Robust Satisficing

Daniel Zhuoyu Long
Department of Systems Engineering & Engineering Management, The Chinese University of Hong Kong,
zylong@se.cuhk.edu.hk

Melvyn Sim
Department of Analytics & Operations (DAO), NUS Business School, National University of Singapore
melvynsim@gmail.com

Minglong Zhou
Department of Management Science, School of Management, Fudan University
minglong_zhou@u.nus.edu

We present a general framework for *robust satisficing* that favors solutions for which a risk-aware objective function would best attain an acceptable target even when the actual probability distribution deviates from the empirical distribution. The satisficing decision maker specifies an acceptable target, or loss of optimality compared to the empirical optimization model, as a trade off for the model's ability to withstand greater uncertainty. We axiomatize the decision criterion associated with robust satisficing, termed as the *fragility measure*, and present its representation theorem. Focusing on Wasserstein distance measure with ℓ_1-norm, we present tractable robust satisficing models for risk-based linear optimization, combinatorial optimization, and linear optimization problems with recourse. Serendipitously, the insights to the approximation of the linear optimization problems with recourse also provide a recipe for approximating solutions for hard stochastic optimization problems without relatively complete recourse. We perform numerical studies on a portfolio optimization problem and a network lot-sizing problem. We show that the solutions to the robust satisficing models are more effective in improving the out-of-sample performance evaluated on a variety of metrics, hence alleviating the *Optimizer's Curse* (Smith and Winkler 2006).

Key words: Robust optimization, robust satisficing, data-driven, discrete optimization, stochastic optimization, fragility measure

History: October 26, 2021.

1. Introduction

"Contentment is the Greatest Wealth." – The Buddha

Optimization under uncertainty, despite its importance and ubiquity in real-world problems, has been a perennial difficulty in prescriptive analytics. Although empirical distribution from historical data may be available as input to the optimization model, depending on the length of the planning horizon, the sample size could be rather limited. For instance, for a one year worth of historical

data, an optimization problem with a one-week planning horizon may avail to at most 52 empirical samples, which may be insufficient for characterizing the uncertain outcomes in future weeks to some desired accuracy. Conceivably, the problem would be aggravated in a non-stationary uncertain environment, such as in events of pandemic, where some parts of the history may no longer be relevant for describing future uncertainty. The aphorism in statistics,

"All models are wrong" – Box (1976)

underscores the limitation of statistical models. Despite the best statistical machinery available, the fact remains that we are still incapable of obtaining the true distribution that Nature uses to generate the historical samples and future ones. Hence, in real-world data-driven optimization problems, we must live with the fact that the actual realized objective function may not necessarily be the same as the objective function that the model is optimizing. In fact, if we solve an optimization problem using an empirical distribution from a dataset and test the out-of-sample outcome on another, we should also expect inferior results, a phenomenon observed by Smith and Winkler (2006), who coin the term *Optimizer's Curse*.

Similar bias in the model's objective has been observed in the context of approximating stochastic optimization problems via Sample Average Approximation (SAA) (Kleywegt et al. 2002). Stochastic optimization typically assumes knowledge of the data generating model, and hence, we can increase the accuracy of the problem simply by having more samples. In data-driven optimization, we do not have the luxury of having an arbitrary number of samples to help us improve the precision of the risk-aware objective function that the model is optimizing. In this paper, we are not focusing on the tractability issues, granted that if we have sufficient samples, SAA would converge well to the optimal solution, at least for a two-stage problem with relatively complete recourse (see, for instance, Shapiro and Nemirovski 2005). Problems without relatively complete recourse are much harder to solve and they are usually not discussed in the literature. In data-driven optimization, we are more concerned with alleviating the Optimizer's Curse, which is a more pertinent issue than tractability because of the potential scarcity of useful data to describe future uncertainty.

To address this issue, data-driven based robust optimization models have recently been proposed (see, for instance, Gao and Kleywegt 2016, Gao et al. 2017, Bertsimas et al. 2018a,b, Mohajerin Esfahani and Kuhn 2018). Instead of evaluating the risk-aware objective using the empirical distribution, the robust optimization model evaluates the worst-case risk-aware objective over a restricted *ambiguity set* of probability distributions within the vicinity of the empirical distribution. Mohajerin Esfahani and Kuhn (2018) show that with an appropriate sized ambiguity set characterized by the Wassersetin distance metric, the ambiguity set is able to capture the true distribution with high level of confidence. If the data-generation model is known, robust optimization

can also be used effectively to improve policies in dynamic optimization problems (see, for instance, Sturt 2021). In practice, the size of the ambiguity set is not usually determined by the theoretical confidence bound of containing the true distribution. Instead, if sufficient samples are available, the size parameter determined by cross-validation techniques would provide better out-of-sample performance.

In this paper, we introduce a new target-oriented model for robust data-driven optimization termed as *robust satisficing*. The philosophical concept of robust satisficing has been discussed in Schwartz et al. (2011), where, in contrast to utility maximization, the goal of robust satisficing is to maximize the robustness to uncertainty of achieving a satisfactory target. Unlike robust optimization, our robust satisficing model do not restrict distributions to an ambiguity set, but allow Nature to take its course and, as much as possible, limit the impact of ambiguity on the risk-aware objective should the true distribution deviates from the empirical distribution. The decision maker specifies a target, or an acceptable loss of optimality compared to the empirical optimization model, as a trade-off for the model's ability to withstand greater uncertainty. As articulated by Simon (1959), target satisficing, as opposed to utility maximizing, is prevalent in human decision making, especially in complex situations facing risks and uncertainty (see, *e.g*, Mao 1970, Chen and Tang 2019). Hence, we believe that articulating preference for robustness from a target-driven perspective is also more interpretable for the decision maker, and may well apply in situations when data availability is limited for cross validation.

We axiomatize the decision criterion associated with the robust satisficing model, termed as the *fragility measure*, which relates to the maximum level of model infeasibility that may occur relative to the magnitude of deviation from the empirical distribution. A fragility measure satisfies the properties of *monotonicity, positive homogeneity, subadditivity, pro-robustness,* and *anti-fragility* and it belongs to a class of *satisficing* measures (Brown and Sim 2009, Brown et al. 2012). We provide a representation theorem and connect with known satisficing measures in the literature including the riskiness index of Aumann and Serrano (2008). Fragility measure is related to convex measure of risks characterized by Follmer and Schied (2004), among others. Ben-Tal et al. (2006) have explored the non-stochastic version of convex risk measure, which they term as *comprehensive robust optimization* or more recently as *globalized robust optimization* by Ben-Tal et al. (2017). Incidentally, to provide the parameters of the globalized robust optimization model, Ben-Tal et al. (2017) introduce a *GRC-sum* model, which coincides with our robust satisficing model under Wasserstein distance measure for the case of only one sample.

Focusing on Wasserstein distance measure with ℓ_1-norm, we present tractable robust satisficing models for risk-based linear optimization, combinatorial optimization, and linear optimization problems with recourse. Interestingly, the insights to the approximation of the linear optimization

problems with recourse also provide a recipe for solving hard stochastic optimization problems approximately. In our computational study, we illustrate in a portfolio optimization problem and a network lot-sizing problem on how we can set targets in the robust satisficing model, which can be more intuitive and effective than specifying the hyper-parameter used in a robust optimization model. The numerical studies show that the solutions to the robust satisficing models are more effective in alleviating the Optimizer's Curse, and they yield solutions with superior out-of-sample performance when evaluated on a variety of metrics.

Notation. We use boldface lowercase letters for vectors (*e.g.*, $\boldsymbol{\theta}$), and calligraphic letters for sets (*e.g.*, $\mathcal{X}$). We use $[N]$ to denote the running index $\{1, 2, 3, \ldots, N\}$ for N a known integer. We adopt the convention that $\inf \emptyset = +\infty$, where $\emptyset$ is the empty set. A random variable $\tilde{v}$ is denoted with a tilde sign such as $\tilde{v} \sim \mathbb{P}, \mathbb{P} \in \mathcal{P}_0$, where $\mathcal{P}_0$ to represent the set of all possible distributions. For $\tilde{v}_1, \tilde{v}_2$, we use $\tilde{v}_1 \geq \tilde{v}_2$ to imply $\tilde{v}_1$ state-wise dominates $\tilde{v}_2$, *i.e.*, $\mathbb{P}[\tilde{v}_1 \geq \tilde{v}_2] = 1$ for all $\mathbb{P} \in \mathcal{P}_0$. For a multivariate random variable, we use $\mathcal{P}_0(\mathcal{Z})$ to represent the set of all distributions for the multivariate random variable that has support $\mathcal{Z} \subseteq \mathbb{R}^N$. Specifically, we use $\tilde{\boldsymbol{z}} \sim \mathbb{P}$, $\mathbb{P} \in \mathcal{P}_0(\mathcal{Z})$ to define $\tilde{\boldsymbol{z}}$ as a multivariate random variable with support $\mathcal{Z}$ and distribution $\mathbb{P}$. We use $\mathbb{E}_{\mathbb{P}}[\tilde{v}]$ to denote expectation of a random variable, $\tilde{v} \sim \mathbb{P}$ over its distribution.

