Globalized Distributionally Robust Counterpart

Feng Liu[1], Zhi Chen[2], and Shuming Wang[1]
[1]School of Economics and Management, University of Chinese Academy of Sciences
liufeng21@mails.ucas.ac.cn, wangshuming@ucas.edu.cn
[2]Department of Management Sciences, College of Business, City University of Hong Kong
zhi.chen@cityu.edu.hk

We extend the notion of globalized robustness to consider distributional information beyond the support of the ambiguous probability distribution. We propose the globalized distributionally robust counterpart that disallows any (resp., allows limited) constraint violation for distributions residing (resp., not residing) in the ambiguity set. By varying its inputs, our proposal recovers several existing perceptions of parameter uncertainty. Focusing on the type-1 Wasserstein distance, we show that the globalized distributionally robust counterpart has an insightful interpretation in terms of *shadow price of globalized robustness*, and it can be seamlessly integrated with many popular optimization models under uncertainty without incurring any extra computational cost. Such computational attractiveness also holds for other ambiguity sets, including the ones based on probability metric, optimal transport, ϕ-divergences, or moment conditions. Numerical studies on an adaptive network lot-sizing problem demonstrate the modeling flexibility of our proposal and its emphases on globalized robustness to constraint violation.

Key words: Robust and distributionally robust optimization; robust satisficing; globalized robustness.
History: September 13, 2022

1. Introduction

Many powerful modeling paradigms have been proposed to inform decisions to real-world problems subject to ubiquitous parameter uncertainty. Although typically modeling uncertainty as a (possibly multi-dimensional) random variable governed by some probability distribution, these models have their own distinctive perception when facing uncertainty.

The modeling paradigm of *distributionally robust optimization* considers an ambiguity set—a family of probability distributions with limited yet common distributional information—and evaluates the decision's performance according to its worst-case expected performance with respect to any distribution that may arise from the ambiguity set. When the ambiguity set contains only the support information, it reduces to an uncertainty set and distributionally robust optimization's perception of uncertainty recovers that of traditional *robust optimization* modeling paradigm (see, *e.g.*, Ben-Tal and Nemirovski 1998, Bertsimas and Sim 2004, El Ghaoui et al. 1998, Soyster 1973), which focuses on the worst-case scenario in the uncertainty set. When the ambiguity set is a singleton, the perception coincides with that of *stochastic programming* modeling paradigm, which takes a specific probability distribution as the ground truth and focuses on the expected performance under it. Apart from the two mentioned, many other types of distributional information can be characterized in the ambiguity set, including moment conditions (Chen et al. 2019, Delage and Ye 2010, Goh and Sim 2010, Wiesemann et al. 2014), structural properties (Popescu 2005, Hanasusanto et al. 2015, Li et al. 2019), and statistical metrics such as ϕ-divergences (Ben-Tal et al. 2013, Gotoh et al. 2018, Jiang and Guan 2018, Wang et al. 2016) and the Wasserstein distance (Blanchet and Murthy 2019, Gao and Kleywegt 2022, Mohajerin Esfahani and Kuhn 2018, Pflug and Wozabal 2007).

All aforementioned traditional robust and distributionally robust modeling paradigms, as well as the stochastic programming one, impose strict constraint satisfaction (*i.e.*, no constraint violation) for any probability distribution in the ambiguity set. In stark contrast, the emerging *robust satisficing* modeling paradigm (Long et al. 2022, Ramachandra et al. 2021) softly allows possible constraint violation as long as it could be controlled by a tolerance level that is to be optimized. One may conceptually view such controllable constraint violation to share the same spirit of *globalized robust optimization* (see the pioneering works of Ben-Tal et al. 2006 and Ben-Tal et al. 2017), wherein the tolerance level is pre-specified. One of the most key features of the robust satisficing modeling paradigm is to limited constraint violation also for distributions beyond the ambiguity set. This, however, comes with bearing possible constraint violation for distributions in the ambiguity set.

In this paper, we are motivated by a unified framework to combine perceptions of (*i*) as in distributionally robust optimization, strictly no constraint violation for distributions in the ambiguity set and (*ii*) as in globalized robust optimization and/or robust satisficing, softly controllable constraint violation for distributions not in the ambiguity set. In particular, we extend the notion of globalized robustness to include distributional information beyond support information and propose a *globalized distributionally robust counterpart* with two parameters—one is to size the ambiguity set wherein no constraint violation is allowed and another is to control the tolerance level of acceptable constraint violation when not in the ambiguity set. Quite notably, the globalized distributionally robust counterpart usually comes with no extra computational cost and is readily to be integrated with several optimization models for addressing decision-making under uncertainty.

The key contributions of our paper may be summarized as follows.

1. We propose a globalized distributionally robust counterpart that guarantees no constraint violation for any distribution in a pre-specific ambiguity set while at the same time, allowing (in a controllable manner) possible constraint violation for distributions that are not necessarily

in the ambiguity set. By varying its inputs, the globalized distributionally robust counterpart recovers several existing perceptions for addressing parameter uncertainty, including the traditional robust and distributionally robust counterparts, stochastic counterpart, robust satisficing counterpart, as well as the original globalized robust counterpart based solely on the support information.

2. Focusing on the type-1 Wasserstein distance and leveraging strong duality to obtain a dual reformulation, we discover that the globalized distributionally robust counterpart merely further imposes an upper bound on the single-dimensional decision variable in the dual. Hence, the notion of globalized robustness condenses to a *shadow price of globalized robustness* on the dual decision variable.
3. At the same level of computational complexity, the globalized distributionally robust counterpart can be seamlessly incorporated into many popular optimization models under uncertainty, including distributionally robust optimization, robust satisficing, chance constrained program, as well as two-stage adaptive linear optimization.
4. We also extend the globalized distributionally robust counterpart to other ambiguity sets based on the probability metric, optimal transport, ϕ-divergences, and moment conditions. Compared to the corresponding distributionally robust counterpart, our proposed globalized variant typically does not incur any additional computational cost.

The reminder of this paper is organized as follows. Section 2 formally introduces the globalized distributionally robust counterpart and offers observations on how it can recover perceptions taken by the existing modeling paradigms. Section 3 focuses on the type-1 Wasserstein distance and shows a dual representation that offers an insightful interpretation of the notion of the globalized robustness and leads to a tractable reformulation under mild conditions. We present finite-sample performance guarantees of the globalized distributionally robust counterpart and demonstrate its potential applications to several optimization problems under uncertainty. Section 4 extends the globalized distributionally robust counterpart to other ambiguity sets based on and Section 5 presents numerical experiments on a lot-sizing problem. All proofs are delegated to Appendix EC.1.

Notation. We use $\tilde{\boldsymbol{\xi}} \sim \mathbb{P}$ to denote a random variable governed by the probability distribution $\mathbb{P}$, probability and expectation with respect to whom are $\mathbb{P}[\cdot]$ and $\mathbb{E}_{\mathbb{P}}[\cdot]$, respectively. The set of probability distributions supported on $\Xi \subseteq \mathbb{R}^I$ is $\mathcal{P}(\Xi)$. The shorthand $[N] = \{1, \ldots, N\}$ stands for the set of all positive integers up to N. We denote by $\|\cdot\|_*$ the dual norm of a general norm $\|\cdot\|$. The convex conjugate of a proper function $f : \mathbb{R}^I \mapsto \mathbb{R}$ is defined as $f^*(\boldsymbol{x}) = \sup\{\boldsymbol{x}^\top \boldsymbol{\xi} - f(\boldsymbol{\xi}) \mid \boldsymbol{\xi} \in \mathbb{R}^I\}$, and we use $f^{\bar{*}}(\boldsymbol{x}, \boldsymbol{\xi})$ with a bar sign to emphasize the partial convex conjugate with respect to the second part of arguments, $\boldsymbol{\xi}$. The indicator function of a set Ξ is defined through $\delta(\boldsymbol{\xi} \mid \Xi) = 0$ if $\boldsymbol{\xi} \in \Xi$; $= \infty$ otherwise, and the support function is defined as $\delta^*(\boldsymbol{x} \mid \Xi) = \sup\{\boldsymbol{x}^\top \boldsymbol{\xi} \mid \boldsymbol{\xi} \in \Xi\}$.

2. Model

In this paper we focus on parameter uncertainty in a (possibly nonlinear) constraint

$$f(\boldsymbol{x},\boldsymbol{\xi}) \leq 0,$$

where $\boldsymbol{x} \in \mathcal{X} \subseteq \mathbb{R}^K$ is the vector of decision variables and $\boldsymbol{\xi} \in \Xi \subseteq \mathbb{R}^I$ is the vector of uncertain parameters. For any fixed decision $\boldsymbol{x}$, the cost function $f(\boldsymbol{x},\boldsymbol{\xi})$ is proper with respect to the uncertainty $\boldsymbol{\xi}$ that follows an unknown probability distribution. In particular, we propose the a *globalized distributionally robust counterpart*

$$\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})] \leq \gamma \cdot \min_{\mathbb{Q}\in\mathcal{F}} d(\mathbb{P},\mathbb{Q}) \quad \forall \mathbb{P} \in \mathcal{P}(\Xi), \tag{1}$$

which involves the following inputs: (*i*) a tolerance level[1] $\gamma \geq 0$, (*ii*) a nonnegative statistical metric d that measures the proximity between probability distributions and satisfies $d(\mathbb{P},\mathbb{Q}) = 0$ if $\mathbb{P} = \mathbb{Q}$, and (*iii*) a prescribed ambiguity set $\mathcal{F} \subseteq \mathcal{P}(\Xi)$. Here, the ambiguity set $\mathcal{F}$, when properly defined, could be sized with a parameter $\theta \geq 0$ to achieve the adjustable robustness. Decomposing the globalized distributionally robust counterpart (1) into

$$\begin{cases} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})] \leq 0 & \forall \mathbb{P} \in \mathcal{F} \\ \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})] \leq \gamma \cdot \min\limits_{\mathbb{Q}\in\mathcal{F}} d(\mathbb{P},\mathbb{Q}) & \forall \mathbb{P} \in \mathcal{P}(\Xi) \setminus \mathcal{F}, \end{cases}$$

it is then clear that for any distribution within the ambiguity set $\mathcal{F}$, the the model guarantees *strict inner robustness* to no constraint violation; on the other hand, for distributions not residing in $\mathcal{F}$, the tolerance level γ ensures *soft outer robustness* that limits possible constraint violation to a proportion of the probability distribution's distance to the ambiguity set $\mathcal{F}$. Integrating these two, the globalized distributionally robust counterpart (1) provides, for *all* distribution $\mathbb{P} \in \mathcal{P}(\Xi)$, globalized robustness in terms of limited constraint violation.

It is worth noting that by varying its inputs, the globalized distributionally robust counterpart (1) recovers several existing perceptions when facing parameter uncertainty, summarized in the following observations.

Observation 1. *When $\gamma = 0$, constraint* (1) *reduces to the traditional robust counterpart:*

$$f(\boldsymbol{x},\boldsymbol{\xi}) \leq 0 \quad \forall \boldsymbol{\xi} \in \Xi.$$

Observation 2. *When $\gamma \to \infty$, constraint* (1) *becomes the distributionally robust counterpart:*

$$\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})] \leq 0 \quad \forall \mathbb{P} \in \mathcal{F}.$$

[1] As we shall see subsequently, the tolerance level can be interpreted as a shadow price of globalized robustness.

OBSERVATION 3. *When* $\gamma \to \infty$ *and* $\mathcal{F} = \{\hat{\mathbb{P}}\}$, *constraint* (1) *reduces to the stochastic counterpart under the reference distribution* $\hat{\mathbb{P}}$:

$$\mathbb{E}_{\hat{\mathbb{P}}}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] \le 0.$$

OBSERVATION 4. *When* $\mathcal{F} = \{\hat{\mathbb{P}}\}$, *constraint* (1) *coincides with the robust satisficing counterpart proposed by Long et al. (2022):*

$$\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] \le \gamma \cdot d(\mathbb{P}, \hat{\mathbb{P}}) \quad \forall \mathbb{P} \in \mathcal{P}(\Xi).$$

OBSERVATION 5. *With proper specifications*[2] *of its inputs, constraint* (1) *recovers the original globalized robust counterpart advocated by Ben-Tal et al. (2006) and Ben-Tal et al. (2017):*

$$f(\boldsymbol{x}, \boldsymbol{\xi}) \le \gamma \cdot \min_{\boldsymbol{\zeta} \in \mathcal{U}_1} c_1(\boldsymbol{\xi}, \boldsymbol{\zeta}) \quad \forall \boldsymbol{\xi} \in \mathcal{U}_2,$$

where $\mathcal{U}_1 \subseteq \mathcal{U}_2$ *and* $c_1(\boldsymbol{\xi}, \boldsymbol{\zeta})$ *measures the nonnegative distance between the parameters* $\boldsymbol{\xi}$ *and* $\boldsymbol{\zeta}$.

These observations showcase the rich modeling flexibility of the globalized distributionally robust counterpart (1). Apart from such expressiveness, the globalized distributionally robust counterpart, with proper choices of the statistical metric d and ambiguity set $\mathcal{F}$, also admits an attractive tractable reformulation; see more detail in the coming sections. In the remainder of this section, we materialize the globalized robustness of model (1) via the following two illustrative examples.[3]

EXAMPLE 1. *Consider a distributionally robust optimization problem*

$$\begin{array}{ll} \min & x \\ \text{s.t.} & \mathbb{E}_{\mathbb{P}}[\ln \tilde{\xi} - x] \le 0 \quad \forall \mathbb{P} \in \mathcal{F}_{\mathrm{W}}(\theta) \\ & x \in \mathbb{R} \end{array} \tag{2}$$

with a Wasserstein ball $\mathcal{F}_{\mathrm{W}}(\theta)$[4] *centering at* $\hat{\mathbb{P}} = \delta_{0.5}$ *and* $\theta = 0.5$. *The optimal value is* 0 *and is attained at* $x^\star = 0$. *Consider next the globalized variant of* (2) *with some* $0 < \gamma \le 1$:

$$\begin{array}{ll} \min & x \\ \text{s.t.} & \mathbb{E}_{\mathbb{P}}[\ln \tilde{\xi} - x] \le \gamma \cdot d_{\mathrm{W}}(\mathbb{P}, \mathbb{Q}) \quad \forall \mathbb{P} \in \mathcal{P}(\mathbb{R}_{++}),\ \mathbb{Q} \in \mathcal{F}_{\mathrm{W}}(\theta) \\ & x \in \mathbb{R}. \end{array} \tag{3}$$

Here, d_{W} *is the type-*1 *Wasserstein distance that will be formally introduced in Section 3.1. The optimal value of problem* (3) *is* $\gamma - \ln\gamma - 1$ *with* $x^\star = \gamma - \ln\gamma - 1$. *Comparing the two problems, it is clear that (i) when* $\gamma \ge 1$, *problem* (3) *coincides with problem* (2), *attaining the same optimal value of zero; (ii) as* γ *gradually decreases, an additional cost of* $-\ln\gamma + \gamma - 1$ *is paid for the globalized robustness; (iii) as* γ *approaches to* 0, *problem* (3) *approaches to a traditional robust optimization model that imposes* $x \ge \ln \xi\ \forall \xi > 0$, *which results in a cost of infinity.*

[2] Detailed derivations are relegated to Appendix EC.2.

[3] Detailed derivations are relegated to Appendix EC.3.

[4] Please refer to Section 3.1 for the formal definition.

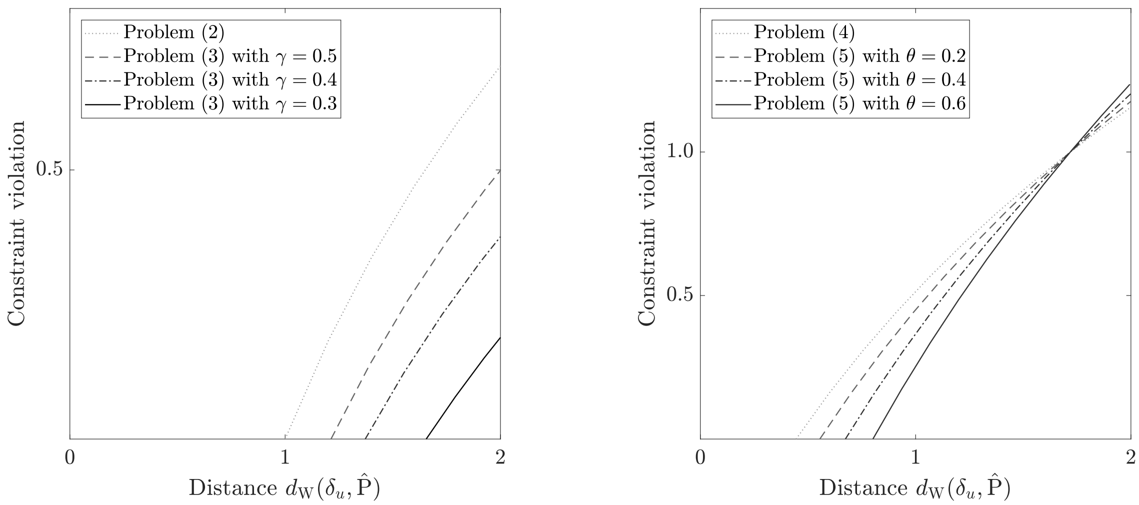

Figure 1 Constraint violation under δ_u that moves away from the center $\hat{\mathbb{P}}$: comparison between problems (2) and (3) on the left panel, and between problems (4) and (5) on the right panel.

To showcase the merit of globalized robustness of problem (3)*, in the left panel of Figure 1, we plot for both optimal solutions to problems* (2) *and* (3)*, the constraint violation under a Dirac distribution* δ_u *that moves away from the center* $\hat{\mathbb{P}} = \delta_{0.5}$*. Here, the Wasserstein distance between* δ_u *and* $\hat{\mathbb{P}}$ *amounts to* $d_{\mathrm{W}}(\delta_u, \hat{\mathbb{P}}) = |u - 0.5|$*. It can be seen that problem* (3) *yields an optimal solution with no constraint violation for a wider range of Dirac distributions* δ_u *and has a smaller constraint violation (if any).*

Example 2. *Consider the following robust satisficing problem:*

$$\begin{array}{ll} \min & \gamma \\ \text{s.t.} & \mathbb{E}_{\mathbb{P}}[1 - x + x \ln \tilde{\xi}] \leq \gamma \cdot d_{\mathrm{W}}(\mathbb{P}, \hat{\mathbb{P}}) \quad \forall \mathbb{P} \in \mathcal{P}(\mathbb{R}_{++}) \\ & \gamma \geq 0,\ x \geq 1, \end{array} \tag{4}$$

where $\hat{\mathbb{P}} = \delta_1$ *is a Dirac distribution and the target is* $\tau = 0$*. We can derive that the optimal value of problem* (4) *is* $1/(\mathrm{e} - 1)$*, attaining at* $x^\star = \mathrm{e}\gamma$ *and* $\gamma^\star = 1/(\mathrm{e} - 1)$*. Consider next the globalized variant of* (4) *with some* $0 \leq \theta < \mathrm{e} - 1$*:*

$$\begin{array}{ll} \min & \gamma \\ \text{s.t.} & \mathbb{E}_{\mathbb{P}}[1 - x + x \ln \xi] \leq \gamma \cdot d_{\mathrm{W}}(\mathbb{P}, \mathbb{Q}) \quad \forall \mathbb{P} \in \mathcal{P}(\mathbb{R}_{++}),\ \mathbb{Q} \in \mathcal{F}_{\mathrm{W}}(\theta) \\ & \gamma \geq 0,\ x \geq 1. \end{array} \tag{5}$$

We can also derive that the optimal value of problem (5) *is* $1/(\mathrm{e} - 1 - \theta)$*, attaining at* $x^\star = \mathrm{e}\gamma^\star$ *and* $\gamma^\star = 1/(\mathrm{e} - 1 - \theta)$*.*

In the right panel of Figure 1 we plot, for both optimal solutions to problems (4) *and* (5)*, the constraint violation under a Dirac distribution* δ_u *with* $u \geq 0$ *that moves away from the center*

$\hat{\mathbb{P}} = \delta_1$. *Here,* $d_{\mathrm{W}}(\delta_u, \hat{\mathbb{P}}) = |u-1|$. *Problem* (5) *yields an optimal solution with no constraint violation for a wider range of Dirac distributions* δ_u*, however as a trade-off, bearing a larger constraint violation if* δ_u *moving too far away from the center.*

3. Dual Reformulation and Potential Applications

In this section, we focus on the type-1 Wasserstein distance (i) as the statistical metric d and (ii) in the definition of a data-driven ambiguity set $\mathcal{F}$ around an empirical distribution motivated by Mohajerin Esfahani and Kuhn (2018). Extensions to other statistical metrics and ambiguity sets are presented in Section 4. For the corresponding globalized distributionally robust counterpart (1), we derive an insightful dual reformulation that is tractable and possesses finite-sample performance guarantees, and we also illustrate how it can be integrated with popular types of optimization models under uncertainty.

3.1. Reformulation: The Shadow Price of Globalized Robustness

The type-1 Wasserstein distance $d_{\mathrm{W}} : \mathcal{P}(\Xi) \times \mathcal{P}(\Xi) \to \mathbb{R}$, equipped with a norm $\|\cdot\|$, is defined as

$$d_{\mathrm{W}}(\mathbb{P}_1, \mathbb{P}_2) = \inf_{\pi \in \mathcal{Q}(\mathbb{P}_1, \mathbb{P}_2)} \int_{\Xi \times \Xi} \|\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2\| \, \mathrm{d}\pi(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2),$$

where $\mathcal{Q}(\mathbb{P}_1, \mathbb{P}_2)$ is the set of joint probability distributions of $(\tilde{\boldsymbol{\xi}}_1, \tilde{\boldsymbol{\xi}}_2)$ with marginals $\mathbb{P}_1$ and $\mathbb{P}_2$, respectively. The data-driven Wasserstein ambiguity set is then given by

$$\mathcal{F}_{\mathrm{W}}(\theta) = \{\mathbb{P} \in \mathcal{P}(\Xi) \mid d_{\mathrm{W}}(\mathbb{P}, \hat{\mathbb{P}}) \le \theta\}, \tag{6}$$

which is a ball of size $\theta \ge 0$ around the empirical distribution $\hat{\mathbb{P}} = \frac{1}{N}\sum_{n \in [N]} \delta_{\hat{\boldsymbol{\xi}}_n}$ that is uniformly supported on N empirical realizations/samples $\hat{\boldsymbol{\xi}}_1, \ldots, \hat{\boldsymbol{\xi}}_N$ of the uncertainty. The corresponding globalized distributionally robust counterpart (1) can then be reexpressed as

$$\sup_{\mathbb{P} \in \mathcal{P}(\Xi),\, \mathbb{Q} \in \mathcal{F}_{\mathrm{W}}(\theta)} \left\{ \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] - \gamma \cdot d_{\mathrm{W}}(\mathbb{P}, \mathbb{Q}) \right\} \le 0. \tag{7}$$

Here, the left-hand side can be represented as the following

$$\max\left\{ \sup_{\mathbb{P} \in \mathcal{F}_{\mathrm{W}}(\theta)} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})], \sup_{\mathbb{P} \in \mathcal{P}(\Xi) \setminus \mathcal{F}_{\mathrm{W}}(\theta)} \left\{ \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] - \gamma \cdot \min_{\mathbb{Q} \in \mathcal{F}_{\mathrm{W}}(\theta)} d_{\mathrm{W}}(\mathbb{P}, \mathbb{Q}) \right\} \right\},$$

illuminating the essence of the globalized robustness: it considers the worst case between (i) the hard worst-case expected cost within $\mathcal{F}_{\mathrm{W}}(\theta)$, which is typically considered in distributionally robust optimization, and (ii) a soft worst-case expected cost over $\mathcal{P}(\Xi) \setminus \mathcal{F}_{\mathrm{W}}(\theta)$, which offsets a distribution's distance to $\mathcal{F}_{\mathrm{W}}(\theta)$ at a unit subsidy of possible constraint violation that is equal to the tolerance level γ. Furthermore, the left-hand side of (7) admits the following dual reformulation.

Lemma 1 (Strong duality). *For any fixed decision $\boldsymbol{x} \in \mathcal{X}$, given the Wasserstein ambiguity set $\mathcal{F}_{\mathrm{W}}(\theta)$ in (6), we have*

$$\sup_{\mathbb{P}\in\mathcal{P}(\Xi),\, \mathbb{Q}\in\mathcal{F}_{\mathrm{W}}(\theta)} \left\{ \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})] - \gamma \cdot d_{\mathrm{W}}(\mathbb{P},\mathbb{Q}) \right\} = \inf_{t\in[0,\gamma]} \left\{ \theta t + \frac{1}{N}\sum_{n\in[N]} \sup_{\boldsymbol{\xi}\in\Xi} \left\{ f(\boldsymbol{x},\boldsymbol{\xi}) - t\|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}_n\| \right\} \right\}. \quad (8)$$

It is now well known that the distributionally robust counterpart satisfies

$$\sup_{\mathbb{P}\in\mathcal{F}_{\mathrm{W}}(\theta)} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})] = \inf_{t\geq 0} \left\{ \theta t + \frac{1}{N}\sum_{n\in[N]} \sup_{\boldsymbol{\xi}\in\Xi} \left\{ f(\boldsymbol{x},\boldsymbol{\xi}) - t\|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}_n\| \right\} \right\}. \quad (9)$$

Comparing this with the right-hand side dual reformulation given in (8), it is interesting to note that the globalized distributionally robust counterpart (1) pays

"*a shadow price of globalized robustness* γ on the dual decision variable t"

for ensuring limited constraint violation even when the unknown probability distribution $\mathbb{P} \notin \mathcal{F}_{\mathrm{W}}(\theta)$. If there exists a finite minimizer $t^\star$ to the right-hand minimization reformulation of (9), then letting $\gamma \geq t^\star$ would be sufficient for the globalized distributionally robust counterpart (1) to recover the distributionally robust counterpart. In other words, $\gamma < t^\star$ would make the globalized distributionally robust counterpart effective.[5]

Another notable implication is that the globalized distributionally robust counterpart preserves the tractability of the corresponding distributionally robust counterpart because the only additional constraint is an upper bound $t \leq \gamma$. Hence, its tractable deterministic reformulation (which we present to keep our paper self-contained) follows straightforwardly from combining the observation $t \in [0,\gamma]$ with existing results (*e.g.*, Mohajerin Esfahani and Kuhn 2018, theorem 4.2).

Proposition 1. *Suppose that the support set Ξ is convex and closed, and for any fixed $\boldsymbol{x} \in \mathcal{X}$, the cost function takes the form $f(\cdot,\boldsymbol{\xi}) = \max_{k\in[K]} f_k(\cdot,\boldsymbol{\xi})$ with constituent functions $f_k(\cdot,\boldsymbol{\xi})$ being proper, upper semicontinuous, and concave in $\boldsymbol{\xi}$. Then for any fixed decision $\boldsymbol{x} \in \mathcal{X}$, given the Wasserstein ambiguity set $\mathcal{F}_{\mathrm{W}}(\theta)$ in (6), we have that*

$$\sup_{\mathbb{P}\in\mathcal{P}(\Xi),\, \mathbb{Q}\in\mathcal{F}_{\mathrm{W}}(\theta)} \left\{ \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})] - \gamma \cdot d_{\mathrm{W}}(\mathbb{P},\mathbb{Q}) \right\}$$

equals the optimal value of the finite-dimensional convex program

$$\begin{array}{lll} \inf & \theta t + \dfrac{1}{N}\displaystyle\sum_{n\in[N]} s_n & \\ \text{s.t.} & [-f_k]^{\bar{*}}(\boldsymbol{x}, \boldsymbol{u}_{nk} - \boldsymbol{w}_{nk}) + \delta^*(\boldsymbol{u}_{nk} \mid \Xi) - \hat{\boldsymbol{\xi}}_n^\top \boldsymbol{w}_{nk} \leq s_n & \forall n \in [N],\, k \in [K] \\ & \|\boldsymbol{w}_{nk}\|_* \leq t & \forall n \in [N],\, k \in [K] \\ & \boldsymbol{u}_{nk},\, \boldsymbol{w}_{nk} \in \mathbb{R}^I & \forall n \in [N],\, k \in [K] \\ & \boldsymbol{s} \in \mathbb{R}^N,\, t \in [0,\gamma]. & \end{array} \quad (10)$$

[5] See Appendix EC.4 for some conditions on the support set Ξ and the cost function f that would guarantee the finiteness of $t^\star$.

Remark 1. *Indeed, there exists a pair of worst-case distributions* $(\mathbb{P}^\star, \mathbb{Q}^\star)$ *for the globalized distributionally robust counterpart, recalling that* $\mathbb{Q}^\star$ *arises from the ambiguity set* $\mathcal{F}_{\mathrm{W}}(\theta)$ *while* $\mathbb{P}^\star$ *can take globally, i.e.,* $\mathbb{P}^\star \in \mathcal{P}(\Xi)$. *We formalize this result in Appendix EC.5.*

Following from the finite-sample guarantee that $\mathcal{F}_{\mathrm{W}}(\theta)$ covers the true data-generating distribution $\mathbb{P}_0$ (see, *e.g.*, Fournier and Guillin 2015), the globalized distributionally robust counterpart possesses probabilistic guarantees on the constraint violation for *all* distributions residing in $\mathcal{P}(\Xi)$, generalizing, in a globalized fashion, theorem 3.4 of Mohajerin Esfahani and Kuhn (2018).

