

Conditional Risk Minimization with Side Information: A Tractable, Universal Optimal Transport Framework

Xinqiao Xie, Jonathan Yu-Meng Li

Telfer School of Management
University of Ottawa
Ottawa, Ontario, Canada K1N 6N5
xxie087@uottawa.ca, jonathan.li@telfer.uottawa.ca

September 30, 2025

Abstract

Conditional risk minimization arises in high-stakes decisions where risk must be assessed in light of side information—such as stressed economic conditions, specific customer profiles, or other contextual covariates. Constructing reliable conditional distributions from limited data is notoriously difficult, motivating a series of optimal-transport-based proposals that address this uncertainty in a distributionally robust manner. Yet these approaches remain fragmented, each constrained by its own limitations: some rely on point estimates or restrictive structural assumptions, others apply only to narrow classes of risk measures, and their structural connections are unclear. We introduce a universal framework for distributionally robust conditional risk minimization, built on a novel union-ball formulation in optimal transport. This framework offers three key advantages: interpretability, by subsuming existing methods as special cases and revealing their deep structural links; tractability, by yielding convex reformulations for virtually all major risk functionals studied in the literature; and scalability, by supporting cutting-plane algorithms for large-scale conditional risk problems. Applications to portfolio optimization with rank-dependent expected utility highlight the practical effectiveness of the framework, with conditional models converging to optimal solutions where unconditional ones clearly do not.

Key words: Contextual Optimization, Distributionally robust optimization, Optimal Transport, Risk Measures

1 Introduction

Leveraging contextual side information—such as customer profiles, economic indicators, or individualized health records—has become integral to modern data-driven decision-making across a wide range of business

domains. These problems are classically formulated through conditional expectation optimization (e.g., Bertsimas and Kallus (2020)):

$$\min_{\alpha \in \mathcal{A}} \mathbb{E}_{\mathcal{Q}} [\ell(Y, \alpha) \mid X \in \mathcal{N}], \quad (1)$$

where $Y \in \mathbb{R}^m$ denotes an uncertain quantity of interest, $X \in \mathbb{R}^n$ represents covariates encoding side information, and $\alpha \in \mathcal{A}$ is a decision that minimizes the mean loss of $\ell(Y, \alpha)$, conditioned on the covariates X lying in the region $\mathcal{N} \subseteq \mathbb{R}^n$. While analytically convenient, the conditional mean often fails to capture the realities of high-stakes decision-making, where performance is rarely judged by averages alone. Financial institutions use measures such as conditional value-at-risk (CoVaR) to monitor systemic fragility (see Adrian and Brunnermeier (2011) for more details of CoVaR); healthcare systems prioritize robustness to adverse scenarios (e.g., demand surges or rare severe outcomes); and personalized operations seek guarantees that account for tail risk beyond average performance. To address such diverse risk tasks, we formalize the notion of conditional risk minimization (CRM):

$$\min_{\alpha \in \mathcal{A}} \rho_{\mathcal{Q}} [\ell(Y, \alpha) \mid X \in \mathcal{N}], \quad (2)$$

where ρ denotes a general risk-aware decision criterion that subsumes expectation as a special case and captures a broad spectrum of nonlinear objectives tailored to different risk preferences.

A fundamental challenge, already present in the classical expectation setting (1), is that the conditional distribution is not directly observable; in practice, only samples from the joint distribution are available. To bridge this gap, methods to approximate the conditional law range from kernel regression (Ban and Rudin 2019), k -nearest neighbors, local-linear estimators, and tree-based methods (Bertsimas and Kallus 2020) to covariate-SAA frameworks that embed prediction models within sample-average approximations (Kannan et al. 2025). These techniques differ in how side information is incorporated—either by reweighting observed samples or by reshaping the support set—yet in all cases the resulting estimates remain vulnerable to modeling bias and statistical error. Consequently, decisions based on these approximations can undermine out-of-sample reliability, underscoring the need for a more principled framework that directly addresses ambiguity in the conditional distribution.

The success of distributionally robust optimization (DRO) in ensuring reliable out-of-sample performance in unconditional settings has inspired analogous efforts to robustify the conditional expectation problem (1) using optimal transport-based ambiguity sets. Two main modeling approaches have emerged. The first places ambiguity directly over the space of conditional distributions, a regime that we refer to as *predict-then-robustify*: a conditional distribution is first estimated using machine-learning techniques, and a Wasserstein ball is then formed around the estimate (see Wang et al. (2021), Wang et al. (2024), Kannan et al. (2024)). For example, Wang et al. (2021) use a kernel-regression-based conditional distribution as the ball’s center, while Kannan et al. (2024) employ a residual-based construction as the center; Wang et al. (2024) further consider the intersection of two Wasserstein balls to capture covariate shifts. While these formulations largely retain the tractability of unconditional Wasserstein DRO and can, in principle, accommodate general risk criteria, their reliance on conditional distribution estimates inherits the attendant bias and error, and they remain

largely restricted to singleton conditions, $\mathcal{N} = \{x_0\}$, rather than arbitrary covariate sets.

The second approach bypasses the prediction step and constructs ambiguity sets directly on the joint distribution, typically centered at the empirical law. Along this line, Nguyen et al. (2020) and Nguyen et al. (2024) employ optimal transport balls augmented with constraints that guarantee positive probability mass on the conditioning set. Because conditional expectations are nonlinear in the joint distribution, the resulting problems are more challenging to solve, and extensions to risk criteria lead only to conservative approximations. Esteban-Perez and Morales (2022) takes a different angle by restricting the admissible joint distributions to those supported on the conditioning set in the covariates. While this makes the conditional expectation linear in the distribution, it requires the use of partial mass transport metrics to construct ambiguity sets from the empirical law, whose interpretation remains less transparent. Despite the comparison made in Esteban-Perez and Morales (2022) with the formulation of Nguyen et al. (2020), the two models are difficult to relate beyond relative conservativeness—each carrying its own interpretation, with structural connections and limitations not yet fully understood.

Despite their promise, both paradigms fall short of providing a universal, robust, and tractable framework for the conditional risk minimization problem (2). Predict-then-robustify methods retain the flexibility and tractability of unconditional DRO but are brittle, as their performance hinges on the quality of the prediction step. Joint-distribution approaches avoid this dependence but remain narrow in scope: they are largely confined to conditional expectation problems (1), and their extensions to richer risk criteria typically yield only conservative relaxations. For instance, Nguyen et al. (2024) extend their framework to risk measures such as mean–variance and mean–conditional value-at-risk (CVaR), but their formulation applies only to those admitting a finite-dimensional expectation representation—a condition not necessarily satisfied by certain important coherent risk measures. Moreover, their tractable reformulations provide no guarantees of tightness. These challenges underscore the difficulty of extending methods designed for the conditional expectation problem (1) to the broader CRM problem (2). More fundamentally, existing approaches are built on structurally distinct formulations, leaving open how they can be compared, related, or generalized. These limitations motivate our development of a universal and tractable framework for conditional risk minimization—one that recovers existing paradigms as special cases and enables principled extensions to a broader class of risk criteria.

Our key innovation lies in uncovering a previously unexplored ambiguity-set structure in DRO: the union of optimal transport balls. While other constructions—such as intersections of balls—have been explored, the use of a union as an ambiguity set remains, to our knowledge, entirely absent. Our formulation arises naturally from a core limitation of the predict-then-robustify paradigm, which fixes both the reference distribution and robustness radius despite uncertainty in each. We instead treat both as variables and define the ambiguity set as the union of balls over all admissible (distribution, radius) pairs. This union-ball structure yields three key breakthroughs: (1) it eliminates the brittleness of predict-then-robustify methods; (2) it provides a unified framework that subsumes existing paradigms for conditional optimization as special cases, enabling exact and transparent comparisons through their corresponding union-ball representations; and (3) most importantly, it yields a structural reformulation of worst-case conditional risk in CRM as a maximization over unconditional counterparts, which, as shown throughout this paper, can in turn be reformulated into

tractable convex programs for a broad class of risk measures—including those previously beyond reach.

We demonstrate the universality and tractability of our union-ball formulation through a comprehensive suite of conditional risk minimization problems spanning a wide range of risk measures. This includes problems studied in Nguyen et al. (2024), such as mean–variance and mean–CVaR, where their approach yields only conservative approximations, while ours delivers exact and simpler solutions. More broadly, our framework applies to risk criteria entirely outside the scope of existing methods, including those lacking finite-dimensional expectation representations. In particular, we establish a tractable distributionally robust conditional formulation of the celebrated rank-dependent expected utility (RDEU) from prospect theory—an important case for which no such formulation was previously available. Finally, we show how our universal framework enables scalable cutting-plane algorithms to solve large-scale conditional RDEU problems with practical efficiency.

1.1 Related Literature

Beyond the methods already discussed in the introduction, two additional strands of work are worth noting. The first is the *predict-then-optimize* regime, which does not attempt to estimate the full conditional distribution but instead predicts conditional statistics—such as conditional expectations or risk measures—from side information, and then uses these predictions as inputs to an optimization problem. Examples include Donti and Kolter (2017), Elmachou and Grigas (2022), and Muñoz et al. (2022), with further analysis of generalization and stability properties in El Balghiti et al. (2019) and Hu et al. (2022). While effective in certain settings, these methods inherit the fragility of the prediction step, provide no explicit safeguard against distributional ambiguity, and are difficult to extend to general risk minimization problems that require a full distributional specification.

A second, related but distinct, line of work transforms conditional expectation problems into unconditional ones by embedding side information directly into the decision rules. Notably, Zhang et al. (2024) and Yang et al. (2022) adopt this paradigm, formulating objectives as expectations under the joint distribution rather than conditional criteria. In particular, Yang et al. (2022) introduce a causal transport distance within this decision-rule framework. While these approaches provide an alternative perspective, they shift the problem away from the CRM formulation studied here.

2 A Universal OT Framework for Conditional Optimization with Side Information

2.1 Preliminaries and Existing OT-Based Approaches

To set the stage for our universal framework, we first recap the three existing optimal transport (OT)-based approaches through their precise formulations, along with the minimal preliminaries needed to state them formally. Recall the following definition of optimal transport cost:

Definition 2.1 (Optimal transport cost). *Let $\Xi \subseteq \mathbb{R}^d$ be a Polish space for some integer d , and let $\mathbb{D} : \Xi \times \Xi \rightarrow \mathbb{R}_+$ be a continuous cost function. The optimal transport cost between two probability distributions*

$\mathbb{Q}_1$ and $\mathbb{Q}_2$ supported on Ξ is defined as

$$\mathcal{W}(\mathbb{Q}_1, \mathbb{Q}_2) \triangleq \min_{\pi \in \Pi(\mathbb{Q}_1, \mathbb{Q}_2)} \mathbb{E}_\pi [\mathbb{D}(\xi_1, \xi_2)],$$

where $\Pi(\mathbb{Q}_1, \mathbb{Q}_2)$ denotes the set of all couplings of $\mathbb{Q}_1$ and $\mathbb{Q}_2$ on $\Xi \times \Xi$.¹

An OT ball is defined as

$$\mathbb{B}_\delta(\widehat{\mathbb{P}}) = \left\{ \mathbb{Q} \in \mathcal{P}(\Xi) : \mathcal{W}(\mathbb{Q}, \widehat{\mathbb{P}}) \leq \delta \right\},$$

where $\widehat{\mathbb{P}}$ is the reference distribution and $\delta \geq 0$ is the ball radius. It consists of all distributions within δ -distance from $\widehat{\mathbb{P}}$ under the cost $\mathcal{W}$.

We begin with the *Predict-then-Robustify* approach as a baseline, given its conceptual simplicity, followed by the more sophisticated formulations developed in Nguyen et al. (2024) and Esteban-Perez and Morales (2022). For a random pair (X, Y) on $\mathcal{X} \times \mathcal{Y}$ with joint distribution $\mathbb{P}$, we denote the conditional law of Y given X by $\mathbb{P}_{Y|X}$. To emphasize it as a distribution of Y corresponding to a fixed covariate value $X = x$, we write $\mathbb{P}^x|_Y$. More generally, we use $\mathbb{P}|_Y$ to denote the distribution of Y under a conditioning event on X .

(1) Predict-then-Robustify. This baseline approach can readily accommodate solving the conditional risk minimization (CRM) problem (2) for general risk criteria, given its plug-and-play nature. Given N samples $(\widehat{X}, \widehat{Y}) = \{(\widehat{x}_i, \widehat{y}_i)\}_{i=1}^N$ drawn from the joint distribution, one first applies a predictive model (e.g., kernel regression) to estimate the conditional distribution $\widehat{\mathbb{P}}_{Y|X}$, and then solves a standard DRO problem with a OT ball centered at the estimate $\widehat{\mathbb{P}}^{x_0}|_Y$:

$$\min_{\alpha \in \mathcal{A}} \sup_{\mathbb{Q}|_Y \in \mathbb{B}_{\widehat{\delta}}(\widehat{\mathbb{P}}^{x_0}|_Y)} \rho_{\mathbb{Q}|_Y} [\ell(Y, \alpha)], \quad (3)$$

where $\mathbb{B}_{\widehat{\delta}}(\widehat{\mathbb{P}}^{x_0}|_Y)$ is an OT ball of radius $\widehat{\delta}$ centered at $\widehat{\mathbb{P}}^{x_0}|_Y$. However, this setup inherits a modeling bias tied to the choice of predictive method (e.g., kernel type, bandwidth), which propagates through the decision process. Moreover, it is typically limited to the singleton case $\mathcal{N} = \{x_0\}$ in (2), since most predictive models are not designed to output conditional distributions over non-singleton covariate sets.

(2) Full Optimal Transport. The approach of Nguyen et al. (2024) circumvents the need to estimate a conditional distribution by placing the ambiguity set over joint distributions $\mathbb{Q}$ supported on $\mathcal{X} \times \mathcal{Y}$. Let $\mathbb{D}_\mathcal{X} : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}_+$ denote a distance matrix on $\mathcal{X}$. For $x_0 \in \mathcal{X}$ and $\gamma \geq 0$, define the covariate region $\mathcal{N}_\gamma(x_0) \triangleq \{x \in \mathcal{X} : \mathbb{D}_\mathcal{X}(x, x_0) \leq \gamma\}$. Given such a region $\mathcal{N} := \mathcal{N}_\gamma(x_0)$, they propose the following

¹Here we adopt the minimization form not the infimum form to define the optimal transport cost. This choice is appropriate because, when the underlying space Ξ is Polish, the existence of an optimal joint distribution is guaranteed by Theorem 4.1 of Villani (2008).

distributionally robust counterpart to problem (1):

$$\min_{\alpha \in \mathcal{A}} \sup_{\substack{\mathbb{Q} \in \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}) \\ \mathbb{Q}(X \in \mathcal{N}_\gamma(x_0)) \in \mathcal{N}_\omega}} \mathbb{E}_{\mathbb{Q}} [\ell(Y, \alpha) \mid X \in \mathcal{N}_\gamma(x_0)], \quad (4)$$

where

$$\mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}) = \left\{ \mathbb{Q} \in \mathcal{M}(\mathcal{X} \times \mathcal{Y}) : \mathcal{W}(\mathbb{Q}, \widehat{\mathbb{P}}) \leq \delta_0 \right\},$$

with $\delta_0 > 0$, $\widehat{\mathbb{P}} = \frac{1}{N} \sum_{i=1}^N \mathbf{I}_{(\widehat{x}_i, \widehat{y}_i)}$, and $\mathcal{N}_\omega = [\omega, 1]$ for $\omega > 0$ or $\mathcal{N}_\omega = (\omega, 1]$ for $\omega = 0$. This formulation evaluates the worst-case conditional expectation over distributions $\mathbb{Q}$ lying within a Wasserstein ball centered at $\widehat{\mathbb{P}}$. The constraint $\mathbb{Q}(X \in \mathcal{N}_\gamma(x_0)) \in \mathcal{N}_\omega$ specifies an admissible range for the probability mass placed on the conditioning region.

(3) Partial Mass Transport. The approach of Esteban-Perez and Morales (2022) likewise circumvents the need to estimate a conditional distribution, but instead places the ambiguity set directly on joint distributions $\mathbb{Q}$ supported within a covariate region $\mathcal{N} := \mathcal{N}_\gamma(x_0)$. They construct this ambiguity set using the partial optimal transport distance, which measures the proximity between a candidate distribution $\mathbb{Q}$ and a trimmed version of the empirical distribution $\widehat{\mathbb{P}}$:

$$\mathcal{W}(\mathbb{Q}, \mathcal{R}_{1-\beta}(\widehat{\mathbb{P}})) \triangleq \min_{\widetilde{\mathbb{P}} \in \mathcal{R}_{1-\beta}(\widehat{\mathbb{P}})} \mathcal{W}(\mathbb{Q}, \widetilde{\mathbb{P}}),$$

where the trimming set is defined as

$$\mathcal{R}_{1-\beta}(\widehat{\mathbb{P}}) \triangleq \left\{ \mathbb{Q} \ll \widehat{\mathbb{P}} : \frac{d\mathbb{Q}}{d\widehat{\mathbb{P}}} \leq \frac{1}{\beta} \right\}, \quad \beta \in (0, 1].$$

The resulting distributionally robust counterpart to problem (1) is:

$$\min_{\alpha \in \mathcal{A}} \sup_{\substack{\mathbb{Q} \in \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}; \beta) \\ \mathbb{Q}(X \in \mathcal{N}_\gamma(x_0)) = 1}} \mathbb{E}_{\mathbb{Q}} [\ell(Y, \alpha)], \quad (5)$$

where $\beta \in (0, 1]$ is a given constant and $\mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}; \beta)$ is the ambiguity set defined via partial mass transport:

$$\mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}; \beta) \triangleq \left\{ \mathbb{Q} \in \mathcal{M}(\mathcal{X} \times \mathcal{Y}) : \mathcal{W}(\mathbb{Q}, \mathcal{R}_{1-\beta}(\widehat{\mathbb{P}})) \leq \delta_0 \right\}.$$

Despite their individual merits, it remains unclear how to choose among these three approaches, how to compare them on equal footing, and to what extent they can accommodate broader classes of risk minimization problems. Their differing structural assumptions and fixed modeling choices also raise concerns about hidden biases. These limitations motivate the need for a more universal formulation—one that systematically integrates these modeling elements within a single optimal transport framework.

2.2 A Universal Framework via Union-Ball Formulation

To introduce our framework, we begin by revisiting the Predict-then-Robustify paradigm and its vulnerability to modeling bias—stemming both from reliance on point estimates of conditional distributions and from fixed choices of the radius. A natural way to mitigate this bias is to explicitly account for the uncertainty in both the reference distribution and the radius, by treating them as variables ranging over an admissible set. Concretely, we consider discrete distributions on $\mathcal{Y}$ supported on empirical outcomes, i.e., distributions of the form $\sum_{i=1}^N p_i \mathbb{I}_{\hat{y}_i}$ for some $p_i \geq 0$, and define $\mathcal{V}$ as a convex set of all admissible pairs (p, δ) of distribution weights and radius values. The resulting ambiguity set takes the form of a union:

$$\bigcup_{(p, \delta) \in \mathcal{V}} \mathbb{B}_\delta \left(\sum_{i=1}^N p_i \mathbb{I}_{\hat{y}_i} \right),$$

capturing the joint uncertainty over both the center and the size of the OT ball.

This leads to the following distributionally robust formulation:

$$\min_{\alpha \in \mathcal{A}} \sup_{\mathbb{Q}|_Y \in \bigcup_{(p, \delta) \in \mathcal{V}} \mathbb{B}_\delta \left(\sum_{i=1}^N p_i \mathbb{I}_{\hat{y}_i} \right)} \rho_{\mathbb{Q}|_Y} [\ell(Y, \alpha)], \quad (6)$$

which serves as a robust counterpart to the Predict-then-Robustify formulation (3), with $(\hat{\mathbb{P}}^{x_0}|_Y, \hat{\delta}) \in \mathcal{V}$.

As we will show, this natural—yet previously unexplored—DRO formulation proves remarkably powerful: it unifies and generalizes existing approaches, reveals hidden tractability, uncovers deep structural connections, and offers full flexibility to encode modeling preferences and prior knowledge. Throughout the paper, we adopt the following definition of the optimal transport metric over the joint space $\Xi = \mathcal{X} \times \mathcal{Y}$, as used in Nguyen et al. (2024).

Definition 2.2 (Separable Transport Cost). *We endow the joint space $\mathcal{X} \times \mathcal{Y}$ with a separable cost function of the form*

$$\mathbb{D}((x, y), (x', y')) = \mathbb{D}_{\mathcal{X}}(x, x') + \mathbb{D}_{\mathcal{Y}}(y, y'),$$

where $\mathbb{D}_{\mathcal{X}}$ and $\mathbb{D}_{\mathcal{Y}}$ are symmetric, continuous, and nonnegative ground costs on $\mathcal{X} \times \mathcal{X}$ and $\mathcal{Y} \times \mathcal{Y}$, respectively, and satisfy $\mathbb{D}((x, y), (x', y')) = 0$ if and only if $(x, y) = (x', y')$.

To best illustrate the expressive power of our formulation in (6), we begin with the following generalization of the full optimal transport model in Nguyen et al. (2024):

$$\min_{\alpha \in \mathcal{A}} \sup_{\substack{\mathbb{Q} \in \mathbb{B}_{\delta_0}(\hat{\mathbb{P}}) \\ \mathbb{Q}(X \in \mathcal{N}_\gamma(x_0)) \in \mathcal{N}_\omega}} \rho_{\mathbb{Q}} [\ell(Y, \alpha) \mid X \in \mathcal{N}_\gamma(x_0)], \quad (7)$$

with $\delta_0 > 0$, $\gamma \geq 0$, and a convex set $\mathcal{N}_\omega \subset (0, 1]$, which means that $\mathcal{N}_\omega$ is an interval contained in $(0, 1]$. This formulation extends Nguyen et al. (2024) in two significant ways. First, it accommodates arbitrary decision criteria ρ , thus enabling a fully general conditional risk minimization framework. In contrast, Nguyen et al. (2024) focus primarily on expectations or risk measures admitting a minimization

form $\rho(Z) = \inf_{t \in \mathbb{R}^k} \mathbb{E}[\ell(Z, t)]$. Second, it allows flexible specification of the admissible mass range $\mathcal{N}_\omega$, offering greater control over the conditioning event when needed.

For the remainder of this section, we assume without loss of generality that the admissible mass interval is $\mathcal{N}_\omega = [\omega_1, \omega_2]$ with $0 < \omega_1 \leq \omega_2 \leq 1$. We now show that the general conditional risk minimization problem in (7) can be written in the form of the union-ball formulation (6). We first state the necessary assumptions for the neighborhood set $\mathcal{N}_\gamma(x_0)$.

Assumption 2.1. (*Regularity conditions*). *The following assumptions hold.*

- (i) *Projection:* For any $i = 1, \dots, N$, there exists a projection $\widehat{x}_i^p \in \partial \mathcal{N}_\gamma(x_0)$ of $\widehat{x}_i$ onto $\partial \mathcal{N}_\gamma(x_0)$ under the cost $\mathbb{D}_X$, that is,

$$0 \leq \min_{x \in \partial \mathcal{N}_\gamma(x_0)} \mathbb{D}_X(x, \widehat{x}_i) = \mathbb{D}_X(\widehat{x}_i^p, \widehat{x}_i) := d_i.$$

- (ii) *Vicinity:* For any $x \in \partial \mathcal{N}_\gamma(x_0)$ and for any radius $r > 0$, the neighborhood set

$$\{x' \in X \setminus \mathcal{N}_\gamma(x_0) : \mathbb{D}_X(x', x) \leq r\}$$

around the boundary point x is non-empty.

Without loss of generality, assume that $(\widehat{x}_i, \widehat{y}_i) \notin \{(x, y) : x \in \mathcal{N}_\gamma(x_0)\}$ for $i = 1, \dots, m$, and $(\widehat{x}_i, \widehat{y}_i) \in \{(x, y) : x \in \mathcal{N}_\gamma(x_0)\}$ for $i = m+1, \dots, N$. As a first step, we establish the necessary and sufficient condition for feasibility of problem (7), presented in the following proposition.

Proposition 1. *Let*

$$\delta_{\min} = \min_{\substack{v_i \in [0,1], i=1,\dots,N, \\ 1/N \sum_{i=1}^N v_i \in \mathcal{N}_\omega}} \frac{1}{N} \left(\sum_{i=1}^m v_i d_i + \sum_{i=m+1}^N (1 - v_i) d_i \right),$$

with $\mathcal{N}_\omega = [\omega_1, \omega_2] \subset (0, 1]$. For the radius δ_0 given in problem (7), we have

- (i) For $\frac{N-m}{N} > \omega_2$, problem (7) is feasible if and only if $\delta_0 > \delta_{\min}$.
- (ii) For $\frac{N-m}{N} \leq \omega_2$, problem (7) is feasible if and only if $\delta_0 \geq \delta_{\min}$. Specifically, if $\frac{N-m}{N} \in [\omega_1, \omega_2]$, $\delta_{\min} = 0$.

Remark 2. If $\omega_2 = 1$, then $\delta_{\min}$ coincides with the minimum radius given in Proposition 2.5 of Nguyen et al. (2024). For the more general case of $\mathcal{N}_\omega$, however, the radius in Proposition 2.5 of Nguyen et al. (2024) no longer applies, whereas $\delta_{\min}$ yields the correct value.

Before presenting the main result, we first formalize the admissible conditional distributions of Y , induced by joint laws $\mathbb{Q} \in \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}})$ satisfying $\mathbb{Q}(X \in \mathcal{N}_\gamma(x_0)) \in \mathcal{N}_\omega$ —as this structure will play a key role in what

follows. Specifically,

$$\mathcal{K}(\mathcal{N}_\gamma(x_0), \mathcal{N}_\omega, \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}})) = \left\{ \mathbb{Q}|_Y : \mathbb{Q}|_Y(A) = \frac{\mathbb{Q}(\mathcal{N}_\gamma(x_0) \times A)}{\mathbb{Q}(\mathcal{N}_\gamma(x_0) \times \mathcal{Y})} \text{ for any measurable set } A \subset \mathcal{Y}, \right. \\ \left. \text{where } \mathbb{Q} \in \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}) \text{ satisfies that } \mathbb{Q}(X \in \mathcal{N}_\gamma(x_0)) \in \mathcal{N}_\omega \right\}.$$

We impose the following mild assumptions on the risk measure ρ and the loss function ℓ . As shown in Lemma 9 in the next section, virtually all major risk functionals studied in the literature satisfy them.

Assumption 2.2. (i) ρ is law-invariant, that is, if $Y_1 \stackrel{d}{=} Y_2$, $\rho(Y_1) = \rho(Y_2)$, where $Y_1 \stackrel{d}{=} Y_2$ means that Y_1 and Y_2 have the same distribution. Assuming the distribution of Y is $\mathbb{Q}$, then we can denote $\rho(Y)$ as $\rho(\mathbb{Q})$.

(ii) Denote $\rho[\ell(Y, \alpha)] = \rho \circ \ell(Y, \alpha) = \rho \circ \ell(\mathbb{Q}, \alpha)$, where $\mathbb{Q}$ is the distribution of Y . $\rho \circ \ell$ is OT-continuity, that is, for any distribution $\mathbb{Q}$, if there exists a series of distribution $\{\mathbb{Q}_n\}$ such that $\mathcal{W}(\mathbb{Q}_n, \mathbb{Q}) \rightarrow 0$ as $n \rightarrow \infty$, then $\rho \circ \ell(\mathbb{Q}_n, \alpha) \rightarrow \rho \circ \ell(\mathbb{Q}, \alpha)$ for any $\alpha \in \mathcal{A}$.

Theorem 3. Suppose the neighborhood set $\mathcal{N}_\gamma(x_0)$ satisfies Assumption 2.1, the radius δ_0 satisfies the feasibility condition of problem (7) in Proposition 1, and the loss function ℓ and risk measure ρ satisfy Assumption 2.2. Then problem (7), equivalently

$$\min_{\alpha \in \mathcal{A}} \sup_{\mathbb{Q}|_Y \in \mathcal{K}(\mathcal{N}_\gamma(x_0), \mathcal{N}_\omega, \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}))} \rho_{\mathbb{Q}|_Y}[\ell(Y, \alpha)], \quad (8)$$

admits the union-ball reformulation (6) with

$$\mathcal{V} := \{(p, \delta) : \exists \varepsilon \text{ such that } (p, \delta, \varepsilon) \in \mathcal{V}_0\},$$

where

$$\mathcal{V}_0 := \left\{ (p, \delta, \varepsilon) : \frac{1}{\varepsilon} \in \mathcal{N}_\omega, p_i \in \left[0, \frac{\varepsilon}{N}\right], i = 1, \dots, N, \sum_{i=1}^N p_i = 1, \right. \\ \left. \delta = \varepsilon \left(\delta_0 - \frac{1}{N} \sum_{i=m+1}^N d_i \right) - \sum_{i=1}^m p_i d_i + \sum_{i=m+1}^N p_i d_i \right\}.$$

We next show that the partial optimal transport model of Esteban-Perez and Morales (2022)—which relies on the partial optimal transport distance rather than the standard Wasserstein formulation—can likewise be recast in the form of (6). This is established through the following generalization of their model:

$$\min_{\alpha \in \mathcal{A}} \sup_{\substack{\mathbb{Q} \in \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}; \mathcal{N}_\omega) \\ \mathbb{Q}(X \in \mathcal{N}_\gamma(x_0)) = 1}} \rho_{\mathbb{Q}}[\ell(Y, \alpha)], \quad (9)$$

where $\mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}; \mathcal{N}_\omega) = \left\{ \mathbb{Q} \in \mathcal{M}(\mathcal{X} \times \mathcal{Y}) : \exists \beta \in \mathcal{N}_\omega, \mathcal{W}(\mathbb{Q}, \mathcal{R}_{1-\beta}(\widehat{\mathbb{P}})) \leq \delta \right\}$ with a convex set $\mathcal{N}_\omega \subset (0, 1]$. This formulation extends Esteban-Perez and Morales (2022) in two ways: it accommodates arbitrary decision criteria ρ , and it allows flexible specification of the admissible mass range $\mathcal{N}_\omega$. Based on the definition of trimming set, one can verify that the constraint $\exists \beta \in \mathcal{N}_\omega, \mathcal{W}(\mathbb{Q}, \mathcal{R}_{1-\beta}(\widehat{\mathbb{P}})) \leq \delta_0$ is equivalent to $\mathcal{W}(\mathbb{Q}, \mathcal{R}_{1-\varepsilon}(\widehat{\mathbb{P}})) \leq \delta_0$, where $\varepsilon := \min_{\beta \in \mathcal{N}_\omega} \beta$. As a first step, we establish the necessary and sufficient condition for feasibility of problem (9), presented in the following proposition.

Proposition 4. *Let*

$$\delta_{min} = \min_{\substack{v_i \in [0,1], i=1,\dots,N, \\ \sum_{i=1}^N v_i = N\varepsilon}} \frac{1}{N\varepsilon} \sum_{i=1}^m v_i d_i,$$

with $\varepsilon := \min_{\beta \in \mathcal{N}_\omega} \beta$. For the radius δ_0 given in problem (9), we have problem (9) is feasible if and only if $\delta_0 \geq \delta_{min}$.

Remark 5. In Definition 2 of Esteban-Perez and Morales (2022), they also give the feasible condition to problem (5). That is, problem (5) is feasible if and only if

$$\delta_0 \geq \min_{\substack{v_i \in [0,1], i=1,\dots,N, \\ \sum_{i=1}^N v_i = N\beta}} \frac{1}{N\beta} \sum_{i=1}^m v_i d_i,$$

which is consistent with the result of Proposition 4 when $\mathcal{N}_\omega = \{\beta\}$ with $\beta \in (0, 1]$.

Next, we again formalize the admissible conditional distributions of Y , induced by feasible joint distributions $\mathbb{Q}$ satisfying $\mathbb{Q} \in \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}; \mathcal{N}_\omega)$ and $\mathbb{Q}(X \in \mathcal{N}_\gamma(x_0)) = 1$. Specifically,

$$\mathcal{K}(\mathcal{N}_\gamma(x_0), \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}; \mathcal{N}_\omega)) = \left\{ \mathbb{Q}|_Y : \mathbb{Q}|_Y \text{ is the marginal distribution of } \mathbb{Q} \text{ on } \mathcal{Y}, \right. \\ \left. \text{where } \mathbb{Q} \in \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}; \mathcal{N}_\omega) \text{ satisfies that } \mathbb{Q}(X \in \mathcal{N}_\gamma(x_0)) = 1 \right\}.$$

Theorem 6. *The problem (9), equivalently*

$$\min_{\alpha \in \mathcal{A}} \sup_{\mathbb{Q}|_Y \in \mathcal{K}(\mathcal{N}_\gamma(x_0), \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}; \mathcal{N}_\omega))} \rho_{\mathbb{Q}|_Y}[\ell(Y, \alpha)], \quad (10)$$

admits the union-ball reformulation (6) with

$$\mathcal{V} := \left\{ (p, \delta) : \varepsilon = \min_{\beta \in \mathcal{N}_\omega} \beta, p_i \in [0, \frac{1}{N\varepsilon}], i = 1, \dots, N, \sum_{i=1}^N p_i = 1, \delta = \delta_0 - \sum_{i=1}^m p_i d_i \right\}.$$

Juxtaposing the admissible sets in Theorems 3 and 6 reveals a shared structural backbone between models (7) and (9)—despite their distinct premises—while also pinpointing the subtle yet consequential ways

they diverge. The key difference lies in how they handle transport costs across the boundary of $\mathcal{N}_\gamma(x_0)$. Model (7) permits bidirectional transport, accounting for the movement of both in-region and out-of-region samples—i.e., the points $(\hat{x}_i, \hat{y}_i)$ for $i = 1, \dots, m$ and $i = m + 1, \dots, N$ —to determine the radius:

$$\delta = \varepsilon \left(\delta_0 - \frac{1}{N} \sum_{i=m+1}^N d_i \right) - \sum_{i=1}^m p_i d_i + \sum_{i=m+1}^N p_i d_i.$$

By contrast, model (9) restricts the budget to inward transport only—considering just $(\hat{x}_i, \hat{y}_i)$ for $i = 1, \dots, m$ —and ignores any offsetting movement out of the region. This structural asymmetry makes the partial model more conservative. Notably, the two models become exactly equivalent in two common cases: (1) when the covariate neighborhood $\mathcal{N}_\gamma(x_0)$ reduces to a singleton², and (2) when the empirical distribution $\hat{\mathbb{P}}$ places no mass within $\mathcal{N}_\gamma(x_0)$. In either case, the transport cost across the boundary vanishes, and the models coincide. This equivalence, previously unnoticed, is practically important: it shows that either model can be applied interchangeably in these cases, which, to the best of our knowledge, had not been recognized before.

The union-ball characterization also enables direct and explicit comparison between the joint-distribution models (7) and (9) and the Predict-then-Robustify model (3). In particular, the joint formulations can be viewed as robust counterparts of (3) whenever the estimated pair $(\hat{\mathbb{P}}^{x_0} |_{\mathcal{Y}}, \hat{\delta})$ from (3) lies within the admissible set $\mathcal{V}$ characterized in Theorems 3 and 6. Verifying this condition reduces to checking the linear (in)equalities that define $\mathcal{V}$ —a straightforward task. Moreover, as long as the conditional distribution $\hat{\mathbb{P}}^{x_0} |_{\mathcal{Y}}$ satisfies the upper-bound constraints encoded in $\mathcal{V}$, one can always calibrate the joint model’s transport budget δ_0 to ensure it is at least as robust as (3). For instance, under the setup of Theorem 6, setting

$$\delta_0 = \hat{\delta} + \sum_{i=1}^m \hat{p}_i d_i$$

guarantees that $(\hat{\mathbb{P}}^{x_0} |_{\mathcal{Y}}, \hat{\delta}) \in \mathcal{V}$, and thus ensures that the joint model dominates in terms of conservativeness. These comparisons, opaque in the original formulations, become transparent, verifiable, and practically calibratable through the lens of the union-ball characterization.

A New Robust Counterpart to Predict-then-Robustify

Having shown that models (7) and (9) can be interpreted as robust counterparts to the Predict-then-Robustify scheme (3), we conclude by illustrating how the framework can also be used to design a new formulation via direct specification of the admissible set $\mathcal{V}$. Given an estimated conditional probability distribution $\hat{\mathbb{P}}_{\mathcal{Y}|\mathcal{X}}$ with a fixed covariate value $X = x_0$, it is natural to account for the uncertainty in using it as a reference by first constructing an uncertainty set around it:

$$\mathcal{C} := \left\{ p \in [0, 1]^N : \sum_{i=1}^N p_i = 1, d(p, \hat{\mathbb{P}}^{x_0} |_{\mathcal{Y}}) \leq \gamma \right\},$$

²According to the proof of Theorem 3, this equivalence holds if the risk measure ρ and the loss function ℓ satisfy Assumption 2.2.

where d denotes a “distance” metric between finite-dimensional probability vectors, such as a ϕ -divergence, and $\gamma \geq 0$ is the radius. To avoid the over-conservatism of a fixed-radius union ball—and inspired by Theorems 3 and 6—we allow the radius δ to vary with the choice of reference distribution p : the farther p is from the estimate $\widehat{\mathbb{P}}_{Y|X}$, the smaller the radius. This trade-off can be modeled through the admissible set

$$\mathcal{V} := \left\{ (p, \delta) : p \in \mathcal{C}, 0 \leq \delta \leq \bar{h} \left(d(p, \widehat{\mathbb{P}}^{x_0} |_{\mathcal{Y}}) \right) \right\},$$

where $\bar{h} : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is assumed to be concave and decreasing, with $\bar{h}(0) = \widehat{\delta}$ and $\bar{h}(\gamma) = 0$. This construction ensures that the union ball includes, as a member, the ball underlying problem (3): when $d(p, \widehat{\mathbb{P}}^{x_0} |_{\mathcal{Y}}) = 0$, the bound $\delta \leq \widehat{\delta}$ holds, consistent with the Predict-then-Robustify scheme.

3 Unified Tractability for Distributionally Robust CRM

A key strength of the union-ball characterization lies not only in its versatility, but in its ability to systematically yield exact and tractable reformulations across a wide spectrum of distributionally robust CRM problems. As demonstrated throughout this section, it accommodates a broad range of risk functionals and modeling choices. Notably, for cases previously treated by Nguyen et al. (2024) via optimal transport—where only conservative approximations were available—our framework provides exact and significantly simpler reformulations.

Let $\mathcal{K}$ denote the set of conditional distributions $\mathbb{Q}|_{\mathcal{Y}}$ arising in a distributionally robust CRM problem (e.g. $\mathcal{K}(\mathcal{N}_{\gamma}(x_0), \mathcal{N}_{\omega}, \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}))$ in Theorem 3 and $\mathcal{K}(\mathcal{N}_{\gamma}(x_0), \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}; \mathcal{N}_{\omega}))$ in Theorem 6). The union-ball characterization reveals that the inner worst-case conditional risk reduces to a maximization over standard optimal transport-based problems. This yields the following reformulation of the distributionally robust CRM problem:

$$\min_{\alpha \in \mathcal{A}} \sup_{\mathbb{Q}|_{\mathcal{Y}} \in \mathcal{K}} \rho_{\mathbb{Q}|_{\mathcal{Y}}} [\ell(Y, \alpha)] \quad (11)$$

$$= \min_{\alpha \in \mathcal{A}} \sup_{\mathbb{Q}|_{\mathcal{Y}} \in \bigcup_{(p, \delta) \in \mathcal{V}} \mathbb{B}_{\delta} \left(\sum_{i=1}^N p_i \mathbb{I}_{\mathcal{Y}_i} \right)} \rho_{\mathbb{Q}|_{\mathcal{Y}}} [\ell(Y, \alpha)] \quad (12)$$

$$= \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}} g(\alpha, (p, \delta)), \quad (13)$$

where

$$g(\alpha, (p, \delta)) := \left\{ \sup_{\mathbb{Q}|_{\mathcal{Y}} \in \mathbb{B}_{\delta} \left(\sum_{i=1}^N p_i \mathbb{I}_{\mathcal{Y}_i} \right)} \rho_{\mathbb{Q}|_{\mathcal{Y}}} [\ell(Y, \alpha)] \right\}. \quad (14)$$

This decomposition offers two key advantages for tackling problem (11):

1. For each fixed (p, δ) , the inner subproblem (14) becomes a well-structured (unconditional) Optimal Transport DRO centered at a finitely supported distribution. This facilitates tractable reformulations by leveraging structural properties inherited from Wasserstein DRO formulations—namely, with $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}^q$ for any $y_1, y_2 \in \mathcal{Y}$ and $q \in [1, \infty)$.