2. Data-driven Optimization Models

We consider an optimization problem with decision variable $\boldsymbol{x} \in \mathcal{X}$, and objective function $f(\boldsymbol{x}, \boldsymbol{z}) : \mathcal{X} \times \mathcal{Z} \mapsto \mathbb{R}$, where the input to the second argument is subject to uncertainty. The uncertain parameters of the problem are collectively denoted by the random variable $\tilde{\boldsymbol{z}}$ over the support $\mathcal{Z} \subseteq \mathbb{R}^N$, which is generally convex and bounded. Ideally, we should solve the following risk-neutral optimization problem,

$$
\begin{aligned}
Z^* = \min \ & \mathbb{E}_{\mathbb{P}^*}[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})] \\
\text{s.t.} \ & \boldsymbol{x} \in \mathcal{X},
\end{aligned}
\tag{1}
$$

where $\mathbb{P}^* \in \mathcal{P}_0(\mathcal{Z})$, $\tilde{\boldsymbol{z}} \sim \mathbb{P}^*$ is the true distribution that characterizes future outcomes. In management decision problems, it is ubiquitous to consider risk-neural objective function, though we can also consider other common risk-aware criteria such as optimized certainty equivalent of Ben-Tal and Teboulle (2007) and expected utilities.

Empirical optimization

In practice, we generally do not know $\mathbb{P}^*$, though we have some information on how the uncertain parameters have evolved from historical records. Let $\hat{\mathbb{P}} \in \mathcal{P}_0(\mathcal{Z})$ denote the empirical distribution constructed from historical data, *i.e.*, $\hat{\mathbb{P}}[\tilde{\boldsymbol{z}} = \hat{\boldsymbol{z}}_s] = 1/S$ and $\hat{\boldsymbol{z}}_s \in \mathbb{R}^N$ represents a realized data record under scenario $s \in [S]$. We first consider the following empirical optimization problem,

$$
\begin{aligned}
Z_0 = \min \ & \mathbb{E}_{\hat{\mathbb{P}}}[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})] \\
\text{s.t.} \ & \boldsymbol{x} \in \mathcal{X}.
\end{aligned}
\tag{2}
$$

Here, we distinguish the empirical optimization problem from the stochastic optimization model, although they coincide when the empirical distribution is the same as the true distribution. In practice, we would not know the true distribution $\mathbb{P}^* \in \mathcal{P}_0(\mathcal{Z})$ and the distribution $\mathbb{P}^\dagger \in \mathcal{P}_0(\mathcal{Z})$ used in stochastic optimization is restricted to a particular family whose parameters are estimated from historical data, *e.g.*, the empirical distribution, $\hat{\mathbb{P}} \in \mathcal{P}_0(\mathcal{Z})$. For example, if the distribution is restricted to the Gaussian family, then $\mathbb{P}^\dagger = \mathcal{N}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$, where the mean $\hat{\boldsymbol{\mu}}$ and the covariance matrix $\hat{\boldsymbol{\Sigma}}$ are estimated from the historical data. Stochastic optimization solves for the following optimization problem

$$
\begin{aligned}
Z^\dagger = \min \; & \mathbb{E}_{\mathbb{P}^\dagger}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right] \\
\text{s.t.} \; & \boldsymbol{x} \in \mathcal{X},
\end{aligned}
\tag{3}
$$

which does not necessarily yield the optimal risk-aware objective value, since the true distribution $\mathbb{P}^*$ would likely deviate from $\mathbb{P}^\dagger$. Nevertheless, the benefit of stochastic optimization is the ability of leveraging advanced statistical tools to determine $\mathbb{P}^\dagger$ so that the true distribution $\mathbb{P}^*$ would be statistically closer to $\mathbb{P}^\dagger$ than to the empirical distribution $\hat{\mathbb{P}}$. Nevertheless, in the special case where $f(\boldsymbol{x}, \boldsymbol{z})$ is an affine function of $\boldsymbol{z}$, and $\mathbb{E}_{\mathbb{P}^\dagger}[\tilde{\boldsymbol{z}}] = \mathbb{E}_{\hat{\mathbb{P}}}[\tilde{\boldsymbol{z}}] = \hat{\boldsymbol{\mu}}$, then the empirical optimization problem would coincide with the stochastic optimization model.

Unlike the empirical optimization problem, it may be significantly harder to evaluate the risk-aware objective function in stochastic optimization (see, for example, Nemirovski and Shapiro 2007, Hanasusanto et al. 2016). Nevertheless, for many practical problems, stochastic optimization can be well approximated via Sample Average Approximation (SAA), which is the same format as the empirical optimization problem, though we can arbitrarily increase the number of samples generated from the assumed distribution $\mathbb{P}^\dagger$ in stochastic optimization to improve the approximation.

The Optimizer's Curse

It has well been known that solving the optimization model with the empirical distribution may yield solutions that perform poorly against the true distribution (see, for instance, Smith and Winkler 2006, Kleywegt et al. 2002). To see this, let $\mathbb{P}^*$ represent the true data generating distribution for which the empirical distribution is obtained, *i.e.*, $\tilde{\boldsymbol{z}}_s \sim \mathbb{P}^*$, $s \in [S]$. The empirical optimization problem would yield a random objective function as follows

$$
Z_0(\tilde{\boldsymbol{z}}_1, \ldots, \tilde{\boldsymbol{z}}_S) = \min_{\boldsymbol{x} \in \mathcal{X}} \frac{1}{S} \sum_{s \in [S]} f(\boldsymbol{x}, \tilde{\boldsymbol{z}}_s).
$$

Denote $\mathbb{P}^S$ as the joint distribution of $\tilde{\boldsymbol{z}}_1, \ldots, \tilde{\boldsymbol{z}}_S$. Note that for any $\boldsymbol{x} \in \mathcal{X}$, observe that

$$
\mathbb{E}_{\mathbb{P}^S}\left[Z_0(\tilde{\boldsymbol{z}}_1, \ldots, \tilde{\boldsymbol{z}}_S)\right] \le \mathbb{E}_{\mathbb{P}^S}\left[\frac{1}{S} \sum_{s \in [S]} f(\boldsymbol{x}, \tilde{\boldsymbol{z}}_s)\right] = \frac{1}{S} \sum_{s \in [S]} \mathbb{E}_{\mathbb{P}^*}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}}_s)\right] = \mathbb{E}_{\mathbb{P}^*}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right].
$$

Hence,

$$\mathbb{E}_{\mathbb{P}^S}\left[Z_0(\tilde{\boldsymbol{z}}_1,\ldots,\tilde{\boldsymbol{z}}_S)\right] \leq \min_{\boldsymbol{x}\in\mathcal{X}}\mathbb{E}_{\mathbb{P}^*}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right],$$

though the gap may diminish with larger number of samples (see Shapiro and Nemirovski 2005). Nevertheless, under limited availability of data, the expected model's objective value may be significantly biased, reflecting an over-optimistic objective value that would never be attainable at all under the true distribution. As the Optimizer's Curse (Smith and Winkler 2006) states, the objective function reflected by the empirical optimization problem is unreliable, and we expect the actual outcome to be worse off.

Robust optimization

To address this issue, robust optimization aims to reduce over-fitting to the historical dataset by featuring a probability-distance-based ambiguity set of possible distributions, $\mathcal{B}(r)$, defined as

$$\mathcal{B}(r) := \left\{\mathbb{P}\in\mathcal{P}_0(\mathcal{Z}) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}}\sim\mathbb{P} \\ \Delta(\mathbb{P},\hat{\mathbb{P}})\leq r \end{array}\right\},$$

where $\Delta(\mathbb{P},\hat{\mathbb{P}})$ is a probability distance of the actual distribution $\mathbb{P}$ from the empirical distribution $\hat{\mathbb{P}}$ as defined in the following.

DEFINITION 1 (PROBABILITY DISTANCE FUNCTION). A probability distance function, $\Delta(\mathbb{P},\hat{\mathbb{P}})$ is a nonnegative function on the domain of probability distributions such that $\Delta(\mathbb{P},\hat{\mathbb{P}})=0$ if $\mathbb{P}=\hat{\mathbb{P}}$.