Proposition 2 (Finite-sample performance guarantee). *Suppose that the true data-generating distribution* $\mathbb{P}_0$ *is a light-tailed distribution such that*

$$A = \mathbb{E}_{\mathbb{P}_0}[\exp(\|\tilde{\boldsymbol{\xi}}\|^a)] = \int_\Xi \exp(\|\boldsymbol{\xi}\|^a)\mathrm{d}\mathbb{P}_0(\boldsymbol{\xi}) < \infty$$

for some $a > 1$. *Let* $\mathbb{P}_0^N$ *be the* N*-fold product of* $\mathbb{P}_0$ *that governs the* N *independent random samples drawn from* $\mathbb{P}_0$. *Then for any decision* $\boldsymbol{x} \in \mathcal{X}$ *satisfying* (1), *we have*

$$\mathbb{P}_0^N[\mathbb{E}_{\mathbb{P}_0}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] > \gamma(\eta - \theta)] \leq \begin{cases} c_1 \exp(-c_2 N \eta^{\max\{I,2\}}) & \text{if } 0 < \eta \leq 1 \\ c_1 \exp(-c_2 N \eta^a) & \text{if } \eta > 1, \end{cases}$$

for all $N \geq 1$, $I \neq 2$, *and* $\eta \geq \theta > 0$, *where* $c_1, c_2 > 0$ *are constants that only depend on* a, A, *and* I.

Letting $\eta = d_{\mathrm{W}}(\mathbb{P}_0, \hat{\mathbb{P}})$, for the true data-generating distribution $\mathbb{P}_0 \in \mathcal{F}_{\mathrm{W}}(\theta)$ such that $d_{\mathrm{W}}(\mathbb{P}_0, \hat{\mathbb{P}}) \leq \theta$, we can bound the probability of the undesired event $\mathbb{P}_0^N[\mathbb{E}_{\mathbb{P}_0}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] > 0]$, while for $\mathbb{P}_0 \in \mathcal{P}(\Xi) \setminus \mathcal{F}_{\mathrm{W}}(\theta)$ such that $d_{\mathrm{W}}(\mathbb{P}_0, \hat{\mathbb{P}}) > \theta$, we can also have an upper bound on $\mathbb{P}_0^N[\mathbb{E}_{\mathbb{P}_0}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] > \gamma \cdot (d_{\mathrm{W}}(\mathbb{P}_0, \hat{\mathbb{P}}) - \theta)]$—the probability that the constraint violation is larger than the shadow price of globalized robustness times the magnitude that the distance between the empirical distribution and the true distribution deviates from the allowable θ.

3.2. Potential Applications to Optimization Problems under Uncertainty

In this section we discuss how to incorporate the globalized distributionally robust counterpart into existing optimization models for decision-making under uncertainty, including the (static) distributionally robust optimization (Mohajerin Esfahani and Kuhn 2018, Gao and Kleywegt 2022), robust satisficing (Long et al. 2022), distributionally robust chance constraint (Chen et al. 2022, Xie 2019), and adaptive distributionally robust optimization under the Wasserstein ambiguity set (Bertsimas et al. 2019, 2022, Chen et al. 2020, Hanasusanto and Kuhn 2018).

Globalized Distributionally Robust Optimization. The popular distributionally robust optimization can be presented in a canonical form

$$\begin{array}{ll} \min & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} & \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] \leq 0 \quad \forall \mathbb{P} \in \mathcal{F}_{\mathrm{W}}(\theta) \\ & \boldsymbol{x} \in \mathcal{X}, \end{array} \tag{DRO}$$

wherein the expectation constraint is satisfied under all possible distributions in the ambiguity set $\mathcal{F}_{\mathrm{W}}(\theta)$. Accordingly, the globalized distributionally robust optimization model represents as

$$\begin{array}{ll} \min & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} & \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] \leq \gamma \cdot \min\limits_{\mathbb{Q} \in \mathcal{F}_{\mathrm{W}}(\theta)} d_{\mathrm{W}}(\mathbb{P}, \mathbb{Q}) \quad \forall \mathbb{P} \in \mathcal{P}(\Xi) \\ & \boldsymbol{x} \in \mathcal{X}. \end{array} \tag{G-DRO}$$

The globalized robustness of G-DRO model over the DRO model, as motivated in the Example 2, will be further justified in the numerical experiments in Section 5.1.

A direct implication of Proposition 1, provided assumptions on the function f therein, gives the following finite-dimensional convex reformulation of the G-DRO model:

$$\begin{array}{lll} \min & \boldsymbol{c}^\top \boldsymbol{x} & \\ \text{s.t.} & \theta N t + \boldsymbol{e}^\top \boldsymbol{s} \leq 0 & \\ & [-f_k]^{\circledast}(\boldsymbol{x}, \boldsymbol{u}_{nk} - \boldsymbol{w}_{nk}) + \delta^*(\boldsymbol{u}_{nk} \mid \Xi) - \hat{\boldsymbol{\xi}}_n^\top \boldsymbol{w}_{nk} \leq s_n & \forall n \in [N],\ k \in [K] \\ & \|\boldsymbol{w}_{nk}\|_* \leq t & \forall n \in [N],\ k \in [K] \\ & \boldsymbol{u}_{nk},\ \boldsymbol{w}_{nk} \in \mathbb{R}^I & \forall n \in [N],\ k \in [K] \\ & \boldsymbol{s} \in \mathbb{R}^N,\ t \in [0, \gamma],\ \boldsymbol{x} \in \mathcal{X}, & \end{array}$$

which, after replacing the constraint $t \in [0, \gamma]$ with $t \geq 0$ (or equivalently, letting $\gamma \to \infty$), becomes the reformulation of the corresponding DRO model.

Globalized Robust Satisficing. The emerging robust satisficing model (Long et al. 2022)

$$\begin{array}{ll} \min & \gamma \\ \text{s.t.} & \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] - \tau \leq \gamma \cdot d_{\mathrm{W}}(\mathbb{P}, \hat{\mathbb{P}}) \quad \forall \mathbb{P} \in \mathcal{P}(\Xi) \\ & \gamma \geq 0,\ \boldsymbol{x} \in \mathcal{X}, \end{array} \tag{RS}$$

which seeks an optimal decision that tolerates the least violation in achieving the targeted cost τ under *all* distributions in $\mathcal{P}(\Xi)$, can be generalized to a globalized robust satisficing model

$$\begin{array}{ll} \min & \gamma \\ \text{s.t.} & \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] - \tau \leq \gamma \cdot \min\limits_{\mathbb{Q} \in \mathcal{F}_{\mathrm{W}}(\theta)} d_{\mathrm{W}}(\mathbb{P}, \mathbb{Q}) \quad \forall \mathbb{P} \in \mathcal{P}(\Xi) \\ & \gamma \geq 0,\ \boldsymbol{x} \in \mathcal{X}. \end{array} \tag{G-RS}$$

For any fixed value of θ, the G-RS model is feasible as long as $\tau \geq \min\limits_{\boldsymbol{x} \in \mathcal{X}} \sup\limits_{\mathbb{P} \in \mathcal{F}_{\mathrm{W}}(\theta)} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})]$, and it reduces to RS when $\theta = 0$ (*i.e.*, $\mathcal{F}_{\mathrm{W}}(0) = \{\hat{\mathbb{P}}\}$).

The constraint in the G-RS model indicates

$$\begin{cases} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})]-\tau\leq 0 & \forall\mathbb{P}\in\mathcal{F}_{\mathrm{W}}(\theta) \\ \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})]-\tau\leq\gamma\cdot\min\limits_{\mathbb{Q}\in\mathcal{F}_{\mathrm{W}}(\theta)} d_{\mathrm{W}}(\mathbb{P},\mathbb{Q})\leq\gamma\cdot d_{\mathrm{W}}(\mathbb{P},\hat{\mathbb{P}}) & \forall\mathbb{P}\in\mathcal{P}(\Xi)\setminus\mathcal{F}_{\mathrm{W}}(\theta), \end{cases}$$

while the constraint in the RS model implies

$$\begin{cases} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})]-\tau\leq\gamma\cdot d_{\mathrm{W}}(\mathbb{P},\hat{\mathbb{P}}) & \forall\mathbb{P}\in\mathcal{F}_{\mathrm{W}}(\theta) \\ \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})]-\tau\leq\gamma\cdot d_{\mathrm{W}}(\mathbb{P},\hat{\mathbb{P}}) & \forall\mathbb{P}\in\mathcal{P}(\Xi)\setminus\mathcal{F}_{\mathrm{W}}(\theta). \end{cases}$$

Hence, the G-RS model not only limits the constraint violation under distributions outside $\mathcal{F}_{\mathrm{W}}(\theta)$ to the minimum by optimizing γ, but also ensures strict feasibility for all distributions within $\mathcal{F}_{\mathrm{W}}(\theta)$—this, however, is not necessarily guaranteed in the RS model. As shown in Example 2 and subsequent numerical experiments in Section 5.2, the G-RS model may sacrifice certain performance in terms of controlling the constraint violation for distributions in $\mathcal{P}(\Xi)\setminus\mathcal{F}_{\mathrm{W}}(\theta)$.

Evoking again Proposition 1, the G-RS model under conditions therein admits the following reformulation as a finite-dimensional convex program:

$$\begin{array}{lll} \min & \gamma & \\ \text{s.t.} & \theta N t+\boldsymbol{e}^{\top}\boldsymbol{s}\leq N\tau & \\ & [-f_k]^{\bar{*}}(\boldsymbol{x},\boldsymbol{u}_{nk}-\boldsymbol{w}_{nk})+\delta^*(\boldsymbol{u}_{nk}\mid\Xi)-\hat{\boldsymbol{\xi}}_n^{\top}\boldsymbol{w}_{nk}\leq s_n & \forall n\in[N],\ k\in[K] \\ & \|\boldsymbol{w}_{nk}\|_*\leq t & \forall n\in[N],\ k\in[K] \\ & \boldsymbol{u}_{nk},\ \boldsymbol{w}_{nk}\in\mathbb{R}^I & \forall n\in[N],\ k\in[K] \\ & \boldsymbol{s}\in\mathbb{R}^N,\ t\in[0,\gamma],\ \gamma\geq 0,\ \boldsymbol{x}\in\mathcal{X}, & \end{array} \tag{11}$$

which becomes the reformulation of the RS model when relaxing $t\in[0,\gamma]$ to $t\geq 0$.

Globalized Distributionally Robust Chance Constraint. Constraints $\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})]\leq 0$ and $\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})]-\tau\leq\gamma\cdot d_{\mathrm{W}}(\mathbb{P},\hat{\mathbb{P}})$ in the aforementioned DRO and RS models hold in the sense of expectation (with some acceptable violation captured by $\gamma\cdot d_{\mathrm{W}}(\mathbb{P},\hat{\mathbb{P}})$ in the latter). Alternatively, it is possible to require them to hold in a probabilistic sense. Indeed, following from the spirit of chance constraint, one could consider instead $\mathbb{P}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})>0]\leq\varepsilon$ and $\mathbb{P}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})>\tau]\leq\varepsilon+\gamma\cdot d_{\mathrm{W}}(\mathbb{P},\hat{\mathbb{P}})$, which state that the probability of violating the cost constraint is bounded from above by some risk threshold $\varepsilon\in(0,1)$ (plus some additional tolerance $\gamma\cdot d_{\mathrm{W}}(\mathbb{P},\hat{\mathbb{P}})$ in the RS model).

Formally, a Wasserstein distributionally robust chance constraint (DRCC) is given by

$$\left[\mathbb{P}[\tilde{\boldsymbol{\xi}}\in\bar{\mathcal{S}}(\boldsymbol{x})]\leq\varepsilon\iff\mathbb{E}_{\mathbb{P}}[\mathbb{I}\{\tilde{\boldsymbol{\xi}}\in\bar{\mathcal{S}}(\boldsymbol{x})\}-\varepsilon]\leq 0\right]\quad\forall\mathbb{P}\in\mathcal{F}_{\mathrm{W}}(\theta), \tag{DRCC}$$

where $\bar{\mathcal{S}}(\boldsymbol{x})\subseteq\Xi$ is a decision-dependent unsafe set that describes unfavourable scenarios. It is then natural to consider the globalized distributionally robust chance constraint

$$\begin{bmatrix} \mathbb{P}[\tilde{\boldsymbol{\xi}}\in\bar{\mathcal{S}}(\boldsymbol{x})]\leq\varepsilon+\gamma\cdot\min\limits_{\mathbb{Q}\in\mathcal{F}_{\mathrm{W}}(\theta)} d_{\mathrm{W}}(\mathbb{P},\mathbb{Q}) \\ \Updownarrow \\ \mathbb{E}_{\mathbb{P}}[\mathbb{I}\{\tilde{\boldsymbol{\xi}}\in\bar{\mathcal{S}}(\boldsymbol{x})\}-\varepsilon]\leq\gamma\cdot\min\limits_{\mathbb{Q}\in\mathcal{F}_{\mathrm{W}}(\theta)} d_{\mathrm{W}}(\mathbb{P},\mathbb{Q}) \end{bmatrix}\quad\forall\mathbb{P}\in\mathcal{P}(\Xi). \tag{G-DRCC}$$

Note that the G-DRCC remains effective for all distributions satisfying

$$\mathbb{P} \in \mathcal{P}(\Xi) \setminus \mathcal{F}_{\mathrm{W}}(\theta): \min_{\mathbb{Q} \in \mathcal{F}_{\mathrm{W}}(\theta)} d_{\mathrm{W}}(\mathbb{P}, \mathbb{Q}) < \frac{1-\varepsilon}{\gamma},$$

because otherwise $\mathbb{P}[\tilde{\boldsymbol{\xi}} \in \bar{\mathcal{S}}(\boldsymbol{x})] \le 1$ would be redundant. A smaller value of γ not only articulates a lower tolerance towards additional violation of the chance constraint, but also characterizes a larger neighbourhood centered around $\hat{\mathbb{P}}$, on which the G-DRCC is imposed.

PROPOSITION 3. *The G-DRCC is satisfiable if and only if there exist $\boldsymbol{s} \ge \boldsymbol{0}$ and $t \ge 1/\gamma$ such that the following constraint system is feasible:*

$$\begin{cases} \varepsilon N t - \boldsymbol{e}^\top \boldsymbol{s} \ge \theta N \\ \mathbf{dist}(\hat{\boldsymbol{\xi}}_n, \bar{\mathcal{S}}(\boldsymbol{x})) \ge t - s_n \quad \forall n \in [N], \end{cases}$$

where the distance from a point $\boldsymbol{\xi}$ to the unsafe set $\bar{\mathcal{S}}(\boldsymbol{x})$, with respect to a norm $\|\cdot\|$, is defined as $\mathbf{dist}(\boldsymbol{\xi}, \bar{\mathcal{S}}(\boldsymbol{x})) = \min\{\|\boldsymbol{\xi} - \boldsymbol{\zeta}\| \mid \boldsymbol{\zeta} \in \bar{\mathcal{S}}(\boldsymbol{x})\}$.

Proposition 3 naturally generalizes the well known representation result for the Wasserstein DRCC (see, *e.g.*, Chen et al. 2022, theorem 3):

$$\mathbb{P}[\tilde{\boldsymbol{\xi}} \in \bar{\mathcal{S}}(\boldsymbol{x})] \le \varepsilon \quad \forall \mathbb{P} \in \mathcal{F}_{\mathrm{W}}(\theta) \quad \Longleftrightarrow \quad \exists\, \boldsymbol{s} \ge \boldsymbol{0},\, t \ge 0: \begin{cases} \varepsilon N t - \boldsymbol{e}^\top \boldsymbol{s} \ge \theta N \\ \mathbf{dist}(\hat{\boldsymbol{\xi}}_n, \bar{\mathcal{S}}(\boldsymbol{x})) \ge t - s_n \quad \forall n \in [N]. \end{cases}$$

The only difference is that in the globalized variant, the lower bound on the decision variable t is tightened from 0 to $1/\gamma$. That is, one pays a price of $1/\gamma$ for globalized robustness in the G-DRCC, without increasing any computational complexity. With Proposition 3, established deterministic mixed-integer reformulations for the DRCC can be straightforwardly extended to their globalized variants. Examples include (i) individual chance constraints where the unsafe set is defined as $\bar{\mathcal{S}}(\boldsymbol{x}) = \{\boldsymbol{\xi} \in \mathbb{R}^K \mid (\boldsymbol{A}\boldsymbol{\xi} + \boldsymbol{a})^\top \boldsymbol{x} \ge \boldsymbol{b}^\top \boldsymbol{\xi} + b\}$ and (ii) joint chance constraints with right-hand side uncertainty such that $\bar{\mathcal{S}}(\boldsymbol{x}) = \cup_{m \in [M]} \{\boldsymbol{\xi} \in \mathbb{R}^K \mid \boldsymbol{a}_m^\top \boldsymbol{x} \ge \boldsymbol{b}_m^\top \boldsymbol{\xi} + b_m\}$; see propositions 1 and 2 in Chen et al. (2022), respectively.

Globalized Two-Stage Adaptive Linear Optimization. In a two-stage adaptive problem, the here-and-now decision $\boldsymbol{x} \in \mathcal{X}$, incurring a deterministic cost $\boldsymbol{c}^\top \boldsymbol{x}$, needs to be determined before the realization of uncertain parameters $\boldsymbol{\xi} \in \Xi$. Thereafter, the uncertainty materializes, and we decide a wait-and-see recourse decision $\boldsymbol{y} \in \mathbb{R}^L$ by solving the second-stage problem. Considering the randomness of uncertain parameters and ambiguity about the underlying true distribution, a Wasserstein distributionally robust model for this problem is given by

$$\min_{\boldsymbol{x} \in \mathcal{X}} \sup_{\mathbb{P} \in \mathcal{F}_{\mathrm{W}}(\theta)} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})], \qquad \text{(ADRO)}$$

where given $\boldsymbol{x}$ and $\boldsymbol{\xi}$, the objective function[6] is

$$f(\boldsymbol{x},\boldsymbol{\xi}) = \begin{cases} \min\limits_{\boldsymbol{y}} & \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{y} \\ \text{s.t.} & \boldsymbol{A}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{b}(\boldsymbol{\xi}) \\ & \boldsymbol{y} \in \mathbb{R}^L, \end{cases} \tag{12}$$

which solves an optimization problem that is adaptive to each realization $\boldsymbol{\xi}$ given $\boldsymbol{x}$. Here, $\boldsymbol{A}(\boldsymbol{\xi}) = \boldsymbol{A}_0 + \sum_{i\in[I]} \boldsymbol{A}_i \xi_i$ and $\boldsymbol{b}(\boldsymbol{\xi}) = \boldsymbol{b}_0 + \sum_{i\in[I]} \boldsymbol{b}_i \xi_i$ are affine functions of $\boldsymbol{\xi}$ with $\boldsymbol{A}_0, \boldsymbol{A}_1, \ldots, \boldsymbol{A}_I \in \mathbb{R}^{M\times L}$ and $\boldsymbol{b}_0, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_I \in \mathbb{R}^M$. In general, problem (12) could be infeasible. Thus we assume that the second-stage problem has relatively complete recourse to ensure feasibility of the second-stage problem: for any $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{\xi} \in \Xi$, there exists $\boldsymbol{y} \in \mathbb{R}^L$ such that $\boldsymbol{A}(\boldsymbol{\xi})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{b}(\boldsymbol{\xi})$.

Introducing an epigraphical decision variable τ, we can express the adaptive distributionally robust optimization (ADRO) model as follows:

$$\begin{array}{ll} \min & \tau \\ \text{s.t.} & \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})] \leq \tau \quad \forall \mathbb{P} \in \mathcal{F}_{\mathrm{W}}(\theta) \\ & \tau \in \mathbb{R},\ \boldsymbol{x} \in \mathcal{X}. \end{array}$$

This motivates the following globalized adaptive distributionally robust optimization model

$$\begin{array}{ll} \min & \tau \\ \text{s.t.} & \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})] \leq \tau + \gamma \cdot \min\limits_{\mathbb{Q}\in\mathcal{F}_{\mathrm{W}}(\theta)} d_{\mathrm{W}}(\mathbb{P},\mathbb{Q}) \quad \forall \mathbb{P} \in \mathcal{P}(\Xi) \\ & \tau \in \mathbb{R},\ \boldsymbol{x} \in \mathcal{X}. \end{array} \tag{G-ADRO}$$

Recall from Observations 2 and 3, when $\gamma \to \infty$ the G-ADRO model recovers the ADRO model, and if in addition $\theta = 0$, then the G-ADRO model reduces to the two-stage stochastic programming problem with recourse under a known distribution $\hat{\mathbb{P}}$; see, *e.g.*, Birge and Louveaux (2011).

Alternatively, with a targeted level of the expected cost τ in mind, we can also consider a globalized adaptive robust satisficing model as follows

$$\begin{array}{ll} \min & \gamma \\ \text{s.t.} & \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})] - \tau \leq \gamma \cdot \min\limits_{\mathbb{Q}\in\mathcal{F}_{\mathrm{W}}(\theta)} d_{\mathrm{W}}(\mathbb{P},\mathbb{Q}) \quad \forall \mathbb{P} \in \mathcal{P}(\Xi) \\ & \gamma \geq 0,\ \boldsymbol{x} \in \mathcal{X}, \end{array} \tag{G-ARS}$$

which, with $\theta = 0$, reduces to the adaptive robust satisficing (ARS) model (see Long et al. 2022).

In both G-ADRO and G-ARS models above, the recourse decision $\boldsymbol{y}$ can be viewed as an infinite-dimensional functional of the uncertainty realization $\boldsymbol{\xi}$ and it makes the models hard to solve. For the sake of tractability, a popular technique is to restrict $\boldsymbol{y}$ to a piecewise affine function of the primary random variable (and possibly, some auxiliary ones arising from describing the

[6] Here, for a fixed here-and-now decision $\boldsymbol{x}$, we count collectively its deterministic first-stage cost and the optimal second-stage when facing a specific uncertainty realization $\boldsymbol{\xi}$.

ambiguity set). In particular, for our interested type-1 Wasserstein ambiguity set $\mathcal{F}_{\mathrm{W}}(\theta)$, we adopt the following piecewise affine approximation proposed by Chen et al. (2020):

$$\boldsymbol{y}(\boldsymbol{\xi},\zeta,n)=\boldsymbol{y}_{0n}+\sum_{i\in[I]}\boldsymbol{y}_{in}\xi_i+\boldsymbol{y}_{Jn}\zeta \quad \forall(\boldsymbol{\xi},\zeta)\in\bar{\Xi}_n,\ n\in[N], \tag{13}$$

where $\bar{\Xi}_n=\{(\boldsymbol{\xi},\zeta)\in\Xi\times\mathbb{R}\mid\zeta\geq\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}_n\|\}$ for each $n\in[N]$. Here $[N]$, the index set of historical samples, can be interpreted as the discrete support set of an auxiliary random scenario that arises from describing $\mathcal{F}_{\mathrm{W}}(\theta)$ and controls the piecewise affine dependency of $\boldsymbol{y}$ on the primary uncertainty $\boldsymbol{\xi}$ and an auxiliary one-dimensional uncertainty $\tilde{\zeta}$ (see more details in section 5.5, Chen et al. 2020).

4. Extensions to Other Ambiguity Sets

In this section, we consider the globalized distributionally robust counterpart (1) with other ambiguity sets based on probability metrics, optimal transport, ϕ-divergences, and moment conditions, and derive the corresponding dual reformulations.

4.1. Probability-Metric Ambiguity Set

We consider a mapping $d_{\mathcal{L}}:\mathcal{P}(\Xi)\times\mathcal{P}(\Xi)\to\mathbb{R}_+$—which is equipped with a nonempty set $\mathcal{L}$ of bounded real-valued measurable functions defined on Ξ—to quantify the closeness of any two given probability distributions $\mathbb{P}_1$ and $\mathbb{P}_2$ in $\mathcal{P}(\Xi)$:

$$d_{\mathcal{L}}(\mathbb{P}_1,\mathbb{P}_2)=\sup_{\ell\in\mathcal{L}}\left\{\mathbb{E}_{\mathbb{P}_1}[\ell(\tilde{\boldsymbol{\xi}})]-\mathbb{E}_{\mathbb{P}_2}[\ell(\tilde{\boldsymbol{\xi}})]\right\}. \tag{14}$$

It can be shown that $d_{\mathcal{L}}(\mathbb{P}_1,\mathbb{P}_2)$ is jointly convex in $(\mathbb{P}_1,\mathbb{P}_2)$ and it satisfies the triangle inequality, *i.e.*, $d_{\mathcal{L}}(\mathbb{P}_1,\mathbb{P}_2)\leq d_{\mathcal{L}}(\mathbb{P}_1,\mathbb{P}_3)+d_{\mathcal{L}}(\mathbb{P}_3,\mathbb{P}_2)$ for any probability distribution $\mathbb{P}_3\in\mathcal{P}(\Xi)$.

Such a class of mappings is not uncommon and it has been used to construct the ambiguity set in DRO; see, *e.g.*, Shapiro (2017). If the set $\mathcal{L}$ is symmetric, *i.e.*, $\ell\in\mathcal{L}$ implies that $-\ell\in\mathcal{L}$, then the mapping $d_{\mathcal{L}}$ can be rewritten as $d_{\mathcal{L}}(\mathbb{P}_1,\mathbb{P}_2)=\sup_{\ell\in\mathcal{L}}\left|\mathbb{E}_{\mathbb{P}_1}[\ell(\tilde{\boldsymbol{\xi}})]-\mathbb{E}_{\mathbb{P}_2}[\ell(\tilde{\boldsymbol{\xi}})]\right|$, which becomes a probability semi-metric (Rachev 1991) that satisfies symmetry, *i.e.*, $d_{\mathcal{L}}(\mathbb{P}_1,\mathbb{P}_2)=d_{\mathcal{L}}(\mathbb{P}_2,\mathbb{P}_1)$. Note that probability semi-metric $d_{\mathcal{L}}$ is also referred to as the ζ-structure semi-metric (Zolotarev 1976), which has been used in data-driven risk-averse stochastic programs (Zhao and Guan 2015). Furthermore, if $\mathcal{L}$ is taken as the set of Lipschitz continuous functions of modulus one, then $d_{\mathcal{L}}$ reduces to the Kantorovich metric, a.k.a. the type-1 Waaserstein distance discussed in Section 3. We summarize some representative $d_{\mathcal{L}}$ as follows.

- If $\mathcal{L}=\{\ell\mid\|\ell\|_{\mathrm{L}}\leq 1\}$, where $\|\ell\|_{\mathrm{L}}=\sup\{(\ell(\boldsymbol{\xi})-\ell(\boldsymbol{\zeta}))/\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|,\ \boldsymbol{\xi}\neq\boldsymbol{\zeta}\text{ in }\Xi\}$, then $d_{\mathcal{L}}$ becomes the well-known Kantorovich metric—the dual representation of the *type-1 Wasserstein distance* (Villani 2009).
- If $\mathcal{L}=\{\ell\mid\|\ell\|_{\mathrm{L}}\leq 1,\ \|\ell\|_\infty\leq 1\}$, then $d_{\mathcal{L}}$ becomes the *bounded Lipschitz metric* (Dudley 2018).

- If $\mathcal{L} = \{\ell \mid \|\ell\|_\infty \leq 1\}$ where $\|\ell\|_\infty = \sup_{\boldsymbol{\xi}\in\Xi} |\ell(\boldsymbol{\xi})|$, then $d_\mathcal{L}$ becomes the *total variation metric* (Tierney 1996).
- If $\mathcal{L} = \{\ell \mid \|\ell\|_\mathrm{C} \leq 1\}$ where $\|\ell\|_\mathrm{C} = \sup\{(\ell(\boldsymbol{\xi}) - \ell(\boldsymbol{\zeta}))/c(\boldsymbol{\xi},\boldsymbol{\zeta}),\ \boldsymbol{\xi} \neq \boldsymbol{\zeta} \text{ in } \Xi\}$ with $c(\boldsymbol{\xi},\boldsymbol{\zeta}) = \max\{1, \|\boldsymbol{\xi}\|^{p-1}, \|\boldsymbol{\zeta}\|^{p-1}\} \cdot \|\boldsymbol{\xi} - \boldsymbol{\zeta}\|$ for some $p > 1$, then $d_\mathcal{L}$ becomes the *Fortet-Mourier metric* (Dupačová et al. 2003).

Given the mapping $d_\mathcal{L}$, we define the corresponding data-driven ambiguity set of a size $\theta \geq 0$ that is centered around the empirical distribution $\hat{\mathbb{P}} = \frac{1}{N}\sum_{n\in[N]} \delta_{\hat{\boldsymbol{\xi}}_n}$:

$$\mathcal{F}_\mathcal{L}(\theta) = \{\mathbb{P} \in \mathcal{P}(\Xi) \mid d_\mathcal{L}(\mathbb{P}, \hat{\mathbb{P}}) \leq \theta\}.$$

We can then represent the globalized distributionally robust counterpart (1) in a canonical form

$$\sup_{\mathbb{P}\in\mathcal{P}(\Xi),\, \mathbb{Q}\in\mathcal{F}_{\mathcal{L}_2}(\theta)} \left\{\mathbb{E}_\mathbb{P}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] - \gamma \cdot d_{\mathcal{L}_1}(\mathbb{P}, \mathbb{Q})\right\} \leq 0, \tag{15}$$

where $d_{\mathcal{L}_1}$ and $d_{\mathcal{L}_2}$ are two different mappings. The left-hand side of the globalized distributionally robust counterpart (15) admits a dual representation.