2. The outer maximization $\sup_{(p, \delta) \in \mathcal{V}} g(\alpha, (p, \delta))$ broadly resembles a robust optimization problem with uncertain probability weights p , but with additional complexity arising from the uncertain radius δ . As will be shown later in this section, it nevertheless admits tractable reformulations.

In particular, to highlight the tractability of (11) through the min–max reformulation (13), we first state the following structural observation.

Proposition 7. *Suppose $\rho_{\mathcal{Q}|Y}$ is a law-invariant risk functional that is concave in the distribution, and $\rho_{\mathcal{Q}|Y} \circ \ell$ is convex in α . Then $g(\alpha, (p, \delta))$ is jointly concave in (p, δ) and convex in α .*

Observe that $\min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}} g(\alpha, (p, \delta)) = \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}, \delta \geq 0} g(\alpha, (p, \delta))$, since $g(\alpha, (p, \delta)) = -\infty$ whenever $\mathbb{B}_\delta\left(\sum_{i=1}^N p_i \mathbf{I}_{\hat{y}_i}\right)$ is empty, i.e., when $\delta < 0$. For brevity, define

$$\mathcal{V}_+ = \{(p, \delta) : (p, \delta) \in \mathcal{V}, \delta \geq 0\},$$

which will be used hereafter. In the remainder of this section, we focus on affine decision rules, where $\ell(Y, \alpha) = \ell(Y^\top \alpha)$ —an assumption already required in unconditional Wasserstein DRO settings to ensure tractability. Leveraging the decomposition insight and structural properties identified above, we show how a wide class of distributionally robust CRM problems can be reformulated as tractable convex programs. The following definitions of convex and concave conjugate functions will be instrumental in this development.

Definition 3.1. *Assume a convex set $\mathcal{D} \subset \mathbb{R}^n$.*

- (i) *The concave conjugate $f_*(\cdot)$ of a function $f : \mathcal{D} \rightarrow \mathbb{R}$ is defined as a function $f_* : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{-\infty\}$:*

$$f_*(v) = \inf_{p \in \mathcal{D}} \{v^\top p - f(p)\}.$$

- (ii) *The convex conjugate $g^*(\cdot)$ of a function $g : \mathcal{D} \rightarrow \mathbb{R}$ is defined as a function $g^* : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$:*

$$g^*(v) = \sup_{p \in \mathcal{D}} \{v^\top p - g(p)\}.$$

Remark 8. *Let $f(\cdot)$ and $g(\cdot)$ be $-\infty$ and ∞ outside $\mathcal{D}$. Then the aforementioned definition fits the standard definition of convex (concave) conjugate.*

To facilitate navigation of the tractable reformulations developed in this section, Table 1 provides a concise summary of the results. Moreover, the risk functionals presented in Table 1 satisfy Assumption 2.2, that is, the risk functionals is OT-continuity under that the cost function $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}^q$ for any $y_1, y_2 \in \mathcal{Y}$ with $q \in [1, \infty)$. We present the result in the following lemma.

Lemma 9. *The risk functionals presented in Table 1 is OT-continuity under that the cost function $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}^q$ for any $y_1, y_2 \in \mathcal{Y}$ with $q \in [1, \infty)$.*

Table 1: Summary of tractable results by risk functional, norm choice, and loss function ℓ

Risk functional	Order 1 norm ($\ \cdot\ _{\mathcal{Y}}$)	Order q norm ($\ \cdot\ _{\mathcal{Y}}^q$)	
		Specific loss function	General loss function
Expectation ($\mathbb{E}[\ell(Z)]$)	Proposition 10 (i)	Proposition 10 (ii)	Proposition 11
$\inf_{t \in \mathbb{R}^k} \mathbb{E}[\ell(Z, t)]$	Proposition 13 (i)	Proposition 13 (ii)	Proposition 14
$\inf_{t \in \mathbb{R}} \left\{ t + (\mathbb{E}[\ell^q(Z, t)])^{1/q} \right\}$	Proposition 17 (i)	Proposition 17 (i)	Proposition 18
Shortfall risk measure	Proposition 19	Remark 20	Proposition 21
Distortion functional	Proposition 22 (i)	Proposition 22 (ii)	—

3.1 Conditional Expectation

We begin with the base case of the conditional expectation problem under its union-ball formulation:

$$\min_{\alpha \in \mathcal{A}} \sup_{\mathbb{Q}|_{\mathcal{Y}} \in \bigcup_{(p, \delta) \in \mathcal{V}} \mathbb{B}_{\delta} \left(\sum_{i=1}^N p_i I_{\tilde{y}_i} \right)} \mathbb{E}_{\mathbb{Q}|_{\mathcal{Y}}} [\ell(Y^{\top} \alpha)] = \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}} \left\{ \sup_{\mathbb{Q}|_{\mathcal{Y}} \in \mathbb{B}_{\delta} \left(\sum_{i=1}^N p_i I_{\tilde{y}_i} \right)} \mathbb{E}_{\mathbb{Q}|_{\mathcal{Y}}} [\ell(Y^{\top} \alpha)] \right\}. \quad (15)$$

Invoking the regularization reformulation of the inner worst-case subproblem (Wu et al. (2022), see Section EC.2), we obtain a family of robust optimization problems. As we show below, these problems admit convex reformulations.

Proposition 10. (i) Let $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}$ for any $y_1, y_2 \in \mathcal{Y}$ and $\ell : \mathbb{R} \rightarrow \mathbb{R}$ be a convex and Lipschitz continuous function. The problem (15) is equivalent to the robust optimization problem

$$\min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}_+} \sum_{i=1}^N p_i \ell(\tilde{y}_i^{\top} \alpha) + \text{Lip}(\ell) \delta \|\alpha\|_*,$$

and can be solved by the convex program

$$\begin{aligned} \inf_{\alpha \in \mathcal{A}, v} \quad & \sigma^*(v \mid \mathcal{V}_+) \\ \text{s.t.} \quad & v_i \geq \ell(\tilde{y}_i^{\top} \alpha), \quad i = 1, \dots, N, \\ & v_{N+1} \geq \text{Lip}(\ell) \|\alpha\|_*. \end{aligned}$$

(ii) Let $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}^q$ for any $y_1, y_2 \in \mathcal{Y}$ and $q \in (1, \infty)$.

(1) Assume the loss function ℓ takes one of the following two forms, multiplied by $C > 0$:

- (a) $\ell_1(x) = x + b$ or $\ell_1(x) = -x + b$ with some $b \in \mathbb{R}$;
- (b) $\ell_2(x) = |x - b_1| + b_2$ with some $b_1, b_2 \in \mathbb{R}$.

The problem (15) is equivalent to the robust optimization problem

$$\min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}_+} \sum_{i=1}^N p_i \ell(\tilde{y}_i^{\top} \alpha) + C \delta^{\frac{1}{q}} \|\alpha\|_*,$$

and can be solved by the convex program

$$\begin{aligned} \inf_{\alpha \in \mathcal{A}, v} \quad & \sigma^*(v|\mathcal{V}_+) + C_q v_{N+1} \left\| \frac{\alpha}{v_{N+1}} \right\|_*^{\frac{q}{q-1}} \\ \text{s.t.} \quad & v_i \geq \ell(\widehat{y}_i^\top \alpha), \quad i = 1, \dots, N, \\ & v_{N+1} \geq 0, \end{aligned}$$

where $C_q = C^{\frac{q}{q-1}} \left(q^{\frac{1}{1-q}} - q^{\frac{q}{1-q}} \right)$.

(2) Assume the loss function ℓ is of the form $\ell(x) = (C \cdot \ell_i(x))^q$, where $C > 0$, $q \in (1, \infty)$, and ℓ_i is one of the following:

- (a) $\ell_1(x) = (x - b)_+$ with some $b \in \mathbb{R}$;
- (b) $\ell_2(x) = (x - b)_-$ with some $b \in \mathbb{R}$;
- (c) $\ell_3(x) = (|x - b_1| - b_2)_+$ with some $b_1 \in \mathbb{R}$ and $b_2 \geq 0$;
- (d) $\ell_4(x) = |x - b_1| + b_2$ with some $b_1 \in \mathbb{R}$ and $b_2 > 0$.

The problem (15) is equivalent to the robust optimization problem

$$\min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}_+} \left[\left(\sum_{i=1}^N p_i \ell(\widehat{y}_i^\top \alpha) \right)^{1/q} + C \delta^{\frac{1}{q}} \|\alpha\|_* \right]^q,$$

and can be solved by the convex program

$$\begin{aligned} \inf_{\alpha \in \mathcal{A}, v, y \geq 0} \quad & \sigma^*(v|\mathcal{V}_+) + C_q^1 y + C_q v_{N+1} \left\| \frac{\alpha}{v_{N+1}} \right\|_*^{\frac{q}{q-1}} \\ \text{s.t.} \quad & v_i \geq \frac{\ell(\widehat{y}_i^\top \alpha)}{y^{q-1}}, \quad i = 1, \dots, N, \\ & v_{N+1} \geq 0. \end{aligned}$$

where $C_q^1 = q^{\frac{1}{1-q}} - q^{\frac{q}{1-q}}$ and $C_q = C^{\frac{q}{q-1}} \left(q^{\frac{1}{1-q}} - q^{\frac{q}{1-q}} \right)$.

More generally, even without regularization equivalents, the decomposition remains powerful: by isolating each Wasserstein DRO subproblem, it enables us to exploit their projection property (Wu et al. 2022):

$$\sup_{\mathbb{Q}|_Y \in \mathbb{B}_\delta \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i} \right)} \mathbb{E}_{\mathbb{Q}|_Y} [\ell(Y^\top \alpha)] = \sup_{\mathbb{Q}|_Z \in \mathbb{B}_{\delta \|\alpha\|_*} \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i^\top \alpha} \right)} \mathbb{E}_{\mathbb{Q}|_Z} [\ell(Z)],$$

which yields a one-dimensional reformulation that is considerably easier to handle. Leveraging this insight, we derive a final convex program reformulation of the conditional expectation problem (15), applicable to any suitably regular convex loss function ℓ .

Proposition 11. Let $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}^q$ for any $y_1, y_2 \in \mathcal{Y}$ and $q \in (1, \infty)$. Assuming the loss function ℓ is convex and satisfies $\ell(x) - \ell(x_0) \leq L|x - x_0|^q + M, x \in \mathbb{R}$ for some $L, M > 0$, and some $x_0 \in \mathbb{R}$,

the problem (15) can be solved by the convex program

$$\begin{aligned}
& \inf_{\alpha \in \mathcal{A}, v, \eta \geq 0} \sigma^*(v | \mathcal{V}_+) \\
& \text{s.t. } v_i \geq \sup_{z \in \mathbb{R}} \left\{ z \cdot \widehat{y}_i^\top \alpha - \ell^*(z) + \eta \cdot \frac{|z|^{q^*}}{q^* q^{q^*-1}} \right\}, \quad i = 1, \dots, N, \\
& v_{N+1} \geq \|\alpha\|_*^q / \eta^{q-1}.
\end{aligned} \tag{16}$$

Constraints like (16), where the right-hand side involves an optimization problem, commonly arise in dual reformulations of Wasserstein DRO. What sets (16) apart is that the optimization is univariate, which not only simplifies analysis and computation but also enables further problem-specific reductions, as we demonstrate below.

Example 3.1. (i) (Piecewise linear convex loss function) Let $\ell(x) = \sup_{j=1, \dots, k} a_j x + b_j$. The convex conjugate function is

$$\ell^*(z) = \begin{cases} -\max_{j: a_j = z} b_j & \text{if } z \in \{a_1, \dots, a_k\}, \\ +\infty & \text{otherwise.} \end{cases}$$

Without loss of generality, assume that $a_{j_1} \neq a_{j_2}$ for $j_1 \neq j_2$. Thus, for $q > 1$, we have

$$\sup_{z \in \mathbb{R}} \left\{ zt - \ell^*(z) + \eta \cdot \frac{|z|^{q^*}}{q^* q^{q^*-1}} \right\} = \max_{j=1, \dots, k} \left\{ a_j t + b_j + \eta \cdot \frac{|a_j|^{q^*}}{q^* q^{q^*-1}} \right\}.$$

(ii) (Piecewise quadratic convex loss function) Let $\ell(x) = \sup_{j=1, \dots, k} a_j x^2 + b_j x + c_j$ and $a_j > 0$ for $j = 1, \dots, k$. The convex conjugate function is

$$\ell^*(z) = \min_{j=1, \dots, k} \left\{ \frac{(z - b_j)^2}{4a_j} - c_j \right\}.$$

Thus, for $q = 2$, we have

$$\sup_{z \in \mathbb{R}} \left\{ zt - \ell^*(z) + \eta \cdot \frac{|z|^{q^*}}{q^* q^{q^*-1}} \right\} = \max_{j=1, \dots, k} \left\{ \frac{a_j t^2 + b_j t}{1 - a_j \eta} + c_j + \frac{a_j b_j^2 \eta}{4(1 - a_j \eta)} \right\},$$

and $\eta < \min_{j=1, \dots, k} \frac{1}{a_j}$.

(iii) (Power loss function) Let $\ell(x) = \frac{|x|^q}{q}$ for $q > 1$. The convex conjugate function is

$$\ell^*(z) = \frac{|z|^{q^*}}{q^*}.$$

Thus, we have

$$\sup_{z \in \mathbb{R}} \left\{ zt - \ell^*(z) + \eta \cdot \frac{|z|^{q^*}}{q^* q^{q^*-1}} \right\} = \frac{|t|^q}{q} \left(1 - \frac{\eta}{q^{q^*-1}} \right)^{-(q-1)},$$

and $\eta < q^{q^*-1}$.

Thus far, our tractable reformulations have been kept general, applicable to any distributionally robust CRM that admits a union-ball representation with a convex admissible set $\mathcal{V}$. To illustrate how the support function $\sigma^*(v | \mathcal{V}_+)$ is derived, we now apply the admissible sets from Theorems 3 and 6. These examples also enable direct comparison between the final formulations from our union-ball approach and those in existing OT-based methods. Recall that $N_\omega = [\omega_1, \omega_2] \subset (0, 1]$.

Example 3.2. Denote $p' = (p, \delta, \varepsilon)$. Following Theorem 3, the conjugate of the indicator function of $\mathcal{V}_+$ is given by

$$\sigma^*(v | \mathcal{V}_+) = \sup_{(p, \delta) \in \mathcal{V}_+} v^\top (p, \delta) = \begin{cases} \sup v^\top (p, \delta) + 0 \cdot \varepsilon \\ \text{s.t. } Ap' \leq b, p' \geq 0 \end{cases} = \begin{cases} \inf z^\top b \\ \text{s.t. } A^\top z \geq (v, 0), z \geq 0, \end{cases}$$

where $A \in \mathbb{R}^{(N+6) \times (N+2)}$, and $b \in \mathbb{R}^{N+6}$ are defined by

$$A = \begin{pmatrix} 1 & \dots & 1 & 0 & 0 \\ -1 & \dots & -1 & 0 & 0 \\ 0 & \dots & 0 & 0 & 1 \\ 0 & \dots & 0 & 0 & -1 \\ 1 & \dots & 0 & 0 & -\frac{1}{N} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & 1 & 0 & -\frac{1}{N} \\ -a_1 & \dots & -a_N & -a_{N+1} & -a_{N+2} \\ a_1 & \dots & a_N & a_{N+1} & a_{N+2} \end{pmatrix}, \quad \text{with } a = \begin{pmatrix} -d_1 \\ \vdots \\ -d_m \\ d_{m+1} \\ \vdots \\ d_N \\ -1 \\ \delta_0 - \frac{1}{N} \sum_{i=m+1}^N d_i \end{pmatrix}, \quad \text{and } b = \begin{pmatrix} 1 \\ -1 \\ \frac{1}{\omega_1} \\ -\frac{1}{\omega_2} \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Example 3.3. Note that $N_\omega = [\omega_1, \omega_2] \subset (0, 1]$. Then $\varepsilon = \min_{\beta \in N_\omega} \beta = \omega_1$. Following Theorem 6, the conjugate of the indicator function of $\mathcal{V}_+$ is given by

$$\sigma^*(v | \mathcal{V}_+) = \sup_{p' \in \mathcal{V}_+} v^\top p' = \begin{cases} \sup v^\top p' \\ \text{s.t. } A' p' \leq b', p' \geq 0 \end{cases} = \begin{cases} \inf z^\top b' \\ \text{s.t. } (A')^\top z \geq v, z \geq 0, \end{cases}$$

where $A' \in \mathbb{R}^{(N+4) \times (N+1)}$, and $b' \in \mathbb{R}^{N+4}$ are defined by

$$A' = \begin{pmatrix} 1 & \dots & 1 & 0 \\ -1 & \dots & -1 & 0 \\ 1 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 1 & 0 \\ -a'_1 & \dots & -a'_N & -a'_{N+1} \\ a'_1 & \dots & a'_N & a'_{N+1} \end{pmatrix}, \quad \text{with } a' = \begin{pmatrix} d_1 \\ \vdots \\ d_m \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}, \quad \text{and } b' = \begin{pmatrix} 1 \\ -1 \\ \frac{1}{N\omega_1} \\ \vdots \\ \frac{1}{N\omega_1} \\ -\delta_0 \\ \delta_0 \end{pmatrix}.$$

Remark 12. Building on the conjugates in Examples 3.2 and 3.3, we provide full reformulations of the problems in Appendix EC.3. There, we also discuss how our union-ball reformulation substantially simplifies the final form, in comparison with the reformulation given in Nguyen et al. (2024), as highlighted in Remark 31.

3.2 Conditional Risk Functionals in Expectation Form

The decomposition framework naturally extends to a broader class of risk functionals that admit expectation-based representations. In this section, we focus on solving the following problem

$$\min_{\alpha \in \mathcal{A}} \sup_{\mathbb{Q}|_Y \in \bigcup_{(p, \delta) \in \mathcal{V}} \mathbb{B}_\delta\left(\sum_{i=1}^N p_i I_{\hat{y}_i}\right)} \rho_{\mathbb{Q}|_Y}(Y^\top \alpha) = \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}} \left\{ \sup_{\mathbb{Q}|_Y \in \mathbb{B}_\delta\left(\sum_{i=1}^N p_i I_{\hat{y}_i}\right)} \rho_{\mathbb{Q}|_Y}(Y^\top \alpha) \right\}. \quad (17)$$

We begin with risk measures of the form

$$\rho^{(1)}(Z) = \inf_{t \in \mathbb{R}^k} \mathbb{E}[\ell(Z, t)], \quad (18)$$

as considered in Nguyen et al. (2024), where only conservative reformulations were available. In contrast, our approach yields reformulations that are not only provably exact, but also substantially simpler—requiring fewer variables, fewer and simpler constraints, and exhibiting a modular structure that facilitates both analysis and computation. The following property of a function is necessary for the propositions in this section.

Definition 3.2. (Level-bounded) A function $\ell : \mathbb{R}^k \rightarrow \mathbb{R}$ is level bounded if for each $A \in \mathbb{R}$, the sublevel set $\{t \in \mathbb{R}^k : \ell(t) \leq A\}$ is bounded (possibly empty).

As in the case of conditional expectation, we first exploit regularization equivalents to obtain robust optimization reformulations of (17), and then show that these admit convex minimization reformulations.

Proposition 13. (i) Let $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}$ for any $y_1, y_2 \in \mathcal{Y}$. Suppose that $\ell(z, t) : \mathbb{R} \times \mathbb{R}^k \rightarrow \mathbb{R}$ is jointly convex in (z, t) , level-bounded in t for all $z \in \mathbb{R}$, and Lipschitz continuous in z for all $t \in \mathbb{R}^k$ with a uniform Lipschitz constant $\text{Lip}(\ell)$. The problem (17) with $\rho = \rho^{(1)}$ is equivalent to the robust

optimization problem

$$\inf_{\alpha \in \mathcal{A}, t \in \mathbb{R}^k} \sup_{(p, \delta) \in \mathcal{V}_+} \sum_{i=1}^N p_i \ell(\widehat{y}_i^\top \alpha, t) + \text{Lip}(\ell) \delta \|\alpha\|_*,$$

and can be solved by the convex program

$$\begin{aligned} & \inf_{\alpha \in \mathcal{A}, t \in \mathbb{R}^k, v} \sigma^*(v | \mathcal{V}_+) \\ & \text{s.t. } v_i \geq \ell(\widehat{y}_i^\top \alpha, t), i = 1, \dots, N, \\ & \quad v_{N+1} \geq \text{Lip}(\ell) \|\alpha\|_*. \end{aligned}$$

(ii) Let $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}^q$ for any $y_1, y_2 \in \mathcal{Y}$, $q \in (1, \infty)$ and $k = 1$. Assume the loss function ℓ is the form $\ell(z, t) := (C \bar{\ell}_i(z - t))^q$, where $C > 0$, and $\bar{\ell}_i$ is one of the following:

- (a) $\bar{\ell}_1(x) = (|x - b_1| - b_2)_+$ with some $b_1 \in \mathbb{R}$ and $b_2 \geq 0$;
- (b) $\bar{\ell}_2(x) = |x - b_1| + b_2$ with some $b_1 \in \mathbb{R}$ and $b_2 > 0$.

The problem (17) with $\rho = \rho^{(1)}$ is equivalent to the robust optimization problem

$$\inf_{\alpha \in \mathcal{A}, t \in \mathbb{R}} \sup_{(p, \delta) \in \mathcal{V}_+} \left[\left(\sum_{i=1}^N p_i \ell(\widehat{y}_i^\top \alpha, t) \right)^{1/q} + C \delta^{\frac{1}{q}} \|\alpha\|_* \right]^q,$$

and can be solved by the convex program

$$\begin{aligned} & \inf_{\alpha \in \mathcal{A}, t \in \mathbb{R}, y \geq 0, v} \left[\sigma^*(v | \mathcal{V}_+) + C_q^1 y + C_q v_{N+1} \left\| \frac{\alpha}{v_{N+1}} \right\|_*^{\frac{q}{q-1}} \right]^q \\ & \text{s.t. } v_i \geq \frac{\ell(\widehat{y}_i^\top \alpha, t)}{y^{q-1}}, i = 1, \dots, N, \\ & \quad v_{N+1} \geq 0, \end{aligned}$$

where $C_q^1 = q^{\frac{1}{1-q}} - q^{\frac{q}{1-q}}$ and $C_q = C^{\frac{q}{q-1}} \left(q^{\frac{1}{1-q}} - q^{\frac{q}{1-q}} \right)$.

For general convex loss functions ℓ , we obtain the following convex reformulation by leveraging two key ingredients inherited from the unconditional Wasserstein DRO setting: (i) the projection property, and (ii) the validity of a min–max switching argument, established in the appendix (see Lemma 30).

Proposition 14. Let $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}^q$ for any $y_1, y_2 \in \mathcal{Y}$ and $q \in (1, \infty)$. Assume the loss function ℓ is the form $\ell(z, t) = \sum_{q_1 \in \mathcal{I}} C_{q_1} (\bar{\ell}_{q_1}(z, t))^{q_1}$, with a finite index set $\mathcal{I} \subset [1, q]$ and $C_{q_1} \geq 0$, where $\bar{\ell}_{q_1}(z, t) : \mathbb{R} \times \mathbb{R}^k \rightarrow \mathbb{R}$ is nonnegative, jointly convex in $(z, t) \in \mathbb{R}^{k+1}$, level-bounded in $t \in \mathbb{R}^k$ for all $z \in \mathbb{R}$, and Lipschitz continuous in z for all $t \in \mathbb{R}^k$ with a uniform Lipschitz constant $\text{Lip}(\bar{\ell}_{q_1})$, for $q_1 \in \mathcal{I}$. The

problem (17) with $\rho = \rho^{(1)}$ can be solved by the convex program

$$\begin{aligned} & \inf_{\alpha \in \mathcal{A}, t \in \mathbb{R}^k, \eta \geq 0, v} \sigma^*(v | \mathcal{V}_+) \\ & \text{s.t. } v_i \geq \sup_{z \in \mathbb{R}} \left\{ z \cdot \widehat{y}_i^\top \alpha - (\ell(\cdot, t))^*(z) + \eta \cdot \frac{|z|^{q^*}}{q^* q^{q^*-1}} \right\}, \quad i = 1, \dots, N, \\ & v_{N+1} \geq \|\alpha\|_*^q / \eta^{q-1}. \end{aligned}$$

Remark 15. Note that if $\bar{\ell}_{q_1}(z, t) = \bar{\ell}_{q_1}(z)$, C_{q_1} can relax the nonnegative constraint, that is, $C_{q_1} \in \mathbb{R}$, and $\bar{\ell}_{q_1}(z)$ also can relax the level-bounded and convex constraint.

As a demonstrative example, we apply our framework to the mean-variance portfolio allocation problem studied in Nguyen et al. (2024), showcasing its ability to yield not only provably exact but also markedly simpler reformulations.

Example 3.4. The robustified conditional mean-variance portfolio allocation problem in Nguyen et al. (2024) is as follows:

$$\min_{\alpha \in \mathcal{A}} \sup_{\substack{\mathbb{Q} \in \mathcal{B}_{\delta_0}(\widehat{\mathbb{P}}) \\ \mathbb{Q}(X \in \mathcal{N}_\gamma(x_0)) \in \mathcal{N}_\omega}} \text{Variance}_{\mathbb{Q}} [Y^\top \alpha \mid X \in \mathcal{N}_\gamma(x_0)] - \theta \cdot \mathbb{E}_{\mathbb{Q}} [Y^\top \alpha \mid X \in \mathcal{N}_\gamma(x_0)], \quad (19)$$

where $\mathcal{N}_\omega = [\varepsilon_0, 1]$, $\varepsilon_0 \in (0, 1]$ and $\theta \geq 0$. Suppose in addition that $X = \mathbb{R}^{n_1}$, $\mathcal{Y} = \mathbb{R}^{n_2}$, $\mathbb{D}_X(x, \widehat{x}) = \|x - \widehat{x}\|^2$ and $\mathbb{D}_Y(y, \widehat{y}) = \|y - \widehat{y}\|_2^2$. The problem (19) can be solved by the convex program

$$\begin{aligned} & \inf_{\alpha \in \mathcal{A}, 0 \leq \eta \leq 1, t \in \mathbb{R}, z, s} z^\top b - \frac{\theta^2}{4} - t\theta \\ & \text{s.t. } z \geq 0, \quad A^\top z \geq s \\ & s_i \geq \frac{(\widehat{y}_i^\top \alpha - \theta/2 - t)^2}{1 - \eta}, \quad i = 1, \dots, N, \\ & s_{N+1} \geq \|\alpha\|_*^2 / \eta, \quad s_{N+2} \geq 0, \end{aligned}$$

where $A \in \mathbb{R}^{(N+6) \times (N+2)}$, and $b \in \mathbb{R}^{N+6}$ are defined by

$$A = \begin{pmatrix} 1 & \dots & 1 & 0 & 0 \\ -1 & \dots & -1 & 0 & 0 \\ 0 & \dots & 0 & 0 & 1 \\ 0 & \dots & 0 & 0 & -1 \\ 1 & \dots & 0 & 0 & -\frac{1}{N} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & 1 & 0 & -\frac{1}{N} \\ -a_1 & \dots & -a_N & -a_{N+1} & -a_{N+2} \\ a_1 & \dots & a_N & a_{N+1} & a_{N+2} \end{pmatrix}, \quad \text{with } a = \begin{pmatrix} -d_1 \\ \vdots \\ -d_m \\ d_{m+1} \\ \vdots \\ d_N \\ -1 \\ \delta_0 - \frac{1}{N} \sum_{i=m+1}^N d_i \end{pmatrix}, \quad \text{and } b = \begin{pmatrix} 1 \\ -1 \\ \frac{1}{\varepsilon_0} \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Remark 16. One can verify that the constraint $s_i \geq \frac{(\hat{y}_i^\top \alpha - \theta/2 - t)^2}{1 - \eta}$, $i = 1, \dots, N$, can be reformulated as the following second order cone constraints

$$\left\| \begin{bmatrix} 2\hat{y}_i^\top \alpha - 2t - \theta \\ s_i - (1 - \eta) \end{bmatrix} \right\|_2 \leq s_i + 1 + \eta, \quad i = 1, \dots, N.$$

and that the constraint $s_{N+1} \geq \|\alpha\|_*^2 / \eta$ can be reformulated as the following second order cone constraint

$$\left\| \begin{bmatrix} 2\alpha \\ \eta - s_{N+1} \end{bmatrix} \right\|_2 \leq \eta - s_{N+1}.$$

Thus, our reformulation is a second-order cone problem (SOCP), while the reformulation of Proposition 4.5 in Nguyen et al. (2024) is a semi-definite problem (SDP). From this perspective, our formulation has a simpler form and is easy to apply.

Our framework also accommodates other prominent risk measures with expectation-based structure, including those of the form

$$\rho^{(2)}(Z) = \inf_{t \in \mathbb{R}} \left\{ t + (\mathbb{E}[\ell^q(Z, t)])^{1/q} \right\}, \quad (20)$$

for $q \in [1, \infty)$, which lie beyond the scope of both Nguyen et al. (2024) and Esteban-Perez and Morales (2022). The derivation of tractable reformulations for this class proceeds analogously, building on the same decomposition principles introduced earlier.

Proposition 17. (i) Let $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}$ for any $y_1, y_2 \in \mathcal{Y}$ and $q = 1$. Suppose that $\ell(z, t) : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ is nonnegative, jointly convex in $(z, t) \in \mathbb{R}^2$, and Lipschitz continuous in z for all $t \in \mathbb{R}$ with a uniform Lipschitz constant $\text{Lip}(\ell)$. Moreover, assume $t + \ell(z, t)$ is level bounded in t for

all $z \in \mathbb{R}$. The problem (17) with $\rho = \rho^{(2)}$ is equivalent to the robust optimization problem

$$\inf_{\alpha \in \mathcal{A}, t \in \mathbb{R}} \sup_{(p, \delta) \in \mathcal{V}_+} t + \sum_{i=1}^N p_i [\ell(\widehat{y}_i^\top \alpha, t)] + \text{Lip}(\ell) \delta \|\alpha\|_*,$$

and can be solved by the convex program

$$\begin{aligned} & \inf_{\alpha \in \mathcal{A}, t \in \mathbb{R}, v} t + \sigma^*(v | \mathcal{V}_+) \\ & \text{s.t. } v_i \geq \ell(\widehat{y}_i^\top \alpha, t), i = 1, \dots, N, \\ & v_{N+1} \geq \text{Lip}(\ell) \|\alpha\|_*. \end{aligned}$$

(ii) Let $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}^q$ for any $y_1, y_2 \in \mathcal{Y}$ and $q \in (1, \infty)$. Assume the loss function ℓ is the form $\ell(z, t) := C \bar{\ell}_i(z - t)$, where $C > 1$, and $\bar{\ell}_i$ is one of the following:

- (a) $\bar{\ell}_1(x) = (x - b)_+$ with some $b \in \mathbb{R}$;
- (b) $\bar{\ell}_2(x) = (|x - b_1| - b_2)_+$ with some $b_1 \in \mathbb{R}$ and $b_2 \geq 0$;
- (c) $\bar{\ell}_3(x) = |x - b_1| + b_2$ with some $b_1 \in \mathbb{R}$ and $b_2 > 0$,

or $\ell(z, t) = C(|z| - t)_+$. The problem (17) with $\rho = \rho^{(2)}$ is equivalent to the robust optimization problem

$$\inf_{\alpha \in \mathcal{A}, t \in \mathbb{R}} \sup_{(p, \delta) \in \mathcal{V}_+} t + \left(\sum_{i=1}^N p_i [\ell^q(\widehat{y}_i^\top \alpha, t)] \right)^{1/q} + C \delta^{\frac{1}{q}} \|\alpha\|_*,$$

and can be solved by the convex program

$$\begin{aligned} & \inf_{\alpha \in \mathcal{A}, t \in \mathbb{R}, y \geq 0, v} t + \sigma^*(v | \mathcal{V}_+) + C_q^1 y + C_q v_{N+1} \left\| \frac{\alpha}{v_{N+1}} \right\|_*^{\frac{q}{q-1}} \\ & \text{s.t. } v_i \geq \ell^q(\widehat{y}_i^\top \alpha, t) / y^{q-1}, i = 1, \dots, N, \\ & v_{N+1} \geq 0, \end{aligned}$$

where $C_q^1 = q^{\frac{1}{1-q}} - q^{\frac{q}{1-q}}$ and $C_q = C^{\frac{q}{q-1}} \left(q^{\frac{1}{1-q}} - q^{\frac{q}{1-q}} \right)$.

For the general case, we again rely on the two key ingredients—the projection property and the min–max switching argument (see Lemma 30)—to derive tractable convex reformulations.

Proposition 18. Let $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}^q$ for any $y_1, y_2 \in \mathcal{Y}$ and $q \in (1, \infty)$. Assume that $\ell(z, t)$ is nonnegative, jointly convex in $(z, t) \in \mathbb{R}^k$, and Lipschitz continuous in z for all $t \in \mathbb{R}$ with a uniform Lipschitz constant $\text{Lip}(\ell)$. Moreover, assume $t + \ell(z, t)$ is level bounded in t for all $z \in \mathbb{R}$. The problem (17) with

$\rho = \rho^{(2)}$ can be solved by the convex program

$$\begin{aligned} & \inf_{\alpha \in \mathcal{A}, \eta \geq 0, t \in \mathbb{R}, v, w \geq 0} t + \sigma^*(v | \mathcal{V}_+) + C_q^1 w \\ & \text{s.t. } v_i \geq \sup_{z \in \mathbb{R}} \left\{ z \cdot \widehat{y}_i^\top \alpha - \left(\frac{\ell^q(\cdot, t)}{w^{q-1}} \right)^* (z) + \eta \cdot \frac{|z|^{q^*}}{q^* q^{q^*-1}} \right\}, \quad i = 1, \dots, N, \\ & v_{N+1} \geq \|\alpha\|_*^q / \eta^{q-1}. \end{aligned}$$

Our methodology also applies to risk measures defined implicitly through expectation constraints, such as the utility-based shortfall risk measure, defined by

$$\rho^{(3)}(Z) = \inf \{ \kappa \in \mathbb{R} : \mathbb{E} [u(-Z - \kappa)] \leq l \},$$

where u is a convex, increasing function and l is a fixed constant in the interior of the range of u . The same reformulation principles apply.

Proposition 19. Let $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}$ for any $y_1, y_2 \in \mathcal{Y}$ and $u : \mathbb{R} \rightarrow \mathbb{R}$ be a convex, increasing and Lipschitz continuous function with Lipschitz constant $\text{Lip}(u)$. The problem (17) with $\rho = \rho^{(3)}$ is equivalent to the robust optimization problem

$$\begin{aligned} & \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}_+} \inf_{\kappa \in \mathbb{R}} \kappa \\ & \text{s.t. } \sum_{i=1}^N p_i u(-\widehat{y}_i^\top \alpha - \kappa) + \text{Lip}(u) \delta \|\alpha\|_* \leq l \\ & = \inf_{\alpha \in \mathcal{A}, \kappa \in \mathbb{R}} \kappa \\ & \text{s.t. } \sum_{i=1}^N p_i u(-\widehat{y}_i^\top \alpha - \kappa) + \text{Lip}(u) \delta \|\alpha\|_* \leq l \quad \forall (p, \delta) \in \mathcal{V}_+, \end{aligned}$$

and can be solved by the convex program

$$\begin{aligned} & \inf_{\alpha \in \mathcal{A}, \kappa \in \mathbb{R}, v} \kappa \\ & \text{s.t. } \sigma^*(v | \mathcal{V}_+) \leq l, \\ & v_i \geq u(-\widehat{y}_i^\top \alpha - \kappa), \quad i = 1, \dots, N, \\ & v_{N+1} \geq \text{Lip}(u) \|\alpha\|_*. \end{aligned}$$

Remark 20. Note that when $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}^q$ for some $q \in (1, \infty)$, a convex program reformulation is still possible, but—unlike the earlier cases where several forms of u were admissible—it applies only to the linear specification $u(x) = Cx + b$ with $C > 0$, as required by monotonicity. In this setting, problem (17)

admits the equivalent convex program:

$$\begin{aligned} & \inf_{\alpha \in \mathcal{A}, \kappa \in \mathbb{R}, v} \kappa \\ & \text{s.t. } \sigma^*(v | \mathcal{V}_+) + C_q v_{N+1} \left\| \frac{\alpha}{v_{N+1}} \right\|_*^{\frac{q}{q-1}} \leq l, \\ & \quad v_i \geq u(-\widehat{y}_i^\top \alpha - \kappa), \quad i = 1, \dots, N, \\ & \quad v_{N+1} \geq 0, \end{aligned}$$

where $C_q = C^{\frac{q}{q-1}} \left(q^{\frac{1}{1-q}} - q^{\frac{q}{1-q}} \right)$.

Finally, we arrive at the following convex reformulation for the general case of a convex utility function u for higher order q .

Proposition 21. *Let $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}^q$ for any $y_1, y_2 \in \mathcal{Y}$ and $q \in (1, \infty)$. Assuming the function u is convex and increasing and satisfies $u(x) - u(x_0) \leq L|x - x_0|^q + M, x \in \mathbb{R}$ for some $L, M > 0$, and some $x_0 \in \mathbb{R}$, the problem (17) with $\rho = \rho^{(3)}$ can be solved by the convex program*

$$\begin{aligned} & \inf_{\alpha \in \mathcal{A}, \kappa \in \mathbb{R}, \eta \geq 0, v} \kappa \\ & \text{s.t. } \sigma^*(v | \mathcal{V}_+) \leq l, \\ & \quad v_i \geq \sup_{z \in \mathbb{R}} \left\{ z \cdot \widehat{y}_i^\top \alpha - (u(\cdot, \kappa))^*(z) + \eta \cdot \frac{|z|^{q^*}}{q^* q^{q^*-1}} \right\}, \quad i = 1, \dots, N, \\ & \quad v_{N+1} \geq \|\alpha\|_*^q / \eta^{q-1}. \end{aligned}$$

3.3 Conditional Risk Functionals in Distortion Form

We now turn to risk functionals defined through distortion functions, which play a central role in both risk measurement—as distortion risk measures—and decision theory, where they appear as rank-dependent utility functionals. For general distortion functions, these risk measures lack finite-dimensional expectation representations, i.e., (18), rendering them fully intractable under prior approaches such as Nguyen et al. (2024) and Esteban-Perez and Morales (2022). A key highlight of our union-ball framework is that, even in this challenging setting, it yields tractable reformulations via regularization equivalents—offering, to our knowledge, the first viable solution framework for distributionally robust conditional distortion problems. We first give the definition of distortion risk measure.

Definition 3.3. (*distortion risk measure*) *Given a random variable Z and a distortion function $h : [0, 1] \rightarrow [0, 1]$, that is, h is an increasing function satisfies that $h(0) = 0$, and $h(1) = 1$, the distortion risk measure is defined as $\rho_h(Z) = \int_{\mathbb{R}} z \, dh(F_Z(z))$, where F_Z is the distribution of Z . Specifically, if h is a convex function, ρ_h is a convex, thus coherent, risk measure.*

To present our reformulation results, we first introduce the admissible set

$$\mathcal{V}_+^h = \{(p, \delta, \bar{p}) : (p, \delta) \in \mathcal{V}_+, \bar{p} \in \mathcal{V}_p^h\},$$

where

$$\mathcal{V}_p^h = \left\{ \bar{p} : \sum_{i=1}^N \bar{p}_i = 1, \bar{p}_i \geq 0, i = 1, \dots, N, h\left(\sum_{i \in J} p_i\right) \leq \sum_{i \in J} \bar{p}_i, \forall J \subset \{1, \dots, N\} \right\},$$

which is the risk envelope associated with coherent distortion risk measures. Leveraging this set characterization, together with the decomposition principle and regularization equivalents, we obtain the following convex reformulation.