The parameter r controls the size of ambiguity set by limiting the probability distance that can deviates from the empirical distribution. Under this definition, the probability-distance-based ambiguity sets would include ϕ-divergence (Pardo 2006, Ben-Tal et al. 2013, Gotoh et al. 2020, 2018) and Wasserstein metric, also known as Kantorovich–Rubinstein metric (Gao and Kleywegt 2016, Mohajerin Esfahani and Kuhn 2018). The choice of ambiguity sets varies in different applications. In data-driven optimization, a popular ambiguity set is constructed via the Wasserstein metric (of type-1), defined as follows:

$$\Delta_W(\mathbb{P},\hat{\mathbb{P}}) := \inf_{\mathbb{Q}\in\mathcal{P}_0(\mathcal{Z}^2)}\left\{\mathbb{E}_{\mathbb{Q}}\left[\|\tilde{\boldsymbol{z}}-\tilde{\boldsymbol{v}}\|\right] \;\middle|\; (\tilde{\boldsymbol{z}},\tilde{\boldsymbol{v}})\sim\mathbb{Q}, \tilde{\boldsymbol{z}}\sim\mathbb{P}, \tilde{\boldsymbol{v}}\sim\hat{\mathbb{P}}\right\}.$$

For a fixed size, r, the *data-driven robust optimization* model can be written as the following:

$$Z_r = \min_{\mathbb{P}\in\mathcal{B}(r)}\sup \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right] \tag{4}$$
$$\text{s.t. } \boldsymbol{x}\in\mathcal{X}.$$

Under the Wasserstein metric and assuming that $\mathcal{Z}$ is convex and closed, the above robust optimization model admits an equivalent robust optimization formulation (Mohajerin Esfahani and Kuhn 2018):

$$
\begin{aligned}
Z_r = \min \ & kr + \frac{1}{S} \sum_{s \in [S]} y_s \\
\text{s.t.} \ & y_s \geq \sup_{\boldsymbol{z}_s \in \mathcal{Z}} \{ f(\boldsymbol{x}, \boldsymbol{z}_s) - k \|\boldsymbol{z}_s - \hat{\boldsymbol{z}}_s\| \} \quad \forall s \in [S] \\
& \boldsymbol{x} \in \mathcal{X}, \ k \geq 0.
\end{aligned}
\tag{5}
$$

It has well been known that if $f(\boldsymbol{x}, \boldsymbol{z})$ can be expressed as the maximum of a finite set of tractable saddle functions, *i.e*, functions that are convex in $\boldsymbol{x}$ and concave in $\boldsymbol{z}$, then Problem (5) can also be expressed as a concise convex optimization problem via their convex conjugate functions, support function of $\mathcal{Z}$ and the dual norm of $\|\cdot\|$ (see, for instance, Ben-Tal et al. 2015). Mohajerin Esfahani and Kuhn (2018) provide the following explicit example.

PROPOSITION 1. **(Mohajerin Esfahani and Kuhn 2018).** *Suppose the support set is a polyhedral set given by* $\mathcal{Z} = \{\boldsymbol{z} \in \mathbb{R}^N \mid \boldsymbol{C}\boldsymbol{z} \leq \boldsymbol{h}\}$, *the data-driven robust optimization model* (5) *with* $f(\boldsymbol{x}, \boldsymbol{z}) := \max_{i \in [I]} \{a_i \boldsymbol{x}^\top \boldsymbol{z} + b_i\}$ *admits the following equivalent reformulation:*

$$
\begin{aligned}
\min \ & kr + \frac{1}{S} \sum_{s \in [S]} y_s \\
\text{s.t.} \ & y_s \geq a_i \boldsymbol{x}^\top \hat{\boldsymbol{z}}_s + b_i + \boldsymbol{\eta}_{is}^\top \boldsymbol{h} - \boldsymbol{\eta}_{is}^\top \boldsymbol{C} \hat{\boldsymbol{z}}_s, \quad \forall i \in [I], \ s \in [S] \\
& k \geq \|a_i \boldsymbol{x} - \boldsymbol{C}^\top \boldsymbol{\eta}_{is}\|_*, \qquad\qquad \forall i \in [I], \ s \in [S] \\
& \boldsymbol{\eta}_{is} \geq \boldsymbol{0} \qquad\qquad\qquad\qquad \forall i \in [I], \ s \in [S] \\
& \boldsymbol{x} \in \mathcal{X}.
\end{aligned}
$$

Incidentally, if $S = 1$ and $f(\boldsymbol{x}, \boldsymbol{z})$ is a saddle function, then by Sion et al. (1958) minmax principle, Problem (5) becomes the classical stochastic-free robust optimization problem

$$
\begin{aligned}
Z_r = \min \ & \max_{z \in \mathcal{U}(r)} f(\boldsymbol{x}, \boldsymbol{z}) \\
\text{s.t.} \ & \boldsymbol{x} \in \mathcal{X}
\end{aligned}
$$

where $\mathcal{U}(r)$ is a norm-based uncertainty set given by

$$
\mathcal{U}(r) = \{\boldsymbol{z} \in \mathcal{Z} \mid \|\boldsymbol{z} - \hat{\boldsymbol{z}}\| \leq r\},
$$

for some nominal value $\hat{\boldsymbol{z}}$. This is a very well studied problem in the literature of robust optimization (see, Bertsimas and Sim 2004, Ben-Tal and Nemirovski 1998, Ben-Tal et al. 2015, among others).

The rationale of solving robust optimization model is when the true distribution $\mathbb{P}^*$ lies within the ambiguity set $\mathcal{B}(r)$, then

$$
\min_{\boldsymbol{x} \in \mathcal{X}} \mathbb{E}_{\mathbb{P}^*}[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})] \leq Z_r,
$$

so that Z_r reflects an objective value that is at least achievable under the true probability distribution $\mathbb{P}^*$. Hence, for an adequately chosen parameter, r, the decision maker can expect the solution to attain an objective value that is at least as good as the robust optimization objective value, Z_r. However, if r is too large, then Z_r may deviate too far from the attainable objective. Therefore, the performance critically depends on the size parameter r, which may not be easy to interpret and specify by the decision maker. Without other assumptions on the underlying uncertainty, it is usually not intuitive for the decision maker to pin down the value of r for the robust optimization problem. Apart from computational tractability reasons, the consideration of the Wasserstein-based ambiguity set is also motivated by how we could relate r to some confidence guarantees that the true distribution, $\mathbb{P}^*$ would be contained within the ambiguity set, $\mathcal{B}(r)$. Fournier and Guillin (2015) provides such an estimate in the following result.

THEOREM 1. **(Fournier and Guillin 2015).** *Suppose the true data-generating distribution $\mathbb{P}^*$, $\tilde{\boldsymbol{z}} \sim \mathbb{P}^*$ is a light-tailed distribution such that*

$$\delta := \mathbb{E}_{\mathbb{P}^*}[\exp(\|\tilde{\boldsymbol{z}}\|^\alpha)] < \infty$$

for some $\alpha > 1$ and $\mathbb{P}^S$ is the distribution that governs the distribution of independent samples $\hat{\boldsymbol{z}}_1, \ldots, \hat{\boldsymbol{z}}_S$ drawn from $\mathbb{P}^*$. Then

$$\mathbb{P}^S\left[\Delta_W(\mathbb{P}^*, \hat{\mathbb{P}}) > r\right] \leq \begin{cases} c_1 \exp(-c_2 S r^{\max\{N,2\}}) & \text{if } 0 < r \leq 1, \\ c_1 \exp(-c_2 S r^\alpha) & \text{if } r > 1, \end{cases}$$

for some positive constants, c_1 and c_2 that only depend on α, δ, and N.

The bound shows intuitively that as the Wasserstein ambiguity set need not be too large to ensure with high confidence that the true probability distribution $\mathbb{P}^*$ is contained within it. In data-driven optimization, S is fixed, the parameters are hard to estimate and the generic probabilistic bound is expected to be loose. Hence, the bound is typically not used in practice to determine a desired value of r. Instead, the parameter r is a hyper-parameter that is determined via cross-validation, which may be difficult to perform if we only have a small data set.

Robust Satisficing

We now present an alternative model, which we call the *robust satisficing* model,

$$\begin{aligned} \kappa_\tau = \min\ & k \\ \text{s.t.}\ & \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})] - \tau \leq k\Delta(\mathbb{P}, \hat{\mathbb{P}}) \quad \forall \mathbb{P} \in \mathcal{P}_0(\mathcal{Z}) \\ & \boldsymbol{x} \in \mathcal{X}, \quad k \geq 0, \end{aligned} \tag{6}$$

where $\tau \geq Z_0$ represents a targeted level of cost, or an acceptable loss of optimality relative to the empirical optimization model. Since the actual distribution $\mathbb{P}^* \in \mathcal{P}_0(\mathcal{Z})$ is unknown and may deviate

from $\hat{\mathbb{P}}$, we do not expect the actual objective Z_0 to be achievable. Having a target higher than Z_0, provides more leeway to improve the robustness of the model. A violation from the desired target level occurs whenever $\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \boldsymbol{z})] > \tau$. As in the case of robust optimization models, the choice of the probability distance function depends on, among other things, how we could efficiently optimize the robust counterparts and, more importantly, how well the solutions would perform in empirical tests. Intuitively, the robust satisficing model minimizes the fragility of the system. Specifically, the fragility of the system is measured by the worst-case (over all possible distributions $\mathbb{P} \in \mathcal{P}_0(\mathcal{Z})$) magnitude of expected target violation under a distribution normalized by the statistical distance of this distribution from $\hat{\mathbb{P}}$. The optimal proportionality factor κ_τ effectively describes the level of fragility of the model. As κ_τ decreases in value, a lower magnitude of expected target violation could occur under any distribution.

Unlike robust optimization, the robust satisficing model considers all possible distributions in $\mathcal{P}_0(\mathcal{Z})$, while controlling the level of expected target violation under any distribution relative to the distance of this distribution from $\hat{\mathbb{P}}$. In other words, a larger level of expected target violation can be accepted when the distribution is further away from the empirical distribution $\hat{\mathbb{P}}$. Observe that the constraint of Problem (6) implies $\tau \geq \mathbb{E}_{\hat{\mathbb{P}}}[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})]$, ensuring that the cost function will meet the desired target over the empirical distribution.