Theorem 1 (Strong duality). *Given any decision $\boldsymbol{x} \in \mathcal{X}$, given the mappings $d_{\mathcal{L}_1}$ and $d_{\mathcal{L}_2}$ defined in* (14) *as well as an ambiguity set $\mathcal{F}_{\mathcal{L}_2}(\theta)$ with $\theta > 0$, we have*

$$\begin{aligned} &\sup_{\mathbb{P}\in\mathcal{P}(\Xi),\, \mathbb{Q}\in\mathcal{F}_{\mathcal{L}_2}(\theta)} \left\{\mathbb{E}_\mathbb{P}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] - \gamma \cdot d_{\mathcal{L}_1}(\mathbb{P}, \mathbb{Q})\right\} \\ &= \inf_{t\geq 0} \sup_{\mathbb{P}\in\mathcal{P}(\Xi),\, \mathbb{Q}\in\mathcal{P}(\Xi)} \left\{\mathbb{E}_\mathbb{P}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] - \gamma \cdot d_{\mathcal{L}_1}(\mathbb{P}, \mathbb{Q}) - t(d_{\mathcal{L}_2}(\mathbb{Q}, \hat{\mathbb{P}}) - \theta)\right\}. \end{aligned} \tag{16}$$

In particular, if $d_{\mathcal{L}_1} = d_{\mathcal{L}_2} = d_\mathcal{L}$, then

$$\sup_{\mathbb{P}\in\mathcal{P}(\Xi),\, \mathbb{Q}\in\mathcal{F}_\mathcal{L}(\theta)} \left\{\mathbb{E}_\mathbb{P}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] - \gamma \cdot d_\mathcal{L}(\mathbb{P}, \mathbb{Q})\right\} = \inf_{t\in[0,\gamma]} \sup_{\mathbb{P}\in\mathcal{P}(\Xi)} \left\{\mathbb{E}_\mathbb{P}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] - t(d_\mathcal{L}(\mathbb{P}, \hat{\mathbb{P}}) - \theta)\right\}. \tag{17}$$

In Theorem 1, we consider two different mappings $d_{\mathcal{L}_1}$ and $d_{\mathcal{L}_2}$, and each of them could cover the type-1 Wasserstein distance as a special case. Hence, Theorem 1 naturaly generalizes Lemma 1 that considers the type-1 Wassertein distance for both $d_{\mathcal{L}_1}$ and $d_{\mathcal{L}_2}$. As long as $d_{\mathcal{L}_1} = d_{\mathcal{L}_2}$, we have the strong duality in a stronger form (17), sharing the same insightful structure as Lemma 1:

"*the shadow price of globalized robustness γ bounds from above the dual variable t.*"

Leveraging the strong duality in Theorem 1, we can further derive deterministic reformulations of the globalized robust optimization counterpart (15) under different combinations of $d_{\mathcal{L}_1}$ and $d_{\mathcal{L}_2}$. We first illustrate a blending case of total variation and Wasserstein metrics.

Proposition 4 (Total variation and Wasserstein metrics). *Suppose that* $\mathcal{L}_1 = \{\ell \mid \|\ell\|_\infty \leq 1\}$ *and* $\mathcal{L}_2 = \{\ell \mid \|\ell\|_{\mathrm{L}} \leq 1\}$*, and let* $d_{\mathcal{L}_1}(\mathbb{P}, \mathbb{Q}) = +\infty$ *if* $\mathbb{P}$ *is not absolutely continuous with respect to* $\mathbb{Q}$*. Given assumptions in Proposition 1 and that* $f(\boldsymbol{x}, \boldsymbol{\xi})$ *is essentially bounded for a fixed value of* $\boldsymbol{x} \in \mathcal{X}$*, any* $\boldsymbol{x} \in \mathcal{X}$ *satisfies* (15) *if and only if the following convex constraint system is satisfiable:*

$$\begin{cases} v + t\theta + \dfrac{1}{N} \displaystyle\sum_{n \in [N]} s_n \leq 0 & \\ [-f]^{\bar{*}}(\boldsymbol{x}, \boldsymbol{v}_n) + \delta^*(\boldsymbol{v}_n \mid \Xi) \leq \gamma + v & \\ [-f]^{\bar{*}}(\boldsymbol{x}, \boldsymbol{u}_n - \boldsymbol{w}_n) + \delta^*(\boldsymbol{u}_n \mid \Xi) - \hat{\boldsymbol{\xi}}_n^\top \boldsymbol{w}_n \leq s_n + v & \forall n \in [N] \\ s_n + \gamma \geq 0 & \forall n \in [N] \\ \|\boldsymbol{w}_n\|_* \leq t & \forall n \in [N] \\ \boldsymbol{u}_n,\ \boldsymbol{w}_n,\ \boldsymbol{v}_n \in \mathbb{R}^I & \forall n \in [N] \\ t \geq 0,\ \boldsymbol{s} \in \mathbb{R}^N,\ v \in \mathbb{R}. & \end{cases}$$

In Proposition 4, for the sake of tractability we restrict the effective total variation to be absolute continuous (Ben-Tal and Teboulle 2007). We can also derive the reformulation of a blending case of Fortet-Mourier and bounded Lipschitz metrics. For the same technical reason, we assume that all distributions in $\mathcal{P}(\Xi)$ are absolutely continuous with respect to the empirical distribution $\hat{\mathbb{P}}$, as in most examples based on the probability metrics (Shapiro 2017). Reformulations of other combinations of probability metrics can be similiar derived.

Proposition 5 (Fortet-Mourier and bounded Lipschitz metrics). *Suppose that* $\mathcal{L}_1 = \{\ell \mid \|\ell\|_{\mathrm{C}} \leq 1\}$ *and* $\mathcal{L}_2 = \{\ell \mid \|\ell\|_{\mathrm{L}} \leq 1,\ \|\ell\|_\infty \leq 1\}$*, and that all distributions in* $\mathcal{P}(\Xi)$ *are absolutely continuous with respect to* $\hat{\mathbb{P}}$*. Given assumptions in Proposition 1, any* $\boldsymbol{x} \in \mathcal{X}$ *satisfies* (15) *if and only if the following convex constraint system is satisfiable:*

$$\begin{cases} t\theta + \max\limits_{n \in [N]} \left\{ f(\boldsymbol{x}, \hat{\boldsymbol{\xi}}_n) - \gamma u_n \right\} + \max\limits_{n \in [N]} \left\{ \gamma u_n - s_n \right\} + \dfrac{1}{N} \displaystyle\sum_{n \in [N]} s_n \leq 0 & \\ |u_m - u_n| \leq \max\{1, \|\hat{\boldsymbol{\xi}}_m\|^{p-1}, \|\hat{\boldsymbol{\xi}}_n\|^{p-1}\} \cdot \|\hat{\boldsymbol{\xi}}_m - \hat{\boldsymbol{\xi}}_n\| & \forall m,\ n \in [N] \\ |s_m - s_n| \leq t \|\hat{\boldsymbol{\xi}}_m - \hat{\boldsymbol{\xi}}_n\| & \forall m,\ n \in [N] \\ |s_n| \leq t & \forall n \in [N] \\ t \geq 0. & \end{cases}$$

According to (17) in Theorem 1, the globalized distributionally robust counterpart (15) may admit a more concise reformulation when $\mathcal{L}_1 = \mathcal{L}_2 = \mathcal{L}$. We next present an example based on the total variation metric and leave examples of other probability metrics to Appendix EC.6. With an additional price of globalized robustness (captured by $t \in [0, \gamma]$), as $\gamma \to \infty$, our reformulations recovers those of the corresponding distributionally robust counterparts, which have been provided in Zhao and Guan (2015).

Example 3 (Bounded Lipschitz metric). *Suppose that* $\mathcal{L} = \{\ell \mid \|\ell\|_{\mathrm{L}} \le 1,\ \|\ell\|_\infty \le 1\}$ *and that all distributions in* $\mathcal{P}(\Xi)$ *are absolutely continuous with respect to* $\hat{\mathbb{P}}$. *By Theorem 1 and Proposition 5, the corresponding globalized distributionally robust counterpart* (15) *is satisfiable if and only if*

$$\begin{cases} t\theta + \max_{n\in[N]} \left\{ f(\boldsymbol{x}, \hat{\boldsymbol{\xi}}_n) - u_n \right\} + \dfrac{1}{N}\displaystyle\sum_{n\in[N]} u_n \le 0 & \\ |u_m - u_n| \le t\|\hat{\boldsymbol{\xi}}_m - \hat{\boldsymbol{\xi}}_n\| & \forall m,\, n \in [N] \\ |u_n| \le t & \forall n \in [N] \\ t \in [0, \gamma]. & \end{cases}$$

This, again, is a convex constraint system under the assumptions on f as given in Proposition 1.

4.2. Optimal Transport Ambiguity Set

The optimal transport cost (Blanchet and Murthy 2019, Santambrogio 2015), equipped with a generic function c, between any two distributions $\mathbb{P}_1$ and $\mathbb{P}_2$ is defined as

$$d_c(\mathbb{P}_1, \mathbb{P}_2) \;=\; \inf_{\pi \in \mathcal{Q}(\mathbb{P}_1, \mathbb{P}_2)} \int_{\Xi\times\Xi} c(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) \mathrm{d}\pi(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2),$$

which recovers the type-1 Wasserstein distance if $c(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) = \|\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2\|$ for some norm. With optimal transport, we can represent the globalized distributionally robust counterpart (1) as

$$\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] \;\le\; \gamma \cdot \min_{\mathbb{Q}\in\mathcal{F}_{c_2}(\theta)} d_{c_1}(\mathbb{P}, \mathbb{Q}) \quad \forall \mathbb{P} \in \mathcal{P}(\Xi), \tag{18}$$

where d_{c_1} is an optimal transport cost defined by the function c_1 and $\mathcal{F}_{c_2}(\theta)$ is an optimal transport ambiguity set based on another optimal transport cost d_{c_2}, captured by

$$\mathcal{F}_{c_2}(\theta) = \{\mathbb{P} \in \mathcal{P}(\Xi) \mid d_{c_2}(\mathbb{P}, \hat{\mathbb{P}}) \le \theta\}. \tag{19}$$

Note that herein we consider two optimal transport costs appearing in the constraint and ambiguity set, respectively. Moreover, we assume c_1 and c_2 are nonnegative, lower semicoutinuous, jointly convex in both arguments, and $c_1(\boldsymbol{\xi}, \boldsymbol{\xi}) = c_2(\boldsymbol{\xi}, \boldsymbol{\xi}) = 0$ for all $\boldsymbol{\xi} \in \Xi$. The globalized distributionally robust counterpart (18) based on the generic optimal transport also admits an attractive tractable reformulation—a result we summarize as follows.

Theorem 2. *Suppose conditions in Proposition 1 hold. Then any fixed decision* $\boldsymbol{x} \in \mathcal{X}$ *satisfies the globalized distributionally robust counterpart* (18) *if and only if there are* $\boldsymbol{s} \in \mathbb{R}^N$, $t \ge 0$, *and* $\boldsymbol{w}_{nk}, \boldsymbol{v}_{nk}, \boldsymbol{y}_{nk} \in \mathbb{R}^I$, $n \in [N]$ *such that the following constraint system is feasible:*

$$\begin{cases} \theta N t + \boldsymbol{e}^\top \boldsymbol{s} \le 0 \\ [-f_k]^{\bar{*}}(\boldsymbol{x}, \boldsymbol{w}_{nk} + \boldsymbol{v}_{nk}) + \delta^*(\boldsymbol{w}_{nk} \mid \Xi) + \delta^*(\boldsymbol{v}_{nk} - \boldsymbol{y}_{nk} \mid \Xi) \\ \qquad\qquad + \gamma c_1^*(\boldsymbol{v}_{nk}/\gamma, -\boldsymbol{v}_{nk}/\gamma) + t c_2^{\bar{*}}(\hat{\boldsymbol{\xi}}_n, \boldsymbol{y}_{nk}/t) \le s_n \quad \forall n \in [N],\, k \in [K]. \end{cases} \tag{20}$$

Both perspective functions $\gamma c_1^*(\boldsymbol{v}_{nk}/\gamma, -\boldsymbol{v}_{nk}/\gamma)$ and $tc_2^{\bar{*}}(\hat{\boldsymbol{\xi}}_n, \boldsymbol{y}_{nk}/t)$ are convex in $(\boldsymbol{v}_{nk},\gamma)$ and $(\boldsymbol{y}_{nk}, t)$, respectively. Hence, under those conditions in Proposition 1 and together with the term $[-f_k]^{\bar{*}}(\boldsymbol{x}, \boldsymbol{w}_{nk} + \boldsymbol{v}_{nk}) + \delta^*(\boldsymbol{w}_{nk} \mid \Xi) + \delta^*(\boldsymbol{v}_{nk} - \boldsymbol{y}_{nk} \mid \Xi)$, these two perspective functions lead to an attractive tractable reformulation of (20) when c_1 and c_2 are properly chosen. For example, if we choose $c_1 = c_2 = \|\cdot\|^\rho$ for some norm $\|\cdot\|$ and $\rho \in [1,\infty)$, then the optimal transport ambiguity set of size θ in (19) is equivalent to a type-ρ Wasserstein ambiguity set of size $\sqrt[\rho]{\theta}$:

$$\mathcal{F}_\rho(\sqrt[\rho]{\theta}) = \{\mathbb{P} \in \mathcal{P}(\Xi) \mid d_\rho(\mathbb{P}, \hat{\mathbb{P}}) \le \sqrt[\rho]{\theta}\};$$

see, *e.g.*, Chen et al. (2020, Appendix C) and Gao and Kleywegt (2022). Here, $\mathcal{F}_\rho(\sqrt[\rho]{\theta})$ is defined through the type-ρ Wasserstein distance

$$d_\rho(\mathbb{P}_1, \mathbb{P}_2) = \left(\inf_{\pi \in \mathcal{Q}(\mathbb{P}_1,\mathbb{P}_2)} \int_{\Xi\times\Xi} \|\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2\|^\rho \, \mathrm{d}\pi(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) \right)^{1/\rho} = \left(d_{\|\cdot\|^\rho}(\mathbb{P}_1, \mathbb{P}_2) \right)^{1/\rho},$$

where the rightmost $d_{\|\cdot\|^\rho}$ is an optimal transport cost with $c(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) = \|\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2\|^\rho$. Reformulation of the corresponding globalized distributionally robust counterpart is provided as follows.

Theorem 3. *Any fixed decision $\boldsymbol{x} \in \mathcal{X}$ satisfies the globalized distributionally robust counterpart*

$$\mathbb{E}_\mathbb{P}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] \le \gamma \cdot \min_{\mathbb{Q} \in \mathcal{F}_\rho(\sqrt[\rho]{\theta})} d_{\|\cdot\|^\rho}(\mathbb{P}, \mathbb{Q}) \quad \forall \mathbb{P} \in \mathcal{P}(\Xi)$$

if and only if there exists $t \ge 0$ satisfying

$$\theta N t + \sum_{n\in[N]} \sup_{\boldsymbol{\xi},\boldsymbol{\zeta}\in\Xi} \left\{ f(\boldsymbol{x}, \boldsymbol{\xi}) - \gamma\|\boldsymbol{\xi} - \boldsymbol{\zeta}\|^\rho - t\|\boldsymbol{\zeta} - \hat{\boldsymbol{\xi}}_n\|^\rho \right\} \le 0.$$

Indeed, Proposition 1 is a special case of Theorem 3 by first taking $\rho = 1$ and then eliminating the variable γ. We next provide another special case where $\rho = 2$ and $\|\cdot\|$ takes the L_2-norm $\|\cdot\|_2$.

Proposition 6. *Let $\rho = 2$ and $\|\cdot\| = \|\cdot\|_2$ in Theorem 3. Then any fixed decision $\boldsymbol{x} \in \mathcal{X}$ satisfies the globalized distributionally robust counterpart*

$$\mathbb{E}_\mathbb{P}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] \le \gamma \cdot \min_{\mathbb{Q} \in \{\mathbb{P}\in\mathcal{P}(\Xi) \mid d_2(\mathbb{P},\hat{\mathbb{P}}) \le \sqrt{\theta}\}} d_{\|\cdot\|_2^2}(\mathbb{P}, \mathbb{Q}) \quad \forall \mathbb{P} \in \mathcal{P}(\Xi)$$

if and only if there exists $t \ge 0$ satisfying

$$\theta N t + \sum_{n\in[N]} \sup_{\boldsymbol{\xi}\in\Xi} \left\{ f(\boldsymbol{x}, \boldsymbol{\xi}) - \frac{t\gamma}{t+\gamma} \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}_n\|_2^2 \right\} \le 0.$$

4.3. ϕ-Divergence Ambiguity Set

Let us now consider in the globalized distributionally robust counterpart the class of ϕ-divergences (Ben-Tal et al. 2013, Hu and Hong 2013, Bayraksan and Love 2015). In particular, we assume the support set is discrete and finite, given by $\Xi = \{\hat{\boldsymbol{\xi}}_1, \ldots, \hat{\boldsymbol{\xi}}_N\}$. A distribution supported on such a discrete set can then be captured by the probability masses put on these finite support points. The ϕ-divergence of a discrete distribution $\mathbb{P} = \sum_{n\in[N]} p_n \delta_{\hat{\boldsymbol{\xi}}_n}$, relative to another discrete distribution $\mathbb{Q} = \sum_{n\in[N]} q_n \delta_{\hat{\boldsymbol{\xi}}_n}$, is given by

$$d_\phi(\mathbb{P} \,|\, \mathbb{Q}) = \sum_{n\in[N]} q_n \phi\left(\frac{p_n}{q_n}\right),$$

where $\boldsymbol{p}$ and $\boldsymbol{q}$ are probability vectors residing in the probability simplex. Throughout this section, we assume the divergence is based on a function ϕ that satisfies the following: (i) ϕ is nonnegative and convex on $\mathbb{R}_+$, (ii) $\phi(1) = 0$, (iii) $0\phi(0/0) = 0$, and (iv) $0\phi(a/0) = a \lim_{t\to\infty} \phi(t)/t$ for $a > 0$.

The globalized distributionally robust counterpart based on ϕ-divergences is defined as follows:

$$\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] \le \gamma \cdot \min_{\mathbb{Q}\in\mathcal{F}_{\phi_2}(\theta)} d_{\phi_1}(\mathbb{P} \,|\, \mathbb{Q}) \quad \forall \mathbb{P} \in \mathcal{P}(\Xi), \tag{21}$$

where $d_{\phi_1}(\mathbb{P} \,|\, \mathbb{Q})$ is a ϕ-divergence based on the function ϕ_1 and $\mathcal{F}_{\phi_2}(\theta)$ is a ϕ-divergence ambiguity set of size θ based on another function ϕ_2:

$$\mathcal{F}_{\phi_2}(\theta) = \{\mathbb{P} \in \mathcal{P}(\Xi) \,|\, d_{\phi_2}(\mathbb{P} \,|\, \hat{\mathbb{P}}) \le \theta\}.$$

Here, we focus on $\hat{\mathbb{P}} = \frac{1}{N}\sum_{n\in[N]} \delta_{\hat{\boldsymbol{\xi}}_n}$ and for technical convenience we assume there exists some probability vector $\boldsymbol{p}$ that satisfies $\sum_{n\in[N]} \phi_2(Nq_n) < N\theta$.

Theorem 4. *Given $\theta > 0$, any fixed decision $\boldsymbol{x} \in \mathcal{X}$ satisfies the globalized distributionally robust counterpart* (21) *if and only if there are $t \ge 0$, $v \in \mathbb{R}$, $\boldsymbol{s} \in \mathbb{R}^N$ satisfying the following system:*

$$v + \theta t + \max_{n\in[N]} \left\{ s_n + \gamma\phi_1^*\left(\frac{f(\boldsymbol{x}, \hat{\boldsymbol{\xi}}_n) - v}{\gamma}\right)\right\} + \frac{1}{N}\sum_{n\in[N]} t\phi_2^*\left(-\frac{s_n}{t}\right) \le 0. \tag{22}$$

Taking a closer look at the reformulation (22) in Theorem 4, it is clear that $t\phi_2^*(-s_n/t)$ is convex in (s_n, t) and $\gamma\phi_2^*((f(\boldsymbol{x}, \hat{\boldsymbol{\xi}}_n) - v)/\gamma) = \sup_{t\ge0} \left\{t(f(\boldsymbol{x}, \hat{\boldsymbol{\xi}}_n) - v) - \gamma\phi_2(t)\right\}$ is also convex in $(\boldsymbol{x}, v, \gamma)$, provided that $f(\boldsymbol{x}, \hat{\boldsymbol{\xi}}_n)$ is convex in $\boldsymbol{x}$. In the following, we provide some examples based on the popular ϕ-divergences including the Kullback-Leibler divergence and χ^2-distance.

Example 4 (Kullback-Leibler divergence). *Consider $\phi_1(t) = \phi_2(t) = t\log(t) - t + 1$ such that $\phi_1^*(s) = \phi_2^*(s) = e^s - 1$. By Theorem 4, constraint* (21) *based on the Kullback-Leibler divergence ambiguity set $\mathcal{F}_{\mathrm{KL}}(\theta) = \{\mathbb{P} \in \mathcal{P}(\Xi) \,|\, d_{\phi_1}(\mathbb{P} \,|\, \hat{\mathbb{P}}) \le \theta\}$ is equivalent to the constraint system*

$$\begin{cases} v + \theta t - t - \gamma + \max\limits_{n\in[N]} \left\{ s_n + \gamma \exp\left(\dfrac{f(\boldsymbol{x}, \hat{\boldsymbol{\xi}}_n) - v}{\gamma}\right)\right\} + \dfrac{1}{N}\sum\limits_{n\in[N]} t\exp\left(-\dfrac{s_n}{t}\right) \le 0 \\ t \ge 0,\ v \in \mathbb{R},\ \boldsymbol{s} \in \mathbb{R}^N, \end{cases} \tag{23}$$

which, as $\gamma \to \infty$, is expected to recover the equivalent constraint system of

$$\sup_{\mathbb{P} \in \mathcal{F}_{\mathrm{KL}}(\theta)} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] \leq 0. \tag{24}$$

To see this, note that $\lim_{\gamma \to \infty} \gamma(\exp(t/\gamma) - 1) = t$, *leading* (23) *to*

$$\begin{cases} \theta t - t + \max_{n \in [N]} \left\{ s_n + f(\boldsymbol{x}, \hat{\boldsymbol{\xi}}_n) \right\} + \dfrac{1}{N} \sum_{n \in [N]} t \exp\left(-\dfrac{s_n}{t} \right) \leq 0 \\ t \geq 0,\ \boldsymbol{s} \in \mathbb{R}^N, \end{cases}$$

$$\iff \quad \begin{cases} \theta t - t + v + \dfrac{1}{N} \sum_{n \in [N]} t \exp\left(-\dfrac{s_n}{t} \right) \leq 0 \\ s_n + f(\boldsymbol{x}, \hat{\boldsymbol{\xi}}_n) \leq v \quad \forall n \in [N] \\ t \geq 0,\ v \in \mathbb{R},\ \boldsymbol{s} \in \mathbb{R}^N. \end{cases}$$

Since the exponential function is nondecreasing, the system can be further simplified into

$$\begin{cases} \theta t - t + v + \dfrac{1}{N} \sum_{n \in [N]} t \exp\left(\dfrac{f(\boldsymbol{x}, \hat{\boldsymbol{\xi}}_n) - v}{t} \right) \leq 0 \\ t \geq 0,\ v \in \mathbb{R}. \end{cases}$$

For any fixed $t \geq 0$, the optimal solution $v^\star$ to the minimization problem

$$\inf_{v \in \mathbb{R}} \left\{ \theta t - t + v + \frac{1}{N} \sum_{n \in [N]} t \exp\left(\frac{f(\boldsymbol{x}, \hat{\boldsymbol{\xi}}_n) - v}{t} \right) \right\}$$

can be determined by the first-order condition

$$1 - \frac{1}{N} \sum_{n \in [N]} \exp\left(\frac{f(\boldsymbol{x}, \hat{\boldsymbol{\xi}}_n) - v^\star}{t} \right) = 0 \quad \Longrightarrow \quad v^\star = t \ln\left(\frac{1}{N} \sum_{n \in [N]} \exp\left(\frac{f(\boldsymbol{x}, \hat{\boldsymbol{\xi}}_n)}{t} \right) \right).$$

Substituting $v^\star$ then yields that when $\gamma \to \infty$, the constraint system (23) *becomes*

$$\begin{cases} \theta + \ln\left(\dfrac{1}{N} \sum_{n \in [N]} \exp\left(\dfrac{f(\boldsymbol{x}, \hat{\boldsymbol{\xi}}_n)}{t} \right) \right) \leq 0 \\ t \geq 0, \end{cases}$$

which is exactly the reformulation of the distributionally robust counterpart (24).

Example 5 (χ^2-distance). *Consider $\phi_1(t) = \phi_2(t) = (t-1)^2/t$ such that $\phi_1^*(s) = \phi_2^*(s) = 2 - 2\sqrt{1-s}$ with $s \leq 1$. By Theorem 4, constraint* (21) *based on the χ^2-distance ambiguity set $\mathcal{F}_{\chi^2}(\theta) = \{\mathbb{P} \in \mathcal{P}(\Xi) \mid d_{\phi_1}(\mathbb{P} \mid \hat{\mathbb{P}}) \leq \theta\}$ is equivalent to*

$$\begin{cases} v + \theta t + 2t + 2\gamma + \max_{n \in [N]} \left\{ s_n - 2\gamma \sqrt{1 - \dfrac{f(\boldsymbol{x}, \hat{\boldsymbol{\xi}}_n) - v}{\gamma}} \right\} - \dfrac{2}{N} \sum_{n \in [N]} t \sqrt{1 + \dfrac{s_n}{t}} \leq 0 & \\ f(\boldsymbol{x}, \hat{\boldsymbol{\xi}}_n) - v \leq \gamma & \forall n \in [N] \\ s_n \geq -t & \forall n \in [N] \\ t \geq 0,\ v \in \mathbb{R},\ \boldsymbol{s} \in \mathbb{R}^N. & \end{cases}$$

4.4. Moment Ambiguity Set

Let $\boldsymbol{g}:\mathbb{R}^I \to \mathbb{R}^M$ be a function such that the set $\{(\boldsymbol{\xi},\boldsymbol{u}) \in \mathbb{R}^I \times \mathbb{R}^M \mid \boldsymbol{g}(\boldsymbol{\xi}) \preceq_{\mathcal{K}} \boldsymbol{u}\}$ is conic representable via a proper cone $\mathcal{K}$. We then consider moment ambiguity sets that take the following generic format proposed by Wiesemann et al. (2014):

$$\mathcal{F}_{\mathrm{M}} = \{\mathbb{P} \in \mathcal{P}(\mathbb{R}^I) \mid \mathbb{P}[\tilde{\boldsymbol{\xi}} \in \Xi] = 1,\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\xi}}] = \boldsymbol{\mu},\ \mathbb{E}_{\mathbb{P}}[\boldsymbol{g}(\tilde{\boldsymbol{\xi}})] \preceq_{\mathcal{K}} \boldsymbol{h}\}.$$

Here, we assume $\boldsymbol{\mu} \in \mathrm{int}(\Xi)$ and $\boldsymbol{g}(\boldsymbol{\mu}) \prec_{\mathcal{K}} \boldsymbol{h}$. With a moment ambiguity set $\mathcal{F}_{\mathrm{M}}$, we consider the globalized distributionally robust counterpart

$$\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})] \le \gamma \cdot \min_{\mathbb{Q}\in\mathcal{F}_{\mathrm{M}}} d_{\mathrm{W}}(\mathbb{P},\mathbb{Q}) \quad \forall \mathbb{P} \in \mathcal{P}(\Xi). \tag{25}$$

Theorem 5. *Suppose conditions in Proposition 1 hold. Then any fixed decision $\boldsymbol{x} \in \mathcal{X}$ satisfies globalized distributionally robust counterpart* (25) *if and only if it satisfies the following system*

$$\begin{cases} \boldsymbol{\alpha}^\top \boldsymbol{\mu} + \boldsymbol{\beta}^\top \boldsymbol{h} + t \le 0 & \\ [-f_k]^{\bar{*}}(\boldsymbol{x}, \boldsymbol{u}_k - \boldsymbol{w}_k) + \delta^*(\boldsymbol{u}_k \mid \Xi) + \delta^*(-\boldsymbol{w}_k - \boldsymbol{\alpha}, -\boldsymbol{\beta} \mid \bar{\Xi}) \le t & \forall k \in [K] \\ \|\boldsymbol{w}_k\|_* \le \gamma & \forall k \in [K] \\ \boldsymbol{u}_k,\ \boldsymbol{w}_k \in \mathbb{R}^I & \forall k \in [K] \\ t \in \mathbb{R},\ \boldsymbol{\alpha} \in \mathbb{R}^I,\ \boldsymbol{\beta} \in \mathcal{K}_*, & \end{cases}$$

where $\bar{\Xi} = \{(\boldsymbol{\xi},\boldsymbol{v}) \in \mathbb{R}^I \times \mathbb{R}^M \mid \boldsymbol{\xi} \in \Xi,\ \boldsymbol{g}(\boldsymbol{\xi}) \preceq_{\mathcal{K}} \boldsymbol{v}\}$.