Proposition 22. *Assume ρ is a convex distortion risk measure with convex distortion function h . Denote the left-derivative of h on $(0, 1]$ as h'_- . Then we have following results:*

- (i) *Let $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}$ for any $y_1, y_2 \in \mathcal{Y}$ and $\ell : \mathbb{R} \rightarrow \mathbb{R}$ be a convex and Lipschitz continuous function. The problem (12) with $\rho = \rho_h$ is equivalent to the robust optimization problem*

$$\min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}_+} \rho_{\sum_{i=1}^N p_i \mathbb{I}_{\hat{\mathcal{Y}}_i}}^h [\ell(Y^\top \alpha)] + \text{Lip}(\ell) \|h'_-\|_{\infty} \delta \|\alpha\|_*. \quad (21)$$

and can be solved by the convex program

$$\begin{aligned} \inf_{\alpha \in \mathcal{A}, v} \quad & \sigma^*(v | \mathcal{V}_+^h) \\ \text{s.t. } \quad & v_i \geq 0, i = 1, \dots, N, \quad v_{N+1} \geq \text{Lip}(\ell) \|h'_-\|_{\infty} \|\alpha\|_*, \\ & v_{i+N+1} \geq \ell(\hat{\mathcal{Y}}_i^\top \alpha), i = 1, \dots, N. \end{aligned}$$

- (ii) *Let $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}^q$ for any $y_1, y_2 \in \mathcal{Y}$ and $q \in (1, \infty)$, and $\frac{1}{q} + \frac{1}{q^*} = 1$. Assume the function ℓ takes one of the following two forms, multiplied by a constant $C > 0$:*

- (a) $\ell_1(x) = x + b$ or $\ell_1(x) = -x + b$ with some $b \in \mathbb{R}$;
(b) $\ell_2(x) = |x - b_1| + b_2$ with some $b_1, b_2 \in \mathbb{R}$.

The problem (12) with $\rho = \rho_h$ is equivalent to the robust optimization problem

$$\min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}_+} \rho_{\sum_{i=1}^N p_i \mathbb{I}_{\hat{\mathcal{Y}}_i}}^h [\ell(Y^\top \alpha)] + C \|h'_-\|_{q^*} \delta^{\frac{1}{q}} \|\alpha\|_*.$$

and can be solved by the convex program

$$\begin{aligned} \inf_{\alpha \in \mathcal{A}, v} \quad & \sigma^*(v | \mathcal{V}_+^h) + C_{h,q} v_{N+1} \left\| \frac{\alpha}{v_{N+1}} \right\|_*^{\frac{q}{q-1}} \\ \text{s.t. } \quad & v_i \geq 0, i = 1, \dots, N+1, \\ & v_{i+N+1} \geq \ell(\hat{\mathcal{Y}}_i^\top \alpha), i = 1, \dots, N, \end{aligned}$$

$$\text{where } C_{h,q} = \left(q^{\frac{1}{1-q}} - q^{\frac{q}{1-q}} \right) \left(C \|h'_-\|_{q^*} \right)^{\frac{q}{q-1}}.$$

Similar to Examples 3.2 and 3.3, we now apply the admissible sets from Theorems 3 and 6 to illustrate

how the conjugate of the support function $\sigma^*(v|\mathcal{V}_+^h)$ is derived. Recall that $N_\omega = [\omega_1, \omega_2] \subset (0, 1]$, and $\varepsilon = \min_{\beta \in N_\omega} \beta = \omega_1$.

Example 3.5. Following the definition of $\mathcal{V}$ in Theorem 3, the conjugate function $\sigma^*(v|\mathcal{V}_+^h)$ is given by

$$\begin{aligned}
& \inf -\frac{\eta_1}{\omega_2} + \frac{\eta_2}{\omega_1} - \lambda - \mu + \sum_{j=1}^{2^N-2} \zeta_j h^*\left(\frac{-\phi_j}{\zeta_j}\right) \\
& \text{s.t. } \lambda, \mu, \xi_2 \in \mathbb{R}, \{\phi_j\}_{j=1}^{2^N-2} \in \mathbb{R}^{2^N-2}, \{\zeta_j\}_{j=1}^{2^N-2}, \{\tau_i^1\}_{i=1}^N, \{\tau_i^2\}_{i=1}^N, \{\gamma_i\}_{i=1}^N, \eta_1, \eta_2, \xi_1 \geq 0 \\
& -\xi_2(\delta_0 - \frac{1}{N} \sum_{i=m+1}^N d_i) + \frac{1}{N} \sum_{i=1}^N \tau_i^2 + \eta_1 - \eta_2 \leq 0, \\
& v_i + \lambda + \tau_i^1 - \tau_i^2 + c_i + \sum_{j:i \in I_j} \phi_j \leq 0, \quad i = 1, \dots, N, \\
& c_i = \xi_2 d_i, i = 1, \dots, m, \quad c_i = -\xi_2 d_i, i = m+1, \dots, N, \\
& v_{N+1+i} + \mu + \sum_{j:i \in I_j} \zeta_j + \gamma_i \leq 0, \quad i = 1, \dots, N, \\
& v_{N+1} + \xi_1 + \xi_2 \leq 0.
\end{aligned}$$

Example 3.6. Following the definition of $\mathcal{V}$ in Theorem 6, the conjugate function $\sigma^*(v|\mathcal{V}_+^h)$ is given by

$$\begin{aligned}
& \inf -\lambda - \mu + \frac{1}{N\varepsilon} \sum_{i=1}^N \tau_i^2 - \xi_2 \delta_0 + \sum_{j=1}^{2^N-2} \zeta_j h^*\left(\frac{-\phi_j}{\zeta_j}\right) \\
& \text{s.t. } \lambda, \mu, \xi_2 \in \mathbb{R}, \{\phi_j\}_{j=1}^{2^N-2} \in \mathbb{R}^{2^N-2}, \{\zeta_j\}_{j=1}^{2^N-2}, \{\tau_i^1\}_{i=1}^N, \{\tau_i^2\}_{i=1}^N, \{\gamma_i\}_{i=1}^N, \xi_1 \geq 0, \\
& v_i + \lambda + \tau_i^1 - \tau_i^2 + c'_i + \sum_{j:i \in I_j} \phi_j \leq 0, \quad i = 1, \dots, N, \\
& c'_i = \xi_2 d_i, i = 1, \dots, m, \quad c_i = 0, i = m+1, \dots, N, \\
& v_{N+1+i} + \mu + \sum_{j:i \in I_j} \zeta_j + \gamma_i \leq 0, \quad i = 1, \dots, N, \\
& v_{N+1} + \xi_1 + \xi_2 \leq 0.
\end{aligned}$$

The mean-CVaR risk measure considered in the example of Nguyen et al. (2024) is a convex distortion risk measure. While their approach yields only conservative approximations to the associated portfolio allocation problem, we demonstrate below how our framework provides reformulations that are not only provably exact but also markedly simpler, featuring fewer variables, fewer constraints, and a modular structure that streamlines computation.

Example 3.7. The robustified conditional mean-CVaR portfolio allocation problem in Nguyen et al. (2024)

is as follows:

$$\min_{\alpha \in \mathcal{A}} \sup_{\substack{\mathbb{Q} \in \mathcal{B}_{\delta_0}(\widehat{\mathbb{P}}) \\ \mathbb{Q}(X \in \mathcal{N}_\gamma(x_0)) \in \mathcal{N}_\omega}} \text{CVaR}_{\mathbb{Q}}^{1-\kappa} [Y^\top \alpha \mid X \in \mathcal{N}_\gamma(x_0)] - \theta \cdot \mathbb{E}_{\mathbb{Q}} [Y^\top \alpha \mid X \in \mathcal{N}_\gamma(x_0)], \quad (22)$$

where $\mathcal{N}_\omega = [\varepsilon_0, 1]$, $\varepsilon_0 \in (0, 1]$, $\theta \geq 0$, $\kappa \in (0, 1]$,

$$\text{CVaR}_{\mathbb{Q}}^{1-\kappa}(Z) = \frac{1}{\kappa} \int_{1-\kappa}^1 \text{VaR}_{\mathbb{Q}}^\phi(-Z) \, d\phi,$$

and $\text{VaR}_{\mathbb{Q}}^\phi(Z) = \inf \{z \in \mathbb{R} : \mathbb{Q}(Z \leq z) \geq \phi\}$. Suppose in addition that $\mathcal{X} = \mathbb{R}^{n_1}$, $\mathcal{Y} = \mathbb{R}^{n_2}$, $\mathbb{D}_X(x, \widehat{x}) = \|x - \widehat{x}\|^2$ and $\mathbb{D}_Y(y, \widehat{y}) = \|y - \widehat{y}\|_2^2$. The problem (22) can be solved by the convex program

$$\begin{aligned} & \inf_{\alpha \in \mathcal{A}, t \in \mathbb{R}, z \geq 0, v_{N+1}, s} z^\top b + s \\ & \text{s.t. } v_{N+1} \geq 0, A^\top z \geq (\ell(\widehat{y}_1^\top \alpha, t), \dots, \ell(\widehat{y}_N^\top \alpha, t), v_{N+1}, 0)^\top, \\ & s \geq C_{\theta, \kappa} v_{N+1} \left\| \frac{\alpha}{v_{N+1}} \right\|_2^2, \end{aligned}$$

where $\ell(z, t) = t + \frac{1}{\kappa}(-z - t)_+ - \theta z$, $A \in \mathbb{R}^{(N+6) \times (N+2)}$, and $b \in \mathbb{R}^{N+6}$ are defined by

$$A = \begin{pmatrix} 1 & \dots & 1 & 0 & 0 \\ -1 & \dots & -1 & 0 & 0 \\ 0 & \dots & 0 & 0 & 1 \\ 0 & \dots & 0 & 0 & -1 \\ 1 & \dots & 0 & 0 & -\frac{1}{N} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & 1 & 0 & -\frac{1}{N} \\ -a_1 & \dots & -a_N & -a_{N+1} & -a_{N+2} \\ a_1 & \dots & a_N & a_{N+1} & a_{N+2} \end{pmatrix}, \quad \text{with } a = \begin{pmatrix} -d_1 \\ \vdots \\ -d_m \\ d_{m+1} \\ \vdots \\ d_N \\ -1 \\ \delta_0 - \frac{1}{N} \sum_{i=m+1}^N d_i \end{pmatrix}, \quad \text{and } b = \begin{pmatrix} 1 \\ -1 \\ \frac{1}{\varepsilon_0} \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Remark 23. One can verify that the constraint $s \geq C_{\theta, \kappa} v_{N+1} \left\| \frac{\alpha}{v_{N+1}} \right\|_2^2$ can be reformulated as the following SOC constraint

$$\left\| \begin{bmatrix} 2\alpha \\ s - C_{\theta, \kappa} v_{N+1} \end{bmatrix} \right\|_2 \leq s + C_{\theta, \kappa} v_{N+1}.$$

Thus, our reformulation is a SOCP, while the reformulation of Proposition 4.6 in Nguyen et al. (2024) is a SDP. From this perspective, our formulation has a simpler form and is easy to apply.

4 Scalable Solution for Large-Scale Distributionally Robust CRM

We now highlight another key advantage of the union-ball decomposition introduced in the previous section: beyond yielding unified and tractable reformulations, it naturally facilitates the development of scalable solution methods. While our earlier convex programs—for instance, those targeting rank-dependent utility—are exact and structurally elegant, they become computationally prohibitive as the sample size N grows. In this section, we show how the decomposition structure lends itself to a cutting-plane algorithm, effectively addressing large-scale distributionally robust CRM problems—an area largely unexplored in the existing literature.

We base our development on the distributionally robust conditional rank-dependent utility problem (21), which subsumes distributionally robust conditional expectation problem as a special case. Following Proposition 22, we have

$$\min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}_+} \rho_{\sum_{i=1}^N p_i \mathbb{I}_{\widehat{y}_i}}^h [\ell(Y^\top \alpha)] + \text{Lip}(\ell) \|h'_-\|_\infty \delta \|\alpha\|_* \quad (23)$$

$$\begin{aligned} &= \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}_+, \bar{p} \in \mathcal{V}_p^h} \sum_{i=1}^N \bar{p}_i \ell(\widehat{y}_i^\top \alpha) + \text{Lip}(\ell) \|h'_-\|_\infty \delta \|\alpha\|_* \\ &:= \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta, \bar{p}) \in \mathcal{V}_+^h} \sum_{i=1}^N \bar{p}_i \ell(\widehat{y}_i^\top \alpha) + \text{Lip}(\ell) \|h'_-\|_\infty \delta \|\alpha\|_*, \end{aligned} \quad (24)$$

where $\mathcal{V}_+^h = \{(p, \delta, \bar{p}) : (p, \delta) \in \mathcal{V}_+, \bar{p} \in \mathcal{V}_p^h\}$.

The computational burden lies in solving the inner supremum problem where the set $\mathcal{V}_p^h$ involves exponentially many constraints as N increases. The structure of inner maximization problem of (24), which is linear in variable $(\delta, \bar{p})$, for any fixed $\alpha \in \mathcal{A}$, naturally leads to the design of a cutting plane method. The central idea is to iteratively refine the uncertainty set by adding only the most violated constraints, allowing convergence without necessarily enumerating all possibilities.

To efficiently solve (24), we proceed as follows:

1. We begin with an initial single-scenario uncertainty set $\mathcal{U}_1 = \{(p, \delta, \bar{p}) = (p_f, \delta_f, p_f)\}$, where $(p_f, \delta_f) \in \mathcal{V}_+$. The associated scenario p_f clearly lies within $\mathcal{V}_{p_f}^h$. Note that we can always find a feasible solution $(p_f, \delta_f) \in \mathcal{V}_+$ when adopting $\mathcal{V}$ specified in Theorems 3 and 6. To guarantee the feasibility of (23) with $\mathcal{V}$ specified in Theorems 3 and 6, let $\delta_0 = \delta_{\min} + \Delta$, where $\Delta > 0$ and $\delta_{\min}$ is given by Propositions 1 and 4, respectively. Moreover, let $\{v_i\}_{i=1}^N$ be the optimal solution such that the optimization problem in Propositions 1 and 4 attain the optimal value $\delta_{\min}$, respectively. Then for $\mathcal{V}$ specified in Theorem 3, let $p_f = (\frac{v_1}{\sum_{i=1}^N v_i}, \dots, \frac{v_N}{\sum_{i=1}^N v_i})^\top$ and $\delta_f = \frac{N\Delta}{\sum_{i=1}^N v_i}$ and for $\mathcal{V}$ specified in Theorem 6, let $p_f = (\frac{v_1}{N\varepsilon}, \dots, \frac{v_N}{N\varepsilon})^\top$ and $\delta_f = \Delta$. One can verify that $(p_f, \delta_f) \in \mathcal{V}_+$ and thus $(p_f, \delta_f, p_f) \in \mathcal{V}_+^h$. Hence, we can always obtain a feasible scenario in $\mathcal{V}_+^h$ and then $\mathcal{U}_1 \subset \mathcal{V}_+^h$.

Algorithm 1: Cutting-Plane Scheme

Input : tolerance $\psi_{\text{tol}} > 0$; feasible $(p_f, \delta_f) \in \mathcal{V}_+$; hypothesis set $\mathcal{A}$.

Output parameter α^* and an (approximate) worst-case scenario $(p^*, \delta^*, \bar{p}^*)$.

:

1 **Initialization:** Set the scenario pool $\mathcal{U}_1 \leftarrow \{(p_f, \delta_f, \bar{p}_f)\}$ with $\bar{p}_f := p_f$; set $j \leftarrow 1$.

2 **repeat**

3 **Lower bound with uncertainty set $\mathcal{U}_j$:**

4

$$l_j^b \leftarrow \min_{\alpha \in \mathcal{A}} \max_{(p, \delta, \bar{p}) \in \mathcal{U}_j} \left\{ \sum_{i=1}^N \bar{p}_i \ell(\widehat{y}_i^\top \alpha) + \text{Lip}(\ell) \|h'_-\|_\infty \delta \|\alpha\|_* \right\}$$

Let α_j be an optimal solution to obtain the lower bound l_j^b .

5 **Upper bound with α_j :**

6 Form the order statistics of the realized losses

$$r_{(1)} \leq \dots \leq r_{(N)}, \quad r_i := \ell(\widehat{y}_i^\top \alpha_j).$$

Solve

$$u_j^b \leftarrow \sup_{(p, \delta) \in \mathcal{V}_+} \sum_{k=1}^{N-1} h\left(\sum_{i=1}^k p_{(i)}\right) (r_{(k)} - r_{(k+1)}) + r_{(N)} + \text{Lip}(\ell) \|h'_-\|_\infty \delta \|\alpha_j\|_*. \quad (25)$$

Let (p_j^*, δ_j^*) be an optimal solution to obtain the upper bound u_j^b and set

$$\bar{p}_{j(i)}^* := h\left(\sum_{k=1}^i p_{j(k)}^*\right) - h\left(\sum_{k=1}^{i-1} p_{j(k)}^*\right), \quad i = 1, \dots, N.$$

7 **Gap check:** If $u_j^b - l_j^b \leq \psi_{\text{tol}}$, terminate with $\alpha^* \leftarrow \alpha_j$ and $(p^*, \delta^*, \bar{p}^*) \leftarrow (p_j^*, \delta_j^*, \bar{p}_j^*)$.

8 **Pool augmentation:** Otherwise enlarge the scenario pool

9

$$\mathcal{U}_{j+1} \leftarrow \mathcal{U}_j \cup \{(p_j^*, \delta_j^*, \bar{p}_j^*)\},$$

and set $j \leftarrow j + 1$ to continue.

10 **until** $u_j^b - l_j^b \leq \psi_{\text{tol}}$

2. At iteration j , we solve the restricted optimization problem:

$$\min_{\alpha \in \mathcal{A}} \sup_{(p, \delta, \bar{p}) \in \mathcal{U}_j} \sum_{i=1}^N \bar{p}_i \ell(\widehat{y}_i^\top \alpha) + \text{Lip}(\ell) \|h'\|_\infty \delta \|\alpha\|_*.$$

Let the optimal solution and optimal value obtained be (α_j, l_j^b) , which provides a lower bound to (24).

3. Fixing α_j , we solve the full inner maximization of (24) to obtain the worst-case solution (p_j^*, δ_j^*) and the corresponding upper bound u_j^b to the problem (24). One can verify that given α_j , we can actually directly solve the inner supremum problem of problem (23) because the distortion risk measure has explicit form (see (25)) and thereby avoid an exponentially growing number of constraints in (24).
4. We check the convergence criterion by comparing the gap $u_j^b - l_j^b$ to a predefined tolerance $\psi_{\text{tol}} > 0$. If the gap satisfies $u_j^b - l_j^b \leq \psi_{\text{tol}}$, the algorithm terminates, returning α_j and l_j^b as approximate optimal solution and optimal value. Otherwise, the scenario $(p_j^*, \delta_j^*, \bar{p}_j^*)$ is appended to the uncertainty set, forming $\mathcal{U}_{j+1} = \mathcal{U}_j \cup (p_j^*, \delta_j^*, \bar{p}_j^*)$, and the process repeats. A critical observation that significantly enhances efficiency is that, given $\alpha \in \mathcal{A}$, for each worst-case distribution p^* , the optimal associated distortion-weighted distribution $\bar{p}^*$, achieving the supremum, is explicitly computed by:

$$\bar{p}_{(i)}^* = h \left(\sum_{k=1}^i p_{(k)}^* \right) - h \left(\sum_{k=1}^{i-1} p_{(k)}^* \right)$$

with the $(\cdot)$ denoting sorting of scenarios according to $\ell(\widehat{y}_i^\top \alpha)$ in ascending order, i.e. $\ell(\widehat{y}_{(1)}^\top \alpha) \leq \dots \leq \ell(\widehat{y}_{(N)}^\top \alpha)$. Hence, each iteration only needs to incorporate a single scenario $(p_j^*, \delta_j^*, \bar{p}_j^*)$ to the uncertainty set rather than explicitly enumerating exponential constraints, $\{(p_j^*, \delta_j^*, \bar{p}_j) : \bar{p}_j \in \mathcal{V}_{p_j^*}^h\}$, to the uncertainty set.

The fact that $u_j^b - l_j^b$ converges to 0 as j increases is guaranteed by that the lower bound of (24), l_j^b , increases with respect to iteration j and converges to (24) as j increases due to $\mathcal{U}_j \subset \mathcal{U}_{j+1} \subset \dots \subset \mathcal{V}_+^h$.

5 Numerical Experiment

In this section, we demonstrate the practical effectiveness of our unified and tractable optimal transport framework through simulation experiments on a portfolio allocation problem governed by a distributionally robust conditional rank-dependent expected utility (RDEU) criterion. These experiments enable a systematic evaluation of our method in a controlled setting and highlight its advantages over both unconditional models and existing conditional approaches. Leveraging the cutting-plane algorithm developed in the previous section, we solve the resulting problems efficiently and underscore the value of incorporating side information in data-driven decision-making.

We adopt the setting of Proposition 22(i) for the choice of loss function and distance metric, where the loss function ℓ is Lipschitz continuous and convex, and the distance metric is defined as $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}$.

The distributionally robust conditional model is solved using our union-ball formulation, with the admissible set $\mathcal{V}$ specified in Theorem 3.

We summarize below the models considered in the comparison, each of which is solved using the cutting-plane algorithm.

- (i) **Union-Ball Conditional Distributionally Robust Optimization Model (UB-CDRO)**: This is our proposed model, presented as

$$\min_{\alpha \in \mathcal{A}} \sup_{\mathbb{Q}|Y \in \bigcup_{(p, \delta) \in \mathcal{V}} \mathbb{B}_\delta \left(\sum_{i=1}^N p_i \mathbb{I}_{\hat{y}_i} \right)} \rho_{\mathbb{Q}|Y} [\ell(Y, \alpha)].$$

- (ii) **Unconditional Sample Average Approximation (SAA)**: This baseline model uses the empirical distribution of Y without conditioning on X -namely, $\hat{\mathbb{P}}_N = \frac{1}{N} \sum_{i=1}^N \mathbb{I}_{\hat{y}_i}$:

$$\min_{\alpha \in \mathcal{A}} \rho_{\hat{\mathbb{P}}_N}^h [\ell(Y^\top \alpha)].$$

- (iii) **Unconditional Distributionally Robust Optimization Model (UDRO)**: This model solves the unconditional distributionally robust optimization problem, using the empirical distribution of Y as the reference measure. Specifically, this model is presented as

$$\min_{\alpha \in \mathcal{A}} \sup_{\mathbb{Q}_Y \in \mathbb{B}_{\delta_0}(\hat{\mathbb{P}}_N)} \rho_{\mathbb{Q}_Y} [\ell(Y, \alpha)].$$

- (iv) **Conditional Sample Average Approximation (CSAA)**: This model is defined as

$$\min_{\alpha \in \mathcal{A}} \rho_{\hat{\mathbb{P}}_{K_N}}^h [\ell(Y^\top \alpha)],$$

where $\hat{\mathbb{P}}_{K_N}$ denotes the empirical conditional distribution of Y on $X \in \mathcal{N}_\gamma(x_0)$. Specifically, K_N represents the number of samples whose associated covariates fall within the neighborhood $\mathcal{N}_\gamma(x_0)$. Sorting $(\hat{x}_i, \hat{y}_i)$ as $(\hat{x}_{(j)}, \hat{y}_{(j)})$ such that $\hat{x}_{(j)} \in \mathcal{N}_\gamma(x_0)$, for $j = 1, \dots, K_N$ and $\hat{x}_{(j)} \notin \mathcal{N}_\gamma(x_0)$, for $j = K_N + 1, \dots, N$. Then $\hat{\mathbb{P}}_{K_N} = \frac{1}{K_N} \sum_{j=1}^{K_N} \mathbb{I}_{\hat{y}_{(j)}}$.

- (v) **Conditional Distributionally Robust Optimization Model (CDRO)**: This model solves the distributionally robust counterpart of CSAA, using the empirical conditional distribution $\mathbb{P}_{K_N}$ as the reference distribution. Specifically, this model is presented as

$$\min_{\alpha \in \mathcal{A}} \sup_{\mathbb{Q}|Y \in \mathbb{B}_{\delta_0}(\hat{\mathbb{P}}_{K_N})} \rho_{\mathbb{Q}|Y} [\ell(Y, \alpha)].$$

In the experiment, we run 50 times for each sample sizes N , choose $\delta = \delta_{\min} + \frac{1}{10} \frac{\sqrt{\ln 100}}{N}$ to make sure that the model (i) is feasible, and choose $\delta_0 = 100/N$ such that the radius $\delta \rightarrow 0$ as $N \rightarrow \infty$. For

portfolio allocation problem setting, $\alpha \in \mathbb{R}_+^6$ denotes decision vector and $Y \in \mathbb{R}^6$ denotes an uncertain quantity of interest with side information $X \in \mathbb{R}^3$. We assume the joint distribution of (X, Y) is as follows. The marginal distribution of X is $X_1 \rightsquigarrow \mathcal{N}(0.1, 10)$, $X_2 \rightsquigarrow \mathcal{N}(0, 1)$ and $X_3 \rightsquigarrow \mathcal{N}(0, 1)$. Let $\mu^\top = (86.8625, 71.6059, 75.3759, 97.6258, 52.7854, 84.8973)$,

$$L = \begin{bmatrix} 136.687 & 0 & 0 & 0 & 0 & 0 \\ 8.79766 & 142.279 & 0 & 0 & 0 & 0 \\ 16.1504 & 15.0637 & 122.613 & 0 & 0 & 0 \\ 18.4944 & 15.6961 & 26.344 & 139.148 & 0 & 0 \\ 3.41394 & 16.5922 & 14.8795 & 13.9914 & 151.732 & 0 \\ 24.8156 & 18.7292 & 17.1574 & 6.36536 & 24.7703 & 144.672 \end{bmatrix},$$

which is the same as the setting in Section 5.2 of Bertsimas and Van (2017). Moreover, let $\text{gate}(x_2) = \frac{1}{1 + \exp\{-4(x_2 - \frac{1}{2})\}}$, $s(x_2) = \left(1, 1 - \frac{1}{2} \text{gate}(x_2), 1 + 6 \text{gate}(x_2), 1, 1 + 6 \text{gate}(x_2), 1\right)$, and $D(x_2) = \text{diag}(s(x_2))$. The conditional distribution of Y given $X = (x_1, x_2, x_3)$ is defined as

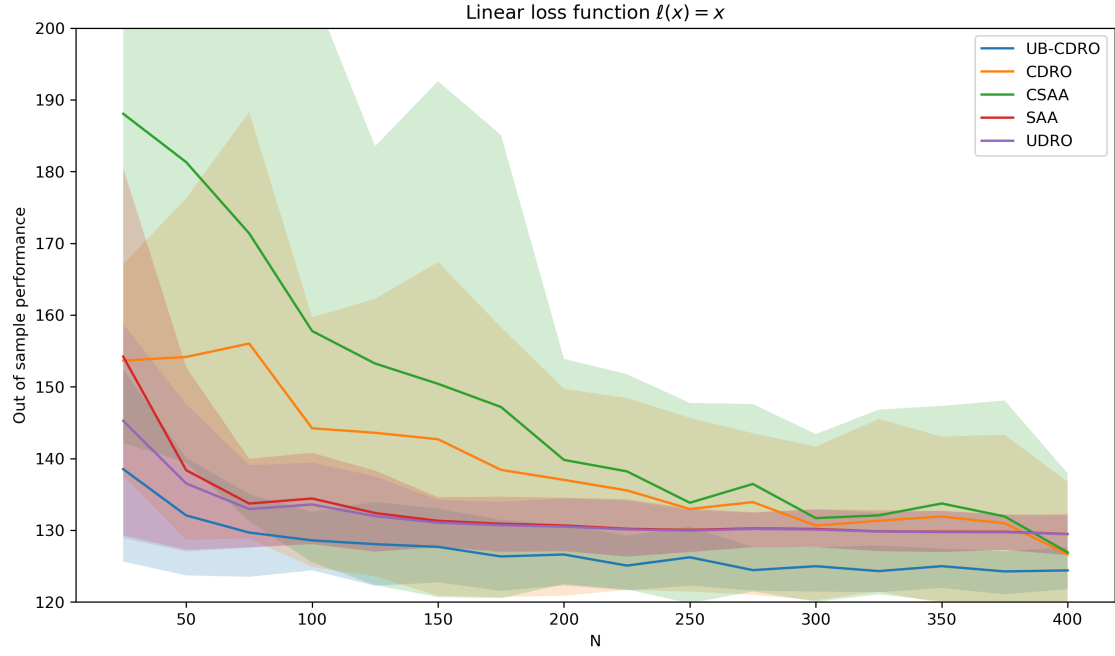
$$Y \mid X = x \rightsquigarrow \mathcal{N}(m(x), \Sigma(x_2)),$$

where

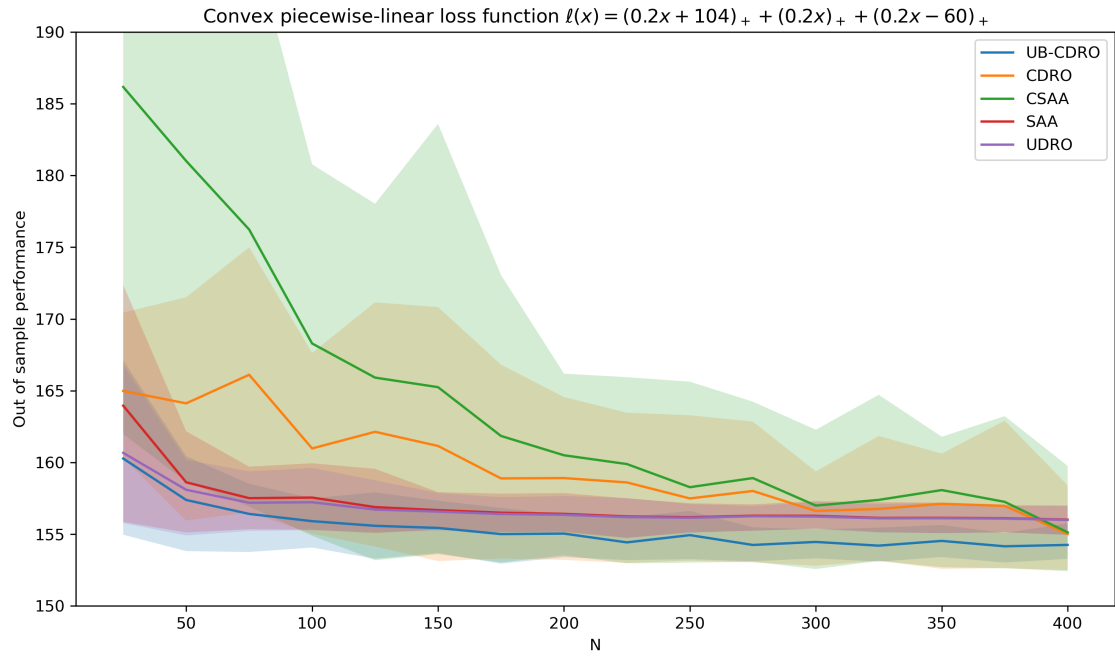
$$m(x) = \mu + x_1 v_1 + x_2 v_2 + \tanh(x_3) v_3, \text{ and } \Sigma(x_2) = L D(x_2)^2 L^\top$$

with $v_1 = (30, -40, 0, 15, -10, 5)^\top$, $v_2 = (-400, 60, 70, 80, 90, 100)^\top$, and $v_3 = (0, 20, -25, 0, 15, -12)^\top$. And we choose the conditional set $\mathcal{N} = \{x : \|x - x_0\|_1 \leq 1.2\}$, where $x_0 = (0.1, 0.1, 0)$.

Figures 1 and 2 report out-of-sample performance for the five models introduced earlier (UB-CDRO, SAA, UDRO, CSAA, CDRO) under various convex distortion functions and Lipschitz-continuous convex loss functions. The plots summarize the distribution of out-of-sample risk values; the shaded regions denote the interquartile range (IQR, 15th-85th percentiles), and the bold colored curves trace the sample means across sample sizes N . The figure indicates that for various distortion functions and loss functions, UB-CDRO consistently outperforms all conditional and unconditional benchmarks: its out-of-sample performance is more stable (smaller IQR), exhibits a lower mean, and approaches the oracle optimal value more rapidly as N increases. Although the conditional methods (CSAA, CDRO) converge to the conditional optimum, the bands of conditional methods remain wide even at $N = 400$, indicating persistent variability. Moreover, the unconditional methods (SAA, UDRO) do not converge to that optimum as N grows. This is expected because the conditional distribution can differ substantially from the marginal distribution; consequently, the decision α that is optimal for the marginal distribution generally differs from the conditional-optimal decision α^* . In fact, SAA and UDRO converge to the marginal-optimal decision rather than to α^* . Thus, unconditional methods (SAA, UDRO) exhibit strong stability but poor convergence, whereas conditional methods (CSAA, CDRO) achieve fast convergence but lack stability. By contrast, UB-CDRO is strong in both stability and convergence, surpassing all conditional and unconditional benchmarks.

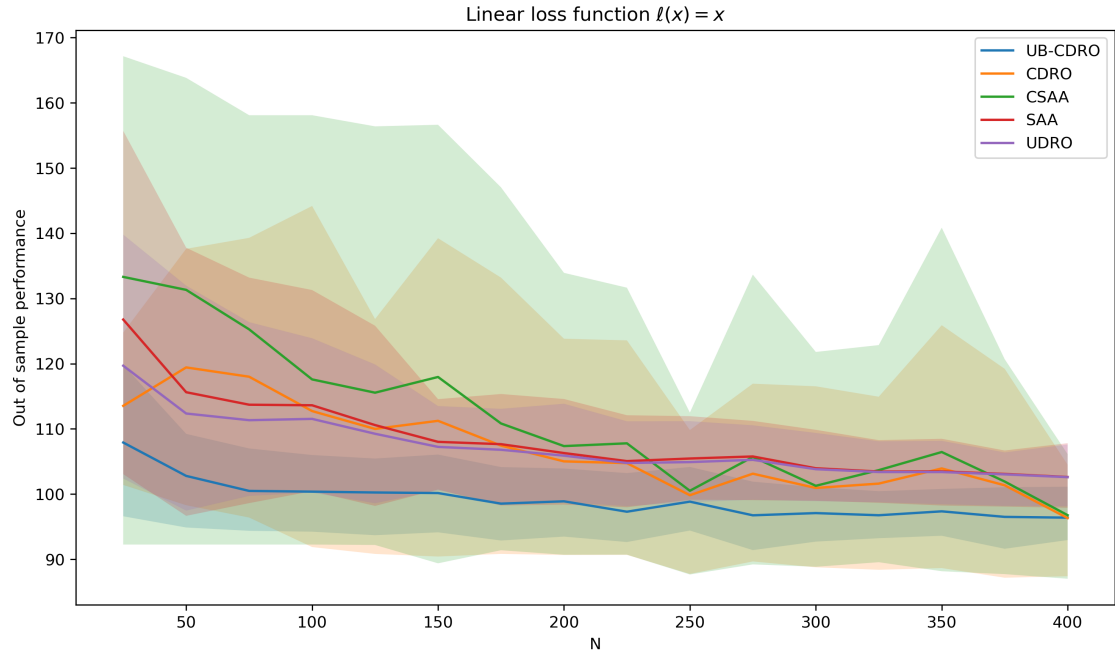


(a)

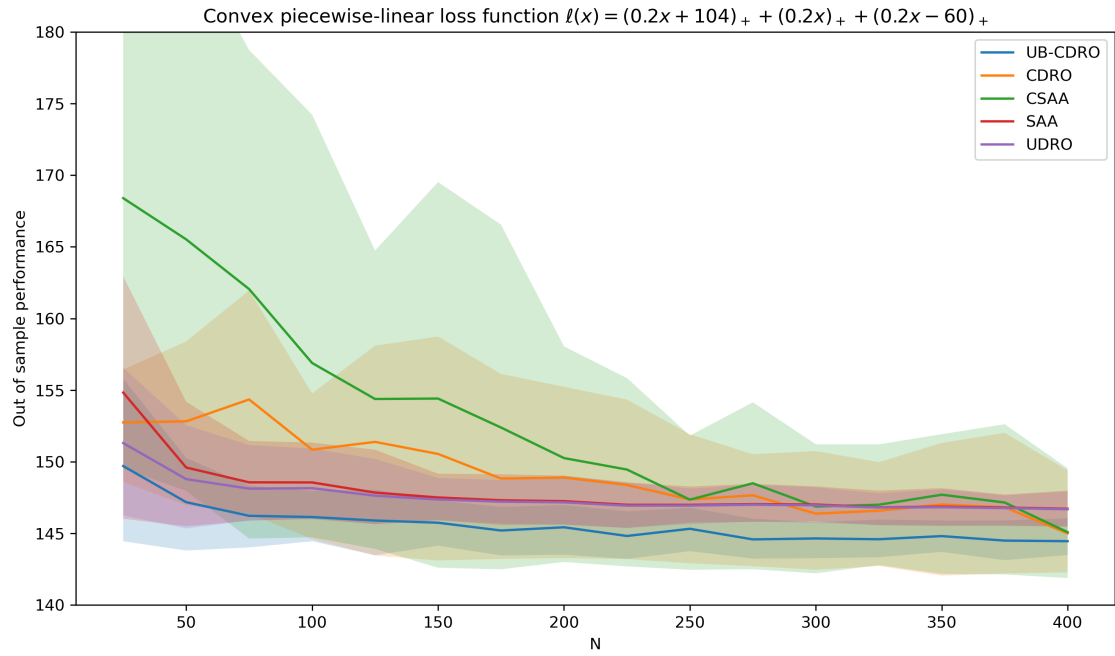


(b)

Figure 1: Comparison of Out-of-Sample Performance Across Models with Distortion Function $h(x) = x^2$



(a)



(b)

Figure 2: Comparison of Out-of-Sample Performance Across Models with Distortion Function $h(x) = \frac{e^x - 1}{e - 1}$

References

- Adrian, T., and Brunnermeier, M. K. (2011). CoVaR (No. w17454). *National Bureau of Economic Research*.
- Ban, G. Y., and Rudin, C. (2019). The big data newsvendor: Practical insights from machine learning. *Operations Research*, **67**(1), 90-108.
- Bertsimas, D., and Kallus, N. (2020). From predictive to prescriptive analytics. *Management Science*, **66**(3), 1025-1044.
- Bertsimas D, Van Parys B (2017) Bootstrap robust prescriptive analytics. *arXiv preprint arXiv:1711.09974v1*.
- Denneberg, D. (2013). Non-additive measure and integral (Vol. 27). *Springer Science and Business Media*.
- Donti, P., Amos, B., and Kolter, J. Z. (2017). Task-based end-to-end model learning in stochastic optimization. *Advances in Neural Information Processing Systems*, **30**.
- El Balghiti, O., Elmachtoub, A. N., Grigas, P., and Tewari, A. (2019). Generalization bounds in the predict-then-optimize framework. *Advances in Neural Information Processing Systems*, **32**.
- Elmachtoub, A. N., and Grigas, P. (2022). Smart “predict, then optimize”. *Management Science*, **68**(1), 9-26.
- Esteban-Perez, A., and Morales, J. M. (2022). Distributionally robust stochastic programs with side information based on trimmings. *Mathematical Programming*, **195**(1), 1069-1105.
- Gao, R., and Kleywegt, A. (2023). Distributionally robust stochastic optimization with Wasserstein distance. *Mathematics of Operations Research*, **48**(2), 603-655.
- Guo, S., and Xu, H. (2019). Distributionally robust shortfall risk optimization model and its approximation. *Mathematical Programming*, **174**, 473–498. <https://doi.org/10.1007/s10107-018-1307-z>.
- Hu, Y., Kallus, N., and Mao, X. (2022). Fast rates for contextual linear optimization. *Management Science*, **68**(6), 4236-4245.
- Jin, G., Laeven, R. J., and Hertog, D. D. (2025). Robust Optimization of Rank-Dependent Models with Uncertain Probabilities. *arXiv preprint arXiv:2502.11780*.
- Kannan, R., Bayraksan, G., and Luedtke, J. R. (2024). Residuals-based distributionally robust optimization with covariate information. *Mathematical Programming*, **207**(1), 369-425.
- Kannan, R., Bayraksan, G., and Luedtke, J. R. (2025). Data-driven sample average approximation with covariate information. *Operations Research*, **0**(0). <https://doi.org/10.1287/opre.2020.0533>.
- Muñoz, M. A., Pineda, S., and Morales, J. M. (2022). A bilevel framework for decision-making under uncertainty with contextual information. *Omega*, **108**, 102575.
- Nguyen, V. A., Zhang, F., Blanchet, J., Delage, E., and Ye, Y. (2020). Distributionally robust local non-parametric conditional estimation. Robustifying conditional portfolio decisions via optimal transport. *Advances in Neural Information Processing Systems*, **33**, 15232-15242.
- Nguyen, V. A., Zhang, F., Wang, S., Blanchet, J., Delage, E., and Ye, Y. (2024). Robustifying conditional portfolio decisions via optimal transport. *Operations Research*, **0**(0). <https://doi.org/10.1287/opre.2021.0243>.
- Sion, M. (1958). On general minimax theorems. *Pacific Journal of Mathematics*, **8**(1), 171–176.
- Villani, C. (2008). Optimal transport: old and new (Vol. 338). *Berlin: springer*.
- Wang, T., Chen, N., and Wang, C. (2021). Distributionally robust prescriptive analytics with wasserstein distance. *arXiv preprint arXiv:2106.05724*.
- Wang, T., Chen, N., and Wang, C. (2024). Contextual optimization under covariate shift: A robust approach by intersecting wasserstein balls. *arXiv preprint arXiv:2406.02426*.