An important difference from the robust optimization model is that the decision maker specifies the parameter τ, as opposed to setting the size of the ambiguity set r for the robust optimization model. The input parameter τ, $\tau \geq Z_0$, is directly related to the objective value of the model the decision maker is addressing, and it can be interpreted as the target objective that she is willing to accept, relative to a reference, *e.g.*, the empirical optimization problem's objective, Z_0. The model will then determine the most robust solution which also achieves the target objective. In practice, one may directly specify the proportionality factor τ/Z_0, naturally leading to a consistent range of model parameters in different model settings and applications. While one may argue that target τ is hard to specify, it should at least be more tangible and intuitive to specify compared to the parameter r used in the robust optimization problem (4). In the presence of sufficient data, we can also determine τ using cross-validation technique.

We emphasize that the robust satisficing model should not be misconstrued as a relaxation of robust optimization. For instance, when $\tau = Z_r$, the robust constraint in Problem (4) indicates

$$\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})] - \tau \leq 0 \qquad \forall \mathbb{P} \in \mathcal{B}(r)$$
$$\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})] - \tau \leq +\infty \quad \forall \mathbb{P} \in \mathcal{P}_0(\mathcal{Z}) \backslash \mathcal{B}(r),$$

while the robust satisficing constraint in Problem (6) implies that

$$\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})] - \tau \leq \kappa_\tau \Delta(\mathbb{P}, \hat{\mathbb{P}}) \quad \forall \mathbb{P} \in \mathcal{B}(r)$$
$$\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})] - \tau \leq \kappa_\tau \Delta(\mathbb{P}, \hat{\mathbb{P}}) \quad \forall \mathbb{P} \in \mathcal{P}_0(\mathcal{Z}) \backslash \mathcal{B}(r).$$

Hence, while the robust satisficing model accepts some additional losses when $\mathbb{P} \in \mathcal{B}(r)$, it can safeguard more severe losses when $\mathbb{P} \in \mathcal{P}_0(\mathcal{Z}) \backslash \mathcal{B}(r)$.

Similar to data-driven robust optimization models, we will mainly focus on using the Wasserstein probability distance, though in Section 3, we demonstrate that a robust satisficing model can be systematically defined using other probability distance functions. Having Wasserstein distance enables us to incorporate available data in the form of empirical distribution and, in many useful instances, formulate the robust satisficing problem as a tractable linear optimization problem. Moreover, unlike probability distance measures such as ϕ-divergence, the Wasserstein distance does not require the distribution in $\mathcal{P}_0(\mathcal{Z})$ to be absolutely continuous with respect to the empirical distribution, $\hat{\mathbb{P}}$. In our model, the robust satisficing constraint

$$\mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right] - \tau \le k\Delta(\mathbb{P}, \hat{\mathbb{P}}) \quad \forall \mathbb{P} \in \mathcal{P}_0(\mathcal{Z})$$

restricts the violation of target τ for all possible distributions on the support $\mathcal{Z}$. If we chose Δ as the ϕ-divergence, then $\Delta(\mathbb{P}, \hat{\mathbb{P}})$ is only well defined for $\mathbb{P}$ that is absolutely continuous with respect to $\hat{\mathbb{P}}$. Hence, we provide no performance guarantee if the true distribution is not absolutely continuous with respect to $\hat{\mathbb{P}}$. The empirical distribution $\hat{\mathbb{P}}$ is discrete and only takes S possible scenarios; hence, the set of distributions that are absolutely continuous with respect to $\hat{\mathbb{P}}$ could be rather limited. As we will show in Proposition 3, the choice of Wasserstein distance in the robust satisficing model can lead to promising statistical guarantee on the probability of target violation for all distributions on support set $\mathcal{Z}$.

Although we focus on the Wasserstein distance, we acknowledge that there are problem instances where other probability distance functions may provide more computationally tractable models than over Wasserstein distance. For instance, when the reference distribution $\hat{\mathbb{P}}$ is associated with a collection of independently distributed random variables, then the KL-divergence may sometimes lead to tractable models. Under the KL-divergence, the fragility measure relates to the riskiness index (Aumann and Serrano 2008), which has been utilized in many recent works (see *e.g.,* Hall et al. 2015, Jaillet et al. 2016). In those problem settings, Wasserstein distance would not provide a tractable optimization because of the exponential sample space.

Observe that the corresponding robust satisficing problem (6) under Wasserstein metric Δ_W can be equivalently written as

$$
\begin{aligned}
\kappa_\tau = \min~ & k \\
\text{s.t.}~ & \mathbb{E}_{\mathbb{Q}}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right] - \tau \le k\mathbb{E}_{\mathbb{Q}}\left[\|\tilde{\boldsymbol{z}} - \tilde{\boldsymbol{u}}\|\right] \quad \forall \mathbb{Q} \in \mathcal{Q} \\
& \boldsymbol{x} \in \mathcal{X}, \quad k \ge 0,
\end{aligned}
\tag{7}
$$

where the ambiguity set of the joint distribution is defined as

$$
\mathcal{Q} := \left\{ \mathbb{Q} \in \mathcal{P}_0(\mathcal{Z}^2) \;\middle|\; \begin{array}{l} (\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \sim \mathbb{Q} \\ \mathbb{Q}\left[\tilde{\boldsymbol{u}} = \hat{\boldsymbol{z}}_s\right] = 1/S, \forall s \in [S] \end{array} \right\}. \tag{8}
$$

When the support set $\mathcal{Z}$ is convex and closed, this model admits the following equivalent robust optimization reformulation:

$$
\begin{aligned}
\kappa_\tau = \min \; & k \\
\text{s.t.} \; & \frac{1}{S} \sum_{s \in [S]} y_s \leq \tau \\
& y_s \geq \sup_{\boldsymbol{z}_s \in \mathcal{Z}} \left\{ f(\boldsymbol{x}, \boldsymbol{z}_s) - k\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}_s\| \right\} \quad \forall s \in [S] \\
& \boldsymbol{x} \in \mathcal{X}, k \geq 0.
\end{aligned} \tag{9}
$$

which is remarkably similar to Problem (5), and both models have similar tractable reformulation if $f(\boldsymbol{x}, \boldsymbol{z})$ can be expressed as the maximum of a finite set of tractable saddle functions. As a comparison with Proposition 1, we have the following result.

PROPOSITION 2. *Suppose the support set is a polyhedral set $\mathcal{Z} = \{\boldsymbol{z} \in \mathbb{R}^N \mid \boldsymbol{C}\boldsymbol{z} \leq \boldsymbol{h}\}$, the robust satisficing model (7) with $f(\boldsymbol{x}, \boldsymbol{z}) := \max_{i \in [I]}\{a_i \boldsymbol{x}^\top \boldsymbol{z} + b_i\}$ admits the following equivalent reformulation:*

$$
\begin{aligned}
\min \; & k \\
\text{s.t.} \; & \frac{1}{S} \sum_{s \in [S]} y_s \leq \tau \\
& y_s \geq a_i \boldsymbol{x}^\top \hat{\boldsymbol{z}}_s + b_i + \boldsymbol{\eta}_{is}^\top \boldsymbol{h} - \boldsymbol{\eta}_{is}^\top \boldsymbol{C} \hat{\boldsymbol{z}}_s & \forall s \in [S], \; i \in [I] \\
& k \geq \|a_i \boldsymbol{x} - \boldsymbol{C}^\top \boldsymbol{\eta}_{is}\|_* & \forall s \in [S], \; i \in [I] \\
& \boldsymbol{\eta}_{is} \geq \boldsymbol{0} & \forall s \in [S], \; i \in [I] \\
& \boldsymbol{x} \in \mathcal{X}, \; k \geq 0.
\end{aligned}
$$

As the consequence of Theorem 1, we can formalize the confidence range of target attainment that can be associated with the objective of the robust satisficing problem as follows.

PROPOSITION 3. *Under the same assumption as in Theorem 1, let $\boldsymbol{x}$ and k be feasible in Problem (7), then*

$$
\mathbb{P}^S\left[\mathbb{E}_{\mathbb{P}^*}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right] > \tau + kr\right] \leq \begin{cases} c_1 \exp(-c_2 S r^{\max\{N,2\}}) & \text{if } 0 < r \leq 1, \\ c_1 \exp(-c_2 S r^\alpha) & \text{if } r > 1, \end{cases}
$$

for some positive constants, c_1 and c_2 that only depend on α, δ, and N.

Therefore, the solution of the robust satisficing problem ensures a high confidence that the true risk neutral objective function is within kr from the specified target, such that the probability of exceeding the confidence range decreases exponentially in r. Hence, regardless of the parameters

c_1, c_2, α and δ, the highest level of robustness is consistent with having the lowest possible k, which is what the robust satisficing model aims to minimize.