When $\gamma \to \infty$, the constraint system in Theorem 5 becomes

$$\begin{cases} \boldsymbol{\alpha}^\top \boldsymbol{\mu} + \boldsymbol{\beta}^\top \boldsymbol{h} + t \le 0 & \\ [-f_k]^{\bar{*}}(\boldsymbol{x}, \boldsymbol{u}_k - \boldsymbol{w}_k) + \delta^*(\boldsymbol{u}_k \mid \Xi) + \delta^*(-\boldsymbol{w}_k - \boldsymbol{\alpha}, -\boldsymbol{\beta} \mid \bar{\Xi}) \le t & \forall k \in [K] \\ \boldsymbol{u}_k, \boldsymbol{w}_k \in \mathbb{R}^I & \forall k \in [K] \\ t \in \mathbb{R},\ \boldsymbol{\alpha} \in \mathbb{R}^I,\ \boldsymbol{\beta} \in \mathcal{K}_*, & \end{cases}$$

which is the dual reformulation of the distributionally robust counterpart

$$\sup_{\mathbb{P}\in\mathcal{F}_{\mathrm{M}}} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})] \le 0.$$

That is to say, the globalized distributionally robust counterpart pays an additional constraint $\max_{k\in[K]} \|\boldsymbol{w}_k\|_* \le \gamma$ as the price of globalized robustness.

The generic format of moment ambiguity sets possesses a powerful modeling flexibility to recover many popular ambiguity sets, as illustrated in the coming two examples.

Example 6 (Covariance information). *To specify covariance information, we can consider*

$$\mathcal{F}_{\mathrm{M}} = \{\mathbb{P} \in \mathcal{P}(\mathbb{R}^I) \mid \mathbb{P}[\tilde{\boldsymbol{\xi}} \in \Xi] = 1,\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\xi}}] = \boldsymbol{\mu},\ \mathbb{E}_{\mathbb{P}}[(\boldsymbol{\xi} - \boldsymbol{\mu})(\boldsymbol{\xi} - \boldsymbol{\mu})^\top] \preceq \boldsymbol{\Sigma}\}$$

such that $\boldsymbol{\Sigma} \succ \mathbf{0}$, $g(\boldsymbol{\xi}) = (\boldsymbol{\xi} - \boldsymbol{\mu})(\boldsymbol{\xi} - \boldsymbol{\mu})^\top$, *and* $\boldsymbol{h}$ *is an identity matrix. Using Theorem 5, we know that any decision* $\boldsymbol{x} \in \mathcal{X}$ *satisfies* (25) *if and only if it satisfies the following constraint system*

$$\begin{cases} \boldsymbol{\alpha}^\top \boldsymbol{\mu} + \langle \boldsymbol{\Sigma}, \boldsymbol{Q} \rangle + s \leq 0 & \\ [-f_k]^{\bar{*}}(\boldsymbol{x}, \boldsymbol{u}_k - \boldsymbol{w}_k) + \delta^*(\boldsymbol{u}_k \mid \Xi) + \delta^*(-\boldsymbol{w}_k - \boldsymbol{\alpha}, -\boldsymbol{Q} \mid \bar{\Xi}) \leq s & \forall k \in [K] \\ \|\boldsymbol{w}_k\|_* \leq \gamma,\ \boldsymbol{u}_k \in \mathbb{R}^I & \forall k \in [K] \\ \boldsymbol{\alpha} \in \mathbb{R}^I,\ \boldsymbol{Q} \succeq \mathbf{0}, & \end{cases}$$

where $\langle \cdot, \cdot \rangle$ *refers to the Frobenius inner product between matrices and*

$$\bar{\Xi} = \left\{ (\boldsymbol{\xi}, \boldsymbol{U}) \in \mathbb{R}^I \times \mathbb{R}^{I \times I} \ \middle|\ \boldsymbol{\xi} \in \Xi,\ \begin{bmatrix} 1 & (\boldsymbol{\xi} - \boldsymbol{\mu})^\top \\ (\boldsymbol{\xi} - \boldsymbol{\mu}) & \boldsymbol{U} \end{bmatrix} \succeq \mathbf{0} \right\}$$

is a conic representable set via the positive semidefinite cone. When $\gamma \to \infty$, *the above system reduces to the reformulation of* $\sup_{\mathbb{P} \in \mathcal{F}_{\mathrm{M}}} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] \leq 0$; *see example 2 in Wiesemann et al. (2014).*

Example 7 (Mahalanobis distance). *The Mahalanobis distance can be used to measure the disturbance of a random variable* $\tilde{\boldsymbol{\xi}}$ *around its mean* $\boldsymbol{\mu}$ *with respect to the covariance matrix* $\boldsymbol{\Sigma}$. *Specifically, we consider a moment ambiguity set*

$$\mathcal{F}_{\mathrm{M}}(\theta) = \left\{ \mathbb{P} \in \mathcal{P}(\mathbb{R}^I) \ \middle|\ \mathbb{P}[\tilde{\boldsymbol{\xi}} \in \Xi] = 1,\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\xi}}] = \boldsymbol{\mu},\ \mathbb{E}_{\mathbb{P}}\left[\sqrt{(\tilde{\boldsymbol{\xi}} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\tilde{\boldsymbol{\xi}} - \boldsymbol{\mu})} \right] \leq \theta \right\}$$

with $\theta \geq 0$, $\boldsymbol{\Sigma} \succ \mathbf{0}$, *and* $g(\boldsymbol{\xi}) = \sqrt{(\boldsymbol{\xi} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\xi} - \boldsymbol{\mu})}$. *To derive a more explicit reformulation using Theorem 5, notice that* $\delta^*(-\boldsymbol{w} - \boldsymbol{\alpha}, -\beta \mid \bar{\Xi}) = \sup_{\boldsymbol{\xi} \in \Xi,\, v \geq g(\boldsymbol{\xi})} \{ (-\boldsymbol{w} - \boldsymbol{\alpha})^\top \boldsymbol{\xi} - \beta v \}$ *tends to* ∞ *when* $\beta < 0$. *On the other hand, when* $\beta \geq 0$ *we have*

$$\begin{aligned} \delta^*(-\boldsymbol{w} - \boldsymbol{\alpha}, -\beta \mid \bar{\Xi}) = \sup_{\boldsymbol{\xi} \in \Xi,\, v \geq g(\boldsymbol{\xi})} \{ (-\boldsymbol{w} - \boldsymbol{\alpha})^\top \boldsymbol{\xi} - \beta v \} &= \sup_{\boldsymbol{\xi} \in \Xi} \left\{ (-\boldsymbol{w} - \boldsymbol{\alpha})^\top \boldsymbol{\xi} - \beta g(\boldsymbol{\xi}) \right\} \\ &= \sup_{\boldsymbol{\xi}} \left\{ (-\boldsymbol{w} - \boldsymbol{\alpha})^\top \boldsymbol{\xi} - \beta g(\boldsymbol{\xi}) - \delta(\boldsymbol{\xi} \mid \Xi) \right\} \\ &= \inf_{\boldsymbol{w},\, \boldsymbol{u}} \left\{ \beta g^*(\boldsymbol{u}/\beta) + \delta^*(-\boldsymbol{w} - \boldsymbol{\alpha} - \boldsymbol{u} \mid \Xi) \right\}, \end{aligned}$$

where the convex conjugate of g *is*

$$\begin{aligned} g^*(\boldsymbol{u}/\beta) = \sup_{\boldsymbol{\xi}} \left\{ \frac{\boldsymbol{u}^\top \boldsymbol{\xi}}{\beta} - \|\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\xi} - \boldsymbol{\mu})\|_2 \right\} &= \sup_{\boldsymbol{\xi}} \left\{ \frac{\boldsymbol{u}^\top \boldsymbol{\xi}}{\beta} - \|\boldsymbol{\Sigma}^{-1/2} \boldsymbol{\xi}\|_2 \right\} + \frac{\boldsymbol{u}^\top \boldsymbol{\mu}}{\beta} \\ &= \sup_{\boldsymbol{\xi}} \left\{ \frac{(\boldsymbol{\Sigma}^{1/2} \boldsymbol{u})^\top \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\xi}}{\beta} - \|\boldsymbol{\Sigma}^{-1/2} \boldsymbol{\xi}\|_2 \right\} + \frac{\boldsymbol{u}^\top \boldsymbol{\mu}}{\beta} \\ &= \begin{cases} \boldsymbol{u}^\top \boldsymbol{\mu}/\beta & \|\boldsymbol{\Sigma}^{1/2} \boldsymbol{u}\|_2 \leq \beta \\ \infty & \text{otherwise.} \end{cases} \end{aligned}$$

Therefore, any decision $\boldsymbol{x} \in \mathcal{X}$ *satisfies* (25) *if and only if it satisfies the constraint system*

$$\begin{cases} \boldsymbol{\alpha}^\top \boldsymbol{\mu} + \beta\theta + s \leq 0 & \\ [-f_k]^{\bar{*}}(\boldsymbol{x}, \boldsymbol{u}_k - \boldsymbol{w}_k) + \delta^*(\boldsymbol{u}_k \mid \Xi) + \boldsymbol{u}_k^\top \boldsymbol{\mu} + \delta^*(-\boldsymbol{w}_k - \boldsymbol{u}_k - \boldsymbol{\alpha} \mid \Xi) \leq s & \forall k \in [K] \\ \|\boldsymbol{w}_k\|_* \leq \gamma,\ \|\boldsymbol{\Sigma}^{1/2} \boldsymbol{u}_k\|_2 \leq \beta & \forall k \in [K] \\ \boldsymbol{\alpha} \in \mathbb{R}^I,\ \beta \in \mathbb{R}. & \end{cases}$$

5. Numerical Experiments

To demonstrate the effectiveness of our proposed globalized distributionally robust counterpart, we conduct numerical experiments on an adaptive network lot-sizing problem in a setting similar to that of Long et al. (2022). Our numerical experiments consist of two parts: (*i*) comparing our proposed globalized distributionally robust approach with the existing distributionally robust approach and (*ii*) comparing the globalized distributionally robust approach with the emerging robust satisficing approach. All results were produced on an Intel Core i9 3.7GHz with 32GB of RAM, using CPLEX 12.10.0 with the RSOME package (Chen et al. 2020) in MATLAB.

Consider I different stores with a random demand ξ_i at each store $i \in [I]$. In the first stage, before realization of random demands, we determine an initial stock allocation $x_i \in [0, \bar{x}_i]$ at a unit ordering cost c_i for different stores $i \in [I]$. In the second stage, after demand realization, we can transport stock y_{ij} from store i to store j at a unit transportation cost $q_{ij} = 2D_{ij}$, proportional to the Euclidean distance D_{ij} between stores i and j. Emergency orders w_i with a unit cost $l_i > c_i$ is also possible. In our simulation, we set $I = 10$ and pick uniformly at random locations of stores from a $[0, I] \times [0, I]$ grid. For all $i \in [I]$, we set $c_i = 10$, $l_i = 30$, and $\bar{x}_i = 40$. We let $N = 20$, *i.e.*, the empirical distribution consists of 20 historical samples, and we take another $M = 2000$ independent samples for testing the out-of-sample performance. Demand samples are generated from the underlying true distribution $\mathbb{P}_0$—a truncated normal distribution $\mathbb{N}(20, 15^2)$ on the support set $\Xi = [0, 40]^I$.

5.1. Comparison with Distributionally Robust Optimization Model

In this section we compare the G-ADRO model and the ADRO model by investigating the trade-off between the objective value and the probability of exceeding the target. We consider a G-ADRO model of the adaptive network lot-sizing problem as follows:

$$
\begin{array}{ll}
\min & \boldsymbol{c}^\top \boldsymbol{x} + \tau \\
\text{s.t.} & \mathbb{E}_{\mathbb{P}}[g(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] - \tau \leq \gamma \cdot \min\limits_{\mathbb{Q} \in \mathcal{F}_{\mathrm{W}}(\theta)} d_{\mathrm{W}}(\mathbb{P}, \mathbb{Q}) \qquad \forall \mathbb{P} \in \mathcal{P}(\Xi) \\
& 0 \leq \boldsymbol{x} \leq \bar{\boldsymbol{x}},
\end{array}
$$

where the second-stage cost function $g(\boldsymbol{x}, \boldsymbol{\xi})$ is given by

$$
g(\boldsymbol{x}, \boldsymbol{\xi}) = \left\{
\begin{array}{lll}
\min & \sum\limits_{i,j \in [I]} q_{ij} y_{ij} + \sum\limits_{i \in [I]} l_i w_i & \\
\text{s.t.} & x_i + w_i + \sum\limits_{j \in [I]} y_{ji} - \sum\limits_{j \in [I]} y_{ij} \geq \xi_i & \forall i \in [I] \\
& w_i \geq 0,\ y_{ij} \geq 0 & \forall i \in [I],\ j \in [I].
\end{array}
\right.
$$

Then we can reformulate the above problem as

$$
\begin{aligned}
\min \quad & \boldsymbol{c}^\top \boldsymbol{x} + \tau \\
\text{s.t.} \quad & \mathbb{E}_{\mathbb{P}}\left[\sum_{i,j\in[I]} q_{ij} y_{ij}(\tilde{\boldsymbol{\xi}}) + \sum_{i\in[I]} l_i w_i(\tilde{\boldsymbol{\xi}})\right] - \tau \le \gamma \cdot \min_{\mathbb{Q}\in\mathcal{F}_{\mathrm{W}}(\theta)} d_{\mathrm{W}}(\mathbb{P},\mathbb{Q}) && \forall \mathbb{P}\in\mathcal{P}(\Xi) \\
& x_i + w_i(\boldsymbol{\xi}) + \sum_{j\in[I]} y_{ji}(\boldsymbol{\xi}) - \sum_{j\in[I]} w_{ij}(\boldsymbol{\xi}) \ge \xi_i && \forall \boldsymbol{\xi}\in\Xi,\ i\in[I] \\
& \boldsymbol{y}(\boldsymbol{\xi}) \ge 0,\ \boldsymbol{w}(\boldsymbol{\xi}) \ge 0 && \forall \boldsymbol{\xi}\in\Xi \\
& 0 \le \boldsymbol{x} \le \bar{\boldsymbol{x}},\ \boldsymbol{y}: \mathbb{R}^I \to \mathbb{R}^{I\times I},\ \boldsymbol{w}: \mathbb{R}^I \to \mathbb{R}^I.
\end{aligned}
$$

Here, by Lemma 1, the globalized distributionally robust counterpart is satisfiable if and only if

$$
\exists\, \boldsymbol{s}\in\mathbb{R}^N,\ t\in[0,\gamma]: \begin{cases} \theta N t + \boldsymbol{e}^\top \boldsymbol{s} \le N\tau \\ \displaystyle\sum_{i,j\in[I]} q_{ij} y_{ij}(\boldsymbol{\xi}) + \sum_{i\in[I]} l_i w_i(\boldsymbol{\xi}) - t\|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}_n\| \le s_n \quad \forall \boldsymbol{\xi}\in\Xi,\ n\in[N]. \end{cases}
$$

In the above reformulation, the recourse functions are defined as infinite-dimensional mappings $\boldsymbol{y}$ and $\boldsymbol{w}$, rendering the problem intractable. To address issue, we adopt the piecewise affine decision rule (13) and solve the following approximate reformulation:

$$
\begin{aligned}
\min \quad & \boldsymbol{c}^\top \boldsymbol{x} + \tau \\
\text{s.t.} \quad & \theta N t + \boldsymbol{e}^\top \boldsymbol{s} \le N\tau \\
& \sum_{i,j\in[I]} q_{ij} y_{ij}(\boldsymbol{\xi},\zeta,n) + \sum_{i\in[I]} l_i w_i(\boldsymbol{\xi},\zeta,n) - \zeta t \le s_n && \forall (\boldsymbol{\xi},\zeta)\in\bar{\Xi}_n,\ n\in[N] \\
& x_i + w_i(\boldsymbol{\xi},\zeta,n) + \sum_{j\in[I]} y_{ji}(\boldsymbol{\xi},\zeta,n) - \sum_{j\in[I]} y_{ij}(\boldsymbol{\xi},\zeta,n) \ge \xi_i && \forall (\boldsymbol{\xi},\zeta)\in\bar{\Xi}_n,\ n\in[N],\ i\in[I] \\
& \boldsymbol{y}(\boldsymbol{\xi},\zeta,n) = \boldsymbol{y}_{0n} + \sum_{i\in[I]} \boldsymbol{y}_{in}\xi_i + \boldsymbol{y}_{Jn}\zeta \ge 0 && \forall (\boldsymbol{\xi},\zeta)\in\bar{\Xi}_n,\ n\in[N] \\
& \boldsymbol{w}(\boldsymbol{\xi},\zeta,n) = \boldsymbol{w}_{0n} + \sum_{i\in[I]} \boldsymbol{w}_{in}\xi_i + \boldsymbol{y}_{Jn}\zeta \ge 0 && \forall (\boldsymbol{\xi},\zeta)\in\bar{\Xi}_n,\ n\in[N] \\
& t\in[0,\gamma],\ 0 \le \boldsymbol{x} \le \bar{\boldsymbol{x}},
\end{aligned} \tag{26}
$$

where for each $n\in[N]$, $\bar{\Xi}_n = \{(\boldsymbol{\xi},\zeta)\in\Xi\times\mathbb{R} \mid \zeta \ge \|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}_n\|\}$. For computational convenience, we take $\|\cdot\|$ as the L_1-norm. Note that when $\gamma\to\infty$, problem (26) reduces to the decision rule approximation (which is based on (13)) of the ADRO model to this lot-sizing problem:

$$
\begin{aligned}
\min \quad & \boldsymbol{c}^\top \boldsymbol{x} + \tau \\
\text{s.t.} \quad & \mathbb{E}_{\mathbb{P}}[g(\boldsymbol{x},\tilde{\boldsymbol{\xi}})] \le \tau \qquad \forall \mathbb{P}\in\mathcal{F}_{\mathrm{W}}(\theta) \\
& 0 \le \boldsymbol{x} \le \bar{\boldsymbol{x}}.
\end{aligned}
$$

Constraint violation	$d_{\mathrm{W}}(\mathbb{P}_{\mathrm{out}}, \hat{\mathbb{P}})$						
	0	1	2	3	4	5	**6**
ADRO ($\gamma = +\infty$)	-14.67%	-9.52%	-4.36%	0.79%	5.94%	11.10%	**16.25%**
G-ADRO ($\gamma = 32$)	-22.51%	-17.51%	-12.51%	-7.50%	-2.50%	2.50%	**7.50%**
G-ADRO ($\gamma = 30$)	-33.40%	-28.81%	-23.81%	-13.81%	-8.81%	-3.81%	**1.19%**

Table 1 Constraint violation under different values of the Wasserstein distance $d_{\mathrm{W}}(\mathbb{P}_{\mathrm{out}}, \hat{\mathbb{P}})$ between the out-of-sample distribution $\mathbb{P}_{\mathrm{out}}$ and the reference distribution $\hat{\mathbb{P}}$.

Globalized Robustness

We first demonstrate the performance of the G-ADRO model in achieving globalized robustness compared to the ADRO model in a stress test. In particular, we consider the ADRO model ($\gamma \to +\infty$) and two G-ADRO models ($\gamma = 32$ and $\gamma = 30$). For all three models, we fix $\theta = 2$ and solve the aforementioned reformulations based on the piecewise affine decision rule (13). For each model, the corresponding nominal expected second-stage cost, given by

$$\frac{1}{N} \sum_{n \in [N]} \texttt{second-stage cost under the n-th historical sample},$$

is taken as a `target` to evaluate the out-of-sample performance. Specifically, for each model we fix its first-stage decision and evaluate its performance on the constraint violation under an out-of-sample distribution $\mathbb{P}_{\mathrm{out}} = \sum_{m \in [M]} p_m \delta_{\hat{\boldsymbol{\xi}}_m}$ supported on at most $M = 2000$ scenarios generated from $\mathbb{P}_0$. The constraint violation is defined by

$$\sum_{m \in [M]} p_m \times \frac{\texttt{second-stage cost under the m-th scenario - target}}{\texttt{target}} \times 100\%.$$

For our experimental purpose, we consider a series of the out-of-sample distributions by varying $d_{\mathrm{W}}(\mathbb{P}_{\mathrm{out}}, \hat{\mathbb{P}})$ from 0 to 6. Specifically, for each $\theta' \in \{0, 1, \ldots, 6\}$, we obtain $\mathbb{P}_{\mathrm{out}}$ by solving a linear program that maximizes the expected second-stage cost of the ADRO solution under $\mathbb{P}_{\mathrm{out}}$ subject to the constraint $d_{\mathrm{W}}(\mathbb{P}_{\mathrm{out}}, \hat{\mathbb{P}}) = \theta'$. The results are summarized in Table 1.

It can be seen that G-ADRO models outperform the ADRO model in mitigating constraint violation. On the one hand, for $\mathbb{P}_{\mathrm{out}}$ inside the ambiguity set $\mathcal{F}_{\mathrm{W}}(\theta)$ (recall that $\theta = 2$ in this experiment), all three models have no constraint violation, yet G-ADRO models exhibit more cost savings. On the other hand, as $\mathbb{P}_{\mathrm{out}}$ moves out of $\mathcal{F}_{\mathrm{W}}(\theta)$, the ADRO model starts to suffer from constraint violation when $d_{\mathrm{W}}(\mathbb{P}_{\mathrm{out}}, \hat{\mathbb{P}}) \geq 3$, while G-ADRO models with $\gamma = 32$ and $\gamma = 30$ are able to ensure no constraint violation until $d_{\mathrm{W}}(\mathbb{P}_{\mathrm{out}}, \hat{\mathbb{P}}) \geq 5$ and $d_{\mathrm{W}}(\mathbb{P}_{\mathrm{out}}, \hat{\mathbb{P}}) \geq 6$, respectively. These observations well justify the globalized robustness equipped with our proposed G-ADRO model and show that a higher level of globalized robustness can be achieved by setting a lower value of γ, which is also consistent with the motivating Example 1.

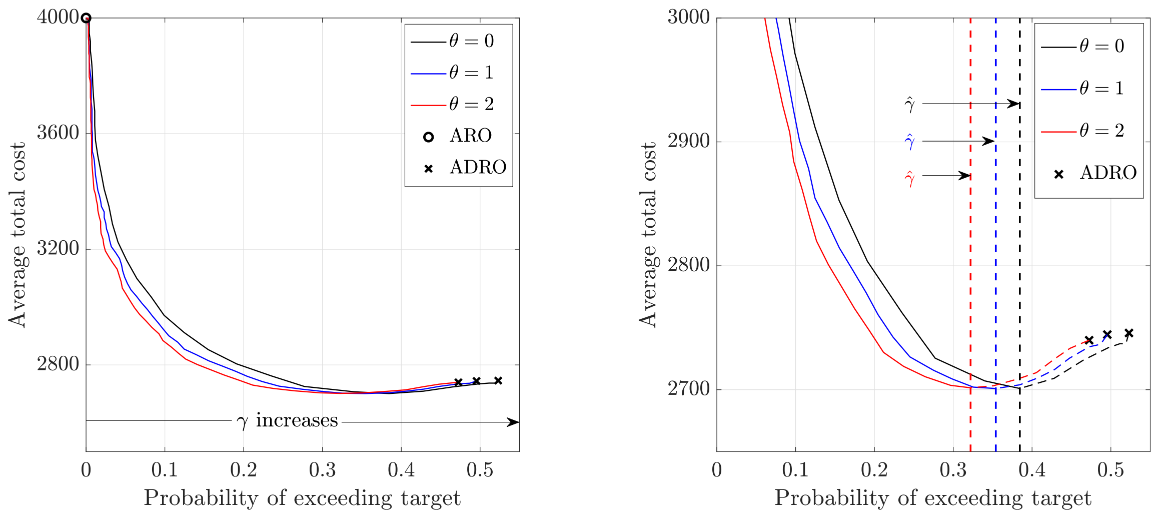

Figure 2 The circle '$\circ$' indicates the ARO solution ($\gamma = 0$) and the cross '$\times$' represents the ADRO solution ($\gamma \to \infty$) with the respective θ. The ADRO solution with $\theta = 0$ is the sample average approximation (SAA) solution. The right panel zooms in the region $[2650, 3000]$ of the average total cost in left panel.

Trade-Off

We also test the out-of-sample performance of the G-ADRO model on a trade-off between the average total cost and the probability of exceeding target. To obtain solutions of different models, for each fixed value of θ in $\{0, 1, 2\}$, we solve the ADRO model for its optimal dual variable $t^{\star}$. We then solve the G-ADRO model under various values of the shadow price of globalized robustness $t \in [0, t^{\star}]$. Note that $t = 0$ and $t = t^{\star}$ corresponds to the traditional robust and distributionally robust approaches, respectively. For each θ in $\{0, 1, 2\}$, we set the nominal second-stage cost of the corresponding ADRO model as the `target`. Then based on the $M = 2000$ testing samples generated from $\mathbb{P}_0$, for each approach we calculate the average total cost,

$$\frac{1}{M} \sum_{m \in [M]} \texttt{total cost under the m-th sample},$$

as well as the probability that the second-stage cost exceeds the target,

$$\frac{1}{M} \sum_{m \in [M]} \mathbb{I}\{\texttt{second-stage cost under the m-th sample exceeds the target}\}.$$

We report the frontier of "average total cost vs. probability of exceeding the target" in Figure 2 and summarize several findings as follows.

(i) As $\gamma \to 0$, all models converge to the adaptive robust optimization (ARO) approach, which has the highest average total cost and appears to be very conservative, albeit with a zero probability of exceeding the target.

(*ii*) Given θ, there is a clear trade-off on the frontier: as γ increases, the average total cost decreases while the probability increases. This well justifies the role of γ—a smaller value of γ indicates a higher level of globalized robustness.

(*iii*) For each given θ, there exists a threshold $\hat{\gamma}$ such that whenever $\gamma > \hat{\gamma}$, both performances with respect to average total cost and probability of exceeding the target deteriorate. In particular, ADRO solutions to different θ's all fall in the inefficient portion of the frontier. This further justifies the importance of incorporating the globalized robustness into the ADRO model by choosing a good γ.

(*iv*) Quite interestingly, the efficient portion of the frontier, $[0, \hat{\gamma}]$ becomes narrower for a larger value of θ, cautioning a higher risk of going beyond the efficient portion when specifying the value of γ.

5.2. Comparison with Robust Satisficing Model

In this subsection we compare the G-ARS model with the ARS model. In particular, we formulate the G-ARS model of the adaptive network lot-sizing problem as follows:

$$
\begin{array}{ll}
\min & \gamma \\
\text{s.t.} & \boldsymbol{c}^\top \boldsymbol{x} + \mathbb{E}_{\mathbb{P}}[g(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] - \tau \le \gamma \cdot \min\limits_{\mathbb{Q} \in \mathcal{F}_{\mathrm{W}}(\theta)} d_{\mathrm{W}}(\mathbb{P}, \mathbb{Q}) \qquad \forall \mathbb{P} \in \mathcal{P}(\Xi) \\
& \gamma \ge 0,\ 0 \le \boldsymbol{x} \le \bar{\boldsymbol{x}}.
\end{array}
$$

with the second-stage cost function $g(\boldsymbol{x}, \boldsymbol{\xi})$ sharing the same expression as in Section 5.1. We focus on the following approximate reformulation based on the piecewise affine decision rule (13):

$$
\begin{array}{lll}
\min & \gamma & \\
\text{s.t.} & \theta N t + N \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{e}^\top \boldsymbol{s} \le N\tau & \\
& \sum\limits_{i,j \in [I]} q_{ij} y_{ij}(\boldsymbol{\xi}, \zeta, n) + \sum\limits_{i \in [I]} l_i w_i(\boldsymbol{\xi}, \zeta, n) - \zeta t \le s_n & \forall (\boldsymbol{\xi}, \zeta) \in \bar{\Xi}_n,\ n \in [N] \\
& x_i + w_i(\boldsymbol{\xi}, \zeta, n) + \sum\limits_{j \in [I]} y_{ji}(\boldsymbol{\xi}, \zeta, n) - \sum\limits_{j \in [I]} y_{ij}(\boldsymbol{\xi}, \zeta, n) \ge \xi_i & \forall (\boldsymbol{\xi}, \zeta) \in \bar{\Xi}_n,\ n \in [N],\ i \in [I] \\
& \boldsymbol{y}(\boldsymbol{\xi}, \zeta, n) = \boldsymbol{y}_{0n} + \sum\limits_{i \in [I]} \boldsymbol{y}_{in} \xi_i + \boldsymbol{y}_{Jn} \zeta \ge 0 & \forall (\boldsymbol{\xi}, \zeta) \in \bar{\Xi}_n,\ n \in [N] \\
& \boldsymbol{w}(\boldsymbol{\xi}, \zeta, n) = \boldsymbol{w}_{0n} + \sum\limits_{i \in [I]} \boldsymbol{w}_{in} \xi_i + \boldsymbol{y}_{Jn} \zeta \ge 0 & \forall (\boldsymbol{\xi}, \zeta) \in \bar{\Xi}_n,\ n \in [N] \\
& \gamma \ge 0,\ t \in [0, \gamma],\ 0 \le \boldsymbol{x} \le \bar{\boldsymbol{x}}. &
\end{array}
\tag{27}
$$

Note that when $\theta = 0$, problem (27) reduces to the decision rule approximation (which is based on (13)) of the ARS model to this lot-sizing problem:

$$
\begin{array}{ll}
\min & \gamma \\
\text{s.t.} & \boldsymbol{c}^\top \boldsymbol{x} + \mathbb{E}_{\mathbb{P}}[g(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] - \tau \le \gamma \cdot d_{\mathrm{W}}(\mathbb{P}, \hat{\mathbb{P}}) \qquad \forall \mathbb{P} \in \mathcal{P}(\Xi) \\
& \gamma \ge 0,\ 0 \le \boldsymbol{x} \le \bar{\boldsymbol{x}}
\end{array}
$$

Constraint violation	$d_W(\mathbb{P}_{out}, \hat{\mathbb{P}})$								
	0	1	2	3	4	**4.8**	**4.9**	**5**	6
ARS ($\theta = 0$)	−4.18%	−3.27%	−2.35%	−1.44%	−0.53%	**0.14%**	**0.23%**	**0.31%**	1.13%
G-ARS ($\theta = 2$)	−5.71%	−4.79%	−3.87%	−2.94%	−2.02%	**-1.29%**	**-1.20%**	**-1.11%**	−0.22%
G-ARS ($\theta = 5$)	−7.56%	−6.63%	−5.69%	−4.76%	−3.82%	**-3.04%**	**-2.94%**	**-2.84%**	−1.85%
Constraint violation	$d_W(\mathbb{P}_{out}, \hat{\mathbb{P}})$								
	7	8	15	20	**24**	25	**26**	28	30
ARS ($\theta = 0$)	1.94%	**2.76%**	8.48%	12.43%	**15.53%**	16.26%	**16.99%**	18.46%	19.92%
G-ARS ($\theta = 2$)	0.68%	**1.57%**	7.80%	12.10%	**15.46%**	16.26%	**17.06%**	18.66%	20.26%
G-ARS ($\theta = 5$)	−0.86%	**0.14%**	6.98%	11.76%	**15.54%**	16.44%	**17.34%**	19.15%	20.96%

Table 2 Constraint violation under different values of the Wasserstein distance $d_W(\mathbb{P}_{out}, \hat{\mathbb{P}})$ between the out-of-sample distribution $\mathbb{P}_{out}$ and the reference distribution $\hat{\mathbb{P}}$.