- Wu, Q., Li, J. Y. M., and Mao, T. (2022). On generalization and regularization via wasserstein distributionally robust optimization. *arXiv preprint arXiv:2212.05716*.
- Yang, J., Zhang, L., Chen, N., Gao, R., and Hu, M. (2022). Decision-making with side information: A causal transport robust approach. *Optimization Online*.
- Zhang, L., Yang, J., and Gao, R. (2024). Optimal robust policy for feature-based newsvendor. *Management Science*, **70**(4), 2315-2329.

Supplement: Technical Appendices

EC.1. Regularization Reformulations for Wasserstein DRO (Wu et al. 2022)

In this section, we first review some regularization results in Wu et al. (2022), which will be used throughout the proofs of Section 3. Let L^q be the set of all random variables with a finite q th moment. For $q \geq 1$, and L^∞ be the set of all bounded random variables, under an arbitrary norm of $\mathcal{Y}$, $\|\cdot\|_{\mathcal{Y}}$. If the cost function on $\mathcal{Y}$ is $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}$ for any $y_1, y_2 \in \mathcal{Y}$, and ℓ is convex and Lipschitz continuous, the unconditional worst-case risk problem (14) is equivalent to

$$\sum_{i=1}^N p_i \ell(\widehat{y}_i^\top \alpha) + \text{Lip}(\ell) \delta \|\alpha\|_*, \quad (26)$$

where $\|\cdot\|_*$ is the dual norm of $\|\cdot\|_{\mathcal{Y}}$, and $\text{Lip}(\ell)$ is the Lipschitz constant of ℓ . Moreover, for $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}^q$ for any $y_1, y_2 \in \mathcal{Y}$, with $q \in (1, \infty]$, Wu et al. (2022) have some regularization results for the unconditional worst-case risk problem (14) with some special loss functions.

Theorem 24. (Wu et al. (2022) Theorem 3) *Let $\ell : \mathbb{R} \rightarrow \mathbb{R}$ be a convex function. For $q \in (1, \infty]$, suppose that $\mathbb{E}[|\ell(Z)|] < \infty$ for all $Z \in L^q$, and $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}^q$, for any $y_1, y_2 \in \mathcal{Y}$. Then the following statements are equivalent.*

- (i) *There exists $C > 0$ such that for any $(p, \delta) \in \mathcal{V}$ with $\delta \geq 0$, it holds that*

$$\sup_{\mathbf{Q} | \mathbf{Y} \in \mathbb{B}_\delta(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i})} \mathbb{E}_{\mathbf{Q} | \mathbf{Y}} [\ell(Y^\top \alpha)] = \sum_{i=1}^N p_i \ell(\widehat{y}_i^\top \alpha) + C \delta^{\frac{1}{q}} \|\alpha\|_*$$

- (ii) *The function ℓ takes one of the following two forms, multiplied by C :*

- (a) $\ell_1(x) = x + b$ or $\ell_1(x) = -x + b$ with some $b \in \mathbb{R}$;
- (b) $\ell_2(x) = |x - b_1| + b_2$ with some $b_1, b_2 \in \mathbb{R}$.

Theorem 25. (Wu et al. (2022) Theorem 4) *Let $\ell : \mathbb{R} \rightarrow \mathbb{R}_+$ be a Lipschitz continuous and convex function. For any $q \in (1, \infty)$, assume $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}^q$, for any $y_1, y_2 \in \mathcal{Y}$. Then the following statements are equivalent.*

(i) There exists $C > 0$ such that for any $(p, \delta) \in \mathcal{V}$ with $\delta \geq 0$, it holds that

$$\sup_{\mathbb{Q}|_Y \in \mathbb{B}_\delta(\sum_{i=1}^N p_i I_{\hat{Y}_i})} \mathbb{E}_{\mathbb{Q}|_Y} [\ell^q(Y^\top \alpha)] = \left[\left(\sum_{i=1}^N p_i \ell^q(\hat{Y}_i^\top \alpha) \right)^{1/q} + C \delta^{\frac{1}{q}} \|\alpha\|_* \right]^q.$$

(ii) The function ℓ takes one of the following four forms, multiplied by C :

- (a) $\ell_1(x) = (x - b)_+$ with some $b \in \mathbb{R}$;
- (b) $\ell_2(x) = (x - b)_-$ with some $b \in \mathbb{R}$;
- (c) $\ell_3(x) = (|x - b_1| - b_2)_+$ with some $b_1 \in \mathbb{R}$ and $b_2 \geq 0$;
- (d) $\ell_4(x) = |x - b_1| + b_2$ with some $b_1 \in \mathbb{R}$ and $b_2 > 0$.

Theorem 26. (Wu et al. (2022) Corollary 1) For any $q \in [1, \infty)$ and $C > 0$, let $\rho^{(1)}$ be defined by (18) with $\ell(z, t) = (C\bar{\ell})^q(z, t)$ and $\bar{\ell}(z, t) = C\bar{\ell}(z - t)$, and $\rho^{(2)}$ be defined by (20) with $\ell(z, t) = \ell(z - t)$, and assume $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}^q$, for any $y_1, y_2 \in \mathcal{Y}$. Take $(p, \delta) \in \mathcal{V}$ with $\delta \geq 0$.

(i) Let the function $\bar{\ell}$ take one of the following two forms:

- (a) $\bar{\ell}_1(x) = (|x - b_1| - b_2)_+$ with some $b_1 \in \mathbb{R}$ and $b_2 \geq 0$;
- (b) $\bar{\ell}_2(x) = |x - b_1| + b_2$ with some $b_1 \in \mathbb{R}$ and $b_2 > 0$.

Then

$$\sup_{\mathbb{Q}|_Y \in \mathbb{B}_\delta(\sum_{i=1}^N p_i I_{\hat{Y}_i})} \rho_{\mathbb{Q}|_Y}^{(1)}(Y^\top \alpha) = \left[\left(\rho_{\sum_{i=1}^N p_i I_{\hat{Y}_i}}^{(1)}(Y^\top \alpha) \right)^{1/q} + C \delta^{1/q} \|\alpha\|_* \right]^q.$$

(ii) For $C > 1$, let the function ℓ take one of the following three forms:

- (a) $\ell_1(x) = (x - b)_+$ with some $b \in \mathbb{R}$;
- (b) $\ell_2(x) = (|x - b_1| - b_2)_+$ with some $b_1 \in \mathbb{R}$ and $b_2 \geq 0$;
- (c) $\ell_3(x) = |x - b_1| + b_2$ with some $b_1 \in \mathbb{R}$ and $b_2 > 0$,

or $\ell(z, t) = C(|z| - t)_+$. It holds that

$$\sup_{\mathbb{Q}|_Y \in \mathbb{B}_\delta(\sum_{i=1}^N p_i I_{\hat{Y}_i})} \rho_{\mathbb{Q}|_Y}^{(2)}(Y^\top \alpha) = \rho_{\sum_{i=1}^N p_i I_{\hat{Y}_i}}^{(2)}(Y^\top \alpha) + C \delta^{1/q} \|\alpha\|_*.$$

Theorem 27. (Wu et al. (2022) Theorem 5) For $q \in [1, \infty]$, let $h : [0, 1] \rightarrow \mathbb{R}$ be an increasing and convex distortion function satisfying $\lim_{x \rightarrow 1^-} h(x) = h(1)$ and $\|h'_-\|_{q^*} \in (0, \infty)$, let $\ell : \mathbb{R} \rightarrow \mathbb{R}$ be a convex function, and let $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}^q$, for any $y_1, y_2 \in \mathcal{Y}$.

(i) Let $q = 1$. If $\text{Lip}(\ell) < \infty$, then for any $(p, \delta) \in \mathcal{V}$ with $\delta \geq 0$, we have

$$\sup_{\mathbb{Q}|_Y \in \mathbb{B}_\delta(\sum_{i=1}^N p_i I_{\hat{Y}_i})} \rho_{\mathbb{Q}|_Y}^h(\ell(Y^\top \alpha)) = \rho_{\sum_{i=1}^N p_i I_{\hat{Y}_i}}^h(\ell(Y^\top \alpha)) + \text{Lip}(\ell) \|h'_-\|_\infty \delta \|\alpha\|_*.$$

(ii) Let $q \in (1, \infty]$. If $\rho_h(|\ell(Z)|) < \infty$ for all $Z \in L^q$, then the following statements are equivalent.

(1) For any $(p, \delta) \in \mathcal{V}$ with $\delta \geq 0$, there exists $C > 0$ such that

$$\sup_{\mathbb{Q}|_Y \in \mathbb{B}_\delta(\sum_{i=1}^N p_i \mathbb{I}_{\tilde{Y}_i})} \rho_{\mathbb{Q}|_Y}^h(\ell(Y^\top \alpha)) = \rho_{\sum_{i=1}^N p_i \mathbb{I}_{\tilde{Y}_i}}^h(\ell(Y^\top \alpha)) + C \|h'_-\|_{q^*} \delta^{1/q} \|\alpha\|_*.$$

(2) The function ℓ takes one of the following two forms, multiplied by C :

(a) $\ell_1(x) = x + b$ or $\ell_1(x) = -x + b$ with some $b \in \mathbb{R}$;

(b) $\ell_2(x) = |x - b_1| + b_2$ with some $b_1, b_2 \in \mathbb{R}$.

EC.2. Main Proofs

EC.2.1. Proofs of Section 2

Proof of Proposition 1. To obtain the sufficient and necessary feasible condition of problem (7), we only need to find the sufficient and necessary feasible condition of problem (8) and thus only need to find the sufficient and necessary condition such that the ambiguity set $\mathcal{K}(\mathcal{N}_\gamma(x_0), \mathcal{N}_\omega, \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}))$ is not empty. We consider the following three cases: (i) $\frac{N-m}{N} > \omega_2$; (ii) $\frac{N-m}{N} \in [\omega_1, \omega_2] = \mathcal{N}_\omega$; and (iii) $\frac{N-m}{N} < \omega_1$. Before the discussion case by case, we first note that

$$\mathcal{K}(\mathcal{N}_\gamma(x_0), \mathcal{N}_\omega, \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}})) = \left(\bigcup_{(p, \delta) \in \mathcal{V}_1} \mathbb{B}_\delta \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i} \right) \right) \cup \left(\bigcup_{(p, \delta) \in \mathcal{V}_2} \mathbb{B}_\delta^\circ \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i} \right) \right) := \mathcal{U}_\mathcal{B}$$

from the proof of Theorem 3, where $\mathcal{V}_1 = \{(p, \delta) : \exists \varepsilon \text{ such that } (p, \delta, \varepsilon) \in \mathcal{V}_0, p_i = \frac{\varepsilon}{N}, \text{ for } i = m+1, \dots, N\}$ and $\mathcal{V}_2 = \{(p, \delta) : \exists \varepsilon \text{ such that } (p, \delta, \varepsilon) \in \mathcal{V}_0, \exists i \in \{m+1, \dots, N\} \text{ such that } p_i < \frac{\varepsilon}{N}\}$. Thus, $\mathcal{K}(\mathcal{N}_\gamma(x_0), \mathcal{N}_\omega, \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}))$ is not empty is equivalent to $\mathcal{U}_\mathcal{B}$ is not empty.

- (i) For $\frac{N-m}{N} > \omega_2$, one can verify that $\mathcal{V}_1 = \emptyset$ due to that for every $(p, \delta) \in \mathcal{V}_1$, we have $\sum_{i=m+1}^N p_i = \frac{N-m}{N} \varepsilon > 1$, which is contradictory to the constraint $\sum_{i=1}^N p_i = 1$. Then, we have

$$\mathcal{K}(\mathcal{N}_\gamma(x_0), \mathcal{N}_\omega, \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}})) = \bigcup_{(p, \delta) \in \mathcal{V}_2} \mathbb{B}_\delta^\circ \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i} \right) = \bigcup_{(p, \delta) \in \mathcal{V}} \mathbb{B}_\delta^\circ \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i} \right),$$

where the second equality follows from that $\mathcal{V} = \mathcal{V}_1 \cup \mathcal{V}_2$ and $\mathcal{V}_1 = \emptyset$. Thus, $\mathcal{K}(\mathcal{N}_\gamma(x_0), \mathcal{N}_\omega, \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}))$ is not empty if and only if there exists $(p, \delta) \in \mathcal{V}$ such that $\delta > 0$, and thus if and only if there exists $(p, \delta, \varepsilon) \in \mathcal{V}_0$ such that $\delta = \varepsilon \left(\delta_0 - \frac{1}{N} \sum_{i=m+1}^N d_i \right) - \sum_{i=1}^m p_i d_i + \sum_{i=m+1}^N p_i d_i > 0$. One can verify that there exists $(p, \delta, \varepsilon) \in \mathcal{V}_0$ such that $\delta > 0$ is equivalent to

$$\max_{\substack{p_i \in [0, \frac{\varepsilon}{N}], i=1, \dots, N, \\ \sum_{i=1}^N p_i = 1, \varepsilon \in [\frac{1}{\omega_2}, \frac{1}{\omega_1}]}} \varepsilon \left(\delta_0 - \frac{1}{N} \sum_{i=m+1}^N d_i \right) - \sum_{i=1}^m p_i d_i + \sum_{i=m+1}^N p_i d_i > 0,$$

and thus equivalent to

$$\begin{aligned} \delta_0 &> \min_{\substack{p_i \in [0, \frac{\varepsilon}{N}], i=1, \dots, N, \\ \sum_{i=1}^N p_i = 1, \varepsilon \in [\frac{1}{\omega_2}, \frac{1}{\omega_1}]}} \frac{1}{N} \left(\sum_{i=1}^m \frac{N p_i}{\varepsilon} d_i + \sum_{i=m+1}^N \left(1 - \frac{N p_i}{\varepsilon} \right) d_i \right) \\ &= \min_{\substack{v_i \in [0, 1], i=1, \dots, N, \\ 1/N \sum_{i=1}^N v_i \in [\omega_1, \omega_2]}} \frac{1}{N} \left(\sum_{i=1}^m v_i d_i + \sum_{i=m+1}^N (1 - v_i) d_i \right) \\ &= \delta_{\min}. \end{aligned}$$

We conclude that in this case problem (7) is feasible if and only if $\delta_0 > \delta_{\min}$.

- (ii) For $\frac{N-m}{N} \in [\omega_1, \omega_2] = \mathcal{N}_\omega$, we find that $\widehat{\mathbb{P}}(X \in \mathcal{N}_\gamma(x_0)) \in [\omega_1, \omega_2] = \mathcal{N}_\omega$ and then $\widehat{\mathbb{P}}|_Y \in \mathcal{K}(\mathcal{N}_\gamma(x_0), \mathcal{N}_\omega, \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}))$. Thus, $\mathcal{K}(\mathcal{N}_\gamma(x_0), \mathcal{N}_\omega, \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}))$ is not empty if and only if $\delta_0 \geq 0$. Moreover, $\min_{v_i \in [0,1], i=1, \dots, N} \frac{1}{N} \left(\sum_{i=1}^m v_i d_i + \sum_{i=m+1}^N (1-v_i) d_i \right) \geq 0$ and can attain lower bound 0 in $v_i = 0$ for $i = 1, \dots, m$, and $v_i = 1$ for $i = m+1, \dots, N$, and we find that $\sum_{i=1}^N v_i = \frac{N-m}{N} \in [\omega_1, \omega_2]$. Thus,

$$\delta_{min} = \min_{\substack{v_i \in [0,1], i=1, \dots, N, \\ 1/N \sum_{i=1}^N v_i \in [\omega_1, \omega_2]}} \frac{1}{N} \left(\sum_{i=1}^m v_i d_i + \sum_{i=m+1}^N (1-v_i) d_i \right) = 0.$$

From above discussion, we conclude that in this case problem (7) is feasible if and only if $\delta_0 \geq \delta_{min}$.

- (iii) For $\frac{N-m}{N} < \omega_1$, if $\delta_0 \geq \delta_{min}$, we consider in the two cases: (iiia) $\delta_0 > \delta_{min}$, and (iiib) $\delta_0 = \delta_{min}$.

- (iiia) If $\delta_0 > \delta_{min}$, there exist $\{v_i\}_{i=1}^N$ satisfies that $v_i \in [0, 1], i = 1, \dots, N, 1/N \sum_{i=1}^N v_i \in [\omega_1, \omega_2]$ such that $\delta_0 > \frac{1}{N} \left(\sum_{i=1}^m v_i d_i + \sum_{i=m+1}^N (1-v_i) d_i \right)$. Given $\varepsilon \in [\frac{1}{\omega_2}, \frac{1}{\omega_1}]$, let $p_i = \frac{\varepsilon v_i}{N} \in [0, \frac{\varepsilon}{N}]$, and $\delta = \varepsilon \left(\delta_0 - \frac{1}{N} \sum_{i=m+1}^N d_i \right) - \sum_{i=1}^m p_i d_i + \sum_{i=m+1}^N p_i d_i$. Then

$$\delta = \varepsilon \left[\delta_0 - \frac{1}{N} \left(\sum_{i=1}^m v_i d_i + \sum_{i=m+1}^N (1-v_i) d_i \right) \right] > 0,$$

and thus there exists $(p, \delta) \in \mathcal{V}$ such that $\delta > 0$. As a result, there exists $(p, \delta) \in \mathcal{V}$ such that $\mathbb{B}_\delta^\circ \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i} \right)$ is not empty. From the construction of $\mathcal{K}(\mathcal{N}_\gamma(x_0), \mathcal{N}_\omega, \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}))$, we conclude that $\mathcal{K}(\mathcal{N}_\gamma(x_0), \mathcal{N}_\omega, \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}))$ is not empty.

- (iiib) If $\delta_0 = \delta_{min}$, from the definition of δ_{min} , we find that there exist $\{v_i^*\}_{i=1}^N$ satisfies that $v_i^* \in [0, 1], i = 1, \dots, N, 1/N \sum_{i=1}^N v_i^* \in [\omega_1, \omega_2]$ such that

$$\delta_0 = \delta_{min} = \frac{1}{N} \left(\sum_{i=1}^m v_i^* d_i + \sum_{i=m+1}^N (1-v_i^*) d_i \right)$$

due to that δ_{min} is a linear problem over a closed and convex set. We claim that $\{v_i^*\}_{i=1}^N$ must satisfy that $v_i^* = 1$ for $i = m+1, \dots, N$. Otherwise, without loss of generality, we assume $v_i^* = 1$ for $i = m+1, \dots, k$ and $v_i^* < 1$ for $i = k+1, \dots, N$. Let $v_i = v_i^* - \Delta_i \geq 0$ for $i = 1, \dots, m$ and $v_i = 1$ for $i = m+1, \dots, N$, where $\Delta_i \geq 0$ and $\sum_{i=1}^m \Delta_i = \sum_{i=k+1}^N (1-v_i^*)$. We find $\{v_i\}_{i=1}^N$ satisfies that $v_i \in [0, 1], i = 1, \dots, N, 1/N \sum_{i=1}^N v_i \in [\omega_1, \omega_2]$ and such $\{v_i\}_{i=1}^N$ must exists due to that $\frac{N-m}{N} < \omega_1 \leq \omega_2$. Thus we have $\sum_{i=1}^m v_i d_i + \sum_{i=m+1}^N (1-v_i) d_i < \sum_{i=1}^m v_i^* d_i + \sum_{i=m+1}^N (1-v_i^*) d_i = \delta_{min}$ and the strict inequality contradicts the definition of δ_{min} . As a result, the claim holds. Given $\varepsilon \in [\frac{1}{\omega_2}, \frac{1}{\omega_1}]$, using the same construction on (iiia), we find that there exists $(p, \delta) \in \mathcal{V}$ such that $\delta = 0$ with $p_i = \frac{\varepsilon}{N}$ for $i = m+1, \dots, N$. Thus, $(p, \delta) \in \mathcal{V}_1$. As a result, there exists $(p, \delta) \in \mathcal{V}_1$ such that $\mathbb{B}_\delta \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i} \right) = \left\{ \sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i} \right\}$ is not empty. From the construction of $\mathcal{K}(\mathcal{N}_\gamma(x_0), \mathcal{N}_\omega, \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}))$, we conclude that $\mathcal{K}(\mathcal{N}_\gamma(x_0), \mathcal{N}_\omega, \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}))$ is not empty.

Combining the above two cases, we find that if $\delta_0 \geq \delta_{min}$, $\mathcal{K}(\mathcal{N}_\gamma(x_0), \mathcal{N}_\omega, \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}))$ is not empty, and

thus problem (7) is feasible. For the “only if” part, if $\delta_0 < \delta_{\min}$, one can verify that for any $(p, \delta) \in \mathcal{V}$,

$$\begin{aligned}
\delta &= \varepsilon \left(\delta_0 - \frac{1}{N} \sum_{i=m+1}^N d_i \right) - \sum_{i=1}^m p_i d_i + \sum_{i=m+1}^N p_i d_i \\
&< \varepsilon \left(\delta_{\min} - \frac{1}{N} \sum_{i=m+1}^N d_i \right) - \sum_{i=1}^m p_i d_i + \sum_{i=m+1}^N p_i d_i \\
&= \min_{\substack{v_i \in [0,1], i=1, \dots, N, \\ 1/N \sum_{i=1}^N v_i \in [\omega_1, \omega_2]}} \sum_{i=1}^m \left(\frac{\varepsilon v_i}{N} - p_i \right) d_i + \sum_{i=m+1}^N \left(p - \frac{\varepsilon v_i}{N} \right) d_i \\
&\leq 0,
\end{aligned}$$

where the last inequality follows from that there exists $\{v_i^0\}_{i=1}^N = \{N p_i / \varepsilon\}_{i=1}^N$ satisfies $v_i^0 \in [0, 1], i = 1, \dots, N$, and $1/N \sum_{i=1}^N v_i^0 \in [\omega_1, \omega_2]$ such that $\sum_{i=1}^m \left(\frac{\varepsilon v_i^0}{N} - p_i \right) d_i + \sum_{i=m+1}^N \left(p - \frac{\varepsilon v_i^0}{N} \right) d_i = 0$. From the construction of $\mathcal{K}(\mathcal{N}_\gamma(x_0), \mathcal{N}_\omega, \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}))$, we conclude that $\mathcal{K}(\mathcal{N}_\gamma(x_0), \mathcal{N}_\omega, \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}))$ is empty. From above discussion, we conclude that in this case problem (7) is feasible if and only if $\delta_0 \geq \delta_{\min}$. $\square$

Proof of Theorem 3. Let $\mathcal{V}_1 = \{(p, \delta) : \exists \varepsilon \text{ such that } (p, \delta, \varepsilon) \in \mathcal{V}_0, p_i = \frac{\varepsilon}{N}, \text{ for } i = m+1, \dots, N\}$ and $\mathcal{V}_2 = \{(p, \delta) : \exists \varepsilon \text{ such that } (p, \delta, \varepsilon) \in \mathcal{V}_0, \exists i \in \{m+1, \dots, N\} \text{ such that } p_i < \frac{\varepsilon}{N}\}$. Then $\mathcal{V} = \mathcal{V}_1 \cup \mathcal{V}_2$. We first prove that

$$\mathcal{K}(\mathcal{N}_\gamma(x_0), \mathcal{N}_\omega, \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}})) = \left(\bigcup_{(p, \delta) \in \mathcal{V}_1} \mathbb{B}_\delta \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{\mathcal{Y}}_i} \right) \right) \cup \left(\bigcup_{(p, \delta) \in \mathcal{V}_2} \mathbb{B}_\delta^\circ \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{\mathcal{Y}}_i} \right) \right),$$

where $\mathbb{B}_\delta \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{\mathcal{Y}}_i} \right)$ is the ambiguity set on $\mathcal{Y}$, defined as

$$\mathbb{B}_\delta \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{\mathcal{Y}}_i} \right) = \left\{ \mathbb{Q}|_Y \in \mathcal{M}(\mathcal{Y}) : \mathcal{W} \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{\mathcal{Y}}_i}, \mathbb{Q}|_Y \right) \leq \delta \right\},$$

and $\mathbb{B}_\delta^\circ \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{\mathcal{Y}}_i} \right)$ is its interior point set, defined as

$$\mathbb{B}_\delta^\circ \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{\mathcal{Y}}_i} \right) = \left\{ \mathbb{Q}|_Y \in \mathcal{M}(\mathcal{Y}) : \mathcal{W} \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{\mathcal{Y}}_i}, \mathbb{Q}|_Y \right) < \delta \right\}.$$

To obtain the set equivalence, we first prove the following containment relation

$$\mathcal{K}(\mathcal{N}_\gamma(x_0), \mathcal{N}_\omega, \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}})) \subset \left(\bigcup_{(p, \delta) \in \mathcal{V}_1} \mathbb{B}_\delta \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{\mathcal{Y}}_i} \right) \right) \cup \left(\bigcup_{(p, \delta) \in \mathcal{V}_2} \mathbb{B}_\delta^\circ \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{\mathcal{Y}}_i} \right) \right).$$

For any $\mathbb{Q} \in \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}})$, $\mathbb{Q}(X \in \mathcal{N}_\gamma(x_0)) \in \mathcal{N}_\omega$, assume that the optimal transport between $\widehat{\mathbb{P}}$ and $\mathbb{Q}$ is π^* . Then

$\mathbb{Q}$ have the following representation by support set and π^* ,

$$\mathbb{Q} = \sum_{i=1}^N p_i \mathbb{Q}_i^1 + \sum_{i=1}^N \left(\frac{1}{N} - p_i \right) \mathbb{Q}_i^2,$$

where $\text{supp } \mathbb{Q}_i^1 \subset \{(x, y) : x \in \mathcal{N}_\gamma(x_0)\}$, $\text{supp } \mathbb{Q}_i^2 \subset \mathcal{X} \times \mathcal{Y} \setminus \{(x, y) : x \in \mathcal{N}_\gamma(x_0)\}$, $\mathbb{Q}_i^1$ and $\mathbb{Q}_i^2$ are transferred from $\mathbb{I}_{(\widehat{x}_i, \widehat{y}_i)}$ under π^* with probability p_i and $\frac{1}{N} - p_i$, respectively, $p_i \in [0, \frac{1}{N}]$ and $\sum_{i=1}^N p_i \in \mathcal{N}_\omega$. Based on this decomposition of $\mathbb{Q}$, if the joint distribution of (X, Y) is $\mathbb{Q}$, the conditional distribution of Y given $X \in \mathcal{N}_\gamma(x_0)$, $\mathbb{Q}|_Y$, can be written as $\mathbb{Q}|_Y = \frac{\sum_{i=1}^N p_i \mathbb{Q}_i^1|_Y}{\sum_{i=1}^N p_i}$. Note that $\text{supp } \mathbb{Q}_i^2 \subset \mathcal{X} \times \mathcal{Y} \setminus \{(x, y) : x \in \mathcal{N}_\gamma(x_0)\}$, $(\widehat{x}_i, \widehat{y}_i) \in \{(x, y) : x \in \mathcal{N}_\gamma(x_0)\}$, $i = m+1, \dots, N$, and $\widehat{x}_i^p = \arg \min_{x \in \partial \mathcal{N}_\gamma(x_0)} \mathbb{D}_\mathcal{X}(x, \widehat{x}_i)$. Then $\int \mathbb{D}_\mathcal{X}(\widehat{x}_i, x) d\mathbb{Q}_i^2(x, y) > d_i$ for $i = m+1, \dots, N$. Due to the continuity of $\mathbb{D}_\mathcal{X}$, there exists $\phi_Q^i > 0$ and $\psi_Q^i > 0$, $i = m+1, \dots, N$ such that $\int \mathbb{D}_\mathcal{X}(\widehat{x}_i, x) d\mathbb{Q}_i^2(x, y) \geq \mathbb{D}_\mathcal{X}(\widehat{x}_i, \widehat{x}_i^p + \phi_Q^i(\widehat{x}_i^p - \widehat{x}_i)) = d_i + \psi_Q^i$. Let $\psi_Q = \min_{i=m+1, \dots, N} \psi_Q^i$. One can verify that under the optimal transport π^* , we have

$$\begin{aligned} \mathcal{W}(\widehat{\mathbb{P}}, \mathbb{Q}) &= \sum_{i=1}^m p_i \int \mathbb{D}_\mathcal{X}(\widehat{x}_i, x) + \mathbb{D}_\mathcal{Y}(\widehat{y}_i, y) d\mathbb{Q}_i^1(x, y) \\ &\quad + \sum_{i=1}^m \left(\frac{1}{N} - p_i \right) \int \mathbb{D}_\mathcal{X}(\widehat{x}_i, x) + \mathbb{D}_\mathcal{Y}(\widehat{y}_i, y) d\mathbb{Q}_i^2(x, y) \\ &\quad + \sum_{i=m+1}^N p_i \int \mathbb{D}_\mathcal{X}(\widehat{x}_i, x) + \mathbb{D}_\mathcal{Y}(\widehat{y}_i, y) d\mathbb{Q}_i^1(x, y) \\ &\quad + \sum_{i=m+1}^N \left(\frac{1}{N} - p_i \right) \int \mathbb{D}_\mathcal{X}(\widehat{x}_i, x) + \mathbb{D}_\mathcal{Y}(\widehat{y}_i, y) d\mathbb{Q}_i^2(x, y) \\ &\geq \sum_{i=1}^m p_i \int d_i + \mathbb{D}_\mathcal{Y}(\widehat{y}_i, y) d\mathbb{Q}_i^1(x, y) \\ &\quad + \sum_{i=1}^m \left(\frac{1}{N} - p_i \right) \int \mathbb{D}_\mathcal{X}(\widehat{x}_i, \widehat{x}_i) + \mathbb{D}_\mathcal{Y}(\widehat{y}_i, \widehat{y}_i) d\mathbb{Q}_i^2(x, y) \\ &\quad + \sum_{i=m+1}^N p_i \int \mathbb{D}_\mathcal{X}(\widehat{x}_i, \widehat{x}_i) + \mathbb{D}_\mathcal{Y}(\widehat{y}_i, y) d\mathbb{Q}_i^1(x, y) \\ &\quad + \sum_{i=m+1}^N \left(\frac{1}{N} - p_i \right) \int \mathbb{D}_\mathcal{X}(\widehat{x}_i, \widehat{x}_i^p + \phi_Q^i(\widehat{x}_i^p - \widehat{x}_i)) + \mathbb{D}_\mathcal{Y}(\widehat{y}_i, \widehat{y}_i) d\mathbb{Q}_i^2(x, y). \end{aligned}$$

Rearranging and combining the equations, and applying the definition of the optimal transport, we obtain

$$\begin{aligned}
\mathcal{W}(\widehat{\mathbb{P}}, \mathbb{Q}) &\geq \sum_{i=1}^N p_i \int \mathbb{D}_{\mathcal{Y}}(\widehat{y}_i, y) d\mathbb{Q}_i^1|_{\mathcal{Y}}(y) + \sum_{i=1}^m p_i d_i + \sum_{i=m+1}^N \left(\frac{1}{N} - p_i\right) d_i + \sum_{i=m+1}^N \left(\frac{1}{N} - p_i\right) \psi_{\mathbb{Q}} \\
&\geq \left(\sum_{j=1}^N p_j\right) \mathcal{W}\left(\sum_{i=1}^N \frac{p_i}{\sum_{j=1}^N p_j} \mathbb{I}_{\widehat{y}_i}, \sum_{i=1}^N \frac{p_i}{\sum_{j=1}^N p_j} \mathbb{Q}_i^1|_{\mathcal{Y}}\right) \\
&\quad + \sum_{i=1}^m p_i d_i + \sum_{i=m+1}^N \left(\frac{1}{N} - p_i\right) d_i + \sum_{i=m+1}^N \left(\frac{1}{N} - p_i\right) \psi_{\mathbb{Q}}.
\end{aligned} \tag{27}$$

Denote $\varepsilon = \frac{1}{\sum_{j=1}^N p_j}$, and $p'_i = \frac{p_i}{\sum_{j=1}^N p_j}$, $i = 1, \dots, N$. Then $\frac{1}{\varepsilon} \in \mathcal{N}_{\omega}$, $p'_i \in [0, \frac{\varepsilon}{N}]$, $i = 1, \dots, N$, $\sum_{i=1}^N p'_i = 1$, and $\mathbb{Q}|_{\mathcal{Y}} = \sum_{i=1}^N p'_i \mathbb{Q}_i^1|_{\mathcal{Y}}$. Let $\widehat{\mathbb{P}}^*|_{\mathcal{Y}} = \sum_{i=1}^N p'_i \mathbb{I}_{\widehat{y}_i}$. Then inequality (27) can be rearranged as

$$\begin{aligned}
\mathcal{W}(\widehat{\mathbb{P}}^*|_{\mathcal{Y}}, \mathbb{Q}|_{\mathcal{Y}}) &\leq \varepsilon \mathcal{W}(\widehat{\mathbb{P}}, \mathbb{Q}) - \sum_{i=1}^m p'_i d_i - \sum_{i=m+1}^N \left(\frac{\varepsilon}{N} - p'_i\right) d_i - \sum_{i=m+1}^N \left(\frac{\varepsilon}{N} - p'_i\right) \psi_{\mathbb{Q}} \\
&\leq \varepsilon \delta_0 - \sum_{i=1}^m p'_i d_i - \sum_{i=m+1}^N \left(\frac{\varepsilon}{N} - p'_i\right) d_i - \sum_{i=m+1}^N \left(\frac{\varepsilon}{N} - p'_i\right) \psi_{\mathbb{Q}}.
\end{aligned} \tag{28}$$

We consider in the following two cases.