We can also extend the framework of robust satisficing to address I multiple risk-based objectives in meeting their desired targets as follows

$$
\begin{aligned}
\min \ & \boldsymbol{w}^\top \boldsymbol{k} \\
\text{s.t.} \ & \mathbb{E}_\mathbb{P}\left[f_i(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right] - \tau_i \leq k_i \Delta(\mathbb{P}, \hat{\mathbb{P}}) \quad \forall \mathbb{P} \in \mathcal{P}_0(\mathcal{Z}), i \in [I] \\
& \boldsymbol{x} \in \mathcal{X}, \quad \boldsymbol{k} \geq \boldsymbol{0},
\end{aligned}
\tag{10}
$$

for some chosen weights parameters w_i, $i \in [I]$. Note that an equivalent model based on the framework of data-driven robust optimization may require the modeler to fiddle with the sizes of different ambiguity sets to obtain the desired performance, which may be more difficult to do so via cross-validation. Alternatively, we may adopt the data-driven joint chance-constrained optimization modeling framework proposed by Chen et al. (2018) and Xie (2019), but this would result in non-convex optimization models that are less computationally scalable.

On stochastic-free robust optimization and Pareto efficiency

Consider the following stochastic-free robust optimization problem,

$$
\begin{aligned}
Z_r = \min \ & \max_{z \in \mathcal{U}(r)} f(\boldsymbol{x}, \boldsymbol{z}) \\
\text{s.t.} \ & \boldsymbol{x} \in \mathcal{X}
\end{aligned}
$$

where $\mathcal{U}(r)$ is a norm-based uncertainty set given by

$$
\mathcal{U}(r) = \left\{ \boldsymbol{z} \in \mathcal{Z} \ \mid \ \|\boldsymbol{z} - \hat{\boldsymbol{z}}\| \leq r \right\},
$$

for some nominal value $\hat{\boldsymbol{z}}$. The corresponding stochastic-free robust satisficing model is given by

$$
\begin{aligned}
\min \ & k \\
\text{s.t.} \ & f(\boldsymbol{x}, \boldsymbol{z}) \leq \tau + k\|\boldsymbol{z} - \hat{\boldsymbol{z}}\| \quad \forall \boldsymbol{z} \in \mathcal{Z} \\
& \boldsymbol{x} \in \mathcal{X}, \quad k \geq 0,
\end{aligned}
$$

which coincides with Problem (9) at $S = 1$ and is a special case of the GRC-sum model of Ben-Tal et al. (2017).

Iancu and Trichakis (2014) observe a major drawback in classical stochastic-free robust optimization models in the potential lack of Pareto optimality in their solutions. Although the corresponding robust satisficing model does not directly address the Pareto optimality issue, in some cases, it could improve the Pareto efficiency compared to the robust optimization model. We illustrate this using the following robust knapsack problem.

EXAMPLE 1 (A KNAPSACK PROBLEM). We consider a stochastic-free robust continuous knapsack problem as follows

$$\max \sum_{i \in [M]} x_i$$
$$\text{s.t. } \boldsymbol{z}^\top \boldsymbol{x} \le M \quad \forall \boldsymbol{z} \in \mathcal{U}(r)$$
$$\boldsymbol{x} \in [0,1]^N,$$

with $N > M$, $N, M \in \mathbb{N}$ and the norm-based uncertainty set is defined as

$$\mathcal{U}(r) = \left\{ \boldsymbol{z} \in [0,1]^N \mid \|\boldsymbol{z}\|_1 \le r \right\}.$$

Observe that there would be multiple optimal solutions to the above problem whenever $r \le M$, with $x_i^* = 1$ for all $i \in [M]$, though the solution with $x_i = 0$ for all $i \in [N] \backslash [M]$ would be the most preferred optimal solution since, in which case, the knapsack constraint will always be feasible for all $\boldsymbol{z} \in \mathcal{Z}$.

Now, consider the corresponding robust satisficing model:

$$\min k$$
$$\text{s.t. } \boldsymbol{z}^\top \boldsymbol{x} \le M + k\|\boldsymbol{z}\|_1 \quad \forall \boldsymbol{z} \in [0,1]^N$$
$$\sum_{i \in [M]} x_i \ge \tau$$
$$\boldsymbol{x} \in [0,1]^N, k \ge 0.$$

The robust satisficing model would not suffer from a similar degeneracy issue. For any $\tau \le M$, the optimal solution to the robust satisficing model must satisfy that $x_i = 0$ for $i \in [N] \backslash [M]$, eliminating the degenerate solutions that are indifferent in the robust optimization framework.

The improvement achieved by the robust satisficing model is not merely because that the robust satisficing model can lead to more Pareto efficient solutions. In general, the family of solutions of the robust satisficing model is different from the family of solutions of the robust optimization model, and the robust satisficing solutions may better protect the system. We illustrate this in the following lot-sizing example with emergency fulfilment.

EXAMPLE 2 (LOT-SIZING WITH EMERGENCY FULFILMENT). We consider a simple robust lot-sizing problem with unit ordering cost and the emergency fulfilment cost being twice the normal ordering cost as follows:

$$Z_r = \min \sum_{i \in [N]} x_i + \max_{\boldsymbol{z} \in \mathcal{U}(r)} \left\{ \sum_{i \in [N]} 2 \max\{z_i - x_i, 0\} \right\}$$
$$\text{s.t. } \boldsymbol{x} \ge 0,$$

where $\boldsymbol{x}$ is the here-and-now ordering decision and the second stage cost is associated with the emergency fulfilment after the true demand $\boldsymbol{z}$ is realized. We consider N locations and the support set of demands is given by $\mathcal{Z} = [0, d]^N$, and the norm-based uncertainty set $\mathcal{U}(r)$ is given by

$$\mathcal{U}(r) = \{\boldsymbol{z} \in \mathcal{Z} \ | \ \|\boldsymbol{z}\|_1 \leq r\}.$$

The robust optimization approach could result in inadequate protection, which is caused by the nature of the uncertainty set rather than by its Pareto inefficiency.

PROPOSITION 4. *For any $r \leq d\lceil N/2 - 1 \rceil$, the optimal here-and-now ordering quantity to the robust lot-sizing problem is zero.*

For a concrete example, consider the case of $d \geq r > 0$ and the uncertainty set $\mathcal{U}(r)$ is a simplex. In the worst case, the demand of r would only occur at one location. If we commit to a here-and-now ordering of r units of inventory at any location, the worst case demand would occur at another location. Hence, the unique robustly optimal solution is to order nothing in the first stage, *i.e.*, $\boldsymbol{x} = \boldsymbol{0}$, resulting in far higher emergency fulfilment costs.

Now, let us focus on the robust satisficing network lot-sizing problem:

$$\begin{aligned} \kappa_\tau = \min \ & k \\ \text{s.t.} \ & \sum_{i \in [N]} x_i + \sum_{i \in [N]} 2\max\{z_i - x_i, 0\} \leq \tau + k\|\boldsymbol{z}\|_1 \quad \forall \boldsymbol{z} \in \mathcal{Z} \\ & \boldsymbol{x} \geq \boldsymbol{0}, \ k \geq 0. \end{aligned}$$

Observe that $Z_0 = 0$ and $Z_\infty = Nd$.

PROPOSITION 5. *For any $\tau \in (0, Nd]$, the optimal here-and-now ordering decision to the above robust satisficing lot-sizing problem with emergency fulfilment is non-zero.*

By Proposition 5, the robust satisficing model would prefer to fully utilize the cost budget to order inventory here-and-now and allocate the inventory in a diversified manner. Specifically, the robust satisficing model would not commit to the no-ordering solution, *i.e., $\boldsymbol{x} = \boldsymbol{0}$*. In our simulation study, we observe that the family of optimal solutions of the robust satisficing model could often yield a better efficient frontier.

As we illustrate above, robust satisficing model leads to a different family of solutions from the robust optimization model. In the two simple examples, the solutions attained by the robust satisficing model could improve Pareto efficiency and provide a better protection to a certain extent compared to the robust optimization model. In the simulation study, we would investigate the model performance more carefully and elucidate that solutions attained by the robust satisficing model remains efficient in complex problems.

3.   Fragility Measure

Note that the decision criterion of robust optimization belongs to a class of coherent risk measures of Artzner et al. (1999) and has a set of salient properties. It is desired that the decision criterion associated with the robust satisficing model should also be characterized. We now justify the decision criterion associated with the robust satisficing model, which we call the *fragility measure*, by laying out some reasonable properties or axioms the criterion is based on. As we will reveal, the fragility measure belongs to a class of satisficing measures proposed by Brown and Sim (2009). Let $(\Omega, \mathcal{F})$ be a measurable space with σ-algebra $\mathcal{F}$. The empirical probability distribution is given by $\hat{\mathbb{P}}$, $\hat{\mathbb{P}} \in \mathcal{P}_0$. Denote $\mathcal{L}$ as the space of measurable real-valued functions under all distributions in $\mathcal{P}_0$ and let $\tilde{v} \in \mathcal{L}$ be the random variable representing uncertain outcomes of $f(\boldsymbol{x}, \tilde{\boldsymbol{z}}) - \tau$ with a fixed $\boldsymbol{x}$.

DEFINITION 2 (FRAGILITY MEASURE). The functional $\rho : \mathcal{L} \mapsto [0, +\infty]$ is a fragility measure associated with the probability distribution $\hat{\mathbb{P}} \in \mathcal{P}_0$ if and only if it has the following representation

$$
\begin{aligned}
\rho(\tilde{v}) = \min \ & k \\
\text{s.t.} \quad & \mathbb{E}_{\mathbb{P}}[\tilde{v}] \leq k\Delta(\mathbb{P}, \hat{\mathbb{P}}) \quad \forall \mathbb{P} \in \mathcal{P}_0 \\
& k \geq 0,
\end{aligned}
\tag{11}
$$

for some probability distance function Δ.