For a more insightful and comprehensive comparison, we benchmark against the SAA approach (see also Long et al. 2022) by normalizing the targets and total costs of G-ARS and ARS models. In particular, we first solve an SAA model under the empirical distribution and obtain a `baseline objective`. Then for each `normalized target`, given by $\tau = \alpha \cdot$ `baseline objective` for some $\alpha \in [1, 1.24]$, we solve G-ARS models (with $\theta = 2$ and 5) and ARS model (*i.e.*, a G-ARS model with $\theta = 0$).

Globalized Robustness

We first compare the G-ARS model with the ARS model in terms of the performance on achieving globalized robustness. In particular, we consider the ARS model ($\theta = 0$) and two G-ARS models ($\theta = 2$ and $\theta = 5$) under the same target $\tau = 1.08 \cdot$ `baseline objective` for all three models, where the coefficient 1.08 is taken to ensure the feasibility of all three models. We solve the reformulations based on the piecewise affine decision rule (13). For each model, we fix its first-stage decision and evaluate, under an out-of-sample distribution $\mathbb{P}_{out} = \sum_{m \in [M]} p_m \delta_{\hat{\boldsymbol{\xi}}_m}$ supported on at most $M = 2000$ scenarios generated from $\mathbb{P}_0$, its performance on the constraint violation defined by

$$\sum_{m \in [M]} p_m \times \frac{\texttt{total cost under the m-th scenario - target}}{\texttt{target}} \times 100\%.$$

Here, we consider a series of the out-of-sample distributions by varying $d_W(\mathbb{P}_{out}, \hat{\mathbb{P}})$ from 0 to 30. Specifically, for each $\theta' \in \{0, 1, \ldots, 30\}$, we obtain $\mathbb{P}_{out}$ by solving a linear program that maximizes the expected second-stage cost of the ARS solution under $\mathbb{P}_{out}$ subject to $d_W(\mathbb{P}_{out}, \hat{\mathbb{P}}) = \theta'$. We summarize the results in Table 2.

On the one hand, G-ARS solutions can provide better protection on constraint violation when $\mathbb{P}_{out}$ is within or not too far from their equipped ambiguity set $\mathcal{F}_W(\theta)$. In particular, as $\mathbb{P}_{out}$ deviates from $\hat{\mathbb{P}}$ but still resides in the ambiguity set $\mathcal{F}_{\mathrm{W}}(2)$, solutions of all three models can ensure no constraint violation; and as $\mathbb{P}_{\mathrm{out}}$ move further away from $\hat{\mathbb{P}}$ yet within the ambiguity set $\mathcal{F}_{\mathrm{W}}(5)$, the ARS model leads to constraint violation when $d_{\mathrm{W}}(\mathbb{P}_{\mathrm{out}}, \hat{\mathbb{P}}) = 4.8, 4.9$ and 5, while the G-ARS model with $\theta = 5$ well guarantees no constraint violation by design. Moreover, the first constraint violations of the G-ARS model with $\theta = 2$ and G-ARS model with $\theta = 5$ occur at $d_{\mathrm{W}}(\mathbb{P}_{\mathrm{out}}, \hat{\mathbb{P}}) = 7$ and $d_{\mathrm{W}}(\mathbb{P}_{\mathrm{out}}, \hat{\mathbb{P}}) = 8$, respectively. It is very interesting to note the following. On the one hand, for $d_{\mathrm{W}}(\mathbb{P}_{\mathrm{out}}, \hat{\mathbb{P}}) \leq 23$, constraint violations of the three models satisfy

$$\text{“G-ARS model with } \theta = 5\text{”} < \text{“G-ARS model with } \theta = 2\text{”} < \text{“ARS model”}.$$

On the other hand, as $\mathbb{P}_{\mathrm{out}}$ keeps moving further away from the reference distribution $\hat{\mathbb{P}}$, G-ARS solutions begin to violate the constraint more than the ARS solution; see, for example, the trend as $d_{\mathrm{W}}(\mathbb{P}_{\mathrm{out}}, \hat{\mathbb{P}}) = 20$, 24, 25 and 26. When $d_{\mathrm{W}}(\mathbb{P}_{\mathrm{out}}, \hat{\mathbb{P}}) \geq 26$, constraint violations of the three models reverse to

$$\text{“G-ARS model with } \theta = 5\text{”} > \text{“G-ARS model with } \theta = 2\text{”} > \text{“ARS model”}.$$

That is to say, the G-ARS model sacrifices the performance in terms of controlling the constraint violation for distributions that are far from the ambiguity set $\mathcal{F}_{\mathrm{W}}(\theta)$ (and the central reference distribution $\hat{\mathbb{P}}$). Such a trade-off is also consistent with observations in Example 2.

Trade-Off

Furthermore, we compute the average total cost as before and evaluate the average deviation from target with $M = 2000$ testing samples generated from $\mathbb{P}_0$, given by

$$\frac{1}{M} \sum_{m \in [M]} (\texttt{total cost under the m-th sample} - \texttt{normalized target}).$$

We then normalize the average total cost and average deviation from target by those obtained from the SAA approach, respectively, and report the results in Figure 3. Some interesting findings are summarized as follows.

(i) For a relatively large value of normalized target, all G-ARS ($\theta = 2, 5$) and ARS ($\theta = 0$) solutions are able to control the total average cost to fall below the target, and G-ARS solutions can achieve a relatively lower cost than the ARS model. In particular, a larger value of θ can save more cost.

(ii) The G-ARS model could attain a wider range of targets than the ARS model: for example, G-ARS solutions with $\theta = 2$ and $\theta = 5$ are able to attain the target when $\alpha = 1.1$, while the ARS solution fails to do so. As a trade-off, when the target is too tight to attained (*e.g.*, $\alpha = 1.05$), the ARS model has a relatively smaller average total cost than the ARS model.

(iii) We note that all G-ARS and ARS solutions can outperform the SAA approach for a relatively tight target, *e.g.*, $\alpha \leq 1.13$. This is consistent with the observations in Sim et al. (2021).

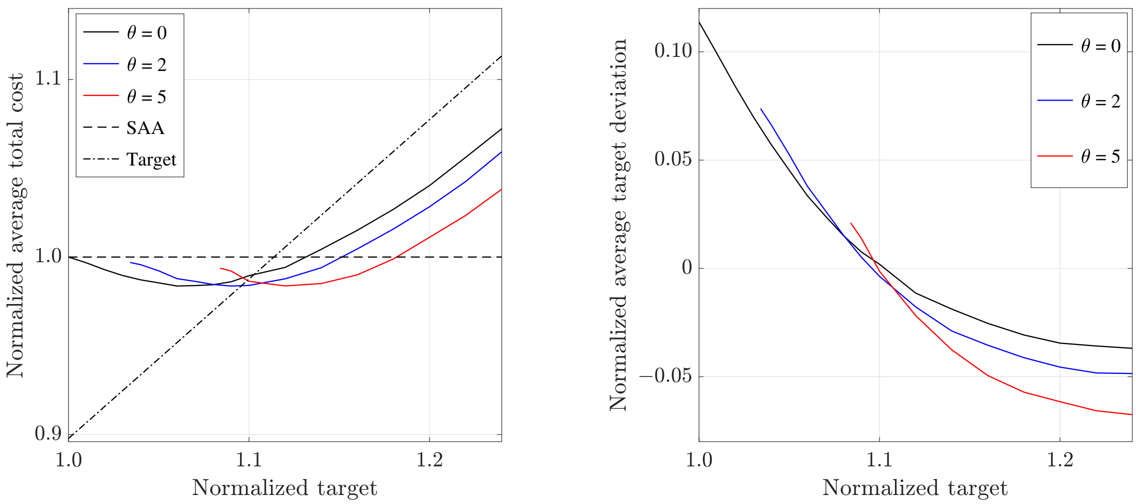

Figure 3 For $\theta = 2$ and $\theta = 5$, those values of normalized target that are less than the starting points of the blue and red lines that correspond to infeasibility.

6. Conclusion

Our proposed globalized distributionally robust counterpart specifies strictly no constraint violation for distributions residing in the ambiguity set, while ensuring soft constraint violation (in a controllable manner) for distributions on the support but not in the ambiguity set. By varying its inputs, the globalized distributionally robust counterpart recovers several existing perceptions of parameter uncertainty, including the traditional robust counterpart, distributionally robust counterpart, stochastic counterpart, globalized robust counterpart (Ben-Tal et al. 2006, Ben-Tal et al. 2017), as well as the more recent robust satisficing counterpart (Long et al. 2022). Besides, under proper specifications of the ambiguity set and cost function, it admits a tractable reformulation without increasing the computational complexity and further possesses attractive performance guarantees if the true distribution is light-tailed. The globalized distributionally robust counterpart extends readily to different types of ambiguity sets and can be easily integrated with many popular optimization models under uncertainty, enriching the pool of alternatives for informing optimal decisions with limited distributional information.

In this paper we chiefly emphasize the rich modeling expressiveness, attractive computational tractablity, and potential applications (as well as trade-offs) of the globalized distributionally robust counterpart. It remains open to calibrate and even to offer strong theoretical justifications on choosing its inputs (*e.g.*, the shadow price of globalized robustness γ) based on data availability. We leave these as future research.

References

Bayraksan, Güzin, David Love. 2015. Data-driven stochastic programming using phi-divergences. *The Operations Research Revolution*. INFORMS, 1–19.

Ben-Tal, Aharon, Stephen Boyd, Arkadi Nemirovski. 2006. Extending scope of robust optimization: comprehensive robust counterparts of uncertain problems. *Mathematical Programming* **107**(1) 63–89.

Ben-Tal, Aharon, Ruud Brekelmans, Dick den Hertog, Jean-Philippe Vial. 2017. Globalized robust optimization for nonlinear uncertain inequalities. *INFORMS Journal on Computing* **29**(2) 350–366.

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

Ben-Tal, Aharon, Arkadi Nemirovski. 1998. Robust convex optimization. *Mathematics of Operations Research* **23**(4) 769–805.

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

Bertsimas, Dimitris, Shimrit Shtern, Bradley Sturt. 2022. Two-stage sample robust optimization. *Operations Research* **70**(1) 624–640.

Bertsimas, Dimitris, Melvyn Sim. 2004. The price of robustness. *Operations Research* **52**(1) 35–53.

Bertsimas, Dimitris, Melvyn Sim, Meilin Zhang. 2019. Adaptive distributionally robust optimization. *Management Science* **65**(2) 604–618.

Birge, John, Francois Louveaux. 2011. *Introduction to stochastic programming*. Springer Science & Business Media.

Blanchet, Jose, Karthyek Murthy. 2019. Quantifying distributional model risk via optimal transport. *Mathematics of Operations Research* **44**(2) 565–600.

Boyd, Stephen, Stephen P Boyd, Lieven Vandenberghe. 2004. *Convex optimization*. Cambridge university press.

Chen, Zhi, Daniel Kuhn, Wolfram Wiesemann. 2022. Data-driven chance constrained programs over wasserstein balls. *Operations Research*.

Chen, Zhi, Melvyn Sim, Peng Xiong. 2020. Robust stochastic optimization made easy with RSOME. *Management Science* **66**(8) 3329–3339.

Chen, Zhi, Melvyn Sim, Huan Xu. 2019. Distributionally robust optimization with infinitely constrained ambiguity sets. *Operations Research* **67**(5) 1328–1344.

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

Dudley, Richard M. 2018. *Real analysis and probability*. CRC Press.

Dupačová, Jitka, Nicole Gröwe-Kuska, Werner Römisch. 2003. Scenario reduction in stochastic programming. *Mathematical programming* **95**(3) 493–511.

El Ghaoui, Laurent, Francois Oustry, Hervé Lebret. 1998. Robust solutions to uncertain semidefinite programs. *SIAM Journal on Optimization* **9**(1) 33–52.

Fournier, Nicolas, Arnaud Guillin. 2015. On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields* **162**(3) 707–738.

Gao, Rui, Anton Kleywegt. 2022. Distributionally robust stochastic optimization with Wasserstein distance. *Mathematics of Operations Research*.

Goh, Joel, Melvyn Sim. 2010. Distributionally robust optimization and its tractable approximations. *Operations Research* **58**(4-part-1) 902–917.

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

Hanasusanto, Grani, Daniel Kuhn. 2018. Conic programming reformulations of two-stage distributionally robust linear programs over Wasserstein balls. *Operations Research* **66**(3) 849–869.

Hanasusanto, Grani, Vladimir Roitch, Daniel Kuhn, Wolfram Wiesemann. 2015. A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Mathematical Programming* **151**(1) 35–62.

Hu, Zhaolin, Liu Jeff Hong. 2013. Kullback-Leibler divergence constrained distributionally robust optimization. *Available at* Optimization Online.

Jiang, Ruiwei, Yongpei Guan. 2018. Risk-averse two-stage stochastic program with distributional ambiguity. *Operations Research* **66**(5) 1390–1405.

Kuhn, Daniel, Peyman Mohajerin Esfahani, Viet Anh Nguyen, Soroosh Shafieezadeh-Abadeh. 2019. Wasserstein distributionally robust optimization: theory and applications in machine learning. *Operations Research & Management Science in the Age of Analytics*. INFORMS, 130–166.

Li, Bowen, Ruiwei Jiang, Johanna Mathieu. 2019. Ambiguous risk constraints with moment and unimodality information. *Mathematical Programming* **173**(1-2) 151–192.

Long, Daniel Zhuoyu, Melvyn Sim, Minglong Zhou. 2022. Robust satisficing. *Operations Research*.

Mohajerin Esfahani, Peyman, Daniel Kuhn. 2018. Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. *Mathematical Programming* **171**(1-2) 115–166.

Pflug, Georg, David Wozabal. 2007. Ambiguity in portfolio selection. *Quantitative Finance* **7**(4) 435–442.

Popescu, Ioana. 2005. A semidefinite programming approach to optimal-moment bounds for convex classes of distributions. *Mathematics of Operations Research* **30**(3) 632–657.

Rachev, Svetlozar T. 1991. *Probability metrics and the stability of stochastic models*, vol. 269. Wiley.

Ramachandra, Arjun, Napat Rujeerapaiboon, Melvyn Sim. 2021. Robust conic satisficing. *Available at* `https://arxiv.org/abs/2107.06714`.

Rockafellar, R Tyrrell. 1970. *Convex analysis*, vol. 18. Princeton university press.

Santambrogio, Filippo. 2015. *Optimal transport for applied mathematicians*. Springer.

Shapiro, Alexander. 2001. On duality theory of conic linear problems. *Semi-infinite programming*. Springer, 135–165.

Shapiro, Alexander. 2017. Distributionally robust stochastic programming. *SIAM Journal on Optimization* **27**(4) 2258–2275.

Sim, Melvyn, Qinshen Tang, Minglong Zhou, Taozeng Zhu. 2021. The analytics of robust satisficing. *Available at* SSRN 3829562.

Soyster, Allen. 1973. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research* **21**(5) 1154–1157.

Tierney, Luke. 1996. Introduction to general state-space markov chain theory. *Markov chain Monte Carlo in practice* 59–74.

Villani, Cédric. 2009. *Optimal transport: old and new*, vol. 338. Springer.

Wang, Zizhuo, Peter W Glynn, Yinyu Ye. 2016. Likelihood robust optimization for data-driven problems. *Computational Management Science* **13**(2) 241–261.

Wiesemann, Wolfram, Daniel Kuhn, Melvyn Sim. 2014. Distributionally robust convex optimization. *Operations Research* **62**(6) 1358–1376.

Xie, Weijun. 2019. On distributionally robust chance constrained programs with Wasserstein distance. *Mathematical Programming* 1–41.

Zhang, Luhao, Jincheng Yang, Rui Gao. 2022. A simple duality proof for Wasserstein distributionally robust optimization. *Available at* `https://arxiv.org/abs/2205.00362`.

Zhao, Chaoyue, Yongpei Guan. 2015. Data-driven risk-averse two-stage stochastic program with ζ-structure probability metrics. *Available on* Optimization Online.

Zolotarev, Vladimir Mikhailovich. 1976. Metric distances in spaces of random variables and their distributions. *Mathematics of the USSR-Sbornik* **30**(3) 373.

E-Companion to "Globalized Distributionally Robust Counterpart"

EC.1. Proofs

Proof of Lemma 1. Before proving strong duality, we first establish weak duality that states

$$\sup_{\mathbb{P}\in\mathcal{P}(\Xi),\,\mathbb{Q}\in\mathcal{F}_{\mathrm{W}}(\theta)} \Big\{\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})]-\gamma\cdot d_{\mathrm{W}}(\mathbb{P},\mathbb{Q})\Big\} \le \inf_{t\in[0,\gamma]}\Big\{\theta t+\frac{1}{N}\sum_{n\in[N]}\sup_{\boldsymbol{\xi}\in\Xi}\big\{f(\boldsymbol{x},\boldsymbol{\xi})-t\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}_n\|\big\}\Big\}. \quad \text{(EC.1)}$$

Exploring the definition of type-1 Wasserstein distance, we rewrite the left-hand side problem as

$$\begin{aligned}
\sup\quad & \int_{\Xi} f(\boldsymbol{x},\boldsymbol{\xi})\,\mathrm{d}\mathbb{P}(\boldsymbol{\xi})-\gamma\int_{\Xi\times\Xi}\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|\,\mathrm{d}\pi(\boldsymbol{\xi},\boldsymbol{\zeta})\\
\text{s.t.}\quad & \int_{\Xi\times\Xi}\|\boldsymbol{\zeta}-\boldsymbol{\eta}\|\,\mathrm{d}\pi'(\boldsymbol{\zeta},\boldsymbol{\eta})\le\theta\\
& \tilde{\boldsymbol{\xi}}\sim\mathbb{P},\ \tilde{\boldsymbol{\zeta}}\sim\mathbb{Q},\ \tilde{\boldsymbol{\eta}}\sim\hat{\mathbb{P}},\ (\tilde{\boldsymbol{\xi}},\tilde{\boldsymbol{\zeta}})\sim\pi,\ (\tilde{\boldsymbol{\zeta}},\tilde{\boldsymbol{\eta}})\sim\pi'\\
& \mathbb{P},\mathbb{Q}\in\mathcal{P}(\Xi),\ \pi\in\mathcal{Q}(\mathbb{P},\mathbb{Q}),\ \pi'\in\mathcal{Q}(\mathbb{Q},\hat{\mathbb{P}}).
\end{aligned}$$

The law of total probability asserts that the joint distribution π' of $(\tilde{\boldsymbol{\zeta}},\tilde{\boldsymbol{\eta}})$ can be constructed from the marginal distribution $\hat{\mathbb{P}}$ of $\tilde{\boldsymbol{\eta}}$, and conditional distributions $\mathbb{Q}_1,\ldots,\mathbb{Q}_N$ of $\tilde{\boldsymbol{\zeta}}$ given the realizations $\hat{\boldsymbol{\xi}}_1,\ldots,\hat{\boldsymbol{\xi}}_N$ of $\tilde{\boldsymbol{\eta}}$. That is to say, we can rewrite the above problem into

$$\begin{aligned}
\sup\ \ & \frac{1}{N}\sum_{n\in[N]}\int_{\Xi} f(\boldsymbol{x},\boldsymbol{\xi})\,\mathrm{d}\mathbb{P}_n(\boldsymbol{\xi})-\frac{\gamma}{N}\sum_{n\in[N]}\int_{\Xi\times\Xi}\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|\,\mathrm{d}\pi_n(\boldsymbol{\xi},\boldsymbol{\zeta})\\
\text{s.t.}\ \ & \frac{1}{N}\sum_{n\in[N]}\int_{\Xi}\|\boldsymbol{\zeta}-\hat{\boldsymbol{\xi}}_n\|\,\mathrm{d}\mathbb{Q}_n(\boldsymbol{\zeta})\le\theta\\
& \tilde{\boldsymbol{\xi}}\sim\frac{1}{N}\sum_{n\in[N]}\mathbb{P}_n,\ \tilde{\boldsymbol{\zeta}}\sim\frac{1}{N}\sum_{n\in[N]}\mathbb{Q}_n,\ (\tilde{\boldsymbol{\xi}},\tilde{\boldsymbol{\zeta}})\sim\frac{1}{N}\sum_{n\in[N]}\pi_n\\
& \mathbb{P}_n,\ \mathbb{Q}_n\in\mathcal{P}(\Xi),\ \pi_n\in\mathcal{Q}(\mathbb{P}_n,\mathbb{Q}_n) \qquad\qquad \forall n\in[N],
\end{aligned} \qquad \text{(EC.2)}$$

whose objective value is not larger than that of its Lagrangian dual, given by

$$\begin{aligned}
&\inf_{t\ge0}\Big\{\theta t+\frac{1}{N}\sum_{n\in[N]}\sup_{\pi_n\in\mathcal{P}(\Xi\times\Xi)}\int_{\Xi\times\Xi}\big(f(\boldsymbol{x},\boldsymbol{\xi})-\gamma\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|-t\|\boldsymbol{\zeta}-\hat{\boldsymbol{\xi}}_n\|\big)\mathrm{d}\pi_n(\boldsymbol{\xi},\boldsymbol{\zeta})\Big\}\\
&=\inf_{t\ge0}\Big\{\theta t+\frac{1}{N}\sum_{n\in[N]}\sup_{\boldsymbol{\xi},\boldsymbol{\zeta}\in\Xi}\big\{f(\boldsymbol{x},\boldsymbol{\xi})-\gamma\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|-t\|\boldsymbol{\zeta}-\hat{\boldsymbol{\xi}}_n\|\big\}\Big\}\\
&=\inf_{t\ge0}\Big\{\theta t+\frac{1}{N}\sum_{n\in[N]}\sup_{\boldsymbol{\xi}\in\Xi}\big\{f(\boldsymbol{x},\boldsymbol{\xi})-\min\{\gamma,t\}\cdot\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}_n\|\big\}\Big\}.
\end{aligned}$$

Here, the last identity follows from the fact that for any values of γ, t, and $\boldsymbol{\xi}$, it holds that

$$\inf_{\boldsymbol{\zeta}\in\Xi}\big\{\gamma\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|+t\|\boldsymbol{\zeta}-\hat{\boldsymbol{\xi}}_n\|\big\}=\begin{cases}\gamma\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}_n\| & \text{if } t>\gamma\\ t\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}_n\| & \text{if } t\le\gamma.\end{cases}$$

Eliminating the inner minimization $\min\{\gamma,t\}$ then yields the right-hand side problem in (EC.1).

To conclude strong duality, we next prove that the inequality in (EC.1) is indeed tight. When $\theta = 0$, the left-hand side of inequality (EC.1) can be reformulated as

$$\begin{aligned}\sup_{\mathbb{P}\in\mathcal{P}(\Xi)}\left\{\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})]-\gamma\cdot d_{\mathrm{W}}(\mathbb{P},\hat{\mathbb{P}})\right\} &= \sup_{\mathbb{P}\in\mathcal{P}(\Xi),\,\pi\in\mathcal{Q}(\mathbb{P},\hat{\mathbb{P}})}\left\{\int_{\Xi} f(\boldsymbol{x},\boldsymbol{\xi})\,\mathrm{d}\mathbb{P}(\boldsymbol{\xi})-\gamma\int_{\Xi\times\Xi}\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|\,\mathrm{d}\pi(\boldsymbol{\xi},\boldsymbol{\zeta})\right\}\\ &= \frac{1}{N}\sum_{n\in[N]}\sup_{\mathbb{P}_n\in\mathcal{P}(\Xi)}\left\{\int_{\Xi}(f(\boldsymbol{x},\boldsymbol{\xi})-\gamma\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}_n\|)\,\mathrm{d}\mathbb{P}_n(\boldsymbol{\xi})\right\}\\ &= \frac{1}{N}\sum_{n\in[N]}\sup_{\boldsymbol{\xi}\in\Xi}\left\{f(\boldsymbol{x},\boldsymbol{\xi})-\gamma\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}_n\|\right\},\end{aligned}$$

coinciding with the right-hand side. It is then sufficient to focus on $\theta > 0$.