(i) If $p'_i = \frac{\varepsilon}{N}$ for $i = m+1, \dots, N$, from inequality (28), we have

$$\begin{aligned}
\mathcal{W}(\widehat{\mathbb{P}}^*|_{\mathcal{Y}}, \mathbb{Q}|_{\mathcal{Y}}) &\leq \varepsilon \delta_0 - \sum_{i=1}^m p'_i d_i - \sum_{i=m+1}^N \left(\frac{\varepsilon}{N} - p'_i\right) d_i - \sum_{i=m+1}^N \left(\frac{\varepsilon}{N} - p'_i\right) \psi_{\mathbb{Q}} \\
&= \varepsilon \delta_0 - \sum_{i=1}^m p'_i d_i = \delta.
\end{aligned} \tag{29}$$

(ii) If there exists $i \in \{m+1, \dots, N\}$ such that $p'_i < \frac{\varepsilon}{N}$, from inequality (28), we have

$$\begin{aligned}
\mathcal{W}(\widehat{\mathbb{P}}^*|_{\mathcal{Y}}, \mathbb{Q}|_{\mathcal{Y}}) &\leq \varepsilon \delta_0 - \sum_{i=1}^m p'_i d_i - \sum_{i=m+1}^N \left(\frac{\varepsilon}{N} - p'_i\right) d_i - \sum_{i=m+1}^N \left(\frac{\varepsilon}{N} - p'_i\right) \psi_{\mathbb{Q}} \\
&< \varepsilon \delta_0 - \sum_{i=1}^m p'_i d_i - \sum_{i=m+1}^N \left(\frac{\varepsilon}{N} - p'_i\right) d_i = \delta.
\end{aligned} \tag{30}$$

We conclude that, for any $\mathbb{Q} \in \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}})$, $\mathbb{Q}(X \in \mathcal{N}_{\gamma}(x_0)) \in \mathcal{N}_{\omega}$, the conditional distribution of Y given $X \in \mathcal{N}_{\gamma}(x_0)$, $\mathbb{Q}|_{\mathcal{Y}}$, satisfies the inequality (29) or inequality (30), where $\frac{1}{\varepsilon} \in \mathcal{N}_{\omega}$, $p'_i \in [0, \frac{\varepsilon}{N}]$, $i = 1, \dots, N$, $\sum_{i=1}^N p'_i = 1$. Thus, we have

$$\mathcal{K}(\mathcal{N}_{\gamma}(x_0), \mathcal{N}_{\omega}, \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}})) \subset \left(\bigcup_{(p, \delta) \in \mathcal{V}_1} \mathbb{B}_{\delta} \left(\sum_{i=1}^N p_i \mathbb{I}_{\widehat{y}_i} \right) \right) \cup \left(\bigcup_{(p, \delta) \in \mathcal{V}_2} \mathbb{B}_{\delta}^{\circ} \left(\sum_{i=1}^N p_i \mathbb{I}_{\widehat{y}_i} \right) \right).$$

Next, we want to prove the inverse inclusion. For any distribution

$$\mathbb{Q}|_Y \in \left(\bigcup_{(p, \delta) \in \mathcal{V}_1} \mathbb{B}_\delta \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i} \right) \right) \cup \left(\bigcup_{(p, \delta) \in \mathcal{V}_2} \mathbb{B}_\delta^\circ \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i} \right) \right),$$

there exists $(p, \delta) \in \mathcal{V}_1$ such that

$$\mathcal{W} \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i}, \mathbb{Q}|_Y \right) \leq \delta,$$

or exists $(p, \delta) \in \mathcal{V}_2$ such that

$$\mathcal{W} \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i}, \mathbb{Q}|_Y \right) < \delta.$$

Assume that the optimal transport between $\sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i}$ and $\mathbb{Q}|_Y$ is $\pi^*|_Y$. Under the optimal transport, $\mathbb{Q}|_Y$ has the following decomposition:

$$\mathbb{Q}|_Y = \sum_{i=1}^N p_i \mathbb{Q}_i|_Y,$$

where $\mathbb{Q}_i|_Y$ is transferred from $\mathbf{I}_{\widehat{y}_i}$ under the optimal transport. One can verify that for $(p, \delta) \in \mathcal{V}$, there exists $\psi_{\mathbb{Q}} > 0$ such that $\mathcal{W}(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i}, \mathbb{Q}|_Y) \leq \delta - \sum_{i=m+1}^N (\frac{\varepsilon}{N} - p_i) \psi_{\mathbb{Q}}$. Specifically, if $(p, \delta) \in \mathcal{V}_1$, we have $p_i = \frac{\varepsilon}{N}$ for $i = m+1, \dots, N$ and $\mathcal{W}(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i}, \mathbb{Q}|_Y) \leq \delta = \delta - \sum_{i=m+1}^N (\frac{\varepsilon}{N} - p_i) \psi_{\mathbb{Q}}$ for any $\psi_{\mathbb{Q}} > 0$. Otherwise, if $(p, \delta) \in \mathcal{V}_2$, we have $\mathcal{W}(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i}, \mathbb{Q}|_Y) \leq \delta - \sum_{i=m+1}^N (\frac{\varepsilon}{N} - p_i) \psi_{\mathbb{Q}}$ for some $\psi_{\mathbb{Q}} > 0$, due to $\mathcal{W}(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i}, \mathbb{Q}|_Y) < \delta$. Then we can construct

$$\begin{aligned} \mathbb{Q} &= \sum_{i=1}^m \frac{p_i}{\varepsilon} \mathbf{I}_{\widehat{x}_i^p} \times \mathbb{Q}_i|_Y + \sum_{i=m+1}^N \frac{p_i}{\varepsilon} \mathbf{I}_{\widehat{x}_i} \times \mathbb{Q}_i|_Y \\ &\quad + \sum_{i=1}^m \left(\frac{1}{N} - \frac{p_i}{\varepsilon} \right) \mathbf{I}_{(\widehat{x}_i, \widehat{y}_i)} + \sum_{i=m+1}^N \left(\frac{1}{N} - \frac{p_i}{\varepsilon} \right) \mathbf{I}_{(\widehat{x}_i^p + \phi_{\mathbb{Q}}^i (\widehat{x}_i^p - \widehat{x}_i), \widehat{y}_i)}, \end{aligned}$$

where $\phi_{\mathbb{Q}}^i$ satisfies that $\mathbb{D}_X(\widehat{x}_i, \widehat{x}_i^p + \phi_{\mathbb{Q}}^i (\widehat{x}_i^p - \widehat{x}_i)) = d_i + \psi_{\mathbb{Q}}^i = d_i + \psi_{\mathbb{Q}}^i$. Note that the conditional distribution of Y given $X \in \mathcal{N}_Y(x_0)$ under the joint distribution $\mathbb{Q}$ is $\mathbb{Q}|_Y$. Moreover, one can verify that $\widehat{\mathbb{P}}$ can be represented as

$$\begin{aligned} \widehat{\mathbb{P}} &= \sum_{i=1}^m \frac{p_i}{\varepsilon} \mathbf{I}_{(\widehat{x}_i, \widehat{y}_i)} + \sum_{i=m+1}^N \frac{p_i}{\varepsilon} \mathbf{I}_{(\widehat{x}_i, \widehat{y}_i)} \\ &\quad + \sum_{i=1}^m \left(\frac{1}{N} - \frac{p_i}{\varepsilon} \right) \mathbf{I}_{(\widehat{x}_i, \widehat{y}_i)} + \sum_{i=m+1}^N \left(\frac{1}{N} - \frac{p_i}{\varepsilon} \right) \mathbf{I}_{(\widehat{x}_i, \widehat{y}_i)}, \end{aligned}$$

and using the definition of optimal transport, we obtain

$$\begin{aligned}
\mathcal{W}(\widehat{\mathbb{P}}, \mathbb{Q}) &\leq \sum_{i=1}^m \frac{p_i}{\varepsilon} (d_i + \int \mathbb{D}_{\mathcal{Y}}(\widehat{y}_i, y) d\mathbb{Q}_i|_Y(y)) \\
&\quad + \sum_{i=m+1}^N \frac{p_i}{\varepsilon} (\mathbb{D}_{\mathcal{X}}(\widehat{x}_i, \widehat{x}_i) + \int \mathbb{D}_{\mathcal{Y}}(\widehat{y}_i, y) d\mathbb{Q}_i|_Y(y)) \\
&\quad + \sum_{i=1}^m \left(\frac{1}{N} - \frac{p_i}{\varepsilon}\right) (\mathbb{D}_{\mathcal{X}}(\widehat{x}_i, \widehat{x}_i) + \mathbb{D}_{\mathcal{Y}}(\widehat{y}_i, \widehat{y}_i)) \\
&\quad + \sum_{i=m+1}^N \left(\frac{1}{N} - \frac{p_i}{\varepsilon}\right) (\mathbb{D}_{\mathcal{X}}(\widehat{x}_i, \widehat{x}_i^p + \phi_{\mathbb{Q}}(\widehat{x}_i^p - \widehat{x}_i)) + \mathbb{D}_{\mathcal{Y}}(\widehat{y}_i, \widehat{y}_i)).
\end{aligned}$$

And thus, we have

$$\begin{aligned}
\mathcal{W}(\widehat{\mathbb{P}}, \mathbb{Q}) &\leq \sum_{i=1}^m \frac{p_i}{\varepsilon} \int \mathbb{D}_{\mathcal{Y}}(\widehat{y}_i, y) d\mathbb{Q}_i|_Y(y) + \sum_{i=m+1}^N \frac{p_i}{\varepsilon} \int \mathbb{D}_{\mathcal{Y}}(\widehat{y}_i, y) d\mathbb{Q}_i|_Y(y) \\
&\quad + \sum_{i=1}^m \frac{p_i}{\varepsilon} d_i + \sum_{i=m+1}^N \left(\frac{1}{N} - \frac{p_i}{\varepsilon}\right) d_i + \sum_{i=m+1}^N \left(\frac{1}{N} - \frac{p_i}{\varepsilon}\right) \psi_{\mathbb{Q}} \\
&= \frac{1}{\varepsilon} \mathcal{W}\left(\sum_{i=1}^N p_i \mathbb{I}_{\widehat{y}_i}, \mathbb{Q}|_Y\right) + \sum_{i=1}^m \frac{p_i}{\varepsilon} d_i + \sum_{i=m+1}^N \left(\frac{1}{N} - \frac{p_i}{\varepsilon}\right) d_i + \sum_{i=m+1}^N \left(\frac{1}{N} - \frac{p_i}{\varepsilon}\right) \psi_{\mathbb{Q}} \\
&\leq \delta_0.
\end{aligned}$$

We conclude that for any

$$\mathbb{Q}|_Y \in \left(\bigcup_{(p, \delta) \in \mathcal{V}_1} \mathbb{B}_{\delta} \left(\sum_{i=1}^N p_i \mathbb{I}_{\widehat{y}_i} \right) \right) \cup \left(\bigcup_{(p, \delta) \in \mathcal{V}_2} \mathbb{B}_{\delta}^{\circ} \left(\sum_{i=1}^N p_i \mathbb{I}_{\widehat{y}_i} \right) \right),$$

there exists $\mathbb{Q} \in \mathbb{B}_{\delta_0}$ satisfies $\mathbb{Q}(X \in \mathcal{N}_{\gamma}(x_0)) \geq \varepsilon$ such that the conditional distribution of Y given $X \in \mathcal{N}_{\gamma}(x_0)$ is $\mathbb{Q}|_Y$. Then we have

$$\left(\bigcup_{(p, \delta) \in \mathcal{V}_1} \mathbb{B}_{\delta} \left(\sum_{i=1}^N p_i \mathbb{I}_{\widehat{y}_i} \right) \right) \cup \left(\bigcup_{(p, \delta) \in \mathcal{V}_2} \mathbb{B}_{\delta}^{\circ} \left(\sum_{i=1}^N p_i \mathbb{I}_{\widehat{y}_i} \right) \right) \subset \mathcal{K}(\mathcal{N}_{\gamma}(x_0), \mathcal{N}_{\omega}, \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}})),$$

and thus

$$\left(\bigcup_{(p, \delta) \in \mathcal{V}_1} \mathbb{B}_{\delta} \left(\sum_{i=1}^N p_i \mathbb{I}_{\widehat{y}_i} \right) \right) \cup \left(\bigcup_{(p, \delta) \in \mathcal{V}_2} \mathbb{B}_{\delta}^{\circ} \left(\sum_{i=1}^N p_i \mathbb{I}_{\widehat{y}_i} \right) \right) = \mathcal{K}(\mathcal{N}_{\gamma}(x_0), \mathcal{N}_{\omega}, \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}})).$$

Denote

$$\mathcal{U}_{\mathcal{B}} = \left(\bigcup_{(p, \delta) \in \mathcal{V}_1} \mathbb{B}_{\delta} \left(\sum_{i=1}^N p_i \mathbb{I}_{\widehat{y}_i} \right) \right) \cup \left(\bigcup_{(p, \delta) \in \mathcal{V}_2} \mathbb{B}_{\delta}^{\circ} \left(\sum_{i=1}^N p_i \mathbb{I}_{\widehat{y}_i} \right) \right).$$

From the above discussion, the problem (7) is equivalent to

$$\min_{\alpha \in \mathcal{A}} \sup_{\mathbb{Q}|_Y \in \mathcal{U}_{\mathcal{B}}} \rho_{\mathbb{Q}|_Y} [\ell(Y, \alpha)].$$

We next prove that given $\alpha \in \mathcal{A}$,

$$\sup_{\mathbb{Q}|_Y \in \mathcal{U}_{\mathcal{B}}} \rho_{\mathbb{Q}|_Y} [\ell(Y, \alpha)] = \sup_{\mathbb{Q}|_Y \in \bigcup_{(p, \delta) \in \mathcal{V}} \mathbb{B}_{\delta} \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i} \right)} \rho_{\mathbb{Q}|_Y} [\ell(Y, \alpha)]. \quad (31)$$

We recall that the risk measure ρ and the loss function ℓ satisfy Assumption 2.2. Based on Assumption 2.2, we only need to prove the following claim.

Claim For any $\mathbb{Q}|_Y \in \mathbb{B}_{\delta} \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i} \right)$ with $(p, \delta) \in \mathcal{V}_2$ and sufficiently large $n \in \mathbb{N}$, denoted by $Y \sim \mathbb{Q}|_Y$ and $Y_0 \sim \sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i}$, there exist $\mathbb{Q}|_{Y_n} \in \mathcal{U}_{\mathcal{B}}$ and $Y_n \sim \mathbb{Q}|_{Y_n}$ such that $\mathcal{W}(Y, Y_n) \leq \frac{1}{n}$.

In fact, if the claim holds, for any $\mathbb{Q}|_Y \in \mathbb{B}_{\delta} \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i} \right)$ with $(p, \delta) \in \mathcal{V}_2$, we find that $\{\mathbb{Q}|_{Y_n}\} \subset \mathcal{U}_{\mathcal{B}}$ satisfies that $\mathcal{W}(\mathbb{Q}|_{Y_n}, \mathbb{Q}|_Y) \rightarrow 0$ as $n \rightarrow \infty$ and then $\rho[\ell(Y_n, \alpha)] \rightarrow \rho[\ell(Y, \alpha)]$ as $n \rightarrow \infty$, due to the OT continuity of $\rho \circ \ell$. Moreover, for any $\mathbb{Q}|_Y \in \mathbb{B}_{\delta} \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i} \right)$ with $(p, \delta) \in \mathcal{V}_1$, let $\mathbb{Q}|_{Y_n} = \mathbb{Q}|_Y$. Then for any $\varepsilon > 0$, we have

$$\sup_{\mathbb{Q}|_Y \in \mathcal{U}_{\mathcal{B}}} \rho_{\mathbb{Q}|_Y} [\ell(Y, \alpha)] - \varepsilon \leq \sup_{\mathbb{Q}|_Y \in \bigcup_{(p, \delta) \in \mathcal{V}} \mathbb{B}_{\delta} \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i} \right)} \rho_{\mathbb{Q}|_Y} [\ell(Y, \alpha)] \leq \sup_{\mathbb{Q}|_Y \in \mathcal{U}_{\mathcal{B}}} \rho_{\mathbb{Q}|_Y} [\ell(Y, \alpha)] + \varepsilon,$$

and thus the equation (31) holds. Next we prove the claim in the two cases: $\delta > 0$ and $\delta = 0$.

- (i) For $\mathbb{Q}|_Y \in \mathbb{B}_{\delta} \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i} \right)$ with $(p, \delta) \in \mathcal{V}_2$ and $\delta > 0$, we find that $\mathbb{B}_{\delta} \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i} \right)$ is the closure of $\mathbb{B}_{\delta}^{\circ} \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i} \right)$ under the cost $\mathcal{W}$. Thus, there exist $\mathbb{Q}|_{Y_n} \in \mathbb{B}_{\delta}^{\circ} \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i} \right) \subset \mathcal{U}_{\mathcal{B}}$ and $Y_n \sim \mathbb{Q}|_{Y_n}$ such that $\mathcal{W}(Y, Y_n) \leq \frac{1}{n}$ for sufficiently large $n \in \mathbb{N}$.
- (ii) For $\mathbb{Q}|_Y \in \mathbb{B}_{\delta} \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i} \right)$ with $(p, \delta) \in \mathcal{V}_2$ and $\delta = 0$, we find that $\mathbb{B}_{\delta} \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i} \right)$ is a single distribution set and $\mathbb{Q}|_Y = \sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i}$. We consider the following two cases.
 - (iia) If $\delta_0 > \delta_{\min}$, from the proof of Proposition 1, there exist $(p^*, \delta^*) \in \mathcal{V}$ such that $\delta^* > 0$. Let $Y \sim \sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i}$ and $Y^* \sim \sum_{i=1}^N p_i^* \mathbf{I}_{\widehat{y}_i}$. One can verify that $\mathcal{W}(Y, Y^*) \leq \max_{i,j=1,\dots,N} \mathbb{D}_{\mathcal{Y}}(\widehat{y}_i, \widehat{y}_j) := D$. For $n \in \mathbb{N}$ satisfies that $nD \geq 1$, let $p^n = (1 - \frac{1}{nD})p + \frac{1}{nD}p^*$ and $\delta^n = (1 - \frac{1}{nD})\delta + \frac{1}{nD}\delta^*$. Due to the convexity of $\mathcal{V}$, we have $(p^n, \delta^n) \in \mathcal{V}$ and thus $\sum_{i=1}^N p_i^n \mathbf{I}_{\widehat{y}_i} \in \mathcal{U}_{\mathcal{B}}$ because $\delta^* > 0$ and then $\delta_n > 0$. Let $Y_n \sim \sum_{i=1}^N p_i^n \mathbf{I}_{\widehat{y}_i}$. Then we find that $\mathcal{W}(Y, Y_n) \leq \frac{1}{nD} \mathcal{W}(Y, Y^*) \leq \frac{1}{n}$.
 - (iib) If $\delta_0 = \delta_{\min}$, we claim that for every $(p, \delta) \in \mathcal{V}_2$, $\delta < 0$. Note that δ_0 can take the smallest radius $\delta_{\min}$ if and only if $\frac{N-m}{N} \leq \omega_2$. In fact, if $(p, \delta) \in \mathcal{V}_2$, there exists ε such that $(p, \delta, \varepsilon) \in \mathcal{V}_0$, and exists $i \in \{m+1, \dots, N\}$ such that $p_i < \frac{\varepsilon}{N}$. Note that $\delta = \varepsilon \left(\delta_0 - \frac{1}{N} \sum_{i=m+1}^N d_i \right) - \sum_{i=1}^m p_i d_i +$

$\sum_{i=m+1}^N p_i d_i$. If there exist $(p, \delta) \in \mathcal{V}_2$ satisfies that $\delta \geq 0$, we have

$$\begin{aligned}
\delta_0 &\geq \sum_{i=1}^m \frac{p_i}{\varepsilon} d_i + \sum_{i=m+1}^N \left(\frac{1}{N} - \frac{p_i}{\varepsilon} \right) d_i \\
&= \frac{1}{N} \left(\sum_{i=1}^m \frac{N p_i}{\varepsilon} d_i + \sum_{i=m+1}^N \left(1 - \frac{N p_i}{\varepsilon} \right) d_i \right) \\
&:= \frac{1}{N} \left(\sum_{i=1}^m v'_i d_i + \sum_{i=m+1}^N (1 - v'_i) d_i \right) \\
&> \min_{\substack{v_i \in [0,1], i=1,\dots,N, \\ 1/N \sum_{i=1}^N v_i \in [\omega_1, \omega_2]}} \frac{1}{N} \left(\sum_{i=1}^m v_i d_i + \sum_{i=m+1}^N (1 - v_i) d_i \right) = \delta_{\min},
\end{aligned}$$

where the strict inequality follows from that we can always find $\{v_i\}_{i=1}^N$ such that $\sum_{i=1}^m v'_i d_i + \sum_{i=m+1}^N (1 - v'_i) d_i > \sum_{i=1}^m v_i d_i + \sum_{i=m+1}^N (1 - v_i) d_i$. In fact, for $(p, \delta) \in \mathcal{V}_2$, we find that v'_i s satisfy $v'_i \in [0, 1], i = 1, \dots, N$, $1/N \sum_{i=1}^N v'_i \in [\omega_1, \omega_2]$, and there exists $i \in \{m+1, \dots, N\}$ such that $v'_i < 1$. Without loss of generality, we assume $v'_i = 1$ for $i = m+1, \dots, k$ and $v'_i < 1$ for $i = k+1, \dots, N$. If $1/N \sum_{i=1}^N v'_i \leq \frac{N-m}{N}$, let $v_i = 0$ for $i = 1, \dots, m$ and $v_i = 1$ for $i = m+1, \dots, N$. Otherwise, let $v_i = v'_i - \Delta_i \geq 0$ for $i = 1, \dots, m$ and $v_i = 1$ for $i = m+1, \dots, N$, where $\Delta_i \geq 0$ and $\sum_{i=1}^m \Delta_i = \sum_{i=k+1}^N (1 - v'_i)$. We find $\{v_i\}_{i=1}^N$ satisfies that $v_i \in [0, 1], i = 1, \dots, N$, $1/N \sum_{i=1}^N v_i \in [\omega_1, \omega_2]$ and such $\{v_i\}_{i=1}^N$ must exists due to that $\frac{N-m}{N} \leq \omega_2$. Under the above construction of $\{v_i\}_{i=1}^N$, we have $\sum_{i=1}^m v_i d_i + \sum_{i=m+1}^N (1 - v_i) d_i < \sum_{i=1}^m v'_i d_i + \sum_{i=m+1}^N (1 - v'_i) d_i$. The above strict inequality contradicts the assumption $\delta_0 = \delta_{\min}$ and thus we have that for every $(p, \delta) \in \mathcal{V}_2$, $\delta < 0$. Thus, $\mathbb{Q}|_Y \in \mathbb{B}_\delta \left(\sum_{i=1}^N p_i I_{\widehat{y}_i} \right)$ is an empty set for every $(p, \delta) \in \mathcal{V}_2$.

Combining the above two cases, we complete the proof of claim and thus complete the proof of Theorem 3. $\square$

Proof of Proposition 4. To obtain the sufficient and necessary feasible condition of problem (9), we only need to find the sufficient and necessary feasible condition of problem (10) and thus only need to find the sufficient and necessary condition such that the ambiguity set $\mathcal{K}(\mathcal{N}_\gamma(x_0), \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}; \mathcal{N}_\omega))$ is not empty. Note that

$$\mathcal{K}(\mathcal{N}_\gamma(x_0), \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}; \mathcal{N}_\omega)) = \bigcup_{(p, \delta) \in \mathcal{V}} \mathbb{B}_\delta \left(\sum_{i=1}^N p_i I_{\widehat{y}_i} \right)$$

from the proof of Theorem 6. Thus, $\mathcal{K}(\mathcal{N}_\gamma(x_0), \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}; \mathcal{N}_\omega))$ is not empty if and only if there exists $(p, \delta) \in \mathcal{V}$ such that $\delta = \delta_0 - \sum_{i=1}^m p_i d_i \geq 0$. One can verify that there exists $(p, \delta) \in \mathcal{V}$ such that $\delta \geq 0$ is

equivalent to

$$\max_{\substack{p_i \in [0, \frac{1}{N\varepsilon}], i=1, \dots, N, \\ \sum_{i=1}^N p_i = 1}} \delta_0 - \sum_{i=1}^m p_i d_i \geq 0,$$

and thus equivalent to

$$\begin{aligned} \delta_0 &\geq \min_{\substack{p_i \in [0, \frac{1}{N\varepsilon}], i=1, \dots, N, \\ \sum_{i=1}^N p_i = 1}} \sum_{i=1}^m p_i d_i \\ &= \min_{\substack{v_i \in [0, 1], i=1, \dots, N, \\ \sum_{i=1}^N v_i = N\varepsilon}} \frac{1}{N\varepsilon} \sum_{i=1}^m v_i d_i = \delta_{\min}. \end{aligned}$$

□

Proof of Theorem 6. For problem (9), note that $\mathbb{Q}$ is the joint distribution of (X, Y) and the objective function is only related to Y . Then we can reformulate the inner problem of problem (9) as

$$\sup_{\mathbb{Q}|_Y \in \mathcal{K}(\mathcal{N}_Y(x_0), \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}; \mathcal{N}_\omega))} \rho_{\mathbb{Q}|_Y} [\ell(Y, \alpha)], \quad (32)$$

where $\mathbb{Q}|_Y$ is the marginal distribution of Y and $\mathcal{K}(\mathcal{N}_Y(x_0), \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}; \mathcal{N}_\omega))$ is the marginal distribution set on $\mathcal{Y}$, which is defined as

$$\begin{aligned} \mathcal{K}(\mathcal{N}_Y(x_0), \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}; \mathcal{N}_\omega)) &= \left\{ \mathbb{Q}|_Y : \mathbb{Q}|_Y \text{ is the marginal distribution of } \mathbb{Q} \text{ on } \mathcal{Y}, \right. \\ &\quad \left. \text{where } \mathbb{Q} \in \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}; \mathcal{N}_\omega) \text{ satisfies that } \mathbb{Q}(X \in \mathcal{N}_Y(x_0)) = 1 \right\}. \end{aligned}$$

Thus, we only need to prove that

$$\mathcal{K}(\mathcal{N}_Y(x_0), \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}; \mathcal{N}_\omega)) = \bigcup_{(p, \delta) \in \mathcal{V}} \mathbb{B}_\delta \left(\sum_{i=1}^N p_i \mathbb{I}_{\widehat{y}_i} \right).$$

For any $\mathbb{Q} \in \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}; \mathcal{N}_\omega)$, based on definition of the trimming set and the aforementioned notifications, we can find that there exists p satisfies that given $\varepsilon = \min_{\beta \in \mathcal{N}_\omega} \beta$, $p_i \in [0, \frac{1}{N\varepsilon}]$, $i = 1, \dots, N$, $\sum_{i=1}^N p_i = 1$, such that $\mathcal{W} \left(\sum_{i=1}^N p_i \mathbb{I}_{(\widehat{x}_i, \widehat{y}_i)}, \mathbb{Q} \right) \leq \delta_0$. Furthermore, if $\mathbb{Q}(X \in \mathcal{N}_Y(x_0)) = 1$, assume that the optimal transport between $\mathbb{Q}$ and $\sum_{i=1}^N p_i \mathbb{I}_{(\widehat{x}_i, \widehat{y}_i)}$ is π . Then we can decompose $\mathbb{Q}$ as $\mathbb{Q} = \sum_{i=1}^N p_i \mathbb{Q}^i$, where $\mathbb{Q}^i$ is transferred

by $I_{(\widehat{x}_i, \widehat{y}_i)}$ under the optimal transport π . Similar to the proof the Theorem 3, we have

$$\begin{aligned}
\mathcal{W}\left(\sum_{i=1}^N p_i I_{(\widehat{x}_i, \widehat{y}_i)}, \mathbb{Q}\right) &= \sum_{i=1}^m p_i \int \mathbb{D}_X(\widehat{x}_i, x) + \mathbb{D}_Y(\widehat{y}_i, y) d\mathbb{Q}_i(x, y) \\
&\quad + \sum_{i=m+1}^N p_i \int \mathbb{D}_X(\widehat{x}_i, x) + \mathbb{D}_Y(\widehat{y}_i, y) d\mathbb{Q}_i(x, y) \\
&\geq \sum_{i=1}^m p_i \int d_i + \mathbb{D}_Y(\widehat{y}_i, y) d\mathbb{Q}_i(x, y) \\
&\quad + \sum_{i=m+1}^N p_i \int \mathbb{D}_X(\widehat{x}_i, \widehat{x}_i) + \mathbb{D}_Y(\widehat{y}_i, y) d\mathbb{Q}_i(x, y) \\
&= \sum_{i=1}^N p_i \int \mathbb{D}_Y(\widehat{y}_i, y) d\mathbb{Q}_i|_Y(y) + \sum_{i=1}^m p_i d_i \\
&\geq \mathcal{W}\left(\sum_{i=1}^N p_i I_{\widehat{y}_i}, \mathbb{Q}|_Y\right) + \sum_{i=1}^m p_i d_i.
\end{aligned}$$

Rearranging the above inequality, we find that the marginal distribution on $\mathcal{Y}$ of $\mathbb{Q}$ satisfies that

$$\mathcal{W}\left(\sum_{i=1}^N p_i I_{\widehat{y}_i}, \mathbb{Q}|_Y\right) \leq \mathcal{W}\left(\sum_{i=1}^N p_i I_{(\widehat{x}_i, \widehat{y}_i)}, \mathbb{Q}\right) - \sum_{i=1}^m p_i d_i \leq \delta.$$

And thus,

$$\mathcal{K}(\mathcal{N}_Y(x_0), \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}; \mathcal{N}_\omega)) \subset \bigcup_{(p, \delta) \in \mathcal{V}} \mathbb{B}_\delta\left(\sum_{i=1}^N p_i I_{\widehat{y}_i}\right).$$

Next, we prove the inverse containment. For any

$$\mathbb{Q}|_Y \in \bigcup_{(p, \delta) \in \mathcal{V}} \mathbb{B}_\delta\left(\sum_{i=1}^N p_i I_{\widehat{y}_i}\right),$$

there exists $(p, \delta) \in \mathcal{V}$ such that $\mathbb{Q}|_Y \in \mathbb{B}_\delta\left(\sum_{i=1}^N p_i I_{\widehat{y}_i}\right)$. Assume the optimal transport between $\mathbb{Q}|_Y$ and $\sum_{i=1}^N p_i I_{\widehat{y}_i}$ is $\pi|_Y$. Then we can decompose it as $\mathbb{Q}|_Y = \sum_{i=1}^N p_i \mathbb{Q}^i|_Y$, where $\mathbb{Q}^i|_Y$ is transferred by $I_{\widehat{y}_i}$ under the optimal transport $\pi|_Y$. Then we can construct a distribution $\widetilde{\mathbb{Q}} = \sum_{i=1}^m p_i I_{\widehat{x}_i^p} \times \mathbb{Q}^i|_Y + \sum_{i=m+1}^N p_i I_{\widehat{x}_i} \times \mathbb{Q}^i|_Y$

and verify that the marginal distribution on $\mathcal{Y}$ of $\tilde{\mathbb{Q}}$ is $\mathbb{Q}|_Y$. Furthermore, we have

$$\begin{aligned}
\mathcal{W}\left(\sum_{i=1}^N p_i \mathbb{I}_{(\hat{x}_i, \hat{y}_i)}, \tilde{\mathbb{Q}}\right) &\leq \sum_{i=1}^m p_i \int \mathbb{D}_{\mathcal{X}}(\hat{x}_i, x) + \mathbb{D}_{\mathcal{Y}}(\hat{y}_i, y) d\mathbb{I}_{\hat{x}_i^p}(x) \times \mathbb{Q}^i|_Y(y) \\
&\quad + \sum_{i=m+1}^N p_i \int \mathbb{D}_{\mathcal{X}}(\hat{x}_i, x) + \mathbb{D}_{\mathcal{Y}}(\hat{y}_i, y) d\mathbb{I}_{\hat{x}_i}(x) \times \mathbb{Q}^i|_Y(y) \\
&= \sum_{i=1}^N p_i \int \mathbb{D}_{\mathcal{Y}}(\hat{y}_i, y) d\mathbb{Q}^i|_Y(y) + \sum_{i=1}^m p_i d_i \\
&= \mathcal{W}\left(\sum_{i=1}^N p_i \mathbb{I}_{\hat{y}_i}, \mathbb{Q}|_Y\right) + \sum_{i=1}^m p_i d_i \\
&\leq \delta_0.
\end{aligned}$$

From the above inequality and the construction of $\tilde{\mathbb{Q}}$, we have $\tilde{\mathbb{Q}} \in \mathbb{B}_{\delta_0}(\hat{\mathbb{P}}; \mathcal{N}_{\omega})$ and $\mathbb{Q}(X \in \mathcal{N}_{\gamma}(x_0)) = 1$. Then we have

$$\bigcup_{(p, \delta) \in \mathcal{V}} \mathbb{B}_{\delta}\left(\sum_{i=1}^N p_i \mathbb{I}_{\hat{y}_i}\right) \subset \mathcal{K}(\mathcal{N}_{\gamma}(x_0), \mathbb{B}_{\delta_0}(\hat{\mathbb{P}}; \mathcal{N}_{\omega})),$$

and thus

$$\mathcal{K}(\mathcal{N}_{\gamma}(x_0), \mathbb{B}_{\delta_0}(\hat{\mathbb{P}}; \mathcal{N}_{\omega})) = \bigcup_{(p, \delta) \in \mathcal{V}} \mathbb{B}_{\delta}\left(\sum_{i=1}^N p_i \mathbb{I}_{\hat{y}_i}\right).$$

□

EC.2.2. Proofs of Section 3

Proof of Proposition 7. Suppose $\rho_{\mathbb{Q}|_Y}$ is a law-invariant risk measure that is concave in distribution, then $g(\alpha, (p, \delta))$ is concave in (p, δ) . In fact, one can verify that

$$\begin{aligned}
g(\alpha, (p, \delta)) &= \sup_{\mathbb{Q}|_Y} \left\{ \rho_{\mathbb{Q}|_Y}[\ell(Y, \alpha)] + \sigma(\mathbb{Q}|_Y, p, \delta) \mathbb{B}_{\delta}\left(\sum_{i=1}^N p_i \mathbb{I}_{\hat{y}_i}\right) \right\} \\
&:= \sup_{\mathbb{Q}|_Y} H(\mathbb{Q}|_Y, p, \delta),
\end{aligned}$$

where $\sigma(\mathbb{Q}|_Y, p, \delta) \mathbb{B}_{\delta}(\sum_{i=1}^N p_i \mathbb{I}_{\hat{y}_i})$ is the indicator function of $\mathbb{B}_{\delta}(\sum_{i=1}^N p_i \mathbb{I}_{\hat{y}_i})$ and $\mathbb{B}_{\delta}(\sum_{i=1}^N p_i \mathbb{I}_{\hat{y}_i})$ is a convex set of $(\mathbb{Q}|_Y, p, \delta)$. Thus, we find that $H(\mathbb{Q}|_Y, p, \delta)$ is jointly concave in $(\mathbb{Q}|_Y, p, \delta)$ and then $g(\alpha, (p, \delta))$ is jointly concave in (p, δ) by taking the supremum of $H(\mathbb{Q}|_Y, p, \delta)$ in $\mathbb{Q}|_Y$. Moreover, suppose $\rho_{\mathbb{Q}|_Y} \circ \ell$ is convex in α . We find that $g(\alpha, (p, \delta)) = \sup_{\mathbb{Q}|_Y \in \mathbb{B}_{\delta}(\sum_{i=1}^N p_i \mathbb{I}_{\hat{y}_i})} \rho_{\mathbb{Q}|_Y}[\ell(Y, \alpha)]$ is a supremum of convex functions with respect to α and thus $g(\alpha, (p, \delta))$ is convex in α . □

Proof of Lemma 9. We begin by introducing an auxiliary lemma that will be used throughout the proof.

Lemma 28. Assume $Y_n \in L^q \sim \mathbb{Q}_n$ for $n \in \mathbb{N}$ and $Y \in L^q \sim \mathbb{Q}$ such that $\mathbb{E} [\mathbb{D}_{\mathcal{Y}}(Y_n, Y)] = \mathcal{W}(\mathbb{Q}_n, \mathbb{Q}) \rightarrow 0$ as $n \rightarrow \infty$, where $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}^q$ for any $y_1, y_2 \in \mathcal{Y}$ with $q \in [1, \infty)$. If ℓ is continuous and satisfies that there exists some $C > 0$ such that $|\ell(x)| \leq C(1 + |x|^q)$, given $\alpha \in \mathcal{A}$, $\mathbb{E} [\ell(Y_n^\top \alpha)] \rightarrow \mathbb{E} [\ell(Y^\top \alpha)]$ as $n \rightarrow \infty$.

Proof of Lemma 28. In the following proof, denote $\mathcal{W}(\mathbb{Q}_n, \mathbb{Q})$ as $\mathcal{W}_q(\mathbb{Q}_n, \mathbb{Q})$, $\|\cdot\|_{\mathcal{Y}}$ as $\|\cdot\|$ and $\|\cdot\|_*$ as the dual norm of $\|\cdot\|$. We first prove that $\{\|Y_n\|^q\}_n$ is uniformly integrable, that is $\lim_{M \rightarrow \infty} \sup_n \mathbb{E} [\|Y_n\|^q \mathbf{I}_{\{\|Y_n\|^q > M\}}] = 0$. From the assumption that $\mathcal{W}_q(\mathbb{Q}_n, \mathbb{Q}) \rightarrow 0$ as $n \rightarrow \infty$, we have these two facts: (i) $\mathbb{Q}_n \implies \mathbb{Q}$ (weak convergence); (ii) $\mathbb{E} [\|Y_n\|^q] \rightarrow \mathbb{E} [\|Y\|^q]$ as $n \rightarrow \infty$. One can verify that $\mathcal{W}_1(\mathbb{Q}_n, \mathbb{Q}) \leq \mathcal{W}_q(\mathbb{Q}_n, \mathbb{Q})$ and thus $\mathcal{W}_1(\mathbb{Q}_n, \mathbb{Q}) \rightarrow 0$ as $n \rightarrow \infty$. Using the Kantorovich–Rubinstein duality for $\mathcal{W}_1$, for any 1-Lipschitz continuous function f , we have $\mathbb{E}_{\mathbb{Q}_n} [f(Y_n)] \rightarrow \mathbb{E}_{\mathbb{Q}} [f(Y)]$. This implies (i). To see (ii), using triangle inequality, we have

$$\left| (\mathbb{E} [\|Y_n\|^q])^{1/q} - (\mathbb{E} [\|Y\|^q])^{1/q} \right| \leq (\mathbb{E} [\|Y_n - Y\|^q])^{1/q} = \mathcal{W}_q(\mathbb{Q}_n, \mathbb{Q}) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Hence, we have $\mathbb{E} [\|Y_n\|^q] \rightarrow \mathbb{E} [\|Y\|^q]$ as $n \rightarrow \infty$ and conclude the proof of (ii). Let $g(y) = \|y\|^q$ and $g_M(y) = \min\{g(y), M\}$ with $M > 0$. Due to that g_M is bounded and continuous, by (i), we have $\mathbb{E} [g_M(Y_n)] \rightarrow \mathbb{E} [g_M(Y)]$ as $n \rightarrow \infty$. Let $T_n(M) = \mathbb{E} [\|Y_n\|^q \mathbf{I}_{\{\|Y_n\|^q > M\}}]$ with $M > 0$. Then $T_n(M) = \mathbb{E} [g(Y_n)] - \mathbb{E} [g_M(Y_n)] + M \mathbb{E} [\mathbf{I}_{\{g(Y_n) > M\}}]$. If

$$\sup_n T_n(M) \rightarrow 0 \text{ as } M \rightarrow \infty, \tag{33}$$

by the definition of uniformly integrable, we conclude that $\{\|Y_n\|^q\}_n$ is uniformly integrable. To prove (33), we only need to prove the following claim.

Claim (i) and (ii) imply that $\sup_n \mathbb{E} [g(Y_n)] - \mathbb{E} [g_M(Y_n)] \rightarrow 0$ as $M \rightarrow \infty$.

In fact, we first note that $M \mathbf{I}_{\{g > M\}} \leq 2(g - g_{\frac{M}{2}})$. If the claim holds, we have

$$\begin{aligned} \sup_n T_n(M) &= \sup_n \left\{ \mathbb{E} [g(Y_n)] - \mathbb{E} [g_M(Y_n)] + M \mathbb{E} [\mathbf{I}_{\{g(Y_n) > M\}}] \right\} \\ &\leq \sup_n \left\{ \mathbb{E} [g(Y_n)] - \mathbb{E} [g_M(Y_n)] + 2 \left(\mathbb{E} [g(Y_n)] - \mathbb{E} \left[g_{\frac{M}{2}}(Y_n) \right] \right) \right\} \\ &\leq \sup_n \left\{ \mathbb{E} [g(Y_n)] - \mathbb{E} [g_M(Y_n)] \right\} + \sup_n \left\{ 2 \left(\mathbb{E} [g(Y_n)] - \mathbb{E} \left[g_{\frac{M}{2}}(Y_n) \right] \right) \right\} \\ &\rightarrow 0 \text{ as } M \rightarrow \infty. \end{aligned}$$

Thus, we have $\{\|Y_n\|^q\}_n$ is uniformly integrable. Next, we prove the claim. For any $n \in \mathbb{N}$ and $M > 0$, we have

$$\begin{aligned} \mathbb{E} [g(Y_n)] - \mathbb{E} [g_M(Y_n)] &\leq |\mathbb{E} [g(Y_n)] - \mathbb{E} [g_M(Y_n)]| \\ &\leq |\mathbb{E} [g(Y_n)] - \mathbb{E} [g(Y)]| + |\mathbb{E} [g(Y)] - \mathbb{E} [g_M(Y)]| + |\mathbb{E} [g_M(Y)] - \mathbb{E} [g_M(Y_n)]|. \end{aligned}$$

By (ii), given $\varepsilon > 0$, there exists N_1 such that for any $n \geq N_1$, we have $|\mathbb{E}[g(Y_n)] - \mathbb{E}[g(Y)]| < \varepsilon/3$. Using (i) and the fact that g_M is bounded and continuous, given $\varepsilon > 0$, for any $M > 0$, there exists $N_2(M)$ such that for any $n \geq N_2(M)$, we have $|\mathbb{E}[g_M(Y)] - \mathbb{E}[g_M(Y_n)]| < \varepsilon/3$. Note that $g_M \nearrow g$ as $M \rightarrow \infty$ and $g \in L^1$ ($Y_n \in L^q$). By monotone convergence theorem, we have $\mathbb{E}[g(Y)] - \mathbb{E}[g_M(Y)] \searrow 0$ as $M \rightarrow \infty$. Thus, given $\varepsilon > 0$, there exists $M_0 > 0$ such that for any $M \geq M_0$, we have $|\mathbb{E}[g(Y)] - \mathbb{E}[g_M(Y)]| < \varepsilon/3$. Combining the above cases, we have the following result. Fix any $M \geq M_0$. Let $N = \max\{N_1, N_2(M)\}$. Then for any $n \geq N$, we have $\mathbb{E}[g(Y_n)] - \mathbb{E}[g_M(Y_n)] < \varepsilon$. Moreover, for any $1 \leq n \leq N$, we have $\mathbb{E}[g(Y_n)] - \mathbb{E}[g_M(Y_n)] \searrow 0$ as $M \rightarrow \infty$ by monotone convergence theorem. Thus, given $\varepsilon > 0$, there exists M_1 such that for $M \geq M_1$, we have $\mathbb{E}[g(Y_n)] - \mathbb{E}[g_M(Y_n)] \leq \varepsilon$ with $1 \leq n \leq N$. Then, for $M \geq \max\{M_0, M_1\}$, we have

$$\begin{aligned} \sup_n \mathbb{E}[g(Y_n)] - \mathbb{E}[g_M(Y_n)] &= \max\left\{ \max_{1 \leq n \leq N} \mathbb{E}[g(Y_n)] - \mathbb{E}[g_M(Y_n)], \sup_{n \geq N} \mathbb{E}[g(Y_n)] - \mathbb{E}[g_M(Y_n)] \right\} \\ &\rightarrow 0 \text{ as } M \rightarrow \infty. \end{aligned}$$

Thus, we conclude the proof of that $\{\|Y_n\|^q\}_n$ is uniformly integrable. In the following, we are going to prove the following two facts.

(a) Given $\alpha \in \mathcal{A}$, $\{\ell(Y_n^\top \alpha)\}_n$ is uniformly integrable.

(b) Given $\alpha \in \mathcal{A}$, $\{\ell(Y_n^\top \alpha)\}_n \rightarrow \ell(Y^\top \alpha)$ in probability.