With this definition, Problem (6) is equivalent to

$$
\begin{aligned}
\min \ & \rho(f(\boldsymbol{x}, \tilde{\boldsymbol{z}}) - \tau) \\
\text{s.t.} \ & \boldsymbol{x} \in \mathcal{X},
\end{aligned}
\tag{12}
$$

and the more general multiple risk-based objective Problem (10) is equivalent to

$$
\begin{aligned}
\min \ & \sum_{i \in [I]} w_i \rho(f_i(\boldsymbol{x}, \tilde{\boldsymbol{z}}) - \tau_i) \\
\text{s.t.} \ & \boldsymbol{x} \in \mathcal{X}.
\end{aligned}
\tag{13}
$$

We next show that the fragility measure is associated with the salient properties that are consistent with coherent decision making for reducing fragility of the solution and attaining model's robustness, while ensuring tractability of the problem when the criterion is to be minimized.

THEOREM 2 (**Axioms of Fragility Measure**). *A fragility measure associated with the probability distribution $\hat{\mathbb{P}} \in \mathcal{P}_0$ is a lower semi-continuous functional $\rho : \mathcal{L} \mapsto [0, +\infty]$ that satisfies the following properties:*

1. **Monotonicity:** *If $\mathbb{P}[\tilde{v}_1 \geq \tilde{v}_2] = 1$ for all $\mathbb{P} \in \mathcal{P}_0$, then $\rho(\tilde{v}_1) \geq \rho(\tilde{v}_2)$.*
2. **Positive homogeneity:** *For any $\lambda \geq 0$, we have $\rho(\lambda\tilde{v}) = \lambda\rho(\tilde{v})$.*
3. **Subadditivity:** $\rho(\tilde{v}_1 + \tilde{v}_2) \leq \rho(\tilde{v}_1) + \rho(\tilde{v}_2)$.

4. **Pro-robustness:** *If $\tilde{v} \leq 0$, then $\rho(\tilde{v}) = 0$.*

5. **Anti-fragility:** *If $\mathbb{E}_{\hat{\mathbb{P}}}[\tilde{v}] > 0$, then $\rho(\tilde{v}) = \infty$.*

Under additional technical conditions on $\mathcal{L}$ (see Theorem 6 in Follmer and Schied 2002), the probability distance Δ associated with ρ can be expressed as

$$\Delta(\mathbb{P}, \hat{\mathbb{P}}) = \sup_{\tilde{v} \in \mathcal{L}} \left\{ \mathbb{E}_{\mathbb{P}}[\tilde{v}] \;\mid\; \rho(\tilde{v}) \leq 1 \right\}. \tag{14}$$

We remark that parts of the technical conditions include restricting $\mathcal{L}$ to a space of bounded measurable real-valued functions under all distributions in $\mathcal{P}_0$. Such an assumption is common in the literature to obtain a useful representation theorem, though it would also rule out the random variable $f(\boldsymbol{x}, \tilde{\boldsymbol{z}})$, $\tilde{\boldsymbol{z}} \sim \mathbb{P}$, $\mathbb{P} \in \mathcal{P}_0(\mathcal{Z})$ for a support $\mathcal{Z}$ with which $\sup_{\boldsymbol{z} \in \mathcal{Z}} f(\boldsymbol{x}, \boldsymbol{z}) = \infty$. As in the representation of convex risk measures, there are also other technical conditions needed to ensure that the underlying separation theorem would go through. For further information on the analysis, we refer interested readers to Follmer and Schied (2002) and the references therein.

The first three properties of the fragility measures coincide with three out of the four axioms of coherent risk measures of by Artzner et al. (1999). The monotonicity property requires that if the uncertain outcomes of the model's constraint are never smaller in values in the direction of infeasibility compared to another random variable for all scenarios, then its fragility measure should not reflect a lower value than the other. Positive homogeneity property dictates that the fragility measure scales accordingly with the underlying uncertainty. Likewise, the property of subadditivity is synonymous with the preference for risk pooling, which is associated with uncertainty aversion. It implies that the collective fragility measure of the combined uncertainty in meeting one aggregate constraint should be smaller than the sum of the fragility measures if the feasibility of the uncertain constraints are considered separately. Incidentally, positive homogeneity and subadditivity imply convexity (Follmer and Schied 2004), which is also an important precursor to obtaining a tractable optimization model when the fragility measure is to be minimized.

The pro-robustness property asserts that if the model's constraint is always feasible, then the corresponding fragility measure should be the lowest value at zero. The anti-fragility property ensures that any solution with finite fragility measure would also be feasible in the model's constraint under the empirical distribution, *i.e.,* $\mathbb{E}_{\hat{\mathbb{P}}}[\tilde{v}] \leq 0$. These two properties would rule out any monetary risk measure as a candidate for fragility measure, since they would violate the translation invariance property (see Definition 3).

Theorem 2 shows that any metric that satisfies the salient properties is a fragility measure, even though it may not have the same explicit probability-distance-based representation. Note that the fragility measure belongs to the class of satisficing measures of Brown and Sim (2009). Similar to the satisficing measures, the fragility measure has a risk-based representation via normalized convex risk measure.

DEFINITION 3. A normalized convex risk measure is a lower semi-continuous functional $\mu : \mathcal{L} \mapsto \mathbb{R}$, that satisfies the following properties:

1. **Monotonicity**: If $\mathbb{P}[\tilde{v}_1 \geq \tilde{v}_2] = 1$ for all $\mathbb{P} \in \mathcal{P}_0$, then $\mu(\tilde{v}_1) \geq \mu(\tilde{v}_2)$.

2. **Translation invariance**: For any $a \in \mathbb{R}$, $\mu(\tilde{v} + a) = \mu(\tilde{v}) + a$.

3. **Convexity**: For any $\lambda \in [0, 1]$, $\mu(\lambda \tilde{v}_1 + (1 - \lambda)\tilde{v}_2) \leq \lambda \mu(\tilde{v}_1) + (1 - \lambda)\mu(\tilde{v}_2)$.

4. **Normalization**: $\mu(0) = 0$.

PROPOSITION 6 (**Risk-based Representation**). *The functional $\rho : \mathcal{L} \mapsto [0, +\infty]$ is a fragility measure associated with the probability distribution $\hat{\mathbb{P}} \in \mathcal{P}_0$ if and only if there exists some normalized convex risk measure $\mu : \mathcal{L} \mapsto \mathbb{R}$, satisfying*

$$\mu(\tilde{v}) \geq \mathbb{E}_{\hat{\mathbb{P}}}[\tilde{v}] \qquad \forall \tilde{v} \in \mathcal{L}, \tag{15}$$

such that

$$\rho(\tilde{v}) = \inf\left\{ k > 0 \;\mid\; k\mu(\tilde{v}/k) \leq 0 \right\}.$$

Note that the property associated with the inequality (15) is implied for convex risk measures that are *law-invariant under $\hat{\mathbb{P}}$* (see Follmer and Schied 2004). Proposition 6 implies that apart from its probability-distance-based representation, we can also construct a fragility measure using convex risk measure, which is well studied in the literature (see, *e.g.,* Follmer and Schied 2002, 2004). Some specific examples of fragility measure include the riskiness index of Aumann and Serrano (2008) and the essential riskiness index of Zhang et al. (2019). We note that several different robust satisficing models have already been introduced in the literature, including, Zhang et al. (2019), where a sum of essential riskiness indices is minimized and Chen et al. (2015), where a weighted sum of riskiness indices are minimized.

As we discussed previously, we focus on the data-driven setting and the robust satisficing problem would incorporate available data in the from of empirical distribution. In the following, we present several robust satisficing models under the Wasserstein distance in the context of common management decision problems.