Let $\mathcal{M}(\Xi)$ be the linear space of finite signed measures on Ξ and $\mathcal{M}_+(\Xi) = \{\mu \in \mathcal{M}(\Xi) \mid \mu \succeq 0\}$—a convex cone in $\mathcal{M}(\Xi)$—be the set of nonnegative measures. Then we can rewrite problem (EC.2) as a conic program for moment problem:

$$\begin{aligned}\sup\quad & \frac{1}{N}\sum_{n\in[N]}\int_{\Xi\times\Xi}(f(\boldsymbol{x},\boldsymbol{\xi})-\gamma\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|)\,\mathrm{d}\pi_n(\boldsymbol{\xi},\boldsymbol{\zeta})\\ \text{s.t.}\quad & \frac{1}{N}\sum_{n\in[N]}\int_{\Xi\times\Xi}\|\boldsymbol{\zeta}-\hat{\boldsymbol{\xi}}_n\|\,\mathrm{d}\pi_n(\boldsymbol{\xi},\boldsymbol{\zeta})\le\theta\\ & \int_{\Xi\times\Xi}\mathrm{d}\pi_n(\boldsymbol{\xi},\boldsymbol{\zeta})=1 && \forall n\in[N]\\ & \pi_n\in\mathcal{M}_+(\Xi)\times\mathcal{M}_+(\Xi) && \forall n\in[N].\end{aligned}\tag{EC.3}$$

By proposition 3.4 in Shapiro (2001), to show strong duality between problem (EC.3) and its dual (*i.e.*, the right-hand side problem in (EC.1)), it is sufficient to show that the vector $(\theta, 1, \dots, 1)$ resides in the interior of the convex cone

$$\left\{(a,\boldsymbol{b})\in\mathbb{R}\times\mathbb{R}^N\;\middle|\;\begin{array}{l}\exists\mu_n\in\mathcal{M}_+(\Xi)\times\mathcal{M}_+(\Xi),\forall n\in[N]:\\ \displaystyle\frac{1}{N}\sum_{n\in[N]}\int_{\Xi\times\Xi}\|\boldsymbol{\zeta}-\hat{\boldsymbol{\xi}}_n\|\,\mathrm{d}\mu_n(\boldsymbol{\xi},\boldsymbol{\zeta})\le a\\ \displaystyle\int_{\Xi\times\Xi}\mathrm{d}\mu_n(\boldsymbol{\xi},\boldsymbol{\zeta})=b_n\quad\forall n\in[N]\end{array}\right\}.$$

To this end, choose any point $(p, \boldsymbol{q}) \in \mathbb{B}_\varepsilon(\theta)\times\mathbb{B}_\varepsilon(1)\times\cdots\times\mathbb{B}_\varepsilon(1)$ for a sufficiently small $\varepsilon > 0$, where $\mathbb{B}_\varepsilon(\theta)$ and $\mathbb{B}_\varepsilon(1)$ are the spherical ε-neighborhoods centered at θ and 1, respectively. Then for each $n \in [N]$, we can always find a nonnegative measure μ_n of $(\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\zeta}})$, which is supported on $\Xi\times\Xi$ with independent marginals $\tilde{\boldsymbol{\xi}} \sim q_n\delta_{\boldsymbol{\xi}_n}$ for some $\boldsymbol{\xi}_n \in \Xi$ and $\tilde{\boldsymbol{\zeta}} \sim \delta_{\hat{\boldsymbol{\xi}}_n}$, and such μ_n satisfies

$$\left\{\begin{array}{l}\displaystyle\frac{1}{N}\sum_{n\in[N]}\int_{\Xi\times\Xi}\|\boldsymbol{\zeta}-\hat{\boldsymbol{\xi}}_n\|\,\mathrm{d}\mu_n(\boldsymbol{\xi},\boldsymbol{\zeta})=0\le p\in\mathbb{B}_\varepsilon(\theta)\\ \displaystyle\int_{\Xi\times\Xi}\mathrm{d}\mu_n(\boldsymbol{\xi},\boldsymbol{\zeta})=q_n\in\mathbb{B}_\varepsilon(1)\quad\forall n\in[N],\end{array}\right.$$

establishing strong duality for problem (EC.3). $\square$

Proof of Proposition 2. We first construct a lower bound of the left-hand side problem in (7) using Lemma 1 and the min-max inequality:

$$\begin{aligned}
&\sup_{\mathbb{P}\in\mathcal{P}(\Xi),\,\mathbb{Q}\in\mathcal{F}_{\mathrm{W}}(\theta)} \left\{\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})]-\gamma\cdot d_{\mathrm{W}}(\mathbb{P},\mathbb{Q})\right\}\\
&=\inf_{0\le t\le\gamma}\left\{\theta t+\frac{1}{N}\sum_{n\in[N]}\sup_{\boldsymbol{\xi}\in\Xi}\left\{f(\boldsymbol{x},\boldsymbol{\xi})-t\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}_n\|\right\}\right\}\\
&=\inf_{0\le t\le\gamma}\left\{\theta t+\frac{1}{N}\sum_{n\in[N]}\sup_{\mathbb{P}_n\in\mathcal{P}(\Xi)}\int_{\Xi}\left(f(\boldsymbol{x},\boldsymbol{\xi})-t\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}_n\|\right)\mathrm{d}\mathbb{P}_n(\boldsymbol{\xi})\right\}\\
&=\inf_{0\le t\le\gamma}\sup_{\mathbb{P}\in\mathcal{P}(\Xi)}\left\{\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})]-td_{\mathrm{W}}(\mathbb{P},\hat{\mathbb{P}})+\theta t\right\}\\
&\ge\sup_{\mathbb{P}\in\mathcal{P}(\Xi)}\inf_{0\le t\le\gamma}\left\{\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})]-td_{\mathrm{W}}(\mathbb{P},\hat{\mathbb{P}})+\theta t\right\}\\
&=\sup_{\mathbb{P}\in\mathcal{P}(\Xi)}\left\{\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})]-\sup_{0\le t\le\gamma}\left\{t(d_{\mathrm{W}}(\mathbb{P},\hat{\mathbb{P}})-\theta)\right\}\right\}\\
&=\sup_{\mathbb{P}\in\mathcal{P}(\Xi)}\left\{\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})]-\gamma\cdot(d_{\mathrm{W}}(\mathbb{P},\hat{\mathbb{P}})-\theta)^+\right\}.
\end{aligned}$$

Here, the third line follows from the identity (see also the proof of Lemma 1),

$$\sup_{\mathbb{P}\in\mathcal{P}(\Xi)}\left\{\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})]-td_{\mathrm{W}}(\mathbb{P},\hat{\mathbb{P}})\right\}=\frac{1}{N}\sum_{n\in[N]}\sup_{\mathbb{P}_n\in\mathcal{P}(\Xi)}\int_{\Xi}\left(f(\boldsymbol{x},\boldsymbol{\xi})-t\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}_n\|\right)\mathrm{d}\mathbb{P}_n(\boldsymbol{\xi}).$$

Hence, for any $\boldsymbol{x}$ feasible to (7), it holds that $\mathbb{E}_{\mathbb{P}_0}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})]\le\gamma\cdot(d_{\mathrm{W}}(\mathbb{P}_0,\hat{\mathbb{P}})-\theta)^+$. Thus for any $\eta\ge\theta$, we have $\mathbb{P}_0^N[\mathbb{E}_{\mathbb{P}_0}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})]>\gamma(\eta-\theta)]\le\mathbb{P}_0^N[d_{\mathrm{W}}(\mathbb{P}_0,\hat{\mathbb{P}})\ge\eta]$. The remaining proof then follows from applying theorem 3.4 of Mohajerin Esfahani and Kuhn (2018). □

Proof of Proposition 3. By Lemma 1, the G-DRCC holds if and only if there is $t\in[0,\gamma]$ satisfying

$$\theta Nt+\sum_{n\in[N]}\sup_{\boldsymbol{\xi}\in\Xi}\left\{\mathbb{I}\{\boldsymbol{\xi}\in\bar{\mathcal{S}}(\boldsymbol{x})\}-\varepsilon-t\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}_n\|\right\}\le 0\iff\theta Nt+\sum_{n\in[N]}\sup_{\boldsymbol{\xi}\in\Xi}\left\{\mathbb{I}\{\boldsymbol{\xi}\in\bar{\mathcal{S}}(\boldsymbol{x})-t\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}_n\|\right\}\le\varepsilon.$$

For each $n\in[N]$, if $\hat{\boldsymbol{\xi}}_n\in\bar{\mathcal{S}}(\boldsymbol{x})$, then by letting $\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}_n$ we have

$$\sup_{\boldsymbol{\xi}\in\Xi}\left\{\mathbb{I}\{\boldsymbol{\xi}\in\bar{\mathcal{S}}(\boldsymbol{x})\}-t\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}_n\|\right\}=1=\left(1-\inf_{\boldsymbol{\xi}\in\Xi:\,\boldsymbol{\xi}\in\bar{\mathcal{S}}(\boldsymbol{x})}t\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}_n\|\right)^+=\left(1-t\cdot\mathbf{dist}(\hat{\boldsymbol{\xi}}_n,\bar{\mathcal{S}}(\boldsymbol{x}))\right)^+;$$

whereas if $\hat{\boldsymbol{\xi}}_n\in\mathcal{S}(\boldsymbol{x})=\Xi\setminus\bar{\mathcal{S}}(\boldsymbol{x})$, we have

$$\begin{aligned}
\sup_{\boldsymbol{\xi}\in\Xi}\left\{\mathbb{I}\{\boldsymbol{\xi}\in\bar{\mathcal{S}}(\boldsymbol{x})\}-t\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}_n\|\right\}&=\max\left\{1-\inf_{\boldsymbol{\xi}\in\Xi:\,\boldsymbol{\xi}\in\bar{\mathcal{S}}(\boldsymbol{x})}t\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}_n\|,\,-\inf_{\boldsymbol{\xi}\in\Xi:\,\boldsymbol{\xi}\in\mathcal{S}(\boldsymbol{x})}t\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}_n\|\right\}\\
&=\max\left\{1-\inf_{\boldsymbol{\xi}\in\Xi:\,\boldsymbol{\xi}\in\bar{\mathcal{S}}(\boldsymbol{x})}t\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}_n\|,\,0\right\}\\
&=\left(1-\inf_{\boldsymbol{\xi}\in\Xi:\,\boldsymbol{\xi}\in\bar{\mathcal{S}}(\boldsymbol{x})}t\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}_n\|\right)^+\\
&=\left(1-t\cdot\mathbf{dist}(\hat{\boldsymbol{\xi}}_n,\bar{\mathcal{S}}(\boldsymbol{x}))\right)^+.
\end{aligned}$$

That is to say, in both scenarios we have

$$\sup_{\boldsymbol{\xi}\in\Xi}\left\{\mathbb{I}\{\boldsymbol{\xi}\in\bar{\mathcal{S}}(\boldsymbol{x})\}-t\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}_n\|\right\}=\Big(1-t\cdot\mathbf{dist}(\hat{\boldsymbol{\xi}}_n,\bar{\mathcal{S}}(\boldsymbol{x}))\Big)^+,$$

implying that the globalized distributionally robust chance constraint holds if and only if

$$\exists\, t\in[0,\gamma]:\ \theta N t+\sum_{n\in[N]}\Big(1-t\cdot\mathbf{dist}(\hat{\boldsymbol{\xi}}_n,\bar{\mathcal{S}}(\boldsymbol{x}))\Big)^+\le\varepsilon N.$$

In fact, it holds that $t>0$ in the above inequality, because otherwise the value of the left-hand side would amount to N and contradict to $\varepsilon N\in(0,N)$. Hence, we can multiply both sides by $1/t$ and let $t\leftarrow 1/t$, then we obtain a constraint system

$$\begin{cases}\theta N-\varepsilon N t+\displaystyle\sum_{n\in[N]}\Big(t-\mathbf{dist}(\hat{\boldsymbol{\xi}}_n,\bar{\mathcal{S}}(\boldsymbol{x}))\Big)^+\le 0\\ \gamma\cdot t\ge 1.\end{cases}$$

Introducing the decision variable $\boldsymbol{s}\ge\mathbf{0}$ and rearranging terms then concludes the proof. $\square$

Proof of Theorem 1. We start with the properties of $\sup\limits_{\mathbb{P}\in\mathcal{P}(\Xi),\,\mathbb{Q}\in\mathcal{F}_{\mathcal{L}_2}(\theta)}\left\{\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})]-\gamma\cdot d_{\mathcal{L}_1}(\mathbb{P},\mathbb{Q})\right\}$ with respect to the radius θ.

LEMMA EC.1. *Given any decision $\boldsymbol{x}\in\mathcal{X}$ and mappings $d_{\mathcal{L}_1}$ and $d_{\mathcal{L}_2}$ defined in* (14)*, the function*

$$L(\theta)=\sup_{\mathbb{P}\in\mathcal{P}(\Xi),\,\mathbb{Q}\in\mathcal{F}_{\mathcal{L}_2}(\theta)}\left\{\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})]-\gamma\cdot d_{\mathcal{L}_1}(\mathbb{P},\mathbb{Q})\right\}$$

is bounded from below by $\mathbb{E}_{\hat{\mathbb{P}}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})]$ and is concave in θ on $\mathbb{R}_+$.

Proof of Lemma EC.1. For any $\theta\in[0,+\infty)$, we have

$$L(\theta)\ge L(0)=\sup_{\mathbb{P}\in\mathcal{P}(\Xi)}\left\{\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})]-\gamma\cdot d_{\mathcal{L}_1}(\mathbb{P},\hat{\mathbb{P}})\right\}\ge\mathbb{E}_{\hat{\mathbb{P}}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})],$$

where the first inequality is due to the fact that $L(\theta)$ increases in θ and the second inequality follows from setting $\mathbb{P}=\hat{\mathbb{P}}$. Since for any $\boldsymbol{x}$ the cost function $f(\boldsymbol{x},\boldsymbol{\xi})$ is proper with respect to $\boldsymbol{\xi}$, we have $\mathbb{E}_{\hat{\mathbb{P}}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})]>-\infty$, thus $L(\theta)$ is bounded from below. It remains to verify the concavity. To this end, we first show that the function $H(\mathbb{P},\theta)=\sup_{\mathbb{Q}\in\mathcal{F}_{\mathcal{L}_2}(\theta)}\{-\gamma\cdot d_{\mathcal{L}_1}(\mathbb{P},\mathbb{Q})\}$ is jointly concave in $(\mathbb{P},\theta)$. We fix any $(\mathbb{P}_1,\theta_1)\in\mathcal{P}(\Xi)\times\mathbb{R}_+$ and $(\mathbb{P}_2,\theta_2)\in\mathcal{P}(\Xi)\times\mathbb{R}_+$. For any $\lambda\in[0,1]$ and $\mathbb{Q}_1,\mathbb{Q}_2$ satisfying $d_{\mathcal{L}_2}(\mathbb{Q}_1,\hat{\mathbb{P}})\le\theta_1$ and $d_{\mathcal{L}_2}(\mathbb{Q}_2,\hat{\mathbb{P}})\le\theta_2$, we define $\mathbb{P}_3=\lambda\mathbb{P}_1+(1-\lambda)\mathbb{P}_2\in\mathcal{P}(\Xi)$, $\theta_3=\lambda\theta_1+(1-\lambda)\theta_2$ and $\mathbb{Q}_3=\lambda\mathbb{Q}_1+(1-\lambda)\mathbb{Q}_2$. Following from the convexity of $d_{\mathcal{L}_2}$, it holds that

$$d_{\mathcal{L}_2}(\mathbb{Q}_3,\hat{\mathbb{P}})\le\lambda d_{\mathcal{L}_2}(\mathbb{Q}_1,\hat{\mathbb{P}})+(1-\lambda)d_{\mathcal{L}_2}(\mathbb{Q}_2,\hat{\mathbb{P}})\le\lambda\theta_1+(1-\lambda)\theta_2=\theta_3.$$

This implies that *(i)* $\mathbb{Q}_3 \in \mathcal{F}_{\mathcal{L}_2}(\theta_3)$ and *(ii)* for any $\mathbb{Q}_1 \in \mathcal{F}_{\mathcal{L}_2}(\theta_1)$ and $\mathbb{Q}_2 \in \mathcal{F}_{\mathcal{L}_2}(\theta_2)$, we have

$$H(\mathbb{P}_3, \theta_3) \geq -\gamma \cdot d_{\mathcal{L}_1}(\mathbb{P}_3, \mathbb{Q}_3) \geq -\gamma\lambda d_{\mathcal{L}_1}(\mathbb{P}_1, \mathbb{Q}_1) - \gamma(1-\lambda) d_{\mathcal{L}_1}(\mathbb{P}_2, \mathbb{Q}_2),$$

where the second inequality follows from the fact that $d_{\mathcal{L}_1}(\mathbb{P}, \mathbb{Q})$ is jointly convex in $\mathbb{P}$ and $\mathbb{Q}$. Taking the supremum over $\mathbb{Q}_1$ and $\mathbb{Q}_2$, we arrive at

$$\begin{aligned} H(\mathbb{P}_3, \theta_3) &\geq \sup_{\mathbb{Q}_1 \in \mathcal{F}_{\mathcal{L}_2}(\theta_1)} \big\{ -\gamma\lambda d_{\mathcal{L}_1}(\mathbb{P}_1, \mathbb{Q}_1) \big\} + \sup_{\mathbb{Q}_2 \in \mathcal{F}_{\mathcal{L}_2}(\theta_2)} \big\{ -\gamma(1-\lambda) d_{\mathcal{L}_1}(\mathbb{P}_2, \mathbb{Q}_2) \big\} \\ &= \lambda H(\mathbb{P}_1, \theta_1) + (1-\lambda) H(\mathbb{P}_2, \theta_2), \end{aligned}$$

proving the joint concavity of $H(\mathbb{P}, \theta)$ in $(\mathbb{P}, \theta)$. Note that $\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] + H(\theta)$ is also jointly concave in $(\mathbb{P}, \theta)$, then $L(\theta)$ is concave in θ since concavity is preserved under maximization over the convex set $\mathcal{P}(\Xi)$ (Boyd et al. 2004). The proof then completes. □

In the remainder of the proof, we adapt the idea of Zhang et al. (2022) to globalized distributionally robust counterpart with probability metrics. For the general case

$$\begin{aligned} &\sup_{\mathbb{P} \in \mathcal{P}(\Xi),\, \mathbb{Q} \in \mathcal{F}_{\mathcal{L}_2}(\theta)} \Big\{ \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] - \gamma \cdot d_{\mathcal{L}_1}(\mathbb{P}, \mathbb{Q}) \Big\} \\ &= \inf_{t \geq 0} \sup_{\mathbb{P} \in \mathcal{P}(\Xi),\, \mathbb{Q} \in \mathcal{P}(\Xi)} \Big\{ \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] - \gamma \cdot d_{\mathcal{L}_1}(\mathbb{P}, \mathbb{Q}) - t(d_{\mathcal{L}_2}(\mathbb{Q}, \hat{\mathbb{P}}) - \theta) \Big\}, \end{aligned}$$

we first focus on the concave function $L(\theta) = \sup_{\mathbb{P} \in \mathcal{P}(\Xi),\, \mathbb{Q} \in \mathcal{F}_{\mathcal{L}_2}(\theta)} \{\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] - \gamma d_{\mathcal{L}_1}(\mathbb{P}, \mathbb{Q})\}$. Taking Legendre transform (Rockafellar 1970) on $L(\theta)$, for any $t > 0$ we have

$$\begin{aligned} L^*(t) = \sup_{\theta > 0} \{L(\theta) - t\theta\} &= \sup_{\theta > 0} \sup_{\mathbb{P} \in \mathcal{P}(\Xi),\, \mathbb{Q} \in \mathcal{P}(\Xi)} \big\{ \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] - \gamma d_{\mathcal{L}_1}(\mathbb{P}, \mathbb{Q}) - t\theta : d_{\mathcal{L}_2}(\mathbb{Q}, \hat{\mathbb{P}}) \leq \theta \big\} \\ &= \sup_{\mathbb{P} \in \mathcal{P}(\Xi),\, \mathbb{Q} \in \mathcal{P}(\Xi)} \sup_{\theta > 0} \big\{ \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] - \gamma d_{\mathcal{L}_1}(\mathbb{P}, \mathbb{Q}) - t\theta : d_{\mathcal{L}_2}(\mathbb{Q}, \hat{\mathbb{P}}) \leq \theta \big\} \\ &= \sup_{\mathbb{P} \in \mathcal{P}(\Xi),\, \mathbb{Q} \in \mathcal{P}(\Xi)} \big\{ \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] - \gamma d_{\mathcal{L}_1}(\mathbb{P}, \mathbb{Q}) - t d_{\mathcal{L}_2}(\mathbb{Q}, \hat{\mathbb{P}}) \big\}. \end{aligned}$$

Since $L(\theta)$ is bounded from below, monotonically increasing and concave on $\mathbb{R}_+$, it is upper semi-continuous. Note that either $L(\theta) < +\infty$ for any $\theta > 0$ or $L(\theta) = +\infty$ for any $\theta > 0$. In the former case, applying Legendre transform again on $L^*(\cdot)$ yields that for any $\theta > 0$,

$$L(\theta) = \inf_{t > 0} \big\{ t\theta + L^*(t) \big\} = \inf_{t > 0} \big\{ t\theta + \sup_{\mathbb{P} \in \mathcal{P}(\Xi),\, \mathbb{Q} \in \mathcal{P}(\Xi)} \big\{ \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] - \gamma d_{\mathcal{L}_1}(\mathbb{P}, \mathbb{Q}) - t d_{\mathcal{L}_2}(\mathbb{Q}, \hat{\mathbb{P}}) \big\} \big\}.$$

In the latter case, $L^*(t) = +\infty$ for any $t > 0$, and the above identity also holds. To obtain the dual reformulation in (16), it remains to show that the function

$$g(t) = t\theta + \sup_{\mathbb{P} \in \mathcal{P}(\Xi),\, \mathbb{Q} \in \mathcal{P}(\Xi)} \big\{ \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] - \gamma d_{\mathcal{L}_1}(\mathbb{P}, \mathbb{Q}) - t d_{\mathcal{L}_2}(\mathbb{Q}, \hat{\mathbb{P}}) \big\}$$

is right continuous at $t = 0$. To see this, it suffices to show $g(0) = \lim_{t \to 0^+} g(t)$. Without loss of optimality, for any $t > 0$ we only need to focus on $\mathbb{Q}$ that satisfies $d_{\mathcal{L}_2}(Q, \hat{\mathbb{P}}) < +\infty$, because otherwise we have $g(t) = -\infty < \mathbb{E}_{\hat{\mathbb{P}}}[f(\boldsymbol{x}, \hat{\boldsymbol{\xi}})]$, which never achieves the optimum. Then for any $\mathbb{Q}$ such that $d_{\mathcal{L}_2}(Q, \hat{\mathbb{P}}) < +\infty$, we have

$$\lim_{t\to 0^+} g(t) = t\theta + \sup_{\mathbb{P}\in\mathcal{P}(\Xi),\, \mathbb{Q}\in\mathcal{P}(\Xi)} \left\{ \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] - \gamma d_{\mathcal{L}_1}(\mathbb{P}, \mathbb{Q}) \right\} = g(0),$$

completing the proof of the first part.

When $\mathcal{L}_1 = \mathcal{L}_2 = \mathcal{L}$, we proceed the proof by showing that

$$\inf_{\mathbb{Q}\in\mathcal{P}(\Xi)} \{\gamma d_{\mathcal{L}}(\mathbb{P}, \mathbb{Q}) + t d_{\mathcal{L}}(\mathbb{Q}, \hat{\mathbb{P}})\} = \min\{t, \gamma\} \cdot d_{\mathcal{L}}(\mathbb{P}, \hat{\mathbb{P}}). \tag{EC.4}$$

If $t \le \gamma$, then by definition of $d_{\mathcal{L}}$ and subadditivity of the supremum operator, for any $\mathbb{Q} \in \mathcal{P}(\Xi)$,

$$\begin{aligned}
\gamma d_{\mathcal{L}}(\mathbb{P}, \mathbb{Q}) + t d_{\mathcal{L}}(\mathbb{Q}, \hat{\mathbb{P}}) &= t(d_{\mathcal{L}}(\mathbb{P}, \mathbb{Q}) + d_{\mathcal{L}}(\mathbb{Q}, \hat{\mathbb{P}})) + (\gamma - t) d_{\mathcal{L}}(\mathbb{P}, \mathbb{Q}) \\
&= t\big(\sup_{\ell\in\mathcal{L}} \{\mathbb{E}_{\mathbb{P}}[\ell] - \mathbb{E}_{\mathbb{Q}}[\ell]\} + \sup_{\ell\in\mathcal{L}} \{\mathbb{E}_{\mathbb{Q}}[\ell] - \mathbb{E}_{\hat{\mathbb{P}}}[\ell]\} \big) + (\gamma - t) d_{\mathcal{L}}(\mathbb{P}, \mathbb{Q}) \\
&\ge t \sup_{\ell\in\mathcal{L}} \{\mathbb{E}_{\mathbb{P}}[\ell] - \mathbb{E}_{\hat{\mathbb{P}}}[\ell]\} = t d_{\mathcal{L}}(\mathbb{P}, \hat{\mathbb{P}}),
\end{aligned}$$

where the equality is attained at $\mathbb{Q} = \mathbb{P}$. Similarly, if $t > \gamma$, then for any $\mathbb{Q} \in \mathcal{P}(\Xi)$,

$$\begin{aligned}
\gamma d_{\mathcal{L}}(\mathbb{P}, \mathbb{Q}) + t d_{\mathcal{L}}(\mathbb{Q}, \hat{\mathbb{P}}) &= \gamma(d_{\mathcal{L}}(\mathbb{P}, \mathbb{Q}) + d_{\mathcal{L}}(\mathbb{Q}, \hat{\mathbb{P}})) + (t - \gamma) d_{\mathcal{L}}(\mathbb{Q}, \hat{\mathbb{P}}) \\
&= \gamma\big(\sup_{\ell\in\mathcal{L}} \{\mathbb{E}_{\mathbb{P}}[\ell] - \mathbb{E}_{\mathbb{Q}}[\ell]\} + \sup_{\ell\in\mathcal{L}} \{\mathbb{E}_{\mathbb{Q}}[\ell] - \mathbb{E}_{\hat{\mathbb{P}}}[\ell]\} \big) + (t - \gamma) d_{\mathcal{L}}(\mathbb{Q}, \hat{\mathbb{P}}) \\
&\ge \gamma \sup_{\ell\in\mathcal{L}} \{\mathbb{E}_{\mathbb{P}}[\ell] - \mathbb{E}_{\hat{\mathbb{P}}}[\ell]\} = \gamma d_{\mathcal{L}}(\mathbb{P}, \hat{\mathbb{P}}),
\end{aligned}$$

where the equality is attained at $\mathbb{Q} = \hat{\mathbb{P}}$. Combing these two parts, we obtain (EC.4). Eliminating the inner minimization $\min\{t, \gamma\}$ then yields the right-hand side problem in (17). □

Proof of Proposition 4. Define $\phi(t) = |t - 1|$. Following theorem 4.2 in Ben-Tal and Teboulle (2007), for any $\mathbb{Q} \in \mathcal{F}_{\mathcal{L}_2}(\theta)$ we then have

$$\sup_{\mathbb{P}\in\mathcal{P}(\Xi)} \left\{ \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] - \gamma d_{\mathcal{L}_1}(\mathbb{P}, \mathbb{Q}) \right\} \;=\; \inf_{v\in\mathbb{R}} \left\{ v + \mathbb{E}_{\mathbb{Q}}[(\gamma\phi)^*(f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) - v)] \right\},$$

where $\phi^*(s) = \max\{-1, s\}$ with domain $(-\infty, 1]$. The left-hand side of (16) then becomes

$$\begin{aligned}
&\sup_{\mathbb{Q}\in\mathcal{F}_{\mathcal{L}_2}(\theta)} \inf_{v\in\mathbb{R}} \left\{ v + \mathbb{E}_{\mathbb{Q}}[(\gamma\phi)^*(f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) - v)] \right\} \\
=\;& \inf_{v\in\mathbb{R}} \left\{ v + \sup_{\mathbb{Q}\in\mathcal{F}_{\mathcal{L}_2}(\theta)} \mathbb{E}_{\mathbb{Q}}[(\gamma\phi)^*(f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}) - v)] \right\} \\
=\;& \inf_{t\ge 0,\, v\in\mathbb{R}} \left\{ v + t\theta + \frac{1}{N} \sum_{n\in[N]} \sup_{\boldsymbol{\xi}\in\Xi} \left\{ (\gamma\phi)^*(f(\boldsymbol{x}, \boldsymbol{\xi}) - v) - t\|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}_n\| \right\} \right\},
\end{aligned}$$

where the first identity follows from the minimax theorem and the second identity is due to the duality result of Wasserstein DRO. Introducing the epigraphical $\boldsymbol{s}$ and plugging ϕ^*, we obtain

$$\begin{array}{lll} \inf & v + t\theta + \dfrac{1}{N}\displaystyle\sum_{n\in[N]} u_n & \\ \text{s.t.} & \displaystyle\sup_{\boldsymbol{\xi}\in\Xi} f(\boldsymbol{x},\boldsymbol{\xi}) \le \gamma + v & \\ & \displaystyle\sup_{\boldsymbol{\xi}\in\Xi} \big\{ f(\boldsymbol{x},\boldsymbol{\xi}) - v - t\|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}_n\| \big\} \le u_n & \forall n\in[N] \\ & \displaystyle\sup_{\boldsymbol{\xi}\in\Xi} \big\{ -\gamma - t\|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}_n\| \big\} \le u_n & \forall n\in[N] \\ & t\ge 0,\ \boldsymbol{s}\in\mathbb{R}^N,\ v\in\mathbb{R}. & \end{array}$$

Here, the third constraint group is equivalent to $u_n + \gamma \ge 0\ \forall n\in[N]$ since for each $n\in[N]$ the optimum is attained at $\boldsymbol{\xi} = \hat{\boldsymbol{\xi}}_n$. The first and second constraint groups can be reformulated as in Proposition 1. Combining the above two parts then completes the proof. □

Proof of Proposition 5. With the assumption that all distributions in $\mathcal{P}(\Xi)$ are absolutely continuous to $\hat{\mathbb{P}}$, $\mathcal{P}(\Xi)$ reduces to the (compact) probability simplex. Exploring the definition of $d_{\mathcal{L}_1}$ and $d_{\mathcal{L}_2}$, the right-hand side problem of (16) can be rewritten as

$$\begin{aligned} &\inf_{t\ge 0} \sup_{\mathbb{P}\in\mathcal{P}(\Xi),\,\mathbb{Q}\in\mathcal{P}(\Xi)} \left\{ t\theta + \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})] - \gamma \sup_{\ell_1\in\mathcal{L}_1} \big\{ \mathbb{E}_{\mathbb{P}}[\ell_1(\tilde{\boldsymbol{\xi}})] - \mathbb{E}_{\mathbb{Q}}[\ell_1(\tilde{\boldsymbol{\xi}})] \big\} - t \sup_{\ell_2\in\mathcal{L}_2} \big\{ \mathbb{E}_{\mathbb{Q}}[\ell_2(\tilde{\boldsymbol{\xi}})] - \mathbb{E}_{\hat{\mathbb{P}}}[\ell_2(\tilde{\boldsymbol{\xi}})] \big\} \right\} \\ = &\inf_{t\ge 0} \inf_{\ell_1\in\mathcal{L}_1,\,\ell_2\in\mathcal{L}_2} \sup_{\mathbb{P}\in\mathcal{P}(\Xi),\,\mathbb{Q}\in\mathcal{P}(\Xi)} \left\{ t\theta + \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})] - \gamma\big(\mathbb{E}_{\mathbb{P}}[\ell_1(\tilde{\boldsymbol{\xi}})] - \mathbb{E}_{\mathbb{Q}}[\ell_1(\tilde{\boldsymbol{\xi}})] \big) - t\big(\mathbb{E}_{\mathbb{Q}}[\ell_2(\tilde{\boldsymbol{\xi}})] - \mathbb{E}_{\hat{\mathbb{P}}}[\ell_2(\tilde{\boldsymbol{\xi}})] \big) \right\} \\ = &\inf_{t\ge 0} \inf_{\ell_1\in\mathcal{L}_1,\,\ell_2\in\mathcal{L}_2} \left\{ t\theta + \max_{n\in[N]} \big\{ f(\boldsymbol{x},\hat{\boldsymbol{\xi}}_n) - \gamma\ell_1(\hat{\boldsymbol{\xi}}_n) \big\} + \max_{n\in[N]} \big\{ \gamma\ell_1(\hat{\boldsymbol{\xi}}_n) - t\ell_2(\hat{\boldsymbol{\xi}}_n) \big\} + \frac{t}{N}\sum_{n\in[N]} \ell_2(\hat{\boldsymbol{\xi}}_n) \right\}, \end{aligned}$$

where the second line follows from the minimax theorem since both $\mathcal{L}_1$ and $\mathcal{L}_2$ are convex sets, and the last line follows from the fact that for any $t\ge 0$, $\ell_1\in\mathcal{L}_1$ and $\ell_2\in\mathcal{L}_2$, we have

$$\begin{aligned} \sup_{\mathbb{P}\in\mathcal{P}(\Xi)} \big\{ \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})] - \gamma\mathbb{E}_{\mathbb{P}}[\ell_1(\tilde{\boldsymbol{\xi}})] \big\} &= \max_{n\in[N]} \big\{ f(\boldsymbol{x},\hat{\boldsymbol{\xi}}_n) - \gamma\ell_1(\hat{\boldsymbol{\xi}}_n) \big\}, \\ \sup_{\mathbb{Q}\in\mathcal{P}(\Xi)} \big\{ \gamma\mathbb{E}_{\mathbb{Q}}[\ell_1(\tilde{\boldsymbol{\xi}})] - t\mathbb{E}_{\mathbb{Q}}[\ell_2(\tilde{\boldsymbol{\xi}})] \big\} &= \max_{n\in[N]} \big\{ \gamma\ell_1(\hat{\boldsymbol{\xi}}_n) - t\ell_2(\hat{\boldsymbol{\xi}}_n) \big\}. \end{aligned}$$