Because $|\ell(x)| \leq C(1 + |x|^q)$ with $C > 0$, we have $|\ell(Y_n^\top \alpha)| \leq C(1 + |Y_n^\top \alpha|^q) \leq C(1 + \|\alpha\|_*^q \|Y_n\|^q)$. Then $|\ell(Y_n^\top \alpha)| \mathbf{I}_{\{|\ell(Y_n^\top \alpha)| > M\}} \leq C(1 + \|\alpha\|_*^q \|Y_n\|^q) \mathbf{I}_{\{C(1 + \|\alpha\|_*^q \|Y_n\|^q) > M\}}$. Thus,

$$\begin{aligned} \sup_n \mathbb{E}[|\ell(Y_n^\top \alpha)| \mathbf{I}_{\{|\ell(Y_n^\top \alpha)| > M\}}] &\leq \sup_n \mathbb{E}[C(1 + \|\alpha\|_*^q \|Y_n\|^q) \mathbf{I}_{\{C(1 + \|\alpha\|_*^q \|Y_n\|^q) > M\}}] \\ &\leq \sup_n C \mathbb{E}[C \mathbf{I}_{\{(1 + \|\alpha\|_*^q \|Y_n\|^q) > M\}}] + \sup_n C \|\alpha\|_*^q \mathbb{E}\left[\|Y_n\|^q \mathbf{I}_{\{\|Y_n\|^q > \frac{M-C}{C\|\alpha\|_*^q}\}}\right] \\ &\rightarrow 0 \text{ as } M \rightarrow \infty. \end{aligned}$$

We complete the proof of (a). To see (b), due to (ii), we find that $Y_n \rightarrow Y$ in L^q . Then $Y_n^\top \alpha \rightarrow Y^\top \alpha$ in L^q and thus in probability. Because of ℓ is continuous, we obtain $\ell(Y_n^\top \alpha) \rightarrow \ell(Y^\top \alpha)$ in probability. Combining (a) and (b), using Vitali's convergence theorem, we have $\mathbb{E}[|\ell(Y_n^\top \alpha) - \ell(Y^\top \alpha)|] \rightarrow 0$ as $n \rightarrow \infty$, and thus $\mathbb{E}[\ell(Y_n^\top \alpha)] \rightarrow \mathbb{E}[\ell(Y^\top \alpha)]$ as $n \rightarrow \infty$. $\square$

We are now ready to complete the proof of Lemma 9.

Continued proof of Lemma 9 Assume $Y_n \in L^q \sim \mathbb{Q}_n$ for $n \in \mathbb{N}$ and $Y \in L^q \sim \mathbb{Q}$ such that $\mathbb{E}[\mathbb{D}_{\mathcal{Y}}(Y_n, Y)] = \mathcal{W}(\mathbb{Q}_n, \mathbb{Q}) \rightarrow 0$ as $n \rightarrow \infty$, where $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}^q$ for any $y_1, y_2 \in \mathcal{Y}$ with $q \in [1, \infty)$. We consider the following risk measure case by case.

(i) $\rho(Y) = \mathbb{E}[\ell(Y^\top \alpha)]$. One can verify that ℓ in Propositions 10 and 11 is continuous and satisfies that there exists $C > 0$ such that $|\ell(x)| \leq C(1 + |x|^q)$. Applying Lemma 28, we have $\mathbb{E}[\ell(Y_n^\top \alpha)] \rightarrow \mathbb{E}[\ell(Y^\top \alpha)]$ as $n \rightarrow \infty$ if $\mathcal{W}(\mathbb{Q}_n, \mathbb{Q}) \rightarrow 0$ as $n \rightarrow \infty$. We conclude that ρ is OT-continuity.

- (ii) $\rho(Y) = \inf_{t \in \mathbb{R}^k} \mathbb{E} [\ell(Y^\top \alpha, t)]$. Because $\ell(z, t) : \mathbb{R} \times \mathbb{R}^k \rightarrow \mathbb{R}$ in Propositions 13 and 14 is convex and level-bounded in t for all $z \in \mathbb{R}$, the infimum is attained. Thus for any $n \in \mathbb{N}$, there exists $t_n \in \mathbb{R}$ such that

$$\left| \inf_{t \in \mathbb{R}^k} \mathbb{E} [\ell(Y_n^\top \alpha, t)] - \inf_{t \in \mathbb{R}^k} \mathbb{E} [\ell(Y^\top \alpha, t)] \right| \leq \left| \mathbb{E} [\ell(Y_n^\top \alpha, t_n)] - \mathbb{E} [\ell(Y^\top \alpha, t_n)] \right|. \quad (34)$$

One can verify that ℓ in Propositions 13 and 14 is continuous and satisfies that there exists $C > 0$ such that $|\ell(x)| \leq C(1 + |x|^q)$. Applying Lemma 28, we have $\mathbb{E} [\ell(Y_n^\top \alpha, t_n)] \rightarrow \mathbb{E} [\ell(Y^\top \alpha, t_n)]$ as $n \rightarrow \infty$ if $\mathcal{W}(\mathbb{Q}_n, \mathbb{Q}) \rightarrow 0$ as $n \rightarrow \infty$. Thus, by inequality (34), we have $\inf_{t \in \mathbb{R}^k} \mathbb{E} [\ell(Y_n^\top \alpha, t)] \rightarrow \inf_{t \in \mathbb{R}^k} \mathbb{E} [\ell(Y^\top \alpha, t)]$ as $n \rightarrow \infty$. We conclude that ρ is OT-continuity.

- (iii) $\rho(Y) = \inf_{t \in \mathbb{R}} \left\{ t + (\mathbb{E} [\ell^q(Y^\top \alpha, t)])^{1/q} \right\}$. Because $t + \ell(z, t) : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ in Propositions 17 and 18 is convex and level-bounded in t for all $z \in \mathbb{R}$, the infimum is attained. This applies that there exists $t_n \in \mathbb{R}$ such that

$$\left| \inf_{t \in \mathbb{R}^k} \left\{ t + \mathbb{E} [\ell^q(Y_n^\top \alpha, t)]^{1/q} \right\} - \inf_{t \in \mathbb{R}^k} \left\{ t + \mathbb{E} [\ell^q(Y^\top \alpha, t)]^{1/q} \right\} \right| \quad (35)$$

$$\leq \left| \mathbb{E} [\ell^q(Y_n^\top \alpha, t_n)]^{1/q} - \mathbb{E} [\ell^q(Y^\top \alpha, t_n)]^{1/q} \right|. \quad (36)$$

One can verify that ℓ^q in Propositions 13 and 14 is continuous and satisfies that there exists $C > 0$ such that $|\ell^q(x)| \leq C(1 + |x|^q)$. Applying Lemma 28, we have $\mathbb{E} [\ell^q(Y_n^\top \alpha, t_n)] \rightarrow \mathbb{E} [\ell^q(Y^\top \alpha, t_n)]$ as $n \rightarrow \infty$ if $\mathcal{W}(\mathbb{Q}_n, \mathbb{Q}) \rightarrow 0$ as $n \rightarrow \infty$. Thus, by inequality (35), we have $\inf_{t \in \mathbb{R}^k} \left\{ t + \mathbb{E} [\ell^q(Y_n^\top \alpha, t)]^{1/q} \right\} \rightarrow \inf_{t \in \mathbb{R}^k} \left\{ t + \mathbb{E} [\ell^q(Y^\top \alpha, t)]^{1/q} \right\}$ as $n \rightarrow \infty$. We conclude that ρ is OT-continuity.

- (iv) $\rho(Y) = \inf \{ \kappa \in \mathbb{R} : \mathbb{E} [u(-Y^\top \alpha - \kappa)] \leq l \}$, where u is convex and increasing and l is a fixed constant in the interior of the range of u . One can verify that u in Propositions 19 and 21, and Remark 20 is continuous and satisfies that there exists $C > 0$ such that $|u(x)| \leq C(1 + |x|^q)$. Applying Lemma 28, for every $\kappa \in \mathbb{R}$, we have $\mathbb{E} [u(-Y_n^\top \alpha - \kappa)] \rightarrow \mathbb{E} [u(-Y^\top \alpha - \kappa)]$ as $n \rightarrow \infty$ if $\mathcal{W}(\mathbb{Q}_n, \mathbb{Q}) \rightarrow 0$ as $n \rightarrow \infty$. Because l is a fixed constant in the interior of the range of u , the sublevel set $\{ \kappa : \mathbb{E} [u(-Y^\top \alpha - \kappa)] \leq l \}$ is nonempty, closed and bounded-below interval, that is $\{ \kappa : \mathbb{E} [u(-Y^\top \alpha - \kappa)] \leq l \} = [\rho(Y), \infty) = [\rho(\mathbb{Q}), \infty)$, where $\rho(Y) = \rho(\mathbb{Q}) \in \mathbb{R}$, and for any $\varepsilon > 0$, we have $\mathbb{E} [u(-Y^\top \alpha - \rho(\mathbb{Q}) - \varepsilon)] < l$. Moreover, by the definition of shortfall risk measure, we have $\mathbb{E} [u(-Y^\top \alpha - \rho(\mathbb{Q}) + \varepsilon)] > l$ for any $\varepsilon > 0$. Because of that $\mathbb{E} [u(-Y_n^\top \alpha - \kappa)] \rightarrow \mathbb{E} [u(-Y^\top \alpha - \kappa)]$ as $n \rightarrow \infty$, for sufficiently large n , we have $\mathbb{E} [u(-Y_n^\top \alpha - \rho(\mathbb{Q}) - \varepsilon)] < l$ and $\mathbb{E} [u(-Y_n^\top \alpha - \rho(\mathbb{Q}) + \varepsilon)] > l$. Thus, $\rho(\mathbb{Q}_n) \in (\rho(\mathbb{Q}) - \varepsilon, \rho(\mathbb{Q}) + \varepsilon)$ for sufficiently large n . Letting ε decrease to 0 yields $\rho(\mathbb{Q}_n) \rightarrow \rho(\mathbb{Q})$ as $n \rightarrow \infty$. We conclude that ρ is OT-continuity.

- (v) $\rho(Y) = \rho^h [\ell(Y^\top \alpha)]$, where h is a convex distortion function. Let $\ell : \mathbb{R} \rightarrow \mathbb{R}$ be a convex and

Lipschitz continuous function. By Hölder inequality and the Lipschitz continuous of ℓ , we have

$$\begin{aligned} |\rho^h [\ell(Y_n^\top \alpha)] - \rho^h [\ell(Y^\top \alpha)]| &\leq \|h'_-\|_{q^*} \mathbb{E} [|\ell(Y_n^\top \alpha) - \ell(Y^\top \alpha)|^q]^{1/q} \\ &\leq \|h'_-\|_{q^*} \text{Lip}(\ell) \|\alpha\|_* \mathbb{E} [\|Y_n - Y\|^q]^{1/q} \\ &= \|h'_-\|_{q^*} \text{Lip}(\ell) \|\alpha\|_* \mathcal{W}(\mathbb{Q}_n, \mathbb{Q}). \end{aligned}$$

Following the above inequality, we have $\mathbb{E} [\ell(Y_n^\top \alpha)] \rightarrow \mathbb{E} [\ell(Y^\top \alpha)]$ as $n \rightarrow \infty$, and thus ρ is OT-continuity.

□

EC.2.2.1. Proofs of Section 3.1

Proof of Proposition 10. (i) Let $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}$, $\forall y_1, y_2 \in \mathcal{Y}$ and $\|\cdot\|_*$ be the dual norm of $\|\cdot\|_{\mathcal{Y}}$. Applying the equation (26), we find that the problem (15) can be reformulated as

$$\min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}, \delta \geq 0} \sup_{\mathbb{Q}|_{\mathcal{Y}} \in \mathbb{B}_{\delta} \left(\sum_{i=1}^N p_i \mathbb{I}_{\widehat{\mathcal{Y}}_i} \right)} \mathbb{E}_{\mathbb{Q}|_{\mathcal{Y}}} [\ell(Y^\top \alpha)] = \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}_+} \sum_{i=1}^N p_i \ell(\widehat{\mathcal{Y}}_i^\top \alpha) + \text{Lip}(\ell) \delta \|\alpha\|_*.$$

Let $p' = (p, \delta)$ and

$$f(p') = f(p, \delta) = \sum_{i=1}^N p_i \ell(\widehat{\mathcal{Y}}_i^\top \alpha) + \text{Lip}(\ell) \delta \|\alpha\|_*.$$

Using the Fenchel duality theory, we can reformulate the above inner supremum problem as

$$\begin{aligned} &\sup_{p' \in \mathcal{V}_+} \sum_{i=1}^N p_i \ell(\widehat{\mathcal{Y}}_i^\top \alpha) + \text{Lip}(\ell) \delta \|\alpha\|_* \\ &= \sup_{p'} \left\{ \sum_{i=1}^N p_i \ell(\widehat{\mathcal{Y}}_i^\top \alpha) + \text{Lip}(\ell) \delta \|\alpha\|_* - \sigma(p' | \mathcal{V}_+) \right\} \\ &= \inf_v \{ \sigma^*(v | \mathcal{V}_+) - f_*(v) \}, \end{aligned}$$

where $\sigma^*(v | \mathcal{V}_+)$ is the convex conjugate function of $\sigma(p' | \mathcal{V}_+)$, and $f_*(v)$ is the concave conjugate function of $f(p')$,

$$f_*(v) = \inf_{p' \geq 0} v^\top p' - v(\alpha)^\top p' = \begin{cases} 0 & \text{if } v - v(\alpha) \geq 0, \\ -\infty & \text{else,} \end{cases}$$

with $v(\alpha)^\top = (\ell(\widehat{y}_1^\top \alpha), \dots, \ell(\widehat{y}_N^\top \alpha), \text{Lip}(\ell)\|\alpha\|_*) \in \mathbb{R}^{N+1}$. Thus, we can reformulate the problem (15) as

$$\begin{aligned} & \inf_{\alpha, v} \sigma^*(v \mid \mathcal{V}_+) \\ & \text{s.t. } \alpha \in \mathcal{A}, v - v(\alpha) \geq 0. \end{aligned}$$

(ii) Let $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}^q$ for any $y_1, y_2 \in \mathcal{Y}$ and $q \in (1, \infty)$.

(1) Applying the result in Theorem 24, we find that the problem (15) can be reformulated as

$$\min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}, \delta \geq 0} \sup_{\mathbb{Q} \mid \mathcal{Y} \in \mathbb{B}_\delta \left(\sum_{i=1}^N p_i \mathbb{I}_{\widehat{y}_i} \right)} \mathbb{E}_{\mathbb{Q} \mid \mathcal{Y}} [\ell(Y^\top \alpha)] = \min_{\alpha \in \mathcal{A}} \sup_{p' \in \mathcal{V}_+} \sum_{i=1}^N p_i \ell(\widehat{y}_i^\top \alpha) + C \delta^{\frac{1}{q}} \|\alpha\|_*.$$

Let $g(p') = \sum_{i=1}^N p_i \ell(\widehat{y}_i^\top \alpha) + C \delta^{\frac{1}{q}} \|\alpha\|_*$. Using the Fenchel duality theory, we can reformulate the above inner supremum problem as

$$\sup_{p' \in \mathcal{V}_+} \sum_{i=1}^N p_i \ell(\widehat{y}_i^\top \alpha) + C \delta^{\frac{1}{q}} \|\alpha\|_* = \inf_v \{ \sigma^*(v \mid \mathcal{V}_+) - g_*(v) \},$$

where $\sigma^*(v \mid \mathcal{V}_+)$ is the convex conjugate function of $\sigma(p' \mid \mathcal{V}_+)$, and $g_*(v)$ is the concave conjugate function of $g(p')$, defined as

$$\begin{aligned} g_*(v) &= \inf_{p' \geq 0} v^\top p' - \sum_{i=1}^N p_i \ell(\widehat{y}_i^\top \alpha) - C \delta^{\frac{1}{q}} \|\alpha\|_* \\ &= \begin{cases} -C_q v_{N+1} \left\| \frac{\alpha}{v_{N+1}} \right\|_*^{\frac{q}{q-1}} & \text{if } v - v'(\alpha) \geq 0, \\ -\infty & \text{else,} \end{cases} \end{aligned}$$

with $C_q = \left(q^{\frac{1}{1-q}} - q^{\frac{q}{1-q}} \right) C^{\frac{q}{q-1}}$, and $v'(\alpha)^\top = (\ell(\widehat{y}_1^\top \alpha), \dots, \ell(\widehat{y}_N^\top \alpha), 0) \in \mathbb{R}^{N+1}$. Thus, we can reformulate the problem (15) as

$$\begin{aligned} & \inf_{\alpha, v} \sigma^*(v \mid \mathcal{V}_+) + C_q v_{N+1} \left\| \frac{\alpha}{v_{N+1}} \right\|_*^{\frac{q}{q-1}} \\ & \text{s.t. } \alpha \in \mathcal{A}, v - v'(\alpha) \geq 0, \end{aligned}$$

(2) Applying the result in Theorem 25, we can reformulate the problem (15) as

$$\begin{aligned}
& \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}, \delta \geq 0} \sup_{\mathbb{Q}|Y \in \mathbb{B}_\delta \left(\sum_{i=1}^N p_i I_{\widehat{y}_i} \right)} \mathbb{E}_{\mathbb{Q}|Y} [\ell^q(Y^\top \alpha)] \\
&= \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}_+} \left(\left(\mathbb{E}_{\sum_{i=1}^N p_i I_{\widehat{y}_i}} [\ell^q(Y, \alpha)] \right)^{1/q} + C \delta^{\frac{1}{q}} \|\alpha\|_* \right)^q \\
&= \min_{\alpha \in \mathcal{A}} \left(\sup_{(p, \delta) \in \mathcal{V}_+} \left(\mathbb{E}_{\sum_{i=1}^N p_i I_{\widehat{y}_i}} [\ell^q(Y, \alpha)] \right)^{1/q} + C \delta^{\frac{1}{q}} \|\alpha\|_* \right)^q \\
&= \min_{\alpha \in \mathcal{A}} \left(\sup_{p' \in \mathcal{V}_+} \left(\mathbb{E}_{\sum_{i=1}^N p_i I_{\widehat{y}_i}} [\ell^q(Y, \alpha)] \right)^{1/q} + C \delta^{\frac{1}{q}} \|\alpha\|_* \right)^q.
\end{aligned}$$

Next, we focus on solving the following problem

$$\sup_{p' \in \mathcal{V}_+} \left(\mathbb{E}_{\sum_{i=1}^N p_i I_{\widehat{y}_i}} [\ell^q(Y, \alpha)] \right)^{1/q} + C \delta^{\frac{1}{q}} \|\alpha\|_*.$$

Let $h(p') = \left(\sum_{i=1}^N p_i \ell^q(\widehat{y}_i^\top \alpha) \right)^{\frac{1}{q}} + C \delta^{\frac{1}{q}} \|\alpha\|_*$. Using the Fenchel duality theory, we can reformulate the above inner supremum problem as

$$\sup_{p'} \{h(p') - \sigma(p' | \mathcal{V}_+)\} = \inf_v \{\sigma^*(v | \mathcal{V}_+) - h_*(v)\},$$

where $\sigma^*(v | \mathcal{V}_+)$ is the convex conjugate function of $\sigma(p' | \mathcal{V}_+)$, and $h_*(v)$ is the concave conjugate function of $h(p')$,

$$\begin{aligned}
h_*(v) &= \inf_{p' \geq 0} v^\top p' - h(p') \\
&= \sup -C_q^1 y^{\frac{1}{1-q}} - C_q v_{N+1}^2 \left\| \frac{\alpha}{v_{N+1}^2} \right\|_*^{\frac{q}{q-1}} \\
&\quad \text{s.t. } v = y \cdot v''(\alpha) + v^2, \quad y \geq 0, \quad v^2 \geq 0, \\
&= \sup -C_q^1 y^{\frac{1}{1-q}} - C_q v_{N+1} \left\| \frac{\alpha}{v_{N+1}} \right\|_*^{\frac{q}{q-1}} \\
&\quad \text{s.t. } v - y \cdot v''(\alpha) \geq 0, \quad y \geq 0,
\end{aligned}$$

where $C_q = C^{\frac{q}{q-1}} \left(q^{\frac{1}{1-q}} - q^{\frac{q}{1-q}} \right)$, $C_q^1 = q^{\frac{1}{1-q}} - q^{\frac{q}{1-q}}$, and

$$v''(\alpha)^\top = (\ell^q(\widehat{y}_1^\top \alpha), \dots, \ell^q(\widehat{y}_N^\top \alpha), 0) \in \mathbb{R}^{N+1}.$$

Thus, we have

$$\begin{aligned}
& \min_{\alpha \in \mathcal{A}} \left(\sup_{p' \in \mathcal{V}_+} \left(\mathbb{E}_{\sum_{i=1}^N p_i \mathbb{I}_{\hat{y}_i}} [\ell^q(Y, \alpha)] \right)^{1/q} + C \delta^{\frac{1}{q}} \|\alpha\|_* \right)^q \\
&= \min_{\alpha \in \mathcal{A}} \left(\inf_v \{ \sigma^*(v | \mathcal{V}_+) - h_*(v) \} \right)^q \\
&= \inf_{\alpha \in \mathcal{A}, v} (\sigma^*(v | \mathcal{V}_+) - h_*(v))^q \\
&= \inf_{\alpha, v, y} \left[\sigma^*(v | \mathcal{V}_+) + C_q^1 y^{\frac{1}{1-q}} + C_q v_{N+1} \left\| \frac{\alpha}{v_{N+1}} \right\|_*^{\frac{q}{q-1}} \right]^q \\
&\quad \text{s.t. } \alpha \in \mathcal{A}, v - y \cdot v''(\alpha) \geq 0, y \geq 0.
\end{aligned}$$

Thus, we obtain the desired result. $\square$

The following result, which reduces the worst-case expectation problem via the projection property of Wu et al. (2022), will be instrumental in simplifying subsequent proofs.

Theorem 29. Let $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}^q$ for any $y_1, y_2 \in \mathcal{Y}$, with $q \in (1, \infty]$. Let $1/q + 1/q^* = 1$. Assume $\ell : \mathbb{R} \rightarrow \mathbb{R}$ is convex and satisfies $\ell(x) - \ell(x_0) \leq L|x - x_0|^q + M, x \in \mathbb{R}$ for some $L, M > 0$, and $x_0 \in \mathbb{R}$, for any distribution F_0 , and $\delta \geq 0$. we have

$$\begin{aligned}
& \sup_{\mathbb{Q}|_Y \in \mathbb{B}_{\delta}(F_0)} \mathbb{E}_{\mathbb{Q}|_Y} [\ell(Y^{\top} \alpha)] \\
&= \inf_{\lambda \geq 0} \mathbb{E}_{F_0} \left[\sup_{y \in \mathbb{R}} \{ \ell(y) - \lambda |y - Y^{\top} \alpha|^q \} \right] + \lambda \delta \|\alpha\|_*^q
\end{aligned} \tag{37}$$

$$= \inf_{\eta \geq 0} \mathbb{E}_{F_0} \left[\sup_{z \in \mathbb{R}} \left\{ z \cdot Y^{\top} \alpha - \ell^*(z) + \eta \cdot \frac{|z|^{q^*}}{q^* q^{q^*-1}} \right\} \right] + \delta \frac{\|\alpha\|_*^q}{\eta^{q-1}}. \tag{38}$$

Moreover, we also find that problems (37) and (38) are convex optimization problems.

Proof of Theorem 29. By the projection result in Proposition 1 of Wu et al. (2022), we have

$$\begin{aligned}
\sup_{\mathbb{Q}|_Y \in \mathbb{B}_{\delta}(F_0)} \mathbb{E}_{\mathbb{Q}|_Y} [\ell(Y^{\top} \alpha)] &= \sup_{\{\mathbb{Q}|_{Y^{\top} \alpha} : \mathbb{Q}|_Y \in \mathbb{B}_{\delta}(F_0)\}} \mathbb{E}_{\mathbb{Q}|_{Y^{\top} \alpha}} [\ell(Y^{\top} \alpha)] \\
&= \sup_{\mathbb{Q}|_{Y^{\top} \alpha} \in \mathbb{B}_{\delta \|\alpha\|_*}(F_0|_{Y^{\top} \alpha})} \mathbb{E}_{\mathbb{Q}|_{Y^{\top} \alpha}} [\ell(Y^{\top} \alpha)].
\end{aligned}$$

Using the strong duality result of Theorem 1 in Gao and Kleywegt (2023), we have that

$$\sup_{\mathbb{Q}|_{Y^{\top} \alpha} \in \mathbb{B}_{\delta \|\alpha\|_*}(F_0|_{Y^{\top} \alpha})} \mathbb{E}_{\mathbb{Q}|_{Y^{\top} \alpha}} [\ell(Y^{\top} \alpha)] = \inf_{\lambda \geq 0} \left\{ \mathbb{E}_{F_0} \left[\sup_{y \in \mathbb{R}} \{ \ell(y) - \lambda |y - Y^{\top} \alpha|^q \} \right] + \lambda \delta \|\alpha\|_*^q \right\}.$$

Thus the first equality holds. Now we focus on the second equality. Since ℓ is a real-valued convex function on $\mathbb{R}$, we have $\ell = \ell^{**}$, i.e., $\ell(y) = \sup_{z \in \mathbb{R}} \{ zy - \ell^*(z) \}$ for all $y \in \mathbb{R}$. Using this relation, (37) can be

reformulated as follows.

$$\begin{aligned}
& \inf_{\lambda \geq 0} \mathbb{E}_{F_0} \left[\sup_{y \in \mathbb{R}} \left\{ \ell(y) - \lambda |y - Y^\top \alpha|^q \right\} \right] + \lambda \delta \|\alpha\|_*^q \\
&= \inf_{\lambda \geq 0} \mathbb{E}_{F_0} \left[\sup_{\Delta \in \mathbb{R}} \left\{ \ell(Y^\top \alpha + \Delta) - \lambda |\Delta|^q \right\} \right] + \lambda \delta \|\alpha\|_*^q \\
&= \inf_{\lambda \geq 0} \mathbb{E}_{F_0} \left[\sup_{\Delta \in \mathbb{R}} \left\{ \sup_{z \in \mathbb{R}} \left\{ z(Y^\top \alpha + \Delta) - \ell^*(z) \right\} - \lambda |\Delta|^q \right\} \right] + \lambda \delta \|\alpha\|_*^q \\
&= \inf_{\lambda \geq 0} \mathbb{E}_{F_0} \left[\sup_{z \in \mathbb{R}} \left\{ z \cdot Y^\top \alpha - \ell^*(z) + \sup_{\Delta \in \mathbb{R}} \left\{ z\Delta - \lambda |\Delta|^q \right\} \right\} \right] + \lambda \delta \|\alpha\|_*^q \\
&= \inf_{\lambda \geq 0} \mathbb{E}_{F_0} \left[\sup_{z \in \mathbb{R}} \left\{ z \cdot Y^\top \alpha - \ell^*(z) + \left(\frac{1}{\lambda q} \right)^{1/(q-1)} \left(1 - \frac{1}{q} \right) |z|^{q/(q-1)} \right\} \right] + \lambda \delta \|\alpha\|_*^q \\
&= \inf_{\lambda \geq 0} \mathbb{E}_{F_0} \left[\sup_{z \in \mathbb{R}} \left\{ z \cdot Y^\top \alpha - \ell^*(z) + \left(\frac{1}{\lambda} \right)^{q^*-1} \frac{|z|^{q^*}}{q^* q^{q^*-1}} \right\} \right] + \lambda \delta \|\alpha\|_*^q.
\end{aligned}$$

By setting $\eta = (1/\lambda)^{1/(q-1)}$, the above equation reduces to

$$\inf_{\eta \geq 0} \mathbb{E}_{F_0} \left[\sup_{z \in \mathbb{R}} \left\{ z \cdot Y^\top \alpha - \ell^*(z) + \eta \cdot \frac{|z|^{q^*}}{q^* q^{q^*-1}} \right\} \right] + \delta \cdot \frac{\|\alpha\|_*^q}{\eta^{q-1}}.$$

Hence, we have verified that (37) and (38) are equivalent. Let us now prove that (38) is a convex optimization problem. The first term in (38) is convex since the mapping $\eta \mapsto z \cdot Y^\top \alpha - \ell^*(z) + \eta |z|^q / (qp^{q-1})$ is convex for any fixed $z \in \mathbb{R}$ and any fixed $\alpha \in \mathcal{A}$. Thus, taking the supremum of z is also convex. Also noting that $\|\alpha\|_*^p / \eta^{p-1}$ is a convex function in $\eta \geq 0$. Therefore, we conclude that (37) is a convex optimization problem. This completes the proof. $\square$

Proof of Proposition 11. Applying Theorem 29, we find that problem (15) can be reformulated as

$$\begin{aligned}
& \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}, \delta \geq 0} \sup_{\mathbb{Q}|_Y \in \mathbb{B}_{\delta(p, \varepsilon)} \left(\sum_{i=1}^N p_i \mathbb{I}_{\widehat{y}_i} \right)} \mathbb{E}_{\mathbb{Q}|_Y} [\ell(Y^\top \alpha)] \\
&= \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}_+, \delta \geq 0} \inf_{\eta \geq 0} \sum_{i=1}^N p_i \cdot \sup_{z \in \mathbb{R}} \left\{ z \cdot \widehat{y}_i^\top \alpha - \ell^*(z) + \eta \cdot \frac{|z|^{q^*}}{q^* q^{q^*-1}} \right\} + \delta \frac{\|\alpha\|_*^q}{\eta^{q-1}} \\
&= \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}_+, \delta \geq 0} \inf_{\eta \geq 0} \sum_{i=1}^N p_i \cdot \sup_{z \in \mathbb{R}} \left\{ z \cdot \widehat{y}_i^\top \alpha - \ell^*(z) + \eta \cdot \frac{|z|^{q^*}}{q^* q^{q^*-1}} \right\} + \delta \frac{\|\alpha\|_*^q}{\eta^{q-1}}.
\end{aligned}$$

One can verify that the objective function of the above optimization problem is linear in $(p, \delta) \in \mathcal{V}_+$ and

convex in $\eta \geq 0$, and that $\mathcal{V}_+$ is a convex compact set. Applying min-max theorem in Sion (1958), we have

$$\begin{aligned} & \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}_+} \inf_{\eta \geq 0} \sum_{i=1}^N p_i \cdot \sup_{z \in \mathbb{R}} \left\{ z \cdot \widehat{y}_i^\top \alpha - \ell^*(z) + \eta \cdot \frac{|z|^{q^*}}{q^* q^{q^*-1}} \right\} + \delta \frac{\|\alpha\|_*^q}{\eta^{q-1}} \\ &= \inf_{\alpha \in \mathcal{A}, \eta \geq 0} \sup_{(p, \delta) \in \mathcal{V}_+} \sum_{i=1}^N p_i \cdot \sup_{z \in \mathbb{R}} \left\{ z \cdot \widehat{y}_i^\top \alpha - \ell^*(z) + \eta \cdot \frac{|z|^{q^*}}{q^* q^{q^*-1}} \right\} + \delta \frac{\|\alpha\|_*^q}{\eta^{q-1}}. \end{aligned}$$

Let $f(p, \delta) = \sum_{i=1}^N p_i \cdot \sup_{z \in \mathbb{R}} \left\{ z \cdot \widehat{y}_i^\top \alpha - \ell^*(z) + \eta \cdot \frac{|z|^{q^*}}{q^* q^{q^*-1}} \right\} + \delta \frac{\|\alpha\|_*^q}{\eta^{q-1}}$. Similar to the proof of Proposition 10, using Fenchel duality, we obtain the desired result. $\square$

EC.2.2.2. Proofs of Section 3.2

In the following, we give a minimax lemma which will be used in the proof of propositions in Section 3.2.

Lemma 30. *Let $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}^q$ for any $y_1, y_2 \in \mathcal{Y}$ and $q \in [1, \infty)$, $\delta \geq 0$, and $\mathbb{Q}_0$ be a distribution on $\mathcal{Y}$. We have the following results.*

- (i) *Assume the loss function $\ell(z, t) : \mathbb{R} \times \mathbb{R}^k \rightarrow \mathbb{R}$ is the form $\ell(z, t) = \sum_{q_1 \in \mathcal{I}} C_{q_1} (\bar{\ell}_{q_1}(z, t))^{q_1}$, with $\mathcal{I} \subset [1, q]$ and $C_{q_1} \geq 0$, where $\bar{\ell}_{q_1}(z, t)$ is nonnegative on $\mathbb{R}^{k+1}$, convex and level bounded in $t \in \mathbb{R}^k$ for all $z \in \mathbb{R}$, and Lipschitz continuous in z for all $t \in \mathbb{R}^k$ with a uniform Lipschitz constant $\text{Lip}(\bar{\ell}_{q_1})$, for $q_1 \in [1, q]$. Then we have*

$$\sup_{\mathbb{Q}|_{\mathcal{Y}} \in \mathbb{B}_{\delta}(\mathbb{Q}_0)} \inf_{t \in \mathbb{R}^k} \mathbb{E}_{\mathbb{Q}|_{\mathcal{Y}}} [\ell(Y^\top \alpha, t)] = \inf_{t \in \mathbb{R}^k} \sup_{\mathbb{Q}|_{\mathcal{Y}} \in \mathbb{B}_{\delta}(\mathbb{Q}_0)} \mathbb{E}_{\mathbb{Q}|_{\mathcal{Y}}} [\ell(Y^\top \alpha, t)].$$

- (ii) *Let $k = 1$. Assume the loss function $\ell(z, t) : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ is nonnegative on $\mathbb{R}^2$, convex in $t \in \mathbb{R}$ for all $z \in \mathbb{R}$, and Lipschitz continuous in z for all $t \in \mathbb{R}$ with a uniform Lipschitz constant $\text{Lip}(\ell)$. Moreover, assume $t + \ell(z, t)$ is level bounded in $t \in \mathbb{R}$ for all $z \in \mathbb{R}$. Then we have*

$$\sup_{\mathbb{Q}|_{\mathcal{Y}} \in \mathbb{B}_{\delta}(\mathbb{Q}_0)} \inf_{t \in \mathbb{R}} \left\{ t + (\mathbb{E}_{\mathbb{Q}|_{\mathcal{Y}}} [\ell^q(Y^\top \alpha, t)])^{1/q} \right\} = \inf_{t \in \mathbb{R}} \sup_{\mathbb{Q}|_{\mathcal{Y}} \in \mathbb{B}_{\delta}(\mathbb{Q}_0)} \left\{ t + (\mathbb{E}_{\mathbb{Q}|_{\mathcal{Y}}} [\ell^q(Y^\top \alpha, t)])^{1/q} \right\}.$$

Proof of Lemma 30. (i) Denoted by $\pi_\ell^1(\mathbb{Q}|_{\mathcal{Y}}, t) = \mathbb{E}_{\mathbb{Q}|_{\mathcal{Y}}} [\ell(Y^\top \alpha, t)]$. First, we show three facts below.

