 + \tilde{\lambda}_+)(C + \delta B)}{B + \delta A}.$$

Also, from the given conditions that $B + \delta A > 0$ together with Eq. (A.24) and Eq. (A.25) for $\tilde{\lambda}_+$, we obtain

$$(1 + \tilde{\lambda}_+)B - \tilde{\theta}_+ A = \frac{\sqrt{G^2 - MH}}{[B + \delta A][(A\delta^2 + 2B\delta + C) - \kappa(\epsilon)^2]} > 0.$$

This means that the optimal solution and optimal value of Eq. (10) are $\frac{\Sigma^{-1}[(1+\tilde{\lambda}_+)\mu - \tilde{\theta}_+ \boldsymbol{e}]}{\boldsymbol{e}^T\Sigma^{-1}[(1+\tilde{\lambda}_+)\mu - \tilde{\theta}_+ \boldsymbol{e}]}$ and $\tilde{\lambda}_+\delta + \tilde{\theta}_+$, respectively, since the conditions (A.20), (A.21), (A.22) and (A.23) are satisfied.

In this case, we can compare the optimal value of problem (10) and the optimal value of problem (9). This is done by considering

$$\begin{aligned} &\tilde{\lambda}_+\delta + \tilde{\theta}_+ - \lambda^* \\ =&\tilde{\lambda}_+\delta + \frac{(1 + \tilde{\lambda}_+)(C + \delta B) - \tilde{\lambda}_+\kappa(\epsilon)^2}{B + \delta A} - \frac{B - \sqrt{\triangle}}{A} \\ =&\frac{\tilde{\lambda}_+(\delta^2 A + 2\delta B + C - \kappa(\epsilon)^2)}{B + \delta A} + \frac{CA - B^2}{A(B + \delta A)} + \frac{\sqrt{\triangle}}{A} > 0. \end{aligned}$$

Case 2: $B > 0$, $B + \delta A < 0$, *and* $\frac{AC - B^2}{A} < \kappa(\epsilon)^2 < \delta^2 A + 2\delta B + C$.

In this case, we check the equation A.23

$$(1 + \tilde{\lambda}_+)B - \tilde{\theta}_+ A = \frac{\sqrt{G^2 - MH}}{[B + \delta A][(A\delta^2 + 2B\delta + C) - \kappa(\epsilon)^2]} < 0.$$

This means that the conditions (A.20), (A.21), (A.22) and (A.23) are not satisfied. We cannot have a local optimal solution.

Next, we prove that under the given conditions, we cannot find $\boldsymbol{x}$ such that $\boldsymbol{x}^T\boldsymbol{e} = 1$ and $\kappa(\epsilon)\sqrt{\boldsymbol{x}^T\Sigma\boldsymbol{x}} - \boldsymbol{x}^T\mu \leq \delta$ are satisfied. We consider

$$\min \kappa(\epsilon)\sqrt{\boldsymbol{x}^T\Sigma\boldsymbol{x}} - \boldsymbol{x}^T\mu \quad \textit{s.t.} \quad \boldsymbol{x}^T\boldsymbol{e} = 1.$$

From the ***Proof of Theorem 3.5*** and the given condition that $\kappa(\epsilon)^2 > \frac{B^2-AC}{A}$, we find that the optimal solution is $\frac{B-\sqrt{B^2-AC+A\kappa(\epsilon)^2}}{A}$. Since $\kappa(\epsilon)^2 < \delta^2 A + 2\delta B + C$ and $B + \delta A < 0$, we know that

$$\begin{aligned}\frac{B-\sqrt{B^2-AC+A\kappa(\epsilon)^2}}{A} &> \frac{B-\sqrt{B^2+2\delta AB+\delta^2A^2}}{A}\\ &= \frac{B-|B+\delta A|}{A} = \frac{2B+\delta A}{A} = \frac{2B}{A} + \delta > \delta\end{aligned}$$

when $B > 0$. This means that no feasible solution in this case exists.

Case 3: $\kappa(\epsilon)^2 < \frac{AC-B^2}{A}$. We check the signs of $\tilde{\lambda}_+$ and $\tilde{\lambda}_-$. $M = [\kappa(\epsilon)^2 - (A\delta^2 + 2B\delta + C)][AC - B^2 - A\kappa(\epsilon)^2] < 0$. Then $0 < G^2 - MH \leq G^2$. As a result,

$$\begin{aligned}\tilde{\lambda}_+ &= \frac{G+\sqrt{G^2-MH}}{M} < 0,\\ \tilde{\lambda}_- &= \frac{G-\sqrt{G^2-MH}}{M} < \frac{G-\sqrt{G^2}}{M} = 0.\end{aligned}$$

This means that (A.20), (A.21), (A.22), and (A.23) are not satisfied. Hence, no local optimal solution for problem (10) exists. □