Then the desired result follows from the definitions of $\mathcal{L}_1$ and $\mathcal{L}_2$ and by eliminating the bilinear terms $t\ell_2(\hat{\boldsymbol{\xi}}_n)$ with a variable substitution $t\ell_2(\hat{\boldsymbol{\xi}}_n) \to s_n$. □

Proof of Theorem 2. For any fixed $\boldsymbol{x}\in\mathcal{X}$, the left-hand side problem in (18) can be rewritten as

$$\begin{array}{lll} \max & \displaystyle\int_\Xi f(\boldsymbol{x},\boldsymbol{\xi})\,\mathrm{d}\mathbb{P}(\boldsymbol{\xi}) - \frac{\gamma}{N}\sum_{n\in[N]} \int_{\Xi\times\Xi} c_1(\boldsymbol{\xi},\boldsymbol{\zeta})\,\mathrm{d}\pi_n(\boldsymbol{\xi},\boldsymbol{\zeta}) & \\ \text{s.t.} & \displaystyle\frac{1}{N}\sum_{n\in[N]} \int_\Xi c_2(\boldsymbol{\zeta},\hat{\boldsymbol{\xi}}_n)\,\mathrm{d}\mathbb{Q}_n(\boldsymbol{\zeta}) \le \theta & \\ & \displaystyle\tilde{\boldsymbol{\xi}} \sim \mathbb{P} = \frac{1}{N}\sum_{n\in[N]} \mathbb{P}_n,\ \tilde{\boldsymbol{\zeta}} \sim \frac{1}{N}\sum_{n\in[N]} \mathbb{Q}_n,\ (\tilde{\boldsymbol{\xi}},\tilde{\boldsymbol{\zeta}}) \sim \frac{1}{N}\sum_{n\in[N]} \pi_n & \\ & \mathbb{P}_n,\ \mathbb{Q}_n \in \mathcal{P}(\Xi),\ \pi_n \in \mathcal{Q}(\mathbb{P}_n,\mathbb{Q}_n) & \forall n\in[N], \end{array}$$

whose Lagrangian dual, after applying similar techniques as in the proof of Lemma 1, is

$$\inf_{t\geq 0}\left\{\theta t+\frac{1}{N}\sum_{n\in[N]}\sup_{\boldsymbol{\xi},\,\boldsymbol{\zeta}\in\Xi}\left\{f(\boldsymbol{x},\boldsymbol{\xi})-\gamma c_1(\boldsymbol{\xi},\boldsymbol{\zeta})-tc_2(\hat{\boldsymbol{\xi}}_n,\boldsymbol{\zeta})\right\}\right\}\leq 0.$$

Here, strong duality can be verified via a similar proof as that of Lemma 1 or Theorem 1. Introducing epigraphical auxiliary variables, the dual representation can be rewritten as follows:

$$\begin{cases}\theta Nt+\boldsymbol{e}^\top\boldsymbol{s}\leq 0\\ \sup\limits_{\boldsymbol{\xi},\boldsymbol{\zeta}\in\Xi}\left\{f_k(\boldsymbol{x},\boldsymbol{\xi})-\gamma c_1(\boldsymbol{\xi},\boldsymbol{\zeta})-tc_2(\hat{\boldsymbol{\xi}}_n,\boldsymbol{\zeta})\right\}\leq s_n\ \forall n\in[N],\,k\in[K]\\ t\geq 0.\end{cases}$$

For any fixed $n\in[N]$ and $k\in[K]$, we consider the inner supremum problem appearing in the left-hand side of the second constraint:

$$\begin{aligned}&\sup_{\boldsymbol{\xi},\boldsymbol{\zeta}\in\Xi}\left\{f_k(\boldsymbol{x},\boldsymbol{\xi})-\gamma c_1(\boldsymbol{\xi},\boldsymbol{\zeta})-tc_2(\hat{\boldsymbol{\xi}}_n,\boldsymbol{\zeta})\right\}\\ =&\sup_{\boldsymbol{\xi},\boldsymbol{\zeta}\in\Xi,\,\boldsymbol{t}=\boldsymbol{\xi},\,\boldsymbol{u}=\boldsymbol{\zeta}}\left\{f_k(\boldsymbol{x},\boldsymbol{\xi})-\gamma c_1(\boldsymbol{t},\boldsymbol{u})-tc_2(\hat{\boldsymbol{\xi}}_n,\boldsymbol{\zeta})\right\}\\ =&\inf_{\boldsymbol{v},\boldsymbol{z}}\sup_{\boldsymbol{\xi},\boldsymbol{\zeta}\in\Xi,\,\boldsymbol{t},\boldsymbol{u}}\left\{f_k(\boldsymbol{x},\boldsymbol{\xi})-\gamma c_1(\boldsymbol{t},\boldsymbol{u})-tc_2(\hat{\boldsymbol{\xi}}_n,\boldsymbol{\zeta})-\boldsymbol{v}^\top(\boldsymbol{u}-\boldsymbol{\zeta})-\boldsymbol{z}^\top(\boldsymbol{t}-\boldsymbol{\xi})\right\}\\ =&\inf_{\boldsymbol{v},\boldsymbol{z}}\left\{\underbrace{\sup_{\boldsymbol{\xi}\in\Xi}\left\{f_k(\boldsymbol{x},\boldsymbol{\xi})+\boldsymbol{z}^\top\boldsymbol{\xi}\right\}}_{g_1(\boldsymbol{z})}+\underbrace{\sup_{\boldsymbol{\zeta}\in\Xi}\left\{\boldsymbol{v}^\top\boldsymbol{\zeta}-tc_2(\hat{\boldsymbol{\xi}}_n,\boldsymbol{\zeta})\right\}}_{g_2(\boldsymbol{v})}+\underbrace{\sup_{\boldsymbol{t},\boldsymbol{u}}\left\{-\gamma c_1(\boldsymbol{t},\boldsymbol{u})-\boldsymbol{v}^\top\boldsymbol{u}-\boldsymbol{z}^\top\boldsymbol{t}\right\}}_{g_3(\boldsymbol{v},\boldsymbol{z})}\right\}.\end{aligned}$$

For the three terms above, we have, respectively,

$$\begin{aligned}g_1(\boldsymbol{z})&=\sup_{\boldsymbol{\xi}\in\Xi}\left\{f_k(\boldsymbol{x},\boldsymbol{\xi})+\boldsymbol{z}^\top\boldsymbol{\xi}\right\}=\sup_{\boldsymbol{\xi}}\left\{f_k(\boldsymbol{x},\boldsymbol{\xi})+\boldsymbol{z}^\top\boldsymbol{\xi}-\delta(\boldsymbol{\xi}\mid\Xi)\right\}=\inf_{\boldsymbol{w}}\left\{[-f_k]^{\bar{*}}(\boldsymbol{x},\boldsymbol{w}-\boldsymbol{z})+\delta^*(\boldsymbol{w}\mid\Xi)\right\}\\ g_2(\boldsymbol{v})&=\sup_{\boldsymbol{\zeta}\in\Xi}\left\{\boldsymbol{v}_n^\top\boldsymbol{\zeta}-tc_2(\hat{\boldsymbol{\xi}}_n,\boldsymbol{\zeta})\right\}=\sup_{\boldsymbol{\zeta}}\left\{\boldsymbol{v}^\top\boldsymbol{\zeta}-tc_2(\hat{\boldsymbol{\xi}}_n,\boldsymbol{\zeta})-\delta(\boldsymbol{\zeta}\mid\Xi)\right\}=\inf_{\boldsymbol{y}}\left\{\delta^*(\boldsymbol{v}-\boldsymbol{y}\mid\Xi)+tc_2^{\bar{*}}(\hat{\boldsymbol{\xi}}_n,\boldsymbol{y}/t)\right\}\\ g_3(\boldsymbol{v},\boldsymbol{z})&=\sup_{\boldsymbol{t},\boldsymbol{u}}\left\{-\gamma c_1(\boldsymbol{t},\boldsymbol{u})-\boldsymbol{v}^\top\boldsymbol{u}-\boldsymbol{z}^\top\boldsymbol{t}\right\}=\gamma c_1^*(-\boldsymbol{z}/\gamma,-\boldsymbol{v}/\gamma).\end{aligned}$$

Thus for each pair of $n\in[N]$ and $k\in[K]$, the second constraint becomes

$$\inf_{\boldsymbol{v},\boldsymbol{w},\boldsymbol{y},\boldsymbol{z}}\left\{[-f_k]^{\bar{*}}(\boldsymbol{x},\boldsymbol{w}-\boldsymbol{z})+\delta^*(\boldsymbol{w}\mid\Xi)+\delta^*(\boldsymbol{v}-\boldsymbol{y}\mid\Xi)+tc_2^{\bar{*}}(\hat{\boldsymbol{\xi}}_n,\boldsymbol{y}/t)+\gamma c_1^*(-\boldsymbol{z}/\gamma,-\boldsymbol{v}/\gamma)\right\}\leq s_n,$$

for which we must have $\boldsymbol{z}=-\boldsymbol{v}$ because otherwise $c_1^*(-\boldsymbol{z}/\gamma,-\boldsymbol{v}/\gamma)$ tends to ∞. The remaining proof then follows from substituting the above class of constraints into the dual representation. □

Proof of Theorem 3. The globalized distributionally robust counterpart herein is equivalent to

$$\sup_{\mathbb{P}\in\mathcal{P}(\Xi),\,\mathbb{Q}\in\mathcal{F}_\rho(\sqrt[\rho]{\theta})}\left\{\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})]-\gamma\cdot d_{\|\cdot\|^\rho}(\mathbb{P},\mathbb{Q})\right\}\leq 0.$$

Using similar tricks as in the proof of Lemma 1, we can rewrite the left-hand side problem into

$$\begin{array}{ll} \sup & \displaystyle\int_{\Xi} f(\boldsymbol{x},\boldsymbol{\xi})\,\mathrm{d}\mathbb{P}(\boldsymbol{\xi}) - \frac{\gamma}{N}\sum_{n\in[N]}\int_{\Xi\times\Xi}\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|^{\rho}\,\mathrm{d}\pi_n(\boldsymbol{\xi},\boldsymbol{\zeta}) \\ \text{s.t.} & \displaystyle\frac{1}{N}\sum_{n\in[N]}\int_{\Xi}\|\boldsymbol{\zeta}-\hat{\boldsymbol{\xi}}_n\|^{\rho}\,\mathrm{d}\mathbb{Q}_n(\boldsymbol{\zeta}) \le \theta \\ & \displaystyle\tilde{\boldsymbol{\xi}} \sim \mathbb{P} = \frac{1}{N}\sum_{n\in[N]}\mathbb{P}_n,\ \tilde{\boldsymbol{\zeta}} \sim \frac{1}{N}\sum_{n\in[N]}\mathbb{Q}_n,\ (\tilde{\boldsymbol{\xi}},\tilde{\boldsymbol{\zeta}}) \sim \frac{1}{N}\sum_{n\in[N]}\pi_n \\ & \mathbb{P}_n,\ \mathbb{Q}_n \in \mathcal{P}(\Xi),\ \pi_n \in \mathcal{Q}(\mathbb{P}_n,\mathbb{Q}_n) \qquad\qquad \forall n\in[N]. \end{array}$$

Objective value of the above problem is not be larger than that of its Lagrangian dual, given by

$$\begin{aligned} &\inf_{t\ge 0}\left\{\theta t + \frac{1}{N}\sum_{n\in[N]}\sup_{\pi_n\in\mathcal{P}(\Xi\times\Xi)}\int_{\Xi\times\Xi}\left(f(\boldsymbol{x},\boldsymbol{\xi}) - \gamma\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|^{\rho} - t\|\boldsymbol{\zeta}-\hat{\boldsymbol{\xi}}_n\|^{\rho}\right)\mathrm{d}\pi_n(\boldsymbol{\xi},\boldsymbol{\eta})\right\} \\ =&\inf_{t\ge 0}\left\{\theta t + \frac{1}{N}\sum_{n\in[N]}\sup_{\boldsymbol{\xi},\boldsymbol{\zeta}\in\Xi}\left\{f(\boldsymbol{x},\boldsymbol{\xi}) - \gamma\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|^{\rho} - t\|\boldsymbol{\zeta}-\hat{\boldsymbol{\xi}}_n\|^{\rho}\right\}\right\}. \end{aligned}$$

Here, strong duality can be established based on similar arguments as in the proof of Lemma 1 or Theorem 1. The remaining proof then follows straightforwardly. □

Proof of Proposition 6. By Theorem 3, it is sufficient to investigate

$$\theta N t + \sum_{n\in[N]}\sup_{\boldsymbol{\xi},\boldsymbol{\zeta}\in\Xi}\left\{f(\boldsymbol{x},\boldsymbol{\xi}) - \gamma\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|_2^2 - t\|\boldsymbol{\zeta}-\hat{\boldsymbol{\xi}}_n\|_2^2\right\} \le 0$$

and for a fixed $\boldsymbol{\xi}\in\Xi$, to work with the infimum problem

$$\inf_{\boldsymbol{\zeta}\in\Xi}\left\{\gamma\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|_2^2 + t\|\boldsymbol{\zeta}-\hat{\boldsymbol{\xi}}_n\|_2^2\right\} = \inf_{\boldsymbol{\zeta}\in\Xi} h_{\boldsymbol{\xi}}(\boldsymbol{\zeta}),$$

where $h_{\boldsymbol{\xi}}(\boldsymbol{\zeta}) = (\gamma+t)\boldsymbol{\zeta}^\top\boldsymbol{\zeta} - 2\boldsymbol{\zeta}^\top(\gamma\boldsymbol{\xi}+t\hat{\boldsymbol{\xi}}_n) + \gamma\boldsymbol{\xi}^\top\boldsymbol{\xi} + t\hat{\boldsymbol{\xi}}_n^\top\hat{\boldsymbol{\xi}}_n$. By the first-order condition, the optimal $\boldsymbol{\zeta}^\star$ must satisfy $2(\gamma+t)\boldsymbol{\zeta}^* - 2(\gamma\boldsymbol{\xi}+t\hat{\boldsymbol{\xi}}_n) = 0$, implying $\boldsymbol{\zeta}^\star = \frac{\gamma}{\gamma+t}\boldsymbol{\xi} + \frac{t}{\gamma+t}\hat{\boldsymbol{\xi}}_n \in \Xi$. This then yields

$$\inf_{\boldsymbol{\zeta}\in\Xi}\left\{\gamma\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|_2^2 + t\|\boldsymbol{\zeta}-\hat{\boldsymbol{\xi}}_n\|_2^2\right\} = \gamma\|\boldsymbol{\xi}-\boldsymbol{\zeta}^\star\|_2^2 + t\|\boldsymbol{\zeta}^\star-\hat{\boldsymbol{\xi}}_n\|_2^2 = \frac{\gamma t}{\gamma+t}\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}_n\|_2^2.$$

The remaining proof follows straightforwardly. □

Proof of Theorem 4. Before deriving the dual reformulation, we first introduce a technical lemma.

LEMMA EC.2. *The dual cone of* $\mathcal{K}_\phi = \mathrm{cl}\{(x,y,z) \mid y\phi(x/y) \le z,\ x\ge 0,\ y>0\}$ *is given by*

$$(\mathcal{K}_\phi)_* = \mathrm{cl}\{(u,v,w) \mid w\phi^*(-u/w) \le v,\ w>0\}.$$

Proof of Lemma EC.2. By definition, any $(u,v,w)\in(\mathcal{K}_\phi)_*$ must satisfy

$$\begin{bmatrix} \inf & ux+vy+wz \\ \text{s.t.} & y\phi(x/y)\le z,\ x\ge 0,\ y>0 \end{bmatrix} \ge 0.$$

Here, we have $w \geq 0$ because otherwise we can always find a sufficiently large z rendering the objective value of the left-hand side problem negative, violating the definition of dual cone. When $w = 0$, it is sufficient to require that $u \geq 0$ and $v \geq 0$. When $w > 0$, the above inequality becomes

$$\inf_{x \geq 0,\, y>0} y \left[u\left(\frac{x}{y}\right) + v + w\phi\left(\frac{x}{y}\right) \right] \geq 0,$$

which is equivalent to $\inf_{x\geq 0} \big\{ux + v + w\phi(x)\big\} \geq 0$. Then we arrive at

$$v \geq \sup_{x\geq 0}\{-ux - w\phi(x)\} = w\phi^*\Big(-\frac{u}{w}\Big).$$

In particular, consider a sequence $\{(u_r, v_r, w_r)\}_{r\in\mathbb{N}}$ that satisfies the above relations and let $w_r \to 0$ as $t \to \infty$, we must have $\lim_{r\to\infty} u_r \geq 0$ and

$$\lim_{r\to\infty} v_r \geq \lim_{r\to\infty} w_r \phi^*\Big(-\frac{u_r}{w_r}\Big) = \lim_{r\to\infty} \sup_{x\geq 0}\{-u_r x - w_r\phi(x)\} = 0.$$

That is, the condition on $\{(u_r, v_r, w_r)\}_{r\in\mathbb{N}}$ in the limiting situation coincides with the condition on (u, v, w) in the case of $w = 0$. In summary, we obtain the desired expression of $(\mathcal{K}_\phi)_*$. □

Consider the following presentation of (21):

$$\sup_{\mathbb{P}\in\mathcal{P}(\Xi),\, \mathbb{Q}\in\mathcal{F}_{\phi_2}(\theta)} \Big\{ \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] - \gamma\cdot d_{\phi_1}(\mathbb{P}\,|\,\mathbb{Q}) \Big\} \leq 0.$$

Then the left-hand side problem can be reformulated as the following optimization problem

$$\begin{array}{ll} \sup & \displaystyle\sum_{n\in[N]} p_n f(\boldsymbol{x}, \hat{\boldsymbol{\xi}}_n) - \gamma \sum_{n\in[N]} q_n \phi_1\left(\frac{p_n}{q_n}\right) \\ \text{s.t.} & \displaystyle\sum_{n\in[N]} \phi_2\left(Nq_n\right) \leq N\theta \\ & \boldsymbol{e}^\top \boldsymbol{p} = 1 \\ & \boldsymbol{e}^\top \boldsymbol{q} = 1 \\ & \boldsymbol{p},\ \boldsymbol{q} \in \mathbb{R}^N_+, \end{array}$$

which can be reformulated as a conic linear optimization problem

$$\begin{array}{ll} \sup & \displaystyle\sum_{n\in[N]} p_n f(\boldsymbol{x}, \hat{\boldsymbol{\xi}}_n) - \gamma \boldsymbol{e}^\top \boldsymbol{\beta} \\ \text{s.t.} & \boldsymbol{e}^\top \boldsymbol{\alpha} \leq \theta \\ & \boldsymbol{e}^\top \boldsymbol{p} = 1 \\ & \boldsymbol{e}^\top \boldsymbol{q} = 1 \\ & (p_n, q_n, \beta_n) \in \mathcal{K}_{\phi_1},\ (q_n, 1/N, \alpha_n) \in \mathcal{K}_{\phi_2} \qquad \forall n \in [N], \end{array}$$

where for $i \in \{1,2\}$, $\mathcal{K}_{\phi_i} = \mathrm{cl}\{(x,y,z) \mid y\phi_i(x/y) \leq z,\ x \geq 0,\ y > 0\}$. Under the condition that there exists some probability vector $\boldsymbol{q}$ such that $\sum_{n\in[N]} \phi_2\left(Nq_n\right) < N\theta$, it is not hard to argue that strong duality holds between the above problem and its dual, which is given as follows:

$$\begin{array}{lll} \inf & \displaystyle v + u + \theta t + \frac{1}{N}\sum_{n\in[N]} w_n & \\ \text{s.t.} & (v - f(\boldsymbol{x}, \hat{\boldsymbol{\xi}}_n), u - s_n, \gamma) \in (\mathcal{K}_{\phi_1})_* & \forall n \in [N] \\ & (s_n, w_n, t) \in (\mathcal{K}_{\phi_2})_* & \forall n \in [N]. \end{array}$$

According to Lemma EC.2, the above problem can be rewritten as

$$\begin{array}{ll}\inf & v+u+\theta t+\dfrac{1}{N}\displaystyle\sum_{n\in[N]} w_n \\ \text{s.t.} & \gamma\phi_1^*\left(\dfrac{f(\boldsymbol{x},\hat{\boldsymbol{\xi}}_n)-v}{\gamma}\right)\le u-s_n \quad \forall n\in[N] \\ & t\phi_2^*\left(-\dfrac{s_n}{t}\right)\le w_n \qquad\qquad \forall n\in[N] \\ & t\ge 0.\end{array}$$

Eliminating the auxiliary variables u and w_n then yields the desired reformulation (22). □

Proof of Theorem 5. The left-hand side problem of (25) can be rewritten as

$$\begin{array}{ll}\sup & \displaystyle\int_\Xi f(\boldsymbol{x},\boldsymbol{\xi})\,\mathrm{d}\mathbb{P}(\boldsymbol{\xi})-\gamma\cdot\int_{\Xi\times\Xi}\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|\,\mathrm{d}\pi(\boldsymbol{\xi},\boldsymbol{\zeta}) \\ \text{s.t.} & \mathbb{E}_\mathbb{Q}[\tilde{\boldsymbol{\zeta}}]=\boldsymbol{\mu} \\ & \mathbb{E}_\mathbb{Q}[\boldsymbol{g}(\tilde{\boldsymbol{\zeta}})]\preceq_\mathcal{K}\boldsymbol{h} \\ & \mathbb{Q}[\tilde{\boldsymbol{\zeta}}\in\Xi]=1 \\ & \mathbb{P}[\tilde{\boldsymbol{\xi}}\in\Xi]=1 \\ & \tilde{\boldsymbol{\xi}}\sim\mathbb{P},\ \tilde{\boldsymbol{\zeta}}\sim\mathbb{Q},\ (\tilde{\boldsymbol{\xi}},\tilde{\boldsymbol{\zeta}})\sim\pi \\ & \mathbb{P}\in\mathcal{P}(\mathbb{R}^I),\ \mathbb{Q}\in\mathcal{P}(\mathbb{R}^I),\ \pi\in\mathcal{Q}(\mathbb{P},\mathbb{Q}).\end{array}$$

Note that the Dirac distribution $\mathbb{Q}^\star=\delta_{\boldsymbol{\mu}}$ with $\boldsymbol{\mu}\in\mathrm{int}(\Xi)$ satisfies $\mathbb{E}_{\mathbb{Q}^\star}[\boldsymbol{g}(\tilde{\boldsymbol{\zeta}})]=\boldsymbol{g}(\boldsymbol{\mu})\prec_\mathcal{K}\boldsymbol{h}$. Hence, strong duality holds. Specifically, the dual can be given by

$$\begin{array}{ll}\inf & t+\boldsymbol{\alpha}^\top\boldsymbol{\mu}+\boldsymbol{\beta}^\top\boldsymbol{h} \\ \text{s.t.} & t+\boldsymbol{\alpha}^\top\boldsymbol{\zeta}+\boldsymbol{\beta}^\top\boldsymbol{g}(\boldsymbol{\zeta})\ge f(\boldsymbol{x},\boldsymbol{\xi})-\gamma\|\boldsymbol{\xi}-\boldsymbol{\zeta}\| \quad \forall\boldsymbol{\xi}\in\Xi,\ \boldsymbol{\zeta}\in\Xi \\ & t\in\mathbb{R},\ \boldsymbol{\alpha}\in\mathbb{R}^I,\ \boldsymbol{\beta}\in\mathcal{K}_*.\end{array}$$

Since $\boldsymbol{\beta}\in\mathcal{K}_*$, introducing an epigraphical decision vector $\boldsymbol{v}$ such that $\boldsymbol{g}(\boldsymbol{\zeta})\prec\boldsymbol{v}$ and exploring the structure of the cost function f, the semi-infinite constraint in the dual becomes

$$\sup_{\boldsymbol{\xi}\in\Xi,\,(\boldsymbol{\zeta},\boldsymbol{v})\in\bar{\Xi}}\left\{f_k(\boldsymbol{x},\boldsymbol{\xi})-\gamma\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|-\boldsymbol{\alpha}^\top\boldsymbol{\zeta}-\boldsymbol{\beta}^\top\boldsymbol{v}\right\}\le t \quad \forall k\in[K].$$

Indeed, for every $k\in[K]$ we have

$$\begin{aligned}&\sup_{\boldsymbol{\xi}\in\Xi,\,(\boldsymbol{\zeta},\boldsymbol{v})\in\bar{\Xi}}\left\{f_k(\boldsymbol{x},\boldsymbol{\xi})-\gamma\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|-\boldsymbol{\alpha}^\top\boldsymbol{\zeta}-\boldsymbol{\beta}^\top\boldsymbol{v}\right\}\\
=&\sup_{(\boldsymbol{\zeta},\boldsymbol{v})\in\bar{\Xi}}\left\{\sup_{\boldsymbol{\xi}}\left\{f_k(\boldsymbol{x},\boldsymbol{\xi})-\gamma\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|-\delta(\boldsymbol{\xi}\,|\,\Xi)\right\}-\boldsymbol{\alpha}^\top\boldsymbol{\zeta}-\boldsymbol{\beta}^\top\boldsymbol{v}\right\}\\
=&\sup_{(\boldsymbol{\zeta},\boldsymbol{v})\in\bar{\Xi}}\left\{\inf_{\boldsymbol{u},\,\|\boldsymbol{w}\|_*\le\gamma}\left\{[-f_k]^{\bar{*}}(\boldsymbol{x},\boldsymbol{u}-\boldsymbol{w})+\delta^*(\boldsymbol{u}\,|\,\Xi)-\boldsymbol{w}^\top\boldsymbol{\zeta}\right\}-\boldsymbol{\alpha}^\top\boldsymbol{\zeta}-\boldsymbol{\beta}^\top\boldsymbol{v}\right\}\\
=&\inf_{\boldsymbol{u},\,\|\boldsymbol{w}\|_*\le\gamma}\left\{[-f_k]^{\bar{*}}(\boldsymbol{x},\boldsymbol{u}-\boldsymbol{w})+\delta^*(\boldsymbol{u}\,|\,\Xi)+\sup_{\boldsymbol{\zeta},\boldsymbol{v}}\left\{(-\boldsymbol{w}-\boldsymbol{\alpha})^\top\boldsymbol{\zeta}-\boldsymbol{\beta}^\top\boldsymbol{v}-\delta(\boldsymbol{\zeta},\boldsymbol{v}\,|\,\bar{\Xi})\right\}\right\}\\
=&\inf_{\boldsymbol{u},\,\|\boldsymbol{w}\|_*\le\gamma}\left\{[-f_k]^{\bar{*}}(\boldsymbol{x},\boldsymbol{u}-\boldsymbol{w})+\delta^*(\boldsymbol{u}\,|\,\Xi)+\delta^*(-\boldsymbol{w}-\boldsymbol{\alpha},-\boldsymbol{\beta}\,|\,\bar{\Xi})\right\}.\end{aligned}$$

Substituting the above expression for each $k \in [K]$ then yields the desired reformulation. $\square$

EC.2. Reduction to Globalized Robust Counterpart

We now consider the globalized distributionally robust counterpart with optimal transport:

$$\sup_{\mathbb{P}\in\mathcal{P}(\Xi),\,\mathbb{Q}\in\mathcal{F}_{c_2}(\theta)} \left\{\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})] - \gamma \cdot d_{c_1}(\mathbb{P},\mathbb{Q})\right\} \le 0, \tag{EC.5}$$

where c_1 and c_2 are defined as in Section 4.2, and $\mathcal{F}_{c_2}(\theta) = \{\mathbb{P} \in \mathcal{P}(\Xi) \mid d_{c_2}(\mathbb{P},\hat{\mathbb{P}}) \le \theta\}$ with $\theta > 0$ is an optimal transport ambiguity set centered around the Dirac reference distribution $\hat{\mathbb{P}} = \delta_{\hat{\boldsymbol{\xi}}}$. Here, we assume $\hat{\boldsymbol{\xi}} \in \operatorname{int}(\Xi)$. Following similar steps as in the proof of Theorem 3, we know that any $\boldsymbol{x} \in \mathcal{X}$ satisfies (EC.5) if and only if it satisfies

$$\begin{bmatrix} \inf & \theta t + \sup\limits_{\boldsymbol{\xi},\boldsymbol{\zeta}\in\Xi} \left\{f(\boldsymbol{x},\boldsymbol{\xi}) - \gamma c_1(\boldsymbol{\xi},\boldsymbol{\zeta}) - t c_2(\boldsymbol{\zeta},\hat{\boldsymbol{\xi}})\right\} \\ \text{s.t.} & t \ge 0 \end{bmatrix} \le 0. \tag{EC.6}$$

Observe that the left-hand side problem is the Lagrangian dual of the following problem

$$\begin{array}{ll} \sup & f(\boldsymbol{x},\boldsymbol{\xi}) - \gamma c_1(\boldsymbol{\xi},\boldsymbol{\zeta}) \\ \text{s.t.} & c_2(\boldsymbol{\zeta},\hat{\boldsymbol{\xi}}) \le \theta \\ & \boldsymbol{\xi},\, \boldsymbol{\zeta} \in \Xi. \end{array}$$

Here, strong duality holds because $\theta > 0$ and $\hat{\boldsymbol{\xi}} \in \operatorname{int}(\Xi)$. Thus, the inequality (EC.6) becomes

$$\begin{bmatrix} \sup & f(\boldsymbol{x},\boldsymbol{\xi}) - \gamma c_1(\boldsymbol{\xi},\boldsymbol{\zeta}) \\ \text{s.t.} & c_2(\boldsymbol{\zeta},\hat{\boldsymbol{\xi}}) \le \theta \\ & \boldsymbol{\xi},\, \boldsymbol{\zeta} \in \Xi. \end{bmatrix} \le 0 \quad \Longleftrightarrow \quad f(\boldsymbol{x},\boldsymbol{\xi}) \le \gamma \cdot \min_{\boldsymbol{\zeta}\in U_2} c_1(\boldsymbol{\xi},\boldsymbol{\zeta}) \quad \forall \boldsymbol{\xi} \in U_1,$$

where we define $\Xi_1 = \Xi$ and $\Xi_2 = \{\boldsymbol{\xi} \in \Xi \mid c_2(\boldsymbol{\xi},\hat{\boldsymbol{\xi}}) \le \theta\}$ for the right-hand side constraint, which is exactly the globalized robust counterpart proposed by Ben-Tal et al. (2006), Ben-Tal et al. (2017).