- (a) $\pi_\ell^1(\mathbb{Q}|_{\mathcal{Y}}, t)$ is concave in $\mathbb{Q}|_{\mathcal{Y}}$ for all $t \in \mathbb{R}^k$; (b) $\pi_\ell^1(\mathbb{Q}|_{\mathcal{Y}}, t)$ is convex in t for all $\mathbb{Q}|_{\mathcal{Y}}$; (c) $\pi_\ell^1(\mathbb{Q}|_{\mathcal{Y}}, t)$ is level bounded in t for all $\mathbb{Q}|_{\mathcal{Y}}$. The fact (a) is trivial. Note that ℓ is nonnegative on $\mathbb{R}^{k+1}$ and convex in $t \in \mathbb{R}^k$ for all $z \in \mathbb{R}$. For any distribution $\mathbb{Q}|_{\mathcal{Y}}$ on $\mathcal{Y}$, $\lambda \in [0, 1]$ and $t_1, t_2 \in \mathbb{R}^k$, it holds that

$$\begin{aligned} \mathbb{E}_{\mathbb{Q}|_{\mathcal{Y}}} [\ell(Y^\top \alpha, \lambda t_1 + (1 - \lambda)t_2)] &\leq \mathbb{E}_{\mathbb{Q}|_{\mathcal{Y}}} [\lambda \ell(Y^\top \alpha, t_1) + (1 - \lambda) \ell(Y^\top \alpha, t_2)] \\ &= \lambda \mathbb{E}_{\mathbb{Q}|_{\mathcal{Y}}} [\ell(Y^\top \alpha, t_1)] + (1 - \lambda) \mathbb{E}_{\mathbb{Q}|_{\mathcal{Y}}} [\ell(Y^\top \alpha, t_2)], \end{aligned}$$

where the inequality from the convexity of ℓ with respect to $t \in \mathbb{R}^k$. This implies (b). We claim that $\pi_\ell^1(\mathbb{Q}|_{\mathcal{Y}}, t)$ is level bounded in t for all $\mathbb{Q}|_{\mathcal{Y}}$. Otherwise, for fixed $\mathbb{Q}|_{\mathcal{Y}}$, there exists $A \in \mathbb{R}$ such that $S_A := \{t \in \mathbb{R}^k : \pi_\ell^1(\mathbb{Q}|_{\mathcal{Y}}, t) \leq A\}$ is unbounded. Then for each integer $n \geq 1$, there exist $t_n \in S_A$

satisfies that $\|t_n\| \geq n$ for arbitrary norm in $\mathbb{R}^k$. For any $q_1 \in \mathcal{I}$, because $\bar{\ell}_{q_1}(z, t)$ is level bounded for any $z \in \mathbb{R}$, we have $\bar{\ell}_{q_1}(z, t_n) \rightarrow \infty$ when $n \rightarrow \infty$. In fact, if there exists $M \in \mathbb{R}$ such that $\bar{\ell}_{q_1}(z, t_n) \leq M$ for any n , one can verify that the set $\{t : \bar{\ell}_{q_1}(z, t) \leq M\}$ is unbounded, which is contradictory to that $\bar{\ell}_{q_1}(z, t)$ is level bounded in t for any z . And then $\bar{\ell}_{q_1}(z, t_n) \rightarrow \infty$ when $n \rightarrow \infty$, implies that $\ell(z, t_n) \rightarrow \infty$ when $n \rightarrow \infty$ for any $z \in \mathbb{R}$ due to the form of ℓ . By nonnegativity of ℓ , using Fatou lemma. we have

$$\liminf_{n \rightarrow \infty} \pi_\ell^1(\mathbb{Q}|_Y, t_n) = \liminf_{n \rightarrow \infty} \mathbb{E}_{\mathbb{Q}|_Y} [\ell(Y^\top \alpha, t_n)] \geq \mathbb{E}_{\mathbb{Q}|_Y} \left[\liminf_{n \rightarrow \infty} \ell(Y^\top \alpha, t_n) \right] = \infty.$$

This is contradictory to that $\liminf_{n \rightarrow \infty} \pi_\ell^1(\mathbb{Q}|_Y, t_n) \leq A$. Hence, we conclude the proof of (c). By (b), we find that $\pi_\ell^1(\mathbb{Q}|_Y, t)$ is lower semicontinuous in t . Combining this fact with (c), and applying Weierstrass theorem, there exists a level set S_{A_0} of $\pi_\ell^1(\mathbb{Q}|_Y, t)$ such that all minimizers of the problem $\inf_{t \in \mathbb{R}} \pi_\ell^1(\mathbb{Q}|_Y, t)$ are contained in it. The set S_{A_0} is closed and bounded because that $\pi_\ell^1(\mathbb{Q}|_Y, t)$ is lower semicontinuous and level bounded, and thus, S_{A_0} is a compact set. Then, the set of all minimizers of the problem $\inf_{t \in \mathbb{R}^k} \pi_\ell^1(\mathbb{Q}|_Y, t)$ is a subset of a compact set. Denote by $t(\mathbb{Q}|_Y) := \arg \inf_{\|t\|} \{\arg \min_t \pi_\ell^1(\mathbb{Q}|_Y, t)\}$. We will show that $\{t(\mathbb{Q}|_Y) : \mathbb{Q}|_Y \in \mathbb{B}_\delta(\mathbb{Q}_0)\}$ is a subset of a compact set. For any $\mathbb{Q}|_Y \in \mathbb{B}_\delta(\mathbb{Q}_0)$ and $t \in \mathbb{R}^k$, let $Y \sim \mathbb{Q}|_Y$ and $Y_0 \sim \mathbb{Q}_0$, and $q_j \in \mathcal{I}$, we have

$$\begin{aligned} & \left| \left(\mathbb{E} \left[C_{q_j} \left(\bar{\ell}_{q_j}(Y^\top \alpha, t) \right)^{q_j} \right] \right)^{1/q_j} - \left(\mathbb{E} \left[C_{q_j} \left(\bar{\ell}_{q_j}(Y_0^\top \alpha, t) \right)^{q_j} \right] \right)^{1/q_j} \right| \\ & \leq C_{q_j} \left(\mathbb{E} \left[\left| \bar{\ell}_{q_j}(Y^\top \alpha, t) - \bar{\ell}_{q_j}(Y_0^\top \alpha, t) \right|^{q_j} \right] \right)^{1/q_j} \\ & \leq C_{q_j} \left(\mathbb{E} \left[\text{Lip}(\bar{\ell}_{q_j}) |(Y - Y_0)^\top \alpha|^{q_j} \right] \right)^{1/q_j} \\ & \leq C_{q_j} \text{Lip}(\bar{\ell}_{q_j}) \left(\mathbb{E} \left[|(Y - Y_0)^\top \alpha|^q \right] \right)^{1/q} \\ & \leq C_{q_j} \text{Lip}(\bar{\ell}_{q_j}) \|\alpha\|_* (\mathbb{E} [\|Y - Y_0\|^q])^{1/q} \leq C_{q_j} \text{Lip}(\bar{\ell}_{q_j}) \|\alpha\|_* \delta^{1/q}, \end{aligned} \quad (39)$$

where the first, third and fourth inequalities from the triangle inequality, Jensen's inequality and Hölder inequality, respectively, and we use the definition of OT ball in the last inequality. From (39), we have that

$$\pi_\ell^1(\mathbb{Q}|_Y, t(\mathbb{Q}_0)) \leq \sum_{j=1,2} \left(\left(\mathbb{E} \left[C_{q_j} \left(\bar{\ell}_{q_j}(Y_0^\top \alpha, t(\mathbb{Q}_0)) \right)^{q_j} \right] \right)^{1/q_j} + C_{q_j} \text{Lip}(\bar{\ell}_{q_j}) \|\alpha\|_* \delta^{1/q} \right)^{q_j}. \quad (40)$$

Note that $\left(\mathbb{E} \left[C_{q_j} \left(\bar{\ell}_{q_j}(Y_0^\top \alpha, t) \right)^{q_j} \right] \right)^{1/q_j} \rightarrow \infty$ as $\|t\| \rightarrow \infty$ for $q_j \in \mathcal{I}$ due to that $\bar{\ell}_{q_j}(z, t)$ is level bounded. There exists $\Delta > 0$ such that

$$\left(\mathbb{E} \left[C_{q_j} \left(\bar{\ell}_{q_j}(Y_0^\top \alpha, t) \right)^{q_j} \right] \right)^{1/q_j} \geq C_{q_j} \text{Lip}(\bar{\ell}_{q_j}) \|\alpha\|_* \delta^{1/q},$$

and

$$\left(\mathbb{E} \left[C_{q_j} \left(\bar{\ell}_{q_j} (Y_0^\top \alpha, t) \right)^{q_j} \right] \right)^{1/q_j} > \left(\mathbb{E} \left[C_{q_j} \left(\bar{\ell}_{q_j} (Y_0^\top \alpha, t(\mathbb{Q}_0)) \right)^{q_j} \right] \right)^{1/q_j} + 2C_{q_j} \text{Lip}(\bar{\ell}_{q_j}) \|\alpha\|_* \delta^{1/q},$$

for all $t \notin \{t : \|t - t(\mathbb{Q}_0)\| \leq \Delta\}$ and $q_j \in \mathcal{I}$. This, combined with (39), imply that

$$\begin{aligned} \pi_\ell^1(\mathbb{Q}|_Y, t) &\geq \sum_{j=1,2} \left(\left(\mathbb{E} \left[C_{q_j} \left(\bar{\ell}_{q_j} (Y_0^\top \alpha, t) \right)^{q_j} \right] \right)^{1/q_j} - C_{q_j} \text{Lip}(\bar{\ell}_{q_j}) \|\alpha\|_* \delta^{1/q} \right)^{q_j} \\ &> \sum_{j=1,2} \left(\left(\mathbb{E} \left[C_{q_j} \left(\bar{\ell}_{q_j} (Y_0^\top \alpha, t(\mathbb{Q}_0)) \right)^{q_j} \right] \right)^{1/q_j} + C_{q_j} \text{Lip}(\bar{\ell}_{q_j}) \|\alpha\|_* \delta^{1/q} \right)^{q_j}, \end{aligned} \quad (41)$$

for all $t \notin \{t : \|t - t(\mathbb{Q}_0)\| \leq \Delta\}$. Applying (40) and (41), we have $\{t(\mathbb{Q}|_Y) : \mathbb{Q}|_Y \in \mathbb{B}_\delta(\mathbb{Q}_0)\} \subseteq \{t : \|t - t(\mathbb{Q}_0)\| \leq \Delta\}$. Using the minimax theorem in Sion (1958), it holds that

$$\begin{aligned} \sup_{\mathbb{Q}|_Y \in \mathbb{B}_\delta(\mathbb{Q}_0)} \inf_{t \in \mathbb{R}^k} \pi_\ell^1(\mathbb{Q}|_Y, t) &= \sup_{\mathbb{Q}|_Y \in \mathbb{B}_\delta(\mathbb{Q}_0)} \inf_{t \in \{t : \|t - t(\mathbb{Q}_0)\| \leq \Delta\}} \pi_\ell^1(\mathbb{Q}|_Y, t) \\ &= \inf_{t \in \{t : \|t - t(\mathbb{Q}_0)\| \leq \Delta\}} \sup_{\mathbb{Q}|_Y \in \mathbb{B}_\delta(\mathbb{Q}_0)} \pi_\ell^1(\mathbb{Q}|_Y, t) \geq \inf_{t \in \mathbb{R}^k} \sup_{\mathbb{Q}|_Y \in \mathbb{B}_\delta(\mathbb{Q}_0)} \pi_\ell^1(\mathbb{Q}|_Y, t). \end{aligned}$$

The converse direction is trivial. Hence, we complete the proof of (i).

(ii) The proof is similar to (i). □

Proof of Proposition 13. (i) Let $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}, \forall y_1, y_2 \in \mathcal{Y}$. Applying Lemma 30, equation (26) and minimax theorem, problem (13) with (14) can be reformulated as

$$\begin{aligned} &\min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}} \sup_{\mathbb{Q}|_Y \in \mathbb{B}_\delta \left(\sum_{i=1}^N p_i I_{\hat{y}_i} \right)} \rho_{\mathbb{Q}|_Y}^{(1)}(Y^\top \alpha) \\ &= \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}} \sup_{\mathbb{Q}|_Y \in \mathbb{B}_\delta \left(\sum_{i=1}^N p_i I_{\hat{y}_i} \right)} \inf_{t \in \mathbb{R}^k} \mathbb{E}_{\mathbb{Q}|_Y} [\ell(Y^\top \alpha, t)] \\ &= \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}} \inf_{t \in \mathbb{R}^k} \sup_{\mathbb{Q}|_Y \in \mathbb{B}_\delta \left(\sum_{i=1}^N p_i I_{\hat{y}_i} \right)} \mathbb{E}_{\mathbb{Q}|_Y} [\ell(Y^\top \alpha, t)] \\ &= \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}, \delta \geq 0} \inf_{t \in \mathbb{R}^k} \mathbb{E}_{\sum_{i=1}^N p_i I_{\hat{y}_i}} [\ell(Y^\top \alpha, t)] + \text{Lip}(\ell) \delta \|\alpha\|_* \\ &= \min_{\alpha \in \mathcal{A}} \inf_{t \in \mathbb{R}^k} \sup_{(p, \delta) \in \mathcal{V}, \delta \geq 0} \mathbb{E}_{\sum_{i=1}^N p_i I_{\hat{y}_i}} [\ell(Y^\top \alpha, t)] + \text{Lip}(\ell) \delta \|\alpha\|_* \\ &= \min_{\alpha \in \mathcal{A}} \inf_{t \in \mathbb{R}^k} \sup_{(p, \delta) \in \mathcal{V}_+} \mathbb{E}_{\sum_{i=1}^N p_i I_{\hat{y}_i}} [\ell(Y^\top \alpha, t)] + \text{Lip}(\ell) \delta \|\alpha\|_*. \end{aligned}$$

Applying the Fenchel duality, we have the desired result.

(ii) Applying Corollary 1 of Wu et al. (2022) and minimax theorem, we find that problem (13) with (14) □

can be reformulated as

$$\begin{aligned}
& \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}} \sup_{\mathbb{Q}|Y \in \mathbb{B}_\delta \left(\sum_{i=1}^N p_i I_{\hat{y}_i} \right)} \rho_{\mathbb{Q}|Y}^{(1)} (Y^\top \alpha) \\
&= \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}, \delta \geq 0} \left(\left(\rho_{\sum_{i=1}^N p_i I_{\hat{y}_i}}^{(1)} (Y^\top \alpha) \right)^{1/q} + C \delta^{\frac{1}{q}} \|\alpha\|_* \right)^q \\
&= \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}_+} \left(\left(\inf_{t \in \mathbb{R}} \mathbb{E}_{\sum_{i=1}^N p_i I_{\hat{y}_i}} [\ell(Y^\top \alpha, t)] \right)^{1/q} + C \delta^{\frac{1}{q}} \|\alpha\|_* \right)^q \\
&= \min_{\alpha \in \mathcal{A}} \left(\sup_{(p, \delta) \in \mathcal{V}_+} \inf_{t \in \mathbb{R}} \left(\mathbb{E}_{\sum_{i=1}^N p_i I_{\hat{y}_i}} [\ell(Y^\top \alpha, t)] \right)^{1/q} + C \delta^{\frac{1}{q}} \|\alpha\|_* \right)^q \\
&= \min_{\alpha \in \mathcal{A}} \left(\inf_{t \in \mathbb{R}} \sup_{(p, \delta) \in \mathcal{V}_+} \left(\mathbb{E}_{\sum_{i=1}^N p_i I_{\hat{y}_i}} [\ell(Y^\top \alpha, t)] \right)^{1/q} + C \delta^{\frac{1}{q}} \|\alpha\|_* \right)^q \\
&= \inf_{\alpha \in \mathcal{A}, t \in \mathbb{R}} \left(\sup_{(p, \delta) \in \mathcal{V}_+} \left(\mathbb{E}_{\sum_{i=1}^N p_i I_{\hat{y}_i}} [\ell(Y^\top \alpha, t)] \right)^{1/q} + C \delta^{\frac{1}{q}} \|\alpha\|_* \right)^q.
\end{aligned}$$

Thus we only need to consider the inner problem

$$\sup_{(p, \delta) \in \mathcal{V}_+} \left(\mathbb{E}_{\sum_{i=1}^N p_i I_{\hat{y}_i}} [\ell(Y^\top \alpha, t)] \right)^{1/q} + C \delta^{\frac{1}{q}} \|\alpha\|_*,$$

for $q \in (1, \infty)$. Similar to the proof of Proposition 10, we can apply the Fenchel duality theory to the above problem and then obtain the desired result. $\square$

Proof of Proposition 14. Applying Lemma 30, Theorem 29 and minimax theorem in Sion (1958), problem (13) with (14) can be reformulated as

$$\begin{aligned}
& \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}} \sup_{\mathbb{Q}|Y \in \mathbb{B}_\delta \left(\sum_{i=1}^N p_i I_{\hat{y}_i} \right)} \rho_{\mathbb{Q}|Y}^{(1)} (Y^\top \alpha) \\
&= \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}} \inf_{t \in \mathbb{R}^k} \sup_{\mathbb{Q}|Y \in \mathbb{B}_\delta \left(\sum_{i=1}^N p_i I_{\hat{y}_i} \right)} \mathbb{E}_{\mathbb{Q}|Y} [\ell(Y^\top \alpha, t)] \\
&= \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}, \delta \geq 0} \inf_{t \in \mathbb{R}^k, \eta \geq 0} \sum_{i=1}^N p_i \cdot \sup_{z \in \mathbb{R}} \left\{ z \cdot \hat{y}_i^\top \alpha - (\ell(\cdot, t))^*(z) + \eta \cdot \frac{|z|^{q^*}}{q^* q^{q^*-1}} \right\} + \delta \frac{\|\alpha\|_*^q}{\eta^{q-1}} \\
&= \inf_{\alpha \in \mathcal{A}, t \in \mathbb{R}^k, \eta \geq 0} \sup_{(p, \delta) \in \mathcal{V}_+} \sum_{i=1}^N p_i \cdot \sup_{z \in \mathbb{R}} \left\{ z \cdot \hat{y}_i^\top \alpha - (\ell(\cdot, t))^*(z) + \eta \cdot \frac{|z|^{q^*}}{q^* q^{q^*-1}} \right\} + \delta \frac{\|\alpha\|_*^q}{\eta^{q-1}}.
\end{aligned}$$

Applying the Fenchel duality, we have the desired result. $\square$

Proof of Example 3.4. The robustified conditional mean-variance portfolio allocation problem in Nguyen

et al. (2024) is as follows:

$$\min_{\alpha \in \mathcal{A}} \sup_{\substack{\mathbb{Q} \in \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}) \\ \mathbb{Q}(X \in \mathcal{N}_\gamma(x_0)) \in \mathcal{N}_\omega}} \text{Variance}_{\mathbb{Q}} [Y^\top \alpha \mid X \in \mathcal{N}_\gamma(x_0)] - \theta \cdot \mathbb{E}_{\mathbb{Q}} [Y^\top \alpha \mid X \in \mathcal{N}_\gamma(x_0)],$$

where $\mathcal{N}_\omega = [\varepsilon_0, 1]$, $\varepsilon_0 \in (0, 1]$ and $\theta \geq 0$. Suppose in addition that $\mathcal{X} = \mathbb{R}^{n_1}$, $\mathcal{Y} = \mathbb{R}^{n_2}$, $\mathbb{D}_X(x, \widehat{x}) = \|x - \widehat{x}\|^2$ and $\mathbb{D}_Y(y, \widehat{y}) = \|y - \widehat{y}\|_2^2$. By Theorem 3 and the minimization representation of variance, one can verify that the problem (19) can be reformulated as

$$\begin{aligned} & \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}} \sup_{\mathbb{Q}|_Y \in \mathbb{B}_\delta(\sum_{i=1}^N p_i \mathbb{I}_{\widehat{y}_i})} \text{Variance}_{\mathbb{Q}|_Y} [Y^\top \alpha] - \theta \cdot \mathbb{E}_{\mathbb{Q}|_Y} [Y^\top \alpha] \\ &= \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}} \sup_{\mathbb{Q}|_Y \in \mathbb{B}_\delta(\sum_{i=1}^N p_i \mathbb{I}_{\widehat{y}_i})} \inf_{t \in \mathbb{R}} \mathbb{E}_{\mathbb{Q}|_Y} \left[(Y^\top \alpha - t)^2 - \theta \cdot Y^\top \alpha \right] \\ &:= \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}} \sup_{\mathbb{Q}|_Y \in \mathbb{B}_\delta(\sum_{i=1}^N p_i \mathbb{I}_{\widehat{y}_i})} \inf_{t \in \mathbb{R}} \mathbb{E}_{\mathbb{Q}|_Y} [\ell(Y^\top \alpha, t)], \end{aligned}$$

where $\ell(z, t) = (z - t)^2 - \theta z$. By Lemma 30, we have that

$$\sup_{\mathbb{Q}|_Y \in \mathbb{B}_\delta(\sum_{i=1}^N p_i \mathbb{I}_{\widehat{y}_i})} \inf_{t \in \mathbb{R}} \mathbb{E}_{\mathbb{Q}|_Y} [\ell(Y^\top \alpha, t)] = \inf_{t \in \mathbb{R}} \sup_{\mathbb{Q}|_Y \in \mathbb{B}_\delta(\sum_{i=1}^N p_i \mathbb{I}_{\widehat{y}_i})} \mathbb{E}_{\mathbb{Q}|_Y} [\ell(Y^\top \alpha, t)].$$

Due to the above equation, we can reformulate problem (19) as

$$\min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}} \inf_{t \in \mathbb{R}} \sup_{\mathbb{Q}|_Y \in \mathbb{B}_\delta(\sum_{i=1}^N p_i \mathbb{I}_{\widehat{y}_i})} \mathbb{E}_{\mathbb{Q}|_Y} [\ell(Y^\top \alpha, t)].$$

Applying Theorem 29 and min-max theorem, the problem (19) can be reformulated as

$$\begin{aligned} & \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}} \inf_{t \in \mathbb{R}} \sup_{\mathbb{Q}|_Y \in \mathbb{B}_\delta(\sum_{i=1}^N p_i \mathbb{I}_{\widehat{y}_i})} \mathbb{E}_{\mathbb{Q}|_Y} [\ell(Y^\top \alpha, t)] \\ &= \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}, \delta \geq 0} \inf_{t \in \mathbb{R}, \eta \geq 0} \sum_{i=1}^N p_i \cdot \sup_{z \in \mathbb{R}} \left\{ z \cdot \widehat{y}_i^\top \alpha - \ell^*(z, t) + \eta \cdot \frac{|z|^{q^*}}{q^* q^{q^*-1}} \right\} + \delta \frac{\|\alpha\|_*^q}{\eta^{q-1}} \\ &= \inf_{\alpha \in \mathcal{A}, t \in \mathbb{R}, \eta \geq 0} \sup_{(p, \delta) \in \mathcal{V}_+} \sum_{i=1}^N p_i \cdot \sup_{z \in \mathbb{R}} \left\{ z \cdot \widehat{y}_i^\top \alpha - \ell^*(z, t) + \eta \cdot \frac{z^2}{4} \right\} + \delta \frac{\|\alpha\|_*^2}{\eta} \\ &= \inf_{\alpha \in \mathcal{A}, t \in \mathbb{R}, \eta \geq 0} \sup_{(p, \delta) \in \mathcal{V}_+} \sum_{i=1}^N p_i \cdot \sup_{z \in \mathbb{R}} \left\{ z \cdot \widehat{y}_i^\top \alpha - \ell^*(z, t) + \eta \cdot \frac{z^2}{4} \right\} + \delta \frac{\|\alpha\|_*^2}{\eta} \\ &= \inf_{\alpha \in \mathcal{A}, t \in \mathbb{R}, 1 \geq \eta \geq 0} \sup_{(p, \delta) \in \mathcal{V}_+} \sum_{i=1}^N p_i \frac{(\widehat{y}_i^\top \alpha - \theta/2 - t)^2}{1 - \eta} + \delta \frac{\|\alpha\|_*^2}{\eta} - \frac{\theta^2}{4} - t\theta. \end{aligned}$$

Applying the Fenchel duality, the problem (19) can be reformulated as an infimum problem

$$\begin{aligned} & \inf_{\alpha \in \mathcal{A}, 0 \leq \eta \leq 1, t \in \mathbb{R}, v} \sigma^*(v|\mathcal{V}_+) - \frac{\theta^2}{4} - t\theta \\ & \text{s.t. } v_i \geq \frac{(\hat{y}_i^\top \alpha - \theta/2 - t)^2}{1 - \eta}, \quad i = 1, \dots, N, \\ & \quad v_{N+1} \geq \|\alpha\|_*^2 / \eta, \end{aligned}$$

where $\sigma^*(v|\mathcal{V}_+)$, the conjugate function of $\sigma(v|\mathcal{V}_+)$, is given in Example 3.2. Thus, substituting the $\sigma^*(v|\mathcal{V}_+)$ in Example 3.2, the robustified conditional mean-variance portfolio allocation problem (19), proposed in Nguyen et al. (2024), can be reformulated as

$$\begin{aligned} & \inf_{\alpha \in \mathcal{A}, 0 \leq \eta \leq 1, t \in \mathbb{R}, z, s} z^\top b - \frac{\theta^2}{4} - t\theta \\ & \text{s.t. } z \geq 0, A^\top z \geq s \\ & \quad s_i \geq \frac{(\hat{y}_i^\top \alpha - \theta/2 - t)^2}{1 - \eta}, \quad i = 1, \dots, N, \\ & \quad s_{N+1} \geq \|\alpha\|_*^2 / \eta, \quad s_{N+2} \geq 0, \end{aligned}$$

where $A \in \mathbb{R}^{(N+6) \times (N+2)}$, and $b \in \mathbb{R}^{N+6}$ are defined by

$$A = \begin{pmatrix} 1 & \dots & 1 & 0 & 0 \\ -1 & \dots & -1 & 0 & 0 \\ 0 & \dots & 0 & 0 & 1 \\ 0 & \dots & 0 & 0 & -1 \\ 1 & \dots & 0 & 0 & -\frac{1}{N} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & 1 & 0 & -\frac{1}{N} \\ -a_1 & \dots & -a_N & -a_{N+1} & -a_{N+2} \\ a_1 & \dots & a_N & a_{N+1} & a_{N+2} \end{pmatrix}, \quad \text{with } a = \begin{pmatrix} -d_1 \\ \vdots \\ -d_m \\ d_{m+1} \\ \vdots \\ d_N \\ -1 \\ \delta_0 - \frac{1}{N} \sum_{i=m+1}^N d_i \end{pmatrix}, \quad \text{and } b = \begin{pmatrix} 1 \\ -1 \\ \frac{1}{\varepsilon_0} \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

□

Proof of Proposition 17. The proof is similar to Proposition 13.

□

Proof of Proposition 18. The problem problem (13) with (14) with $\rho = \rho^{(2)}$ can be reformulated as

$$\begin{aligned}
& \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}} \sup_{\mathbb{Q}|Y \in \mathbb{B}_\delta \left(\sum_{i=1}^N p_i I_{\hat{y}_i} \right)} \rho_{\mathbb{Q}|Y}^{(2)} (Y^\top \alpha) \\
&= \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}} \inf_{t \in \mathbb{R}} \sup_{\mathbb{Q}|Y \in \mathbb{B}_\delta \left(\sum_{i=1}^N p_i I_{\hat{y}_i} \right)} t + (\mathbb{E}_{\mathbb{Q}|Y} [\ell^q(Y^\top \alpha, t)])^{1/q} \\
&= \inf_{\alpha \in \mathcal{A}, t \in \mathbb{R}} \sup_{(p, \delta) \in \mathcal{V}} t + \sup_{\mathbb{Q}|Y \in \mathbb{B}_\delta \left(\sum_{i=1}^N p_i I_{\hat{y}_i} \right)} (\mathbb{E}_{\mathbb{Q}|Y} [\ell^q(Y^\top \alpha, t)])^{1/q} \\
&= \inf_{\alpha \in \mathcal{A}, t \in \mathbb{R}} \sup_{(p, \delta) \in \mathcal{V}, \delta \geq 0} t + \left(\sup_{\mathbb{Q}|Y \in \mathbb{B}_\delta \left(\sum_{i=1}^N p_i I_{\hat{y}_i} \right)} \mathbb{E}_{\mathbb{Q}|Y} [\ell^q(Y^\top \alpha, t)] \right)^{1/q} \\
&= \inf_{\alpha \in \mathcal{A}, t \in \mathbb{R}} \sup_{(p, \delta) \in \mathcal{V}_+} t + \left(\inf_{\lambda \geq 0} \sum_{i=1}^N p_i \cdot \sup_{z \in \mathbb{R}} \{ \ell^q(z, t) - \lambda |z - \hat{y}_i^\top \alpha|^q \} + \lambda \delta \|\alpha\|_*^q \right)^{1/q} \\
&= \inf_{\alpha \in \mathcal{A}, t \in \mathbb{R}} t + \left(\sup_{(p, \delta) \in \mathcal{V}_+} \inf_{\lambda \geq 0} \sum_{i=1}^N p_i \cdot \sup_{z \in \mathbb{R}} \{ \ell^q(z, t) - \lambda |z - \hat{y}_i^\top \alpha|^q \} + \lambda \delta \|\alpha\|_*^q \right)^{1/q} \\
&= \inf_{\alpha \in \mathcal{A}, t \in \mathbb{R}} t + \left(\inf_{\lambda \geq 0} \sup_{(p, \delta) \in \mathcal{V}_+} \sum_{i=1}^N p_i \cdot \sup_{z \in \mathbb{R}} \{ \ell^q(z, t) - \lambda |z - \hat{y}_i^\top \alpha|^q \} + \lambda \delta \|\alpha\|_*^q \right)^{1/q} \\
&= \inf_{\alpha \in \mathcal{A}, t \in \mathbb{R}, \lambda \geq 0} t + \left(\sup_{(p, \delta) \in \mathcal{V}_+} \sum_{i=1}^N p_i \cdot \sup_{z \in \mathbb{R}} \{ \ell^q(z, t) - \lambda |z - \hat{y}_i^\top \alpha|^q \} + \lambda \delta \|\alpha\|_*^q \right)^{1/q} \\
&= \inf_{\alpha \in \mathcal{A}, t \in \mathbb{R}, \lambda \geq 0} t + \sup_{(p, \delta) \in \mathcal{V}_+} \left(\sum_{i=1}^N p_i \cdot \sup_{z \in \mathbb{R}} \{ \ell^q(z, t) - \lambda |z - \hat{y}_i^\top \alpha|^q \} + \lambda \delta \|\alpha\|_*^q \right)^{1/q},
\end{aligned}$$

where the first equality is from Lemma EC.8 of Wu et al. (2022), the second and sixth equalities are from minimax theorem in Sion (1958), the fourth equality is from Theorem 29, and other equalities hold for the reason that $x^{1/q}$ is a nonnegative and increasing function and thus it does not change optimal solution of the optimization problem. Using the Fenchel duality, we can reformulate the above problem as

$$\begin{aligned}
& \inf_{\alpha \in \mathcal{A}, t \in \mathbb{R}, \lambda \geq 0} t + \sup_{(p, \delta) \in \mathcal{V}_+} \left(\sum_{i=1}^N p_i \cdot \sup_{z \in \mathbb{R}} \{ \ell^q(z, t) - \lambda |z - \hat{y}_i^\top \alpha|^q \} + \lambda \delta \|\alpha\|_*^q \right)^{1/q} \\
&= \inf_{\alpha \in \mathcal{A}, t \in \mathbb{R}, \lambda \geq 0, v, y} t + \sigma^*(v | \mathcal{V}_+) + C_q^1 y^{\frac{1}{1-q}} \\
& \quad \text{s.t. } v_i \geq y \cdot \sup_{z \in \mathbb{R}} \{ \ell^q(z, t) - \lambda |z - \hat{y}_i^\top \alpha|^q \}, i = 1, \dots, N, \\
& \quad v_{N+1} \geq \lambda y \|\alpha\|_*^q, y \geq 0.
\end{aligned}$$

Let $u = \lambda y$ and $w = y^{\frac{1}{1-q}}$. By the convexity of ℓ in z , one can verify that $\frac{\ell^q(z, t)}{w^{q-1}}$ is convex in z , and thus $\left(\frac{\ell^q(\cdot, t)}{w^{q-1}} \right)^{**} = \frac{\ell^q(\cdot, t)}{w^{q-1}}$. Moreover, we find that $\frac{\ell^q(z, t)}{w^{q-1}}$ is jointly convex in (z, t, w) and thus $\left(\frac{\ell^q(\cdot, t)}{w^{q-1}} \right)^*$ is jointly concave in (t, w) according to the definition of concave conjugate function. Let $u = \eta^{1-q}$. Similar to the

proof of Theorem 29, the problem can be further reformulated as

$$\begin{aligned}
& \inf_{\alpha \in \mathcal{A}, t \in \mathbb{R}, u \geq 0, v, y} t + \sigma^*(v|\mathcal{V}_+) + C_q^1 y^{\frac{1}{1-q}} \\
& \text{s.t. } v_i \geq \sup_{z \in \mathbb{R}} \{y \ell^q(z, t) - u |z - \widehat{y}_i^\top \alpha|^q\}, i = 1, \dots, N, \\
& v_{N+1} \geq u \|\alpha\|_*^q, y \geq 0, \\
& = \inf_{\alpha \in \mathcal{A}, t \in \mathbb{R}, u \geq 0, v, w} t + \sigma^*(v|\mathcal{V}_+) + C_q^1 w \\
& \text{s.t. } v_i \geq \sup_{z \in \mathbb{R}} \left\{ \frac{\ell^q(z, t)}{w^{q-1}} - u |z - \widehat{y}_i^\top \alpha|^q \right\}, i = 1, \dots, N, \\
& v_{N+1} \geq u \|\alpha\|_*^q, w \geq 0, \\
& = \inf_{\alpha \in \mathcal{A}, \eta \geq 0, t \in \mathbb{R}, v, w \geq 0} t + \sigma^*(v|\mathcal{V}_+) + C_q^1 w \\
& \text{s.t. } v_i \geq \sup_{x \in \mathbb{R}} \left\{ x \cdot \widehat{y}_i^\top \alpha - \left(\frac{\ell^q(\cdot, t)}{w^{q-1}} \right)^* (x) + \eta \cdot \frac{|x|^{q^*}}{q^* q^{q^*-1}} \right\}, i = 1, \dots, N, \\
& v_{N+1} \geq \|\alpha\|_*^q / \eta^{q-1}.
\end{aligned}$$

□

Proof of Proposition 19. Let $\mathcal{P} = \bigcup_{(p, \delta) \in \mathcal{V}} \mathbb{B}_\delta \left(\sum_{i=1}^N p_i \mathbf{I}_{\widehat{y}_i} \right)$ be a convex set. We first prove that

$$\begin{aligned}
\sup_{\mathbb{Q}|_Y \in \mathcal{P}} \inf \{ \kappa \in \mathbb{R} : \mathbb{E}_{\mathbb{Q}|_Y} [u(-Y^\top \alpha - \kappa)] \leq l \} &= \inf_{\kappa \in \mathbb{R}} \kappa \\
&\text{s.t. } \sup_{\mathbb{Q}|_Y \in \mathcal{P}} \mathbb{E}_{\mathbb{Q}|_Y} [u(-Y^\top \alpha - \kappa)] \leq l.
\end{aligned} \tag{42}$$

If $\sup_{\mathbb{Q}|_Y \in \mathcal{P}} \mathbb{E}_{\mathbb{Q}|_Y} [u(-Y^\top \alpha - \kappa)] \leq l$, we have $\mathbb{E}_{\mathbb{Q}|_Y} [u(-Y^\top \alpha - \kappa)] \leq l$ for any $\mathbb{Q}|_Y \in \mathcal{P}$. Then

$$\left\{ \kappa : \sup_{\mathbb{Q}|_Y \in \mathcal{P}} \mathbb{E}_{\mathbb{Q}|_Y} [u(-Y^\top \alpha - \kappa)] \leq l \right\} \subseteq \left\{ \kappa : \mathbb{E}_{\mathbb{Q}|_Y} [u(-Y^\top \alpha - \kappa)] \leq l \right\}.$$

Thus, we have

$$\inf \{ \kappa \in \mathbb{R} : \mathbb{E}_{\mathbb{Q}|_Y} [u(-Y^\top \alpha - \kappa)] \leq l \} \leq \inf \left\{ \kappa \in \mathbb{R} : \sup_{\mathbb{Q}|_Y \in \mathcal{P}} \mathbb{E}_{\mathbb{Q}|_Y} [u(-Y^\top \alpha - \kappa)] \leq l \right\}, \forall \mathbb{Q}|_Y \in \mathcal{P},$$

and

$$\sup_{\mathbb{Q}|_Y \in \mathcal{P}} \inf \{ \kappa \in \mathbb{R} : \mathbb{E}_{\mathbb{Q}|_Y} [u(-Y^\top \alpha - \kappa)] \leq l \} \leq \inf \left\{ \kappa \in \mathbb{R} : \sup_{\mathbb{Q}|_Y \in \mathcal{P}} \mathbb{E}_{\mathbb{Q}|_Y} [u(-Y^\top \alpha - \kappa)] \leq l \right\}. \tag{43}$$

Let $\kappa^* = \sup_{\mathbb{Q}|_Y \in \mathcal{P}} \inf \{ \kappa \in \mathbb{R} : \mathbb{E}_{\mathbb{Q}|_Y} [u(-Y^\top \alpha - \kappa)] \leq l \}$. We prove the equality (42) in the following three cases: (i) $\kappa^* = \infty$; (ii) $\kappa^* = -\infty$; (iii) κ^* is finite.

(i) From the inequality (43), we directly obtain

$$\infty \geq \inf \left\{ \kappa \in \mathbb{R} : \sup_{Q|Y \in \mathcal{P}} \mathbb{E}_Q [u(-Y^\top \alpha - \kappa)] \leq l \right\} \geq \kappa^* = \infty.$$

Thus, the equality (42) holds.

(ii) If $\kappa^* = -\infty$, for any $Q|Y \in \mathcal{P}$, we have $\mathbb{E}_{Q|Y} [u(-Y^\top \alpha - \kappa)] \leq l$ for any $\kappa \in \mathbb{R}$. Then for any $\kappa \in \mathbb{R}$, we have $\mathbb{E}_{Q|Y} [u(-Y^\top \alpha - \kappa)] \leq l$ for any $Q|Y \in \mathcal{P}$. Thus, for any $\kappa \in \mathbb{R}$, we have $\sup_{Q|Y \in \mathcal{P}} \mathbb{E}_{Q|Y} [u(-Y^\top \alpha - \kappa)] \leq l$. Hence,

$$\inf \left\{ \kappa \in \mathbb{R} : \sup_{Q|Y \in \mathcal{P}} \mathbb{E}_{Q|Y} [u(-Y^\top \alpha - \kappa)] \leq l \right\} = -\infty = \kappa^*.$$

(iii) Note that

$$\kappa^* \geq \inf \left\{ \kappa \in \mathbb{R} : \mathbb{E}_{Q|Y} [u(-Y^\top \alpha - \kappa)] \leq l \right\} \quad \forall Q|Y \in \mathcal{P}.$$

Thus,

$$\mathbb{E}_{Q|Y} [u(-Y^\top \alpha - \kappa^*)] \leq l, \quad \forall Q|Y \in \mathcal{P},$$

which implies that

$$\sup_{Q|Y \in \mathcal{P}} \mathbb{E}_{Q|Y} [u(-Y^\top \alpha - \kappa^*)] \leq l.$$

Then $\kappa^* \in \left\{ \kappa : \sup_{Q|Y \in \mathcal{P}} \mathbb{E}_{Q|Y} [u(-Y^\top \alpha - \kappa)] \leq l \right\}$, and thus

$$\kappa^* \geq \inf \left\{ \kappa \in \mathbb{R} : \sup_{Q|Y \in \mathcal{P}} \mathbb{E}_{Q|Y} [u(-Y^\top \alpha - \kappa)] \leq l \right\}.$$

Combining with inequality (43), we obtain the equality (42).

Combining the above three cases, we obtain the equality (42) in all instances. Note that similar reasoning also appears in Proposition 1 of Guo and Xu (2019), where the finiteness of

$$\inf \left\{ \kappa \in \mathbb{R} : \sup_{Q|Y \in \mathcal{P}} \mathbb{E}_{Q|Y} [u(-Y^\top \alpha - \kappa)] \leq l \right\}$$

is assumed. Here, we establish (42) in the general case to ensure completeness.

Next, we apply the equality (42) to derive the desired result. In particular, for $\rho = \rho^{(3)}$, the problem (17)

becomes

$$\begin{aligned}
& \min_{\alpha \in \mathcal{A}} \sup_{\mathbb{Q}|_Y \in \bigcup_{(p, \delta) \in \mathcal{V}} \mathbb{B}_\delta \left(\sum_{i=1}^N p_i \mathbf{I}_{\hat{y}_i} \right)} \rho_{\mathbb{Q}|_Y} [Y^\top \alpha] \\
&= \min_{\alpha \in \mathcal{A}} \sup_{\mathbb{Q}|_Y \in \bigcup_{(p, \delta) \in \mathcal{V}} \mathbb{B}_\delta \left(\sum_{i=1}^N p_i \mathbf{I}_{\hat{y}_i} \right)} \inf \{ \kappa \in \mathbb{R} : \mathbb{E}_{\mathbb{Q}|_Y} [u(-Y^\top \alpha - \kappa)] \leq l \} \\
&= \min_{\alpha \in \mathcal{A}} \inf_{\kappa \in \mathbb{R}} \kappa \\
&\quad \text{s.t.} \quad \sup_{\mathbb{Q}|_Y \in \bigcup_{(p, \delta) \in \mathcal{V}} \mathbb{B}_\delta \left(\sum_{i=1}^N p_i \mathbf{I}_{\hat{y}_i} \right)} \mathbb{E}_{\mathbb{Q}|_Y} [u(-Y^\top \alpha - \kappa)] \leq l.
\end{aligned}$$

And then using the equation (26), we have

$$\begin{aligned}
& \min_{\alpha \in \mathcal{A}} \inf_{\kappa \in \mathbb{R}} \kappa \\
&\quad \text{s.t.} \quad \sup_{\mathbb{Q}|_Y \in \bigcup_{(p, \delta) \in \mathcal{V}} \mathbb{B}_\delta \left(\sum_{i=1}^N p_i \mathbf{I}_{\hat{y}_i} \right)} \mathbb{E}_{\mathbb{Q}|_Y} [u(-Y^\top \alpha - \kappa)] \leq l \\
&= \min_{\alpha \in \mathcal{A}} \inf_{\kappa \in \mathbb{R}} \kappa \\
&\quad \text{s.t.} \quad \sup_{(p, \delta) \in \mathcal{V}_+} \sup_{\mathbb{Q}|_Y \in \mathbb{B}_\delta \left(\sum_{i=1}^N p_i \mathbf{I}_{\hat{y}_i} \right)} \mathbb{E}_{\mathbb{Q}|_Y} [u(-Y^\top \alpha - \kappa)] \leq l \\
&= \min_{\alpha \in \mathcal{A}} \inf_{\kappa \in \mathbb{R}} \kappa \\
&\quad \text{s.t.} \quad \sup_{(p, \delta) \in \mathcal{V}_+} \sum_{i=1}^N p_i u(-\hat{y}_i^\top \alpha - \kappa) + \text{Lip}(u) \delta \|\alpha\|_* \leq l.
\end{aligned}$$

Using the Fenchel duality, we have

$$\begin{aligned}
& \min_{\alpha \in \mathcal{A}} \inf_{\kappa \in \mathbb{R}} \kappa \\
&\quad \text{s.t.} \quad \sup_{(p, \delta) \in \mathcal{V}_+} \sum_{i=1}^N p_i u(-\hat{y}_i^\top \alpha - \kappa) + \text{Lip}(u) \delta \|\alpha\|_* \leq l \\
&= \min_{\alpha \in \mathcal{A}} \inf_{\kappa \in \mathbb{R}} \kappa \\
&\quad \text{s.t.} \quad \inf_v \{ \sigma^*(v | \mathcal{V}_+) | v - v(\alpha, \kappa) \geq 0 \} \leq l \\
&= \inf_{\alpha, \kappa, v} \kappa \\
&\quad \text{s.t.} \quad \alpha \in \mathcal{A}, \kappa \in \mathbb{R}, \\
&\quad \sigma^*(v | \mathcal{V}_+) \leq l, \quad v - v(\alpha, \kappa) \geq 0,
\end{aligned}$$

where $v(\alpha, \kappa)^\top = (u(-\hat{y}_1^\top \alpha - \kappa), \dots, u(-\hat{y}_N^\top \alpha - \kappa), \text{Lip}(u) \|\alpha\|_*) \in \mathbb{R}^{N+1}$. $\square$

Proof of Proposition 21. Similar to the proof of Proposition 19, applying the equality (42), Theorem 29,

min-max theorem, and the Fenchel duality, we have the desired result. $\square$

EC.2.2.2. Proofs of Section 3.3

Proof of Proposition 22. (i) For $q = 1$ and $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}$, applying Theorem 5 in Wu et al. (2022), we find that problem (13) with (14) can be reformulated as

$$\min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}_+} \rho_{\sum_{i=1}^N p_i \mathbb{I}_{\widehat{y}_i}}^h \left[\ell(Y^\top \alpha) \right] + \text{Lip}(\ell) \|h'_-\|_{\infty} \delta \|\alpha\|_*.$$

According to Proposition 10.3 of Denneberg, D. (2013), we can reformulate the above inner supremum problem as

$$\begin{aligned} & \sup_{(p, \delta) \in \mathcal{V}_+} \sup_{\bar{p} \in \mathcal{V}_p^h} \sum_{i=1}^N \bar{p}_i \ell(\widehat{y}_i^\top \alpha) + \text{Lip}(\ell) \|h'_-\|_{\infty} \delta \|\alpha\|_* \\ &= \sup_{(p, \delta) \in \mathcal{V}_+, \bar{p} \in \mathcal{V}_p^h} \sum_{i=1}^N \bar{p}_i \ell(\widehat{y}_i^\top \alpha) + \text{Lip}(\ell) \|h'_-\|_{\infty} \delta \|\alpha\|_*. \end{aligned}$$

Using the Fenchel duality theory, we have

$$\begin{aligned} & \sup_{(p, \delta) \in \mathcal{V}_+, \bar{p} \in \mathcal{V}_p^h} \sum_{i=1}^N \bar{p}_i \ell(\widehat{y}_i^\top \alpha) + \text{Lip}(\ell) \|h'_-\|_{\infty} \delta \|\alpha\|_* \\ &= \inf_v \left\{ \sigma^*(v | \mathcal{V}_+^h) | v - \nu(\alpha) \geq 0 \right\}, \end{aligned}$$

where

$$\mathcal{V}_+^h = \{p'' = (p, \delta, \bar{p}) : (p, \delta) \in \mathcal{V}, \delta \geq 0, \bar{p} \in \mathcal{V}_p^h\},$$

$$\mathcal{V}_p^h = \left\{ \bar{p} : \sum_{i=1}^N \bar{p}_i = 1, \bar{p}_i \geq 0, i = 1, \dots, N, h \left(\sum_{i \in J} p_i \right) \leq \sum_{i \in J} \bar{p}_i, \forall J \subset \{1, \dots, N\} \right\},$$

and

$$\nu(\alpha)^\top = \left(0, \dots, 0, \text{Lip}(\ell) \|h'_-\|_{\infty} \|\alpha\|_*, \ell(\widehat{y}_1^\top \alpha), \dots, \ell(\widehat{y}_N^\top \alpha) \right) \in \mathbb{R}^{2N+1}.$$

Combing with the outer minimization problem, we obtain the result in (i).