4. Risk-based Linear Optimization

To motivate a tractable risk-based linear optimization model, we first consider an uncertain linear optimization problem, for instance in the context of production planning as follows:

$$
\begin{aligned}
\max \;\; & \boldsymbol{c}(\tilde{\boldsymbol{z}})^\top \boldsymbol{x} \\
\text{s.t.} \;\; & \boldsymbol{a}_i(\tilde{\boldsymbol{z}})^\top \boldsymbol{x} \leq b_i(\tilde{\boldsymbol{z}}) \quad \forall i \in [I] \\
& \boldsymbol{x} \geq \boldsymbol{0},
\end{aligned}
\tag{16}
$$

where the uncertain parameter $a_{ij}(\tilde{\boldsymbol{z}})$ denotes the number of units of ith resource needed to produce a unit of the jth item, $b_i(\tilde{\boldsymbol{z}})$ represents the quantity of the ith resource available for production, and $c_j(\tilde{\boldsymbol{z}})$ is the uncertainty in profit associated with selling a unit of the jth item. These parameters are affinely depend on some exogenous uncertainty $\boldsymbol{z} \in \mathcal{Z}$:

$$\boldsymbol{a}_i(\boldsymbol{z}) = \boldsymbol{a}_{i,0} + \sum_{n \in [N]} \boldsymbol{a}_{i,n} z_n = \boldsymbol{a}_{i,0} + \boldsymbol{A}_i \boldsymbol{z} \qquad \forall i \in [I]$$

$$b_i(\boldsymbol{z}) = b_{i,0} + \sum_{n \in [N]} b_{i,n} z_n = b_{i,0} + \boldsymbol{b}_i^\top \boldsymbol{z} \qquad \forall i \in [I]$$

$$\boldsymbol{c}(\boldsymbol{z}) = \boldsymbol{c}_0 + \sum_{n \in [N]} \boldsymbol{c}_n z_n = \boldsymbol{c}_0 + \boldsymbol{C} \boldsymbol{z}.$$

In such representation, we typically normalize the support set to a regular box $\mathcal{Z} = [-\mathbf{1}, \mathbf{1}]$ (see Bertsimas and Sim 2004), though the results can easily be extended to other polyhedron. The model determines the optimal profit maximizing decision $\boldsymbol{x}$, subject to the feasibility of the constraints under uncertainty. Since $\tilde{\boldsymbol{z}}$ is uncertain, Problem (16) is not well defined. In a chance constrained programming model (Charnes and Cooper 1959), violations of constraints may be tolerated without impacting the objective function, as long as their risks are within an acceptable limit specified by the decision maker. Under the empirical distribution, the chance-constrained programming problem can be expressed as a mixed-integer optimization problem, which can be computationally challenging to solve.

To obtain a tractable empirical optimization model that incorporate risk awareness in the objective function, as well as to the violation of constraints, we consider the following empirical risk-based linear optimization problem,

$$\begin{aligned}
Z_0 = \max \ & -\mathbb{C}_{\hat{\mathbb{P}}}^{\epsilon_0} \left[-\boldsymbol{c}(\tilde{\boldsymbol{z}})^\top \boldsymbol{x} \right] \\
\text{s.t.} \ & \mathbb{C}_{\hat{\mathbb{P}}}^{\epsilon_i} \left[\boldsymbol{a}_i(\tilde{\boldsymbol{z}})^\top \boldsymbol{x} - b_i(\tilde{\boldsymbol{z}}) \right] \leq 0 \ \forall i \in [I] \\
& \boldsymbol{x} \geq \mathbf{0},
\end{aligned} \qquad (17)$$

where $\mathbb{C}_{\mathbb{P}}^{\epsilon}[\tilde{v}]$ denotes the Conditional Value-at-Risk (CVaR) of the random variable $\tilde{v}$, denoting uncertain in costs, shortfalls or losses, at level $1 - \epsilon$ evaluated under a given distribution $\tilde{v} \sim \mathbb{P}$ given by

$$\mathbb{C}_{\mathbb{P}}^{\epsilon}[\tilde{v}] := \inf_{\alpha \in \mathbb{R}} \alpha + \frac{1}{\epsilon} \mathbb{E}_{\mathbb{P}} \left[(\tilde{v} - \alpha)^+ \right].$$

In finance, CVaR, also known as expected shortfall, quantifies the average loss over a specified time period of unlikely scenarios beyond the $1 - \epsilon$ confidence level (Rockafellar and Uryasev 2002). For example, with $\epsilon = 0.05$, a one-day CVaR of \$1 million means that the expected loss of the worst 5% scenarios over a one-day period is \$1 million. Hence, with $\epsilon = 1$, the CVaR criterion reduces to $\mathbb{C}_{\mathbb{P}}^1[\tilde{v}] = \mathbb{E}_{\mathbb{P}}[\tilde{v}]$ and recovers the risk-neutral case. CVaR is a well known convex risk measure

and the safeguarding constraints in Problem (17) would satisfy the following individual chance constraints

$$\hat{\mathbb{P}}\left[\boldsymbol{a}_i(\tilde{\boldsymbol{z}})^\top \boldsymbol{x} \le b_i(\tilde{\boldsymbol{z}})\right] \ge 1 - \epsilon \qquad \forall i \in [I],$$

which are nonlinear and non-convex constraints, and hence harder to incorporate in data-driven optimization models (see Luedtke et al. 2010). In contrast, Problem (17) is equivalent to the following linear optimization problem

$$
\begin{aligned}
Z_0 = \max \; &-t \\
\text{s.t.} \quad &\alpha_i + \frac{1}{S\epsilon_i}\sum_{s\in[S]} y_{is} \le 0 && \forall i \in [I] \\
&y_{is} \ge \boldsymbol{a}_i(\hat{\boldsymbol{z}}_s)^\top \boldsymbol{x} - b_i(\hat{\boldsymbol{z}}_s) - \alpha_i && \forall i \in [I],\ s\in[S] \\
&\alpha_0 + \frac{1}{S\epsilon_0}\sum_{s\in[S]} y_{0s} \le t \\
&y_{0s} \ge -\boldsymbol{c}(\hat{\boldsymbol{z}}_s)^\top \boldsymbol{x} - \alpha_0 && \forall s\in[S] \\
&\boldsymbol{x} \ge \boldsymbol{0} \\
&\alpha_i \in \mathbb{R},\ y_{is} \ge 0 && \forall i\in[I]\cup\{0\}, s\in[S].
\end{aligned}
$$

Given that the true distribution may deviate from the empirical distribution, it is very likely that empirical optimization model may under estimate the underlying risks associated with the model in order to achieve an unattainable risk-adjusted profit. For some acceptable decrease in the target risk adjusted profit over the empirical optimization model, *i.e.*, $\tau < Z_0$, we consider the following robust satisficing model,

$$
\begin{aligned}
\kappa_0 = \min \; &\sum_{i\in[I]\cup\{0\}} w_i k_i \\
\text{s.t.} \quad &\alpha_i + \frac{1}{\epsilon_i}\mathbb{E}_{\mathbb{P}}\left[(\boldsymbol{a}_i(\tilde{\boldsymbol{z}})^\top \boldsymbol{x} - b_i(\tilde{\boldsymbol{z}}) - \alpha_i)^+\right] \le k_i \Delta_W(\mathbb{P}, \hat{\mathbb{P}}) && \forall i\in[I], \mathbb{P}\in\mathcal{P}_0(\mathcal{Z}) \\
&\alpha_0 + \frac{1}{\epsilon_0}\mathbb{E}_{\mathbb{P}}\left[(-\boldsymbol{c}(\tilde{\boldsymbol{z}})^\top \boldsymbol{x} - \alpha_0)^+\right] + \tau \le k_0 \Delta_W(\mathbb{P}, \hat{\mathbb{P}}) && \forall \mathbb{P}\in\mathcal{P}_0(\mathcal{Z}) \\
&\boldsymbol{x} \ge \boldsymbol{0} \\
&\alpha_i \in \mathbb{R},\ k_s \ge 0,\ y_{is} \ge 0 && \forall i\in[I]\cup\{0\}, s\in[S].
\end{aligned}
\tag{18}
$$

where w_i, $i\in[I]\cup\{0\}$ are weights to reflect the relative importance of individual constraints. In the context of production planning, we can choose $w_0 = 1$, and w_i to be the cost needed to replenish one unit of the ith the resource to reflect the relative importance of different resource constraints. We remark that the robust satisficing solutions are feasible in the empirical optimization problem (17). Specifically, the set of I risk-based constraints are feasible under the empirical distribution.

THEOREM 3. *For a box support* $\mathcal{Z} = [-\mathbf{1}, \mathbf{1}]$*, the robust satisficing model* (18) *can be equivalently written as the following convex optimization problem:*

$$
\begin{aligned}
\min \quad & \sum_{i \in [I] \cup \{0\}} w_i k_i \\
\text{s.t.} \quad & \alpha_i + \frac{1}{S\epsilon_i} \sum_{s \in [S]} y_{is} \leq 0 && \forall i \in [I] \\
& y_{is} \geq \boldsymbol{a}_i(\hat{\boldsymbol{z}}_s)^\top \boldsymbol{x} - b_i(\hat{\boldsymbol{z}}_s) + (\boldsymbol{\eta}_{1is} + \boldsymbol{\eta}_{2is})^\top \mathbf{1} - (\boldsymbol{\eta}_{1is} - \boldsymbol{\eta}_{2is})^\top \hat{\boldsymbol{z}}_s - \alpha_i && \forall i \in [I], \ s \in [S] \\
& \alpha_0 + \frac{1}{S\epsilon_0} \sum_{s \in [S]} y_{0s} + \tau \leq 0 \\
& y_{0s} \geq -\boldsymbol{c}(\hat{\boldsymbol{z}}_s)^\top \boldsymbol{x} + (\boldsymbol{\eta}_{10s} + \boldsymbol{\eta}_{20s})^\top \mathbf{1} - (\boldsymbol{\eta}_{10s} - \boldsymbol{\eta}_{20s})^\top \hat{\boldsymbol{z}}_s - \alpha_0 && \forall s \in [S] \\
& \|\boldsymbol{A}_i^\top \boldsymbol{x} - \boldsymbol{b}_i - \boldsymbol{\eta}_{1is} + \boldsymbol{\eta}_{2is}\|_* \leq k_i && \forall i \in [I], \ s \in [S] \\
& \|\boldsymbol{C}^\top \boldsymbol{x} + \boldsymbol{\eta}_{10s} - \boldsymbol{\eta}_{20s}\|_* \leq k_0 && \forall s \in [S] \\
& \boldsymbol{\eta}_{1is} \geq \mathbf{0}, \ \boldsymbol{\eta}_{2is} \geq \mathbf{0} && \forall i \in [I] \cup \{0\}, \ s \in [S] \\
& \boldsymbol{x} \geq \mathbf{0} \\
& \alpha_i \in \mathbb{R}, \ k_i \geq 0, \ y_{is} \geq 0 && \forall i \in [I] \cup \{0\}, s \in [S].
\end{aligned}
$$

Although we focus on a normalized support set $\mathcal{Z} = [-\mathbf{1}, \mathbf{1}]$ for concreteness, this result can be naturally extended to a box support set $\mathcal{Z} = [\underline{\boldsymbol{z}}, \bar{\boldsymbol{z}}]$ or a polyhedral set $\mathcal{Z} = \{\boldsymbol{z} \in \mathbb{R}^N \mid \boldsymbol{C}\boldsymbol{z} \leq \boldsymbol{h}\}$. When the dual norm is ℓ_1-norm or ℓ_∞-norm (*i.e.*, the distance metric in Δ_W is given by ℓ_∞-norm and ℓ_1-norm, respectively), the final model in Theorem 3 is a linear optimization model. In our numerical study, we will apply this model to address a portfolio optimization problem.

5. Combinatorial Optimization

Combinatorial optimization problems, where $\mathcal{X} \subseteq \{0, 1\}^N$, have many applications in operations research and management science such as network design, capital budgeting, among others. They are inherently difficult optimization problems to solve, though there are many celebrated combinatorial optimization problems, such as the shortest path problem and the minimum cost spanning tree problem, that are polynomial time solvable. Extension to uncertainty in the objective function would general lead to a much harder problem to solve. The seminal work of Bertsimas and Sim (2003) shows how we can incorporate uncertainty in the objective function that would retain the tractability of the underlying combinatorial optimization problem. As far as we know, no such results have been extend to data-driven setting, unless for the trivial case of the following empirical combinatorial optimization model

$$
\begin{aligned}
Z_0 = \min \ & \mathbb{E}_{\hat{\mathbb{P}}} \left[\boldsymbol{c}^\top \boldsymbol{x} + \sum_{n \in [N]} d_n \tilde{z}_n x_n \right] \\
\text{s.t.} \ & \boldsymbol{x} \in \mathcal{X},
\end{aligned}
\tag{19}
$$

for some $\boldsymbol{c}, \boldsymbol{d} \geq \boldsymbol{0}$. We assume $\mathbb{E}_{\hat{\mathbb{P}}}[\tilde{\boldsymbol{z}}] = \boldsymbol{0}$, so that the objective function is simply $\boldsymbol{c}^\top \boldsymbol{x}$. We assume a box support set $\mathcal{Z} = [-\underline{\boldsymbol{z}}, \boldsymbol{1}]$ for any $\underline{\boldsymbol{z}} > \boldsymbol{0}$. The setup is the same as the one in Bertsimas and Sim (2003).

To obtain a tractable model for the corresponding robust satisficing problem, we focus on the ℓ_1-norm Wasserstein metric, *i.e.,*

$$\Delta_W(\mathbb{P}, \hat{\mathbb{P}}) := \inf_{\mathbb{Q} \in \mathcal{P}_0(\mathcal{Z}^2)} \left\{ \mathbb{E}_{\mathbb{Q}}\left[\|\tilde{\boldsymbol{z}} - \tilde{\boldsymbol{v}}\|_1 \right] \; \middle| \; (\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{v}}) \sim \mathbb{Q}, \tilde{\boldsymbol{z}} \sim \mathbb{P}, \tilde{\boldsymbol{v}} \sim \hat{\mathbb{P}} \; \right\}.$$

The robust combinatorial optimization in this setting can be written as:

$$Z_r = \min \sup_{\mathbb{P} \in \mathcal{B}(r)} \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{c}^\top \boldsymbol{x} + \sum_{n \in [N]} d_n \tilde{z}_n x_n \right] \tag{20}$$
$$\text{s.t.} \;\; \boldsymbol{x} \in \mathcal{X},$$

for a given Wasserstein-based ambiguity set,

$$\mathcal{B}(r) := \left\{ \mathbb{P} \in \mathcal{P}_0(\mathcal{Z}) \; \middle| \; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \Delta_W(\mathbb{P}, \hat{\mathbb{P}}) \leq r \end{array} \right\}.$$

THEOREM 4. *The robust combinatorial model* (20) *admits the following equivalent reformulation:*

$$Z_r = \min \; \boldsymbol{c}^\top \boldsymbol{x} + kr + \sum_{n \in [N]} x_n (d_n - k)^+$$
$$\text{s.t.} \;\; \boldsymbol{x} \in \mathcal{X}, k \in \{0, d_1, \ldots, d_N\}.$$

Coincidentally, this reformulation is the same as the reformulation of the classic robust combinatorial optimization model in Bertsimas and Sim (2003). Hence, the optimal solution can be obtained by solving $N + 1$ of the following combinatorial optimization problems with different linear objective functions:

$$\min_{\boldsymbol{x} \in \mathcal{X}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} + \sum_{n \in [N]} (d_n - k)^+ x_n \right\}, \tag{21}$$

for $k \in \{0, d_1, \ldots, d_N\}$. Hence, the robust combinatorial model is polynomial-time solvable if the underlying combinatorial problem is also polynomial-time solvable (Bertsimas and Sim 2003, Theorem 3).

For $\tau \geq Z_0$, the combinatorial robust satisficing model would be given by

$$\kappa_\tau = \min \; k$$
$$\text{s.t.} \;\; \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{c}^\top \boldsymbol{x} + \sum_{n \in [N]} d_n \tilde{z}_n x_n \right] \leq \tau + k \Delta_W(\mathbb{P}, \hat{\mathbb{P}}) \quad \forall \mathbb{P} \in \mathcal{P}_0(\mathcal{Z}) \tag{22}$$
$$\boldsymbol{x} \in \mathcal{X}.$$

THEOREM 5. *The combinatorial robust satisficing model* (22) *admits the following equivalent reformulation:*

$$\kappa_\tau = \min k$$
$$\text{s.t.} \ \ \boldsymbol{c}^\top \boldsymbol{x} + \sum_{n \in [N]} x_n (d_n - k)^+ \leq \tau \tag{23}$$
$$\boldsymbol{x} \in \mathcal{X}.$$

Problem (23) *can be solved via a bisection search to any accuracy $\epsilon > 0$ in at most $\lceil \log_2(\|\boldsymbol{d}\|_\infty / \epsilon) \rceil$ iterations, where each iteration solves a problem with the same complexity as the underlying combinatorial problem with a different linear objective function.*

By Theorem 5, the combinatorial robust satisficing optimization model (23) is polynomial-time solvable as long as the underlying combinatorial problems with linear objective is also tractable. In practice, the number of iterations required in the bisection search is small. As a simple comparison, consider a shortest path problem with $N = 1000$ arcs and the largest deviation of travel time along an arc is $\|\boldsymbol{d}\|_\infty = 10$. The robust combinatorial model requires solving $N + 1 = 1001$ standard combinatorial problems with linear objectives, while the combinatorial robust satisficing model only requires solving 20 of such problems to achieve an accuracy of $\epsilon = 10^{-5}$.

Assuming uniqueness in solving the underlying combinatorial optimization problem, from Problem (21), it is interesting to observe that there are at most $N + 1$ different solutions generated by the robust optimization problem. In contrast, the number of different solutions generated by the robust satisficing problem could be much larger. This observation coincides with our numerical studies that solving the robust satisficing problems generally yield a larger family of different solutions compared to solving similar robust optimization problems.

6. Linear Optimization with Recourse

Dantzig (1955) proposes the seminal stochastic optimization problem in a linear optimization framework where recourse decisions adapt to uncertain outcomes, resulting in a risky objective function for which the model minimises. Such models are ubiquitous in operations research such as a lot-sizing problem to determine the level of inventories at various locations to meet uncertain demands in a distribution network. We focus on a linear optimization with recourse. In the first stage, we set the values of here-and-now variables $\boldsymbol{x} \in \mathcal{X}$ before the realization of the random variable $\tilde{\boldsymbol{z}}$. The linear empirical optimization problem with recourse is given by

$$Z_0 = \min \mathbb{E}_{\hat{\mathbb{P}}} \left[\boldsymbol{c}(\tilde{\boldsymbol{z}})^\top \boldsymbol{x} + Q(\boldsymbol{x}, \tilde{\boldsymbol{z}}) \right]$$
$$\text{s.t.} \ \ \boldsymbol{x} \in \mathcal{X}. \tag{24}$$

Here, $Q(\boldsymbol{x}, \boldsymbol{z})$ represents the second-stage objective function:

$$
\begin{aligned}
Q(\boldsymbol{x}, \boldsymbol{z}) = \min \ & \boldsymbol{d}^\top \boldsymbol{y} \\