EC.3. Derivation of Examples 1 and 2

In Example 1, the distributionally robust optimization problem is equivalent to

$$\begin{array}{ll} \min & x \\ \text{s.t.} & x \ge \ln \xi \quad \forall \xi \in (0,1] \\ & x \in \mathbb{R}, \end{array}$$

whose optimal solution (also optimal value) is $x^\star = 0$.

Consider next the globalized distributionally robust variant, wherein the constraint can be rewritten as

$$x \ge \max_{\xi>0,\,0<\zeta\le 1} \{\ln \xi - \gamma|\xi - \zeta|\} = \max\left\{\max_{0<\xi\le 1,\,0<\zeta\le 1} \{\ln \xi - \gamma|\xi - \zeta|\},\ \max_{\xi\ge 1,\,0<\zeta\le 1} \{\ln \xi - \gamma|\xi - \zeta|\}\right\}.$$

Since $\max_{0<\xi\leq 1,\, 0<\zeta\leq 1}\{\ln\xi-\gamma|\xi-\zeta|\}=\max_{0<\xi\leq 1}\ln\xi=0$ and $\max_{\xi\geq 1,\, 0<\zeta\leq 1}\{\ln\xi-\gamma|\xi-\zeta|\}=\max_{\xi\geq 1}\{\ln\xi-\gamma(\xi-1)\}=\gamma-\ln\gamma-1\geq 0$, we can determine that the optimal value of (3) is $\gamma-\ln\gamma-1$ with also $x^\star=\gamma-\ln\gamma-1$.

In Example 2, we consider the robust satisficing problem. Note that for any fixed $\gamma\geq 0$ and $x\geq 1$, its constraint is equivalent to $\max_{\xi>0}\{1-x+x\ln\xi-\gamma|\xi-1|\}\leq 0$. Since it always holds that $\max_{0<\xi\leq 1}\{1-x+x\ln\xi+\gamma(\xi-1)\}=1-x\leq 0$, we next study $\max_{\xi\geq 1}\{1-x+x\ln\xi-\gamma(\xi-1)\}$. If $1\leq x\leq\gamma$, then $x^\star=\gamma^\star=1$ at optimality and the constraint is satisfied; if not, then $\max_{\xi\geq 1}\{1-x+x\ln\xi-\gamma(\xi-1)\}=1-2x+x\ln(x/\gamma)+\gamma$ with the optimum $\xi^\star=x/\gamma$ determined by the first-order condition. It remains to solve $\min\{\gamma\,|1-2x+x\ln(x/\gamma)+\gamma\leq 0,\ x\geq\gamma\geq 0,\ x\geq 1\}$, for which the first-order condition yields $x^\star=\mathrm{e}\gamma$ and leads to $\gamma\geq 1/(\mathrm{e}-1)$. To sum up, the optimal value of problem (4) is $1/(\mathrm{e}-1)$, attaining at $x^\star=\mathrm{e}\gamma$ and $\gamma^\star=1/(\mathrm{e}-1)$.

Consider next the globalized robust satisficing variant, where for any fixed $\gamma\geq 0$ and $x\geq 1$, the constraint can be simplified into

$$\max_{\xi>0,\, 1-\theta\leq\zeta\leq 1+\theta}\{1-x+x\ln\xi-\gamma|\xi-\zeta|\}\leq 0. \tag{EC.7}$$

For any fixed $\xi\in[1-\theta,1+\theta]$, (EC.7) is equivalent to

$$\max_{1-\theta\leq\xi\leq 1+\theta,\, 1-\theta\leq\zeta\leq 1+\theta}\{1-x+x\ln\xi-\gamma|\xi-\zeta|\}=\max_{1-\theta\leq\xi\leq 1+\theta}\{1-x+x\ln\xi\}=1-x+x\ln(\theta+1)\leq 0,$$

implying $x\geq 1/(1-\ln(\theta+1))$; for $0<\xi\leq 1-\theta$, (EC.7) becomes

$$\max_{0<\xi\leq 1-\theta,\, 1-\theta\leq\zeta\leq 1+\theta}1-x+x\ln\xi-\gamma|\xi-\zeta|=\max_{0<\xi\leq 1-\theta}1-x+x\ln\xi-\gamma(1-\xi-\theta)=1-x+x\ln(1-\theta)\leq 0,$$

leading to $x\geq 1/(1-\ln(1-\theta))$; and for $\xi\geq 1+\theta$, (EC.7) is

$$\max_{\xi\geq 1+\theta,\, 1\leq\zeta\leq 1+\theta}\{1-x+x\ln\xi-\gamma|\xi-\zeta|\}=\max_{\xi\geq 1+\theta}\{1-x+x\ln\xi-\gamma(\xi-\theta-1)\},$$

which, after applying the first-order condition, gives

$$\max_{\xi\geq 1+\theta}\{1-x+x\ln\xi-\gamma(\xi-\theta-1)\}=\begin{cases}1-x+x\ln(1+\theta) & x\leq\gamma(1+\theta)\\ 1-2x+x\ln(x/\gamma)+\gamma(\theta-1) & x\geq\gamma(1+\theta).\end{cases}$$

When $1\leq x\leq\gamma(1+\theta)$, the constraint in (5) becomes $\gamma\geq x/(1+\theta)$ with $x\geq(1-\ln(1+\theta))^{-1}$, then $x^\star=(1-\ln(1+\theta))^{-1}$ and $\gamma^\star=((1-\ln(1+\theta))(1+\theta))^{-1}$ at optimality. When $x\geq\gamma(1+\theta)$, we arrive at $\inf_{x\geq\gamma(1+\theta)}\{1-2x+x\ln(x/\gamma)+\gamma(1+\theta)\}\leq 0$, which can be reformulated as $\gamma\geq 1/(\mathrm{e}-1-\theta)$ with $x^\star=\mathrm{e}\gamma$, implying $x^\star=\mathrm{e}/(\mathrm{e}-1-\theta)$ and $\gamma^\star=1/(\mathrm{e}-1-\theta)$ at optimality. In summary, the optimal value of problem (5) is $1/(\mathrm{e}-1-\theta)$, attaining at $x^\star=\mathrm{e}\gamma^\star$ and $\gamma^\star=1/(\mathrm{e}-1-\theta)$.

EC.4. Conditions on Finiteness of $t^\star$

Here, we list some conditions on the support set Ξ and cost function f that would guarantee the finiteness of the optimal dual variable $t^\star$ to the dual reformulation of the worst-case expectation $\sup_{\mathbb{P}\in\mathcal{F}_{\mathrm{W}}(\theta)} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})]$:

$$\inf_{t\geq 0}\left\{\theta t+\frac{1}{N}\sum_{n\in[N]}\sup_{\boldsymbol{\xi}\in\Xi}\left\{f(\boldsymbol{x},\boldsymbol{\xi})-t\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}_n\|\right\}\right\}.$$

(*i*) Suppose the cost function $f(\boldsymbol{x},\cdot)$ is upper semicontinuous and the growth rate κ, defined as

$$\kappa=\limsup_{\|\boldsymbol{\xi}-\boldsymbol{\xi}_0\|\to\infty}\frac{f(\boldsymbol{x},\boldsymbol{\xi})-f(\boldsymbol{x},\boldsymbol{\xi}_0)}{\|\boldsymbol{\xi}-\boldsymbol{\xi}_0\|}<\infty,$$

is finite for some fixed $\boldsymbol{\xi}_0\in\Xi$. According to the proof of theorem 1 in Gao and Kleywegt (2022), the minimizer $t^\star$ then lies in the interval $[\max(0,\kappa),\infty)$. Note that when $t\to\infty$,

$$\theta N t+\sum_{n\in[N]}\sup_{\boldsymbol{\xi}\in\Xi}\left\{f(\boldsymbol{x},\boldsymbol{\xi})-t\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}_n\|\right\}\to\infty,$$

which violates the globalized distributionally robust constraint. Therefore, we can conclude the finiteness of $t^\star$.

(*ii*) Suppose that the cost function $f(\boldsymbol{x},\cdot)$ is bounded in the support set Ξ, *i.e.*, there exists a and b such that $a\leq f(\boldsymbol{x},\boldsymbol{\xi})\leq b$ for any $\boldsymbol{\xi}\in\Xi$. Since

$$\sup_{\mathbb{P}\in\mathcal{F}_{\mathrm{W}}(\theta)}\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})]=\left\{\begin{array}{lll}\inf & \theta t+\dfrac{1}{N}\displaystyle\sum_{n\in[N]}s_n & \\ 0\text{s.t.} & \displaystyle\sup_{\boldsymbol{\xi}\in\Xi}\left\{f(\boldsymbol{x},\boldsymbol{\xi})-t\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}_n\|\right\}\leq s_n & \forall n\in[N]\\ & t\geq 0, & \end{array}\right.$$

substituting $\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}_n$ in each n^{th} constraint, we then have $a\leq f(\boldsymbol{x},\hat{\boldsymbol{\xi}}_n)\leq s_n$. Moreover, by the boundedness of f, we can identify $(t,s_1,\ldots,s_N)=(0,b,\ldots,b)$ as a feasible solution to the above problem. This then implies that the optimal solution $(t^\star,\boldsymbol{s}^\star)$ must satisfy $\theta t^\star+\frac{1}{N}\sum_{n\in[N]}s_n^\star\leq b$. Therefore, we have $t^\star\leq(b-\frac{1}{N}\sum_{n\in[N]}s_n^\star)/\theta\leq(b-a)/\theta$, concluding the finiteness of $t^\star$.

(*iii*) Suppose that $\Xi=\mathbb{R}^I$ and $f(\boldsymbol{x},\boldsymbol{\xi})$ is convex in $\boldsymbol{\xi}$ with a finite Lipschitz constant, $\mathrm{Lip}(f)$. Following from theorem 10 in Kuhn et al. (2019), we know that

$$\sup_{\mathbb{P}\in\mathcal{F}_{\mathrm{W}}(\theta)}\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})]=\mathbb{E}_{\hat{\mathbb{P}}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})]+\theta\cdot\mathrm{Lip}(f),$$

which implies that $t^\star=\mathrm{Lip}(f)$ is finite.

EC.5. Worst-Case Distributions

PROPOSITION EC.1. *Given assumptions in Proposition 1, the optimal value of the problem*

$$\sup_{\mathbb{P}\in\mathcal{P}(\Xi),\,\mathbb{Q}\in\mathcal{F}_{\mathrm{W}}(\theta)} \left\{\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})] - \gamma\cdot d_{\mathrm{W}}(\mathbb{P},\mathbb{Q})\right\} \tag{EC.8}$$

coincides with the optimal value of the following finite-dimensional convex program

$$\begin{array}{lll} \sup & \displaystyle\frac{1}{N}\sum_{n\in[N]}\sum_{k\in[K]}\alpha_{nk}\cdot f_k\Big(\boldsymbol{x},\hat{\boldsymbol{\xi}}_n - \frac{\boldsymbol{v}_{nk}}{\alpha_{nk}}\Big) - \gamma\phi & \\ \text{s.t.} & \displaystyle\frac{1}{N}\sum_{n\in[N]}\sum_{k\in[K]}\|\boldsymbol{v}_{nk}\| \le \theta + \phi & \\ & \displaystyle\sum_{k\in[K]}\alpha_{nk} = 1 & \forall n\in[N] \\ & \displaystyle\hat{\boldsymbol{\xi}}_n - \frac{\boldsymbol{v}_{nk}}{\alpha_{nk}} \in \Xi & \forall n\in[N],\, k\in[K] \\ & \phi\in\mathbb{R}_+,\ \alpha_{nk}\in\mathbb{R}_+,\ \boldsymbol{v}_{nk}\in\mathbb{R}^I & \forall n\in[N],\, k\in[K]. \end{array} \tag{EC.9}$$

Let $\{(\phi^\star(r),\boldsymbol{\alpha}^\star(r),\boldsymbol{v}^\star(r))\}_{r\in\mathbb{N}}$ be a sequence of solutions that attains the supremum of problem (EC.9) *and define correspondingly for each $r\in\mathbb{N}$, $n\in[N]$ and $k\in[K]$:*

$$\boldsymbol{p}^\star_{nk}(r) = \hat{\boldsymbol{\xi}}_n - \frac{\boldsymbol{v}^\star_{nk}(r)}{\alpha^\star_{nk}(r)} \quad \text{and} \quad \boldsymbol{q}^\star_{nk}(r) = \hat{\boldsymbol{\xi}}_n - \left(1 - \frac{N\phi^\star(r)}{\sum\limits_{n\in[N]}\sum\limits_{k\in[K]}\|\boldsymbol{v}^\star_{nk}(r)\|}\right)\cdot\frac{\boldsymbol{v}^\star_{nk}(r)}{\alpha^\star_{nk}(r)}.$$

Then the sequence of discrete distributions $\{(\mathbb{P}^\star(r),\mathbb{Q}^\star(r))\}_{r\in\mathbb{N}}$ with

$$\mathbb{P}^\star(r) = \frac{1}{N}\sum_{n\in[N]}\sum_{k\in[K]}\alpha^\star_{nk}(r)\cdot\delta_{\boldsymbol{p}^\star_{nk}} \quad \text{and} \quad \mathbb{Q}^\star(r) = \frac{1}{N}\sum_{n\in[N]}\sum_{k\in[K]}\alpha^\star_{nk}(r)\cdot\delta_{\boldsymbol{q}^\star_{nk}}$$

attains the supremum of problem (EC.8) *asymptotically. Here, $0/0=0$ and $a/0=\infty$ for any $a\neq 0$.*

Proof of Proposition EC.1. Introducing dual variables ϕ, α_{nk}, and β_{nk}, the Lagrangian dual of (10), after eliminating the original variables $\boldsymbol{s}$ and t, is given by

$$\begin{array}{lll} \sup & \displaystyle\sum_{n\in[N]}\sum_{k\in[K]}\inf_{\boldsymbol{w}_{nk},\boldsymbol{u}_{nk}}\left\{\beta_{nk}\|\boldsymbol{w}_{nk}\|_* + \alpha_{nk}\big([-f_k]^{\bar{*}}(\boldsymbol{x},\boldsymbol{u}_{nk}-\boldsymbol{w}_{nk}) + \delta^*(\boldsymbol{u}_{nk}\,|\,\Xi)\big) - \boldsymbol{w}_{nk}^\top\hat{\boldsymbol{\xi}}_n\big)\right\} - \gamma\phi & \\ \text{s.t.} & \displaystyle\sum_{n\in[N]}\sum_{k\in[K]}\beta_{nk}\le\theta+\phi & \\ & \displaystyle\sum_{k\in[K]}\alpha_{nk} = \frac{1}{N} & \forall n\in[N] \\ & \phi\ge 0,\ \alpha_{nk}\ge 0,\ \beta_{nk}\ge 0 & \forall n\in[N],\, k\in[K]. \end{array} \tag{EC.10}$$

By definition of dual norm and the minimax theorem, the objective of (EC.10) becomes

$$\sum_{n\in[N]}\sum_{k\in[K]}\sup_{\|\boldsymbol{v}_{nk}\|\le\beta_{nk}}\inf_{\boldsymbol{w}_{nk},\boldsymbol{u}_{nk}}\left\{\boldsymbol{w}_{nk}^\top\boldsymbol{v}_{nk} + \alpha_{nk}\big([-f_k]^{\bar{*}}(\boldsymbol{x},\boldsymbol{u}_{nk}-\boldsymbol{w}_{nk}) + \delta^*(\boldsymbol{u}_{nk}\,|\,\Xi)\big) - \boldsymbol{w}_{nk}^\top\hat{\boldsymbol{\xi}}_n\big)\right\} - \gamma\phi.$$

With this objective function and after eliminating variables β_{nk}, problem (EC.10) becomes

$$\begin{array}{ll} \sup & \displaystyle\sum_{n\in[N]}\sum_{k\in[K]} \inf_{\boldsymbol{w}_{nk},\boldsymbol{u}_{nk}} \left\{ \boldsymbol{w}_{nk}^\top \boldsymbol{v}_{nk} + \alpha_{nk}\big([-f_k]^{\bar{*}}(\boldsymbol{x},\boldsymbol{u}_{nk}-\boldsymbol{w}_{nk}) + \delta^*(\boldsymbol{u}_{nk}\mid\Xi) - \boldsymbol{w}_{nk}^\top\hat{\boldsymbol{\xi}}_n\big)\right\} - \gamma\phi \\ \text{s.t.} & \displaystyle\sum_{n\in[N]}\sum_{k\in[K]} \|\boldsymbol{v}_{nk}\| \le \theta+\phi \\ & \displaystyle\sum_{k\in[K]} \alpha_{nk} = \frac{1}{N} \qquad\qquad \forall n\in[N] \\ & \phi\ge 0,\ \alpha_{nk}\ge 0,\ \boldsymbol{v}_{nk}\in\mathbb{R}^I \quad \forall n\in[N],\ k\in[K]. \end{array}$$

For any $n\in[N]$ and $k\in[K]$, the inner minimization in the objective can be further simplified into

$$\begin{aligned} &\inf_{\boldsymbol{w}_{nk}} \left\{ \boldsymbol{w}_{nk}^\top \boldsymbol{v}_{nk} + \alpha_{nk}\big([-f_k]^{\bar{*}}(\boldsymbol{x},\boldsymbol{u}_{nk}-\boldsymbol{w}_{nk}) + \delta^*(\boldsymbol{u}_{nk}\mid\Xi) - \boldsymbol{w}_{nk}^\top\hat{\boldsymbol{\xi}}_n\big)\right\} \\ =\; & -\alpha_{nk}\cdot\sup_{\boldsymbol{w}_{nk}} \left\{ \boldsymbol{w}_{nk}^\top\Big(\hat{\boldsymbol{\xi}}_n - \frac{\boldsymbol{v}_{nk}}{\alpha_{nk}}\Big) - [-f_k]^{\bar{*}}(\boldsymbol{x},\boldsymbol{u}_{nk}-\boldsymbol{w}_{nk}) - \delta^*(\boldsymbol{u}_{nk}\mid\Xi)\right\} \\ =\; & \alpha_{nk}\cdot f_k\Big(\boldsymbol{x},\hat{\boldsymbol{\xi}}_n - \frac{\boldsymbol{v}_{nk}}{\alpha_{nk}}\Big) - \alpha_{nk}\cdot\delta\Big(\hat{\boldsymbol{\xi}}_n - \frac{\boldsymbol{v}_{nk}}{\alpha_{nk}}\;\Big|\;\Xi\Big), \end{aligned}$$

for any $\boldsymbol{u}_{nk}$, where the last equality follows from the fact that the bi-conjugate (*i.e.*, the conjugate of conjugate) of any proper, convex and lower semicontinuous function is the original function. Reexpressing the indicator functions into explicit hard constraints and applying the variable substitutions $(\alpha_{nk},\boldsymbol{v}_{nk})/N \leftarrow (\alpha_{nk},\boldsymbol{v}_{nk})$, we obtain the desired reformulation (EC.9).

We next investigate the sequence of worst-case distributions. Since it is clear that $\mathbb{P}^\star(r)\in\mathcal{P}(\Xi)$, we focus on showing that $\mathbb{Q}^\star(r)\in\mathcal{F}_{\mathrm{W}}(\theta)$, *i.e.*, $d_{\mathrm{W}}(\mathbb{Q}^\star(r),\hat{\mathbb{P}})\le\theta$. Firstly, consider a joint distribution π_r of $(\tilde{\boldsymbol{\xi}},\tilde{\boldsymbol{\zeta}})$ such that $\tilde{\boldsymbol{\xi}}\sim\mathbb{P}^\star(r)$, $\tilde{\boldsymbol{\zeta}}\sim\mathbb{Q}^\star(r)$, and $\pi_r(\boldsymbol{p}_{nk}^\star(r),\boldsymbol{q}_{nk}^\star(r)) = \alpha_{nk}^\star/N$ for all $n\in[N]$ and $k\in[K]$, we then obtain an upper bound of $d_{\mathrm{W}}(\mathbb{P}^\star(r),\mathbb{Q}^\star(r))$ as follows:

$$d_{\mathrm{W}}(\mathbb{P}^\star(r),\mathbb{Q}^\star(r)) \;\le\; \int_{\Xi\times\Xi}\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|\,\mathrm{d}\pi_r(\boldsymbol{\xi},\boldsymbol{\zeta}) \;=\; \frac{1}{N}\sum_{n\in[N]}\sum_{k\in[K]}\alpha_{nk}^\star(r)\|\boldsymbol{p}_{nk}^\star(r)-\boldsymbol{q}_{nk}^\star(r)\| \;=\; \phi^\star(r).$$

Secondly, consider a joint distribution π_r' of $(\tilde{\boldsymbol{\zeta}},\tilde{\boldsymbol{\eta}})$ such that $\tilde{\boldsymbol{\zeta}}\sim\mathbb{Q}^\star(r)$, $\tilde{\boldsymbol{\eta}}\sim\hat{\mathbb{P}}$, and $\pi_r'(\boldsymbol{q}_{nk}^\star(r),\hat{\boldsymbol{\xi}}_n) = \alpha_{nk}^\star/N$ for all $n\in[N]$ and $k\in[K]$, we have

$$\begin{aligned} d_{\mathrm{W}}(\mathbb{Q}^\star(r),\hat{\mathbb{P}}) \le \int_{\Xi\times\Xi}\|\boldsymbol{\zeta}-\boldsymbol{\eta}\|\,\mathrm{d}\pi_r'(\boldsymbol{\zeta},\boldsymbol{\eta}) &= \frac{1}{N}\sum_{n\in[N]}\sum_{k\in[K]}\alpha_{nk}^\star\|\boldsymbol{q}_{nk}^\star(r)-\hat{\boldsymbol{\xi}}_n\| \\ &= \frac{1}{N}\sum_{n\in[N]}\sum_{k\in[K]}\alpha_{nk}^\star\left(\|\boldsymbol{p}_{nk}^\star(r)-\hat{\boldsymbol{\xi}}_n\| - \|\boldsymbol{p}_{nk}^\star(r)-\boldsymbol{q}_{nk}^\star(r)\|\right) \\ &= \frac{1}{N}\sum_{n\in[N]}\sum_{k\in[K]}\|\boldsymbol{u}_{nk}^\star(r)\| - \phi^\star(r) \\ &\le \theta, \end{aligned}$$

where the second identity follows from the definitions of $\boldsymbol{p}_{nk}^\star(r)$ and $\boldsymbol{q}_{nk}^\star(r)$. Hence, $\mathbb{Q}^\star(r) \in \mathcal{F}_{\mathrm{W}}(\theta)$. Finally, it now holds that

$$\begin{aligned}\sup_{\mathbb{P}\in\mathcal{P}(\Xi),\,\mathbb{Q}\in\mathcal{F}_{\mathrm{W}}(\theta)} \left\{\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})] - \gamma\cdot d_{\mathrm{W}}(\mathbb{P},\mathbb{Q})\right\} &\geq \limsup_{r\to\infty}\left\{\mathbb{E}_{\mathbb{P}^\star(r)}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})] - \gamma\cdot d_{\mathrm{W}}(\mathbb{P}^\star(r),\mathbb{Q}^\star(r))\right\}\\ &\geq \limsup_{r\to\infty}\left\{\frac{1}{N}\sum_{n\in[N]}\sum_{k\in[K]}\alpha_{nk}^\star(r)\cdot f(\boldsymbol{x},\boldsymbol{p}_{nk}^\star(r)) - \gamma\phi^\star(r)\right\}\\ &\geq \limsup_{r\to\infty}\left\{\frac{1}{N}\sum_{n\in[N]}\sum_{k\in[K]}\alpha_{nk}^\star(r)\cdot f_k(\boldsymbol{x},\boldsymbol{p}_{nk}^\star(r)) - \gamma\phi^\star(r)\right\}\\ &= \sup_{\mathbb{P}\in\mathcal{P}(\Xi),\,\mathbb{Q}\in\mathcal{F}_{\mathrm{W}}(\theta)}\left\{\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})] - \gamma\cdot d_{\mathrm{W}}(\mathbb{P},\mathbb{Q})\right\},\end{aligned}$$

where the last line follows from the construction of $\{(\phi^\star(r),\, \alpha_{nk}^\star(r),\, \boldsymbol{p}_{nk}^\star(r))\}_{r\in\mathbb{N}}$. □

Intuitively, when $\gamma\to\infty$, the optimal solution in problem (EC.9) shall satisfy $\phi^\star(r)=0$, yielding $\mathbb{Q}^\star(r)=\mathbb{P}^\star(r)=\frac{1}{N}\sum_{n\in[N]}\sum_{k\in[K]}\alpha_{nk}^\star(r)\cdot\delta_{\boldsymbol{p}_{nk}^\star}$. This is consistent with the fact that as $\gamma\to\infty$,

$$\sup_{\mathbb{P}\in\mathcal{P}(\Xi),\,\mathbb{Q}\in\mathcal{F}_{\mathrm{W}}(\theta)}\left\{\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})] - \gamma\cdot d_{\mathrm{W}}(\mathbb{P},\mathbb{Q})\right\}\leq 0 \;\to\; \sup_{\mathbb{P}\in\mathcal{F}_{\mathrm{W}}(\theta)}\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})]\leq 0.$$

In such cases, the discrete distributions $\{\mathbb{P}^\star(r)\}_{r\in\mathbb{N}}$ belong to $\mathcal{F}_{\mathrm{W}}(\theta)$ and asymptotically attain the supremum $\sup_{\mathbb{P}\in\mathcal{F}_{\mathrm{W}}(\theta)}\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{\xi}})]$, recovering theorem 4.4 of Mohajerin Esfahani and Kuhn (2018).

EC.6. Reformulations under Probability Metrics

For the sake of tractability, we also assume here that the space $\mathcal{P}(\Xi)$ being of distributions that are absolutely continuous with respect to the empirical distribution $\hat{\mathbb{P}}$.

Example EC.1 (Fortet-Mourier metric). *Suppose that $\mathcal{L}_1=\mathcal{L}_2=\{\ell \mid \|\ell\|_{\mathrm{C}}\leq 1\}$ and that all distributions in $\mathcal{P}(\Xi)$ are absolutely continuous with respect to $\hat{\mathbb{P}}$. Then the corresponding globalized distributionally robust counterpart* (15) *is satisfiable if and only if*

$$\begin{cases} t\theta + \max_{n\in[N]}\left\{f(\boldsymbol{x},\hat{\boldsymbol{\xi}}_n) - u_n\right\} + \dfrac{1}{N}\displaystyle\sum_{n\in[N]} u_n \leq 0 \\ |u_m - u_n| \leq t\cdot\max\{1, \|\hat{\boldsymbol{\xi}}_m\|^{p-1}, \|\hat{\boldsymbol{\xi}}_n\|^{p-1}\}\cdot\|\hat{\boldsymbol{\xi}}_m - \hat{\boldsymbol{\xi}}_n\| \quad \forall m,\, n\in[N] \\ t\in[0,\gamma]. \end{cases}$$

Example EC.2 (Total variation metric). *Suppose that $\mathcal{L}_1=\mathcal{L}_2=\{\ell \mid \|\ell\|_\infty\leq 1\}$ and $\theta\in(0,2)$*[7]*, and that all distributions in $\mathcal{P}(\Xi)$ are absolutely continuous with respect to $\hat{\mathbb{P}}$. Then the corresponding globalized distributionally robust counterpart* (15) *is satisfiable if and only if*

$$\begin{cases} t\theta + \max_{n\in[N]}\left\{f(\boldsymbol{x},\hat{\boldsymbol{\xi}}_n) - u_n\right\} + \dfrac{1}{N}\displaystyle\sum_{n\in[N]} u_n \leq 0 \\ |u_n| \leq t \qquad \forall n\in[N] \\ t\in[0,\gamma]. \end{cases}$$

[7] We assume that $\theta\in(0,2)$ because the total variance metric between any two probability distributions $\mathbb{P}$ and $\mathbb{Q}$ satisfies $d_{\mathcal{L}_2}(\mathbb{P},\mathbb{Q}) = \sup_{\ell\in\mathcal{L}_2}\left\{\mathbb{E}_{\mathbb{P}}[\ell(\tilde{\boldsymbol{\xi}})] - \mathbb{E}_{\mathbb{Q}}[\ell(\tilde{\boldsymbol{\xi}})]\right\} \leq \sup_{\ell\in\mathcal{L}_2}\mathbb{E}_{\mathbb{P}}[\ell(\tilde{\boldsymbol{\xi}})] + \sup_{\ell\in\mathcal{L}_2}\mathbb{E}_{\mathbb{Q}}[\ell(\tilde{\boldsymbol{\xi}})] = 2$.