(ii) Similar to the case (i), applying Theorem 5 in Wu et al. (2022), Proposition 10.3 of Denneberg, D. (2013), and the Fenchel duality theory, we obtain the result. $\square$

Proof of Example 3.5. For the full optimal transport model in Section 2.1, we have

$$\mathcal{V} := \{(p, \delta) : \exists \varepsilon \text{ such that } (p, \delta, \varepsilon) \in \mathcal{V}_0\},$$

where

$$\mathcal{V}_0 := \left\{ (p, \delta, \varepsilon) : \frac{1}{\varepsilon} \in \mathcal{N}_\omega, p_i \in \left[0, \frac{\varepsilon}{N}\right], i = 1, \dots, N, \sum_{i=1}^N p_i = 1, \right. \\ \left. \delta = \varepsilon \left(\delta_0 - \frac{1}{N} \sum_{i=m+1}^N d_i \right) - \sum_{i=1}^m p_i d_i + \sum_{i=m+1}^N p_i d_i \right\}.$$

Let

$$\mathcal{V}_+ := \{(p, \delta) : (p, \delta) \in \mathcal{V}, \delta \geq 0\},$$

and

$$\mathcal{V}_+^h := \{(p, \delta, \bar{p}) : (p, \delta) \in \mathcal{V}_+, \bar{p} \in \mathcal{V}_p^h\}.$$

Then in the following we obtain the conjugate function of the indicator function of $\mathcal{V}_+^h$ to have the tractable reformulation for the full optimal transport model. The conjugate function of the indicator function of $\mathcal{V}_+^h$,

$$\sigma^*(v \mid \mathcal{V}_+^h) = \sup_{(p, \delta, \bar{p}) \in \mathcal{V}_+^h} v^\top (p, \delta, \bar{p}) = \sup_{(p, \delta, \varepsilon) \in \mathcal{V}_0, \delta \geq 0, \bar{p} \in \mathcal{V}_p^h} v^\top (p, \delta, \bar{p}) + 0 \cdot \varepsilon.$$

Denote $p'' = (p, \delta, \bar{p}, \varepsilon)$. One can verify that the above problem is a convex optimization problem, and thus we can apply strong duality theory. The Lagrangian function is

$$\begin{aligned} \mathcal{L}(p'') &= (v, 0)^\top p'' + \lambda \left(\sum_{i=1}^N p_i - 1 \right) + \mu \left(\sum_{i=1}^N \bar{p}_i - 1 \right) - \sum_{J \subset [N]} \zeta_J \left(h \left(\sum_{i \in J} p_i \right) - \sum_{i \in J} \bar{p}_i \right) \\ &\quad - \sum_{i=1}^N \tau_i^1 (-p_i) - \sum_{i=1}^N \gamma_i (-\bar{p}_i) - \sum_{i=1}^N \tau_i^2 \left(p_i - \frac{\varepsilon}{N} \right) - \eta_1 \left(\frac{1}{R_{\mathcal{N}_\omega}} - \varepsilon \right) - \eta_2 \left(\varepsilon - \frac{1}{L_{\mathcal{N}_\omega}} \right) \\ &\quad - \xi_1 (-\delta) + \xi_2 \left(\delta - \varepsilon \left(\delta_0 - \frac{1}{N} \sum_{i=m+1}^N d_i \right) + \sum_{i=1}^m p_i d_i - \sum_{i=m+1}^N p_i d_i \right), \end{aligned}$$

where $p^\top = (p_1, \dots, p_N) \in [0, 1]^N$, $q^\top = (q_1, \dots, q_N) \in [0, 1]^N$, $\lambda, \mu, \xi_2 \in \mathbb{R}$, and $\zeta_J, \tau_i^1, \gamma_i, \tau_i^2, \eta_1, \eta_2, \xi_1 \geq 0$. Then we simplify the problem $\sup_{p''} \mathcal{L}(p'')$. One can verify that $\sup_{p''} \mathcal{L}(p'')$ can be decomposed into four optimization problems, which are only related to ε , p , $\bar{p}$, and δ , respectively. For the problem related

to ε with constant terms, we have

$$\begin{aligned} & \sup_{\varepsilon \geq 0} \left\{ \varepsilon \left[-\xi_2 \left(\delta_0 - \frac{1}{N} \sum_{i=m+1}^N d_i \right) + \frac{1}{N} \sum_{i=1}^N \tau_i^2 + \eta_1 - \eta_2 \right] - \frac{\eta_1}{R_{N\omega}} + \frac{\eta_2}{L_{N\omega}} - \lambda - \mu \right\} \\ &= \begin{cases} -\frac{\eta_1}{R_{N\omega}} + \frac{\eta_2}{L_{N\omega}} - \lambda - \mu & \text{if } -\xi_2 \left(\delta_0 - \frac{1}{N} \sum_{i=m+1}^N d_i \right) + \frac{1}{N} \sum_{i=1}^N \tau_i^2 + \eta_1 - \eta_2 \leq 0 \\ \infty & \text{else.} \end{cases} \end{aligned}$$

Let

$$c_i = \begin{cases} \xi_2 d_i, & i = 1, \dots, m \\ -\xi_2 d_i, & i = m+1, \dots, N. \end{cases}$$

For the problem related to p , we have

$$\begin{aligned} & \sup_{p \geq 0} \left\{ \sum_{i=1}^N p_i [v_i + \lambda + \tau_i^1 - \tau_i^2 + c_i] - \sum_{J \subset \{1, \dots, N\}} \zeta_J h \left(\sum_{i \in J} p_i \right) \right\} \\ &= \sup_{p \geq 0, w_1, \dots, w_{2^N-2} \geq 0} \left\{ \sum_{i=1}^N p_i [v_i + \lambda + \tau_i^1 - \tau_i^2 + c_i] - \sum_{j=1}^{2^N-2} \zeta_j h(w_j) |w_j = \sum_{k \in I_j} p_k \right\} \\ &= \inf_{\phi_1, \dots, \phi_{2^N-2}} \sup_{p \geq 0, w_1, \dots, w_{2^N-2} \geq 0} \left\{ \sum_{i=1}^N p_i [v_i + \lambda + \tau_i^1 - \tau_i^2 + c_i] \right. \\ & \quad \left. - \sum_{j=1}^{2^N-2} \zeta_j h(w_j) - \sum_{j=1}^{2^N-2} \phi_j (w_j - \sum_{k \in I_j} p_k) \right\} \\ &= \inf_{\phi_1, \dots, \phi_{2^N-2}} \left\{ \sup_{p \geq 0} \left\{ \sum_{i=1}^N p_i [v_i + \lambda + \tau_i^1 - \tau_i^2 + c_i] + \sum_{j=1}^{2^N-2} \phi_j \sum_{k \in I_j} p_k \right\} \right. \\ & \quad \left. + \sup_{w_1, \dots, w_{2^N-2} \geq 0} \left\{ - \sum_{j=1}^{2^N-2} \zeta_j h(w_j) - \sum_{j=1}^{2^N-2} \phi_j w_j \right\} \right\}. \end{aligned}$$

We find that

$$\begin{aligned} & \sup_{p \geq 0} \left\{ \sum_{i=1}^N p_i [v_i + \lambda + \tau_i^1 - \tau_i^2 + c_i] + \sum_{j=1}^{2^N-2} \phi_j \sum_{k \in I_j} p_k \right\} \\ &= \sup_{p \geq 0} \left\{ \sum_{i=1}^N p_i \left[v_i + \lambda + \tau_i^1 - \tau_i^2 + c_i + \sum_{j: i \in I_j} \phi_j \right] \right\} \\ &= \begin{cases} 0 & \text{if } v_i + \lambda + \tau_i^1 - \tau_i^2 + c_i + \sum_{j: i \in I_j} \phi_j \leq 0, i = 1, \dots, N \\ \infty & \text{else,} \end{cases} \end{aligned}$$

and that

$$\sup_{w_1, \dots, w_{2^N-2} \geq 0} \left\{ - \sum_{j=1}^{2^N-2} \zeta_j h(w_j) - \sum_{j=1}^{2^N-2} \phi_j w_j \right\} = \sum_{j=1}^{2^N-2} \zeta_j h^* \left(\frac{-\phi_j}{\zeta_j} \right).$$

Thus,

$$\begin{aligned} & \sup_{p \geq 0} \left\{ \sum_{i=1}^N p_i [v_i + \lambda + \tau_i^1 - \tau_i^2 + c_i] - \sum_{J \subset \{1, \dots, N\}} \zeta_J h \left(\sum_{i \in J} p_i \right) \right\} \\ &= \inf_{\phi_1, \dots, \phi_{2^N-2}} \sum_{j=1}^{2^N-2} \zeta_j h^* \left(\frac{-\phi_j}{\zeta_j} \right) \\ & \quad \text{s.t. } v_i + \lambda + \tau_i^1 - \tau_i^2 + c_i + \sum_{j: i \in I_j} \phi_j \leq 0, \quad i = 1, \dots, N. \end{aligned}$$

For the problem related to $\bar{p}$, we have

$$\begin{aligned} & \sup_{\bar{p} \geq 0} \left\{ \sum_{i=1}^N \bar{p}_i \left[v_{N+1+i} + \mu + \sum_{j: i \in I_j} \zeta_j + \gamma_i \right] \right\} \\ &= \begin{cases} 0 & \text{if } v_{N+1+i} + \mu + \sum_{j: i \in I_j} \zeta_j + \gamma_i \leq 0, \quad i = 1, \dots, N, \\ \infty & \text{else.} \end{cases} \end{aligned}$$

For the problem related to δ , we have

$$\sup_{\delta \geq 0} \{ \delta^* [v_{N+1} + \xi_1 + \xi_2] \} = \begin{cases} 0 & \text{if } v_{N+1} + \xi_1 + \xi_2 \leq 0, \\ \infty & \text{else.} \end{cases}$$

Then the conjugate function of $\sigma((p, \delta, \bar{p}) | \mathcal{V}_+^h)$, $\sigma^*(v | \mathcal{V}_+^h)$, is equivalent to

$$\begin{aligned}
& \inf -\frac{\eta_1}{R_{N_\omega}} + \frac{\eta_2}{L_{N_\omega}} - \lambda - \mu + \sum_{j=1}^{2^N-2} \zeta_j h^*\left(\frac{-\phi_j}{\zeta_j}\right) \\
& \text{s.t. } \lambda, \mu, \xi_2 \in \mathbb{R}, \{\phi_j\}_{j=1}^{2^N-2} \in \mathbb{R}^{2^N-2}, \{\zeta_j\}_{j=1}^{2^N-2}, \{\tau_i^1\}_{i=1}^N, \{\tau_i^2\}_{i=1}^N, \{\gamma_i\}_{i=1}^N, \eta_1, \eta_2, \xi_1 \geq 0 \\
& -\xi_2(\delta_0 - \frac{1}{N} \sum_{i=m+1}^N d_i) + \frac{1}{N} \sum_{i=1}^N \tau_i^2 + \eta_1 - \eta_2 \leq 0, \\
& v_i + \lambda + \tau_i^1 - \tau_i^2 + c_i + \sum_{j:i \in I_j} \phi_j \leq 0, \quad i = 1, \dots, N, \\
& c_i = \xi_2 d_i, \quad i = 1, \dots, m, \quad c_i = -\xi_2 d_i, \quad i = m+1, \dots, N, \\
& v_{N+1+i} + \mu + \sum_{j:i \in I_j} \zeta_j + \gamma_i \leq 0, \quad i = 1, \dots, N, \\
& v_{N+1} + \xi_1 + \xi_2 \leq 0.
\end{aligned}$$

□

Proof of Example 3.6. For the partial optimal transport model in Section 2.1, we have

$$\mathcal{V} := \left\{ (p, \delta) : \varepsilon = \min_{\beta \in \mathcal{N}_\omega} \beta, p_i \in [0, \frac{1}{N\varepsilon}], i = 1, \dots, N, \sum_{i=1}^N p_i = 1, \delta = \delta_0 - \sum_{i=1}^m p_i d_i \right\}.$$

One can verify that ε is a constant given $\mathcal{N}_\omega$. Then

$$\mathcal{V}_+ := \{(p, \delta) : (p, \delta) \in \mathcal{V}, \delta \geq 0\},$$

and

$$\mathcal{V}_+^h := \{p'' = (p, \delta, \bar{p}) : (p, \delta) \in \mathcal{V}_+, \bar{p} \in \mathcal{V}_p^h\},$$

and we want to obtain the conjugate function of the indicator function of $\mathcal{V}_+^h$,

$$\sigma^*(v | \mathcal{V}_+^h) = \sup_{p'' \in \mathcal{V}_+^h} v^\top p''.$$

One can verify that the above problem is a convex optimization problem, and thus we can apply strong duality theory. Let

$$c'_i = \begin{cases} \xi_2 d_i, & i = 1, \dots, m \\ 0, & i = m+1, \dots, N. \end{cases}$$

Similar to the proof of Example 3.5, the conjugate function of $\sigma((p, \delta, \bar{p})|\mathcal{V}_+^h)$, $\sigma^*(v|\mathcal{V}_+^h)$, is equivalent to

$$\begin{aligned}
& \inf -\lambda - \mu + \frac{1}{N\varepsilon} \sum_{i=1}^N \tau_i^2 - \xi_2 \delta_0 + \sum_{j=1}^{2^N-2} \zeta_j h^*\left(\frac{-\phi_j}{\zeta_j}\right) \\
& \text{s.t. } \alpha \in \mathcal{A}, \lambda, \mu, \xi_2 \in \mathbb{R}, \{\phi_j\}_{j=1}^{2^N-2} \in \mathbb{R}^{2^N-2}, \{\zeta_j\}_{j=1}^{2^N-2}, \{\tau_i^1\}_{i=1}^N, \{\tau_i^2\}_{i=1}^N, \{\gamma_i\}_{i=1}^N, \xi_1 \geq 0, \\
& v_i + \lambda + \tau_i^1 - \tau_i^2 + c'_i + \sum_{j:i \in I_j} \phi_j \leq 0, \quad i = 1, \dots, N, \\
& c'_i = \xi_2 d_i, \quad i = 1, \dots, m, \quad c_i = 0, \quad i = m+1, \dots, N, \\
& v_{N+1+i} + \mu + \sum_{j:i \in I_j} \zeta_j + \gamma_i \leq 0, \quad i = 1, \dots, N, \\
& v_{N+1} + \xi_1 + \xi_2 \leq 0.
\end{aligned}$$

□

Proof of Example 3.7. The robustified conditional mean-CVaR portfolio allocation problem in Nguyen et al. (2024) is as follows:

$$\min_{\alpha \in \mathcal{A}} \sup_{\substack{\mathbb{Q} \in \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}) \\ \mathbb{Q}(X \in \mathcal{N}_\gamma(x_0)) \in \mathcal{N}_\omega}} \text{CVaR}_{\mathbb{Q}}^{1-\kappa} [Y^\top \alpha \mid X \in \mathcal{N}_\gamma(x_0)] - \theta \cdot \mathbb{E}_{\mathbb{Q}} [Y^\top \alpha \mid X \in \mathcal{N}_\gamma(x_0)],$$

where $\mathcal{N}_\omega = [\varepsilon_0, 1]$, $\varepsilon_0 \in (0, 1]$, $\theta \geq 0$, $\kappa \in [0, 1]$,

$$\text{CVaR}_{\mathbb{Q}}^{1-\kappa}(Z) = \frac{1}{\kappa} \int_{1-\kappa}^1 \text{VaR}_{\mathbb{Q}}^\phi(-Z) d\phi,$$

and $\text{VaR}_{\mathbb{Q}}^\phi(Z) = \inf \{z \in \mathbb{R} : \mathbb{Q}(Z \leq z) \geq \phi\}$. Suppose in addition that $\mathcal{X} = \mathbb{R}^{n_1}$, $\mathcal{Y} = \mathbb{R}^{n_2}$, $\mathbb{D}_{\mathcal{X}}(x, \widehat{x}) = \|x - \widehat{x}\|^2$ and $\mathbb{D}_{\mathcal{Y}}(y, \widehat{y}) = \|y - \widehat{y}\|_2^2$. One can verify that the objective function can be reformulated as

$$\begin{aligned}
& \text{CVaR}_{\mathbb{Q}}^{1-\kappa} [Y^\top \alpha \mid X \in \mathcal{N}_\gamma(x_0)] - \theta \cdot \mathbb{E}_{\mathbb{Q}} [Y^\top \alpha \mid X \in \mathcal{N}_\gamma(x_0)] \\
& = (1 + \theta) \rho_{\mathbb{Q}}^h [-Y^\top \alpha \mid X \in \mathcal{N}_\gamma(x_0)],
\end{aligned}$$

where

$$h(x) = \begin{cases} \frac{\theta}{1+\theta} x & x \in [0, 1-\kappa], \\ \frac{1+\theta\kappa}{\kappa(1+\theta)} x - \frac{1-\kappa}{\kappa(1+\theta)} & x \in (1-\kappa, 1]. \end{cases}$$

Applying Theorem 5 in Wu et al. (2022), we have

$$\begin{aligned} & \min_{\alpha \in \mathcal{A}} \sup_{\substack{\mathbb{Q} \in \mathbb{B}_{\delta_0}(\hat{\mathbb{P}}) \\ \mathbb{Q}(X \in \mathcal{N}_\gamma(x_0)) \in \mathcal{N}_\omega}} (1 + \theta) \rho_{\mathbb{Q}}^h [-Y^\top \alpha \mid X \in \mathcal{N}_\gamma(x_0)] \\ &= \min_{\alpha \in \mathcal{A}} \sup_{(p, \delta) \in \mathcal{V}_+} (1 + \theta) \left[\rho_{\sum_{i=1}^N p_i \mathbb{I}_{\hat{\mathcal{Y}}_i}}^h (-Y^\top \alpha) + \|h'_-\|_2 \delta(p, \varepsilon)^{\frac{1}{2}} \|\alpha\|_2 \right], \end{aligned}$$

where $\|h'_-\|_2 = \left(1 + \frac{1-\kappa}{\kappa(1+\theta)^2}\right)^{\frac{1}{2}}$. Using the minimization form of CVaR and minimax theorem, we reformulate the inner problem of the above problem as

$$\begin{aligned} & \sup_{(p, \delta) \in \mathcal{V}_+} (1 + \theta) \left[\rho_{\sum_{i=1}^N p_i \mathbb{I}_{\hat{\mathcal{Y}}_i}}^h (-Y^\top \alpha) + \|h'_-\|_2 \delta^{\frac{1}{2}} \|\alpha\|_2 \right] \\ &= \sup_{(p, \delta) \in \mathcal{V}_+} \inf_{t \in \mathbb{R}} \left\{ \mathbb{E}_{\sum_{i=1}^N p_i \mathbb{I}_{\hat{\mathcal{Y}}_i}} \left[t + \frac{1}{\kappa} (-Y^\top \alpha - t)_+ - \theta \cdot Y^\top \alpha \right] + (1 + \theta) \|h'_-\|_2 \delta^{\frac{1}{2}} \|\alpha\|_2 \right\} \\ &= \inf_{t \in \mathbb{R}} \sup_{(p, \delta) \in \mathcal{V}_+} \left\{ \mathbb{E}_{\sum_{i=1}^N p_i \mathbb{I}_{\hat{\mathcal{Y}}_i}} \left[t + \frac{1}{\kappa} (-Y^\top \alpha - t)_+ - \theta \cdot Y^\top \alpha \right] + (1 + \theta) \|h'_-\|_2 \delta^{\frac{1}{2}} \|\alpha\|_2 \right\}. \end{aligned}$$

Applying the Fenchel duality theory, the problem (22) is equivalent to

$$\begin{aligned} & \inf_{\alpha \in \mathcal{A}, t \in \mathbb{R}, v} \sigma^*(v \mid \mathcal{V}_+) + C_{\theta, \kappa} v_{N+1} \left\| \frac{\alpha}{v_{N+1}} \right\|_2^2 \\ & \text{s.t. } v_i \geq \ell(\hat{\mathcal{Y}}_i^\top \alpha, t), i = 1, \dots, N, \\ & v_{N+1} \geq 0, \end{aligned}$$

where $C_{\theta, \kappa} = \frac{1}{4} \left((1 + \theta)^2 + \frac{1-\kappa}{\kappa} \right)$, $\ell(y^\top \alpha, t) = t + \frac{1}{\kappa} (-y^\top \alpha - t)_+ - \theta \cdot y^\top \alpha$, and $\sigma^*(v \mid \mathcal{V}_+)$, the conjugate function of $\sigma(v \mid \mathcal{V}_+)$, is given in Example 3.2. Thus, substituting the $\sigma^*(v \mid \mathcal{V}_+)$ in Example 3.2, the robustified conditional mean-CVaR portfolio allocation problem (22), proposed in Nguyen et al. (2024), can be reformulated as

$$\begin{aligned} & \inf_{\alpha, t, v, z} z^\top b + C_{\theta, \kappa} v_{N+1} \left\| \frac{\alpha}{v_{N+1}} \right\|_2^2 \\ & \text{s.t. } \alpha \in \mathcal{A}, t \in \mathbb{R}, v - v(\alpha, t) \geq 0, A^\top z \geq v, z \geq 0. \end{aligned}$$

And thus, it is equivalent to

$$\begin{aligned} & \inf_{\alpha \in \mathcal{A}, t \in \mathbb{R}, z \geq 0, v_{N+1}, s} z^\top b + s \\ & \text{s.t. } v_{N+1} \geq 0, A^\top z \geq (\ell(\hat{\mathcal{Y}}_1^\top \alpha, t), \dots, \ell(\hat{\mathcal{Y}}_N^\top \alpha, t), v_{N+1}, 0)^\top, \\ & s \geq C_{\theta, \kappa} v_{N+1} \left\| \frac{\alpha}{v_{N+1}} \right\|_2^2, \end{aligned}$$

where $A \in \mathbb{R}^{(N+6) \times (N+2)}$, and $b \in \mathbb{R}^{N+6}$ are defined by

$$A = \begin{pmatrix} 1 & \dots & 1 & 0 & 0 \\ -1 & \dots & -1 & 0 & 0 \\ 0 & \dots & 0 & 0 & 1 \\ 0 & \dots & 0 & 0 & -1 \\ 1 & \dots & 0 & 0 & -\frac{1}{N} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & 1 & 0 & -\frac{1}{N} \\ -a_1 & \dots & -a_N & -a_{N+1} & -a_{N+2} \\ a_1 & \dots & a_N & a_{N+1} & a_{N+2} \end{pmatrix}, \quad \text{with } a = \begin{pmatrix} -d_1 \\ \vdots \\ -d_m \\ d_{m+1} \\ \vdots \\ d_N \\ -1 \\ \delta_0 - \frac{1}{N} \sum_{i=m+1}^N d_i \end{pmatrix}, \quad \text{and } b = \begin{pmatrix} 1 \\ -1 \\ \frac{1}{\varepsilon_0} \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

□

EC.3. Full Reformulation for the Conditional Expectation Case Based on Admissible Sets from Theorems 1 and 2, and Comparison with Nguyen et al. (2024)

In this section, we present full reformulations of the conditional expectation problem under our union-ball framework, leveraging the admissible sets characterized in Theorems 1 and 2. These yield tractable convex programs for the full and partial optimal transport models, respectively. While the primary focus is on developing these reformulations, we also include a brief comparison with the formulation proposed in Nguyen et al. (2024), highlighting key structural differences in terms of tractability, compactness, and interpretability.

(A) Note that full optimal transport model (4) given by Nguyen et al. (2024) is equivalent to problem (15) with set $\mathcal{V}$ given in Theorem 3. Applying the conjugate of the indicator function of $\mathcal{V}_+$ in Example 3.2 to Propositions 10 and 11, we have the following results.

(i) Let $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}$ for any $y_1, y_2 \in \mathcal{Y}$ and $\ell : \mathbb{R} \rightarrow \mathbb{R}$ be a convex and Lipschitz continuous function. The problem (15) can be solved by the convex program

$$\begin{aligned} & \inf_{\alpha \in \mathcal{A}, z \geq 0} z^\top b \\ & \text{s.t. } [A^\top z]_i \geq \ell(\widehat{y}_i^\top \alpha), \quad i = 1, \dots, N, \\ & \quad [A^\top z]_{N+1} \geq \text{Lip}(\ell) \|\alpha\|_*, \\ & \quad [A^\top z]_{N+2} \geq 0. \end{aligned}$$

Note that the objective function of above reformulation is linear. Moreover, given α , this formulation has $N + 6$ variables, $2N + 8$ linear constraints and 1 convex constraint.

(ii) Let $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}^q$ for any $y_1, y_2 \in \mathcal{Y}$ and $q \in (1, \infty)$.

(1) Assume the loss function ℓ takes one of the following two forms, multiplied by $C > 0$:

- (a) $\ell_1(x) = x + b$ or $\ell_1(x) = -x + b$ with some $b \in \mathbb{R}$;
(b) $\ell_2(x) = |x - b_1| + b_2$ with some $b_1, b_2 \in \mathbb{R}$.

The problem (15) can be solved by the convex program

$$\begin{aligned} & \inf_{\alpha \in \mathcal{A}, v_{N+1} \geq 0, z \geq 0, s} z^\top b + s \\ & \text{s.t. } [A^\top z]_i \geq \ell(\widehat{y}_i^\top \alpha), \quad i = 1, \dots, N, \\ & \quad [A^\top z]_{N+1} \geq v_{N+1}, \\ & \quad [A^\top z]_{N+2} \geq 0, \\ & \quad s \geq C_q v_{N+1} \left\| \frac{\alpha}{v_{N+1}} \right\|_*^{\frac{q}{q-1}}, \end{aligned}$$

where $C_q = C^{\frac{q}{q-1}} \left(q^{\frac{1}{1-q}} - q^{\frac{q}{1-q}} \right)$. Note that the objective function of above reformulation is linear. Moreover, given α , this formulation has $N + 8$ variables, $2N + 9$ linear constraints and 1 convex constraint.

- (2) Assume the loss function ℓ is of the form $\ell(x) = (C \cdot \ell_i(x))^q$, where $C > 0$, $q \in (1, \infty)$, and ℓ_i is one of the following:

- (a) $\ell_1(x) = (x - b)_+$ with some $b \in \mathbb{R}$;
(b) $\ell_2(x) = (x - b)_-$ with some $b \in \mathbb{R}$;
(c) $\ell_3(x) = (|x - b_1| - b_2)_+$ with some $b_1 \in \mathbb{R}$ and $b_2 \geq 0$;
(d) $\ell_4(x) = |x - b_1| + b_2$ with some $b_1 \in \mathbb{R}$ and $b_2 > 0$.

The problem (15) can be solved by the convex program

$$\begin{aligned} & \inf_{\alpha \in \mathcal{A}, v_{N+1} \geq 0, y \geq 0, z \geq 0, s} [z^\top b + C_q^1 y + s]^q \\ & \text{s.t. } [A^\top z]_i \geq \frac{\ell(\widehat{y}_i^\top \alpha)}{y^{q-1}}, \quad i = 1, \dots, N, \\ & \quad [A^\top z]_{N+1} \geq v_{N+1}, \\ & \quad [A^\top z]_{N+2} \geq 0, \\ & \quad s \geq C_q v_{N+1} \left\| \frac{\alpha}{v_{N+1}} \right\|_*^{\frac{q}{q-1}}, \end{aligned}$$

where $C_q^1 = q^{\frac{1}{1-q}} - q^{\frac{q}{1-q}}$ and $C_q = C^{\frac{q}{q-1}} \left(q^{\frac{1}{1-q}} - q^{\frac{q}{1-q}} \right)$. And thus the problem (15) can be

solved by the convex program

$$\begin{aligned}
& \inf_{\alpha \in \mathcal{A}, v_{N+1} \geq 0, y \geq 0, z \geq 0, s} z^\top b + C_q^1 y + s \\
& \text{s.t. } [A^\top z]_i \geq \frac{\ell(\widehat{y}_i^\top \alpha)}{y^{q-1}}, \quad i = 1, \dots, N, \\
& [A^\top z]_{N+1} \geq v_{N+1}, \\
& [A^\top z]_{N+2} \geq 0, \\
& s \geq C_q v_{N+1} \left\| \frac{\alpha}{v_{N+1}} \right\|_*^{\frac{q}{q-1}}.
\end{aligned}$$

Note that the objective function of above reformulation is linear. Moreover, given α , this formulation has $N + 9$ variables, $N + 10$ linear constraints and $N + 1$ convex constraint.

- (iii) Let $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}^q$ for any $y_1, y_2 \in \mathcal{Y}$ and $q \in (1, \infty)$. Assuming the loss function ℓ is convex and satisfies $\ell(x) - \ell(x_0) \leq L|x - x_0|^q + M, x \in \mathbb{R}$ for some $L, M > 0$, and some $x_0 \in \mathbb{R}$, the problem (15) can be solved by the convex program

$$\begin{aligned}
& \inf_{\alpha \in \mathcal{A}, \eta \geq 0, z \geq 0} z^\top b \\
& \text{s.t. } [A^\top z]_i \geq \sup_{z \in \mathbb{R}} \left\{ z \cdot \widehat{y}_i^\top \alpha - \ell^*(z) + \eta \cdot \frac{|z|^{q^*}}{q^* q^{q^*-1}} \right\}, \quad i = 1, \dots, N, \\
& [A^\top z]_{N+1} \geq \|\alpha\|_*^q / \eta^{q-1}, \\
& [A^\top z]_{N+2} \geq 0.
\end{aligned} \tag{44}$$

Note that the objective function of above reformulation is linear. Moreover, given α , this formulation has $N + 7$ variables, $N + 8$ linear constraints and $N + 1$ convex constraint.

- (B) Note that partial optimal transport model (5) given by Esteban-Perez and Morales (2022) is equivalent to problem (15) with set $\mathcal{V}$ given in Theorem 6. Applying the conjugate of the indicator function of $\mathcal{V}_+$ in Example 3.3 to Propositions 10 and 11, we have the following results.

- (i) Let $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}$ for any $y_1, y_2 \in \mathcal{Y}$ and $\ell : \mathbb{R} \rightarrow \mathbb{R}$ be a convex and Lipschitz continuous function. The problem (15) can be solved by the convex program

$$\begin{aligned}
& \inf_{\alpha \in \mathcal{A}, z \geq 0} z^\top b' \\
& \text{s.t. } [(A')^\top z]_i \geq \ell(\widehat{y}_i^\top \alpha), \quad i = 1, \dots, N, \\
& [(A')^\top z]_{N+1} \geq \text{Lip}(\ell) \|\alpha\|_*.
\end{aligned}$$

- (ii) Let $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}^q$ for any $y_1, y_2 \in \mathcal{Y}$ and $q \in (1, \infty)$.

- (1) Assume the loss function ℓ takes one of the following two forms, multiplied by $C > 0$:

- (a) $\ell_1(x) = x + b$ or $\ell_1(x) = -x + b$ with some $b \in \mathbb{R}$;

(b) $\ell_2(x) = |x - b_1| + b_2$ with some $b_1, b_2 \in \mathbb{R}$.

The problem (15) can be solved by the convex program

$$\begin{aligned} \inf_{\alpha \in \mathcal{A}, v_{N+1} \geq 0, z \geq 0} \quad & z^\top b' + C_q v_{N+1} \left\| \frac{\alpha}{v_{N+1}} \right\|_*^{\frac{q}{q-1}} \\ \text{s.t.} \quad & [(A')^\top z]_i \geq \ell(\widehat{y}_i^\top \alpha), \quad i = 1, \dots, N, \\ & [(A')^\top z]_{N+1} \geq v_{N+1}, \end{aligned}$$

where $C_q = C^{\frac{q}{q-1}} \left(q^{\frac{1}{1-q}} - q^{\frac{q}{1-q}} \right)$.

(2) Assume the loss function ℓ is of the form $\ell(x) = (C \cdot \ell_i(x))^q$, where $C > 0$, $q \in (1, \infty)$, and ℓ_i is one of the following:

- (a) $\ell_1(x) = (x - b)_+$ with some $b \in \mathbb{R}$;
- (b) $\ell_2(x) = (x - b)_-$ with some $b \in \mathbb{R}$;
- (c) $\ell_3(x) = (|x - b_1| - b_2)_+$ with some $b_1 \in \mathbb{R}$ and $b_2 \geq 0$;
- (d) $\ell_4(x) = |x - b_1| + b_2$ with some $b_1 \in \mathbb{R}$ and $b_2 > 0$.

The problem (15) can be solved by the convex program

$$\begin{aligned} \inf_{\alpha \in \mathcal{A}, v_{N+1} \geq 0, y \geq 0, z \geq 0} \quad & \left[z^\top b' + C_q^1 y + C_q v_{N+1} \left\| \frac{\alpha}{v_{N+1}} \right\|_*^{\frac{q}{q-1}} \right]^q \\ \text{s.t.} \quad & [(A')^\top z]_i \geq \frac{\ell(\widehat{y}_i^\top \alpha)}{y^{q-1}}, \quad i = 1, \dots, N, \\ & [(A')^\top z]_{N+1} \geq v_{N+1}, \end{aligned}$$

where $C_q^1 = q^{\frac{1}{1-q}} - q^{\frac{q}{1-q}}$ and $C_q = C^{\frac{q}{q-1}} \left(q^{\frac{1}{1-q}} - q^{\frac{q}{1-q}} \right)$.

(iii) Let $\mathbb{D}_{\mathcal{Y}}(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}^q$ for any $y_1, y_2 \in \mathcal{Y}$ and $q \in (1, \infty)$. Assuming the loss function ℓ is convex and satisfies $\ell(x) - \ell(x_0) \leq L|x - x_0|^q + M$, $x \in \mathbb{R}$ for some $L, M > 0$, and some $x_0 \in \mathbb{R}$, the problem (15) can be solved by the convex program

$$\begin{aligned} \inf_{\alpha \in \mathcal{A}, \eta \geq 0, z \geq 0} \quad & z^\top b' \\ \text{s.t.} \quad & [(A')^\top z]_i \geq \sup_{z \in \mathbb{R}} \left\{ z \cdot \widehat{y}_i^\top \alpha - \ell^*(z) + \eta \cdot \frac{|z|^{q^*}}{q^* q^{q^*-1}} \right\}, \quad i = 1, \dots, N, \\ & [(A')^\top z]_{N+1} \geq \|\alpha\|_*^q / \eta^{q-1}. \end{aligned}$$

Remark 31. In what follows, we present a more detailed discussion of the full optimal transport model in comparison with the analysis of Nguyen et al. (2024). Specifically, for the case with expectation as the risk measure and $\mathcal{N}_\omega = [\varepsilon, 1]$ for some $\varepsilon \in (0, 1]$, Nguyen et al. (2024) were the first to study this model and

established the following equivalence:

$$\begin{aligned} & \inf_{\alpha \in \mathcal{A}} \sup_{\substack{Q \in \mathbb{B}_{\delta_0}(\widehat{\mathbb{P}}) \\ Q(X \in \mathcal{N}_\gamma(x_0)) \in [\varepsilon, 1]}} \mathbb{E}_Q [\ell(Y, \alpha) \mid X \in \mathcal{N}_\gamma(x_0)] \\ &= \begin{cases} \inf & \phi + (N\varepsilon)^{-1} v^+ - N^{-1} v^- \\ \text{s.t.} & \alpha \in \mathcal{A}, (\lambda, s, v^+, v^-, \phi, \varphi, \psi) \in \mathcal{U} \\ & s_i \geq \sup_{y_i \in \mathcal{Y}} \{\ell(y_i, \alpha) - \lambda_i \mathbb{D}_{\mathcal{Y}}(y_i, \widehat{y}_i)\} \quad i = 1, \dots, N, \end{cases} \end{aligned}$$

where

$$\mathcal{U} \triangleq \left\{ \begin{array}{ll} (\lambda, s, v^+, v^-, \phi, \varphi, \psi) \in \mathbb{R}_+^N \times \mathbb{R}^N \times \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+^N & \text{such that:} \\ \phi + d_i \varphi + \psi_i - s_i \geq 0 & i = 1, \dots, m \\ \phi - d_i \varphi + \psi_i - s_i \geq 0 & i = m + 1, \dots, N \\ v^+ - v^- + \left(\sum_{i=m+1}^N d_i - N\delta_0 \right) \varphi - \sum_{i=1}^N \psi_i \geq 0 & \\ \varphi - \lambda_i \geq 0 & i = 1, \dots, N \end{array} \right\}.$$

Note that the objective function of the above reformulation is linear. Moreover, given α , this formulation has $3N + 4$ variables, $4N + 4$ linear constraints, and N convex constraints. A further comparison with Nguyen et al. (2024) yields the following observations.

- (i) Although our formulations based on the projection property of Wu et al. (2022) also involve the unsolved supremum problem in constraint (44), this problem is one-dimensional, whereas in Nguyen et al. (2024) it is high-dimensional, resulting in substantially greater complexity.
- (ii) For specific loss functions—such as the mean–variance and mean–CVaR loss functions studied in Nguyen et al. (2024)—the N supremum problems in the constraints (44) admit explicit forms, which reduce to SOC constraints. Compared with Nguyen et al. (2024), our reformulation is considerably simpler and more tractable, requiring simpler constraints (SOC constraints) while reformulation in Nguyen et al. (2024) requires more complex constraint (SDP constraints). See Remark 16 for a detailed comparison under the mean–variance loss function with the squared ℓ_2 -norm, and Remark 23 for the corresponding comparison under the mean–CVaR loss function with the squared ℓ_2 -norm.
- (iii) Our reformulations are considerably simpler largely thanks to the union-ball characterization, which leverages the structural properties of standard Wasserstein DRO and requires invoking duality only once, introducing a single multiplier η . By contrast, the analysis in Nguyen et al. (2024), which overlooks this structural property and instead relies on a more cumbersome break-and-conquer approach, applies duality repeatedly—most clearly reflected in the N multipliers $\lambda \in \mathbb{R}_+^N$ that grow with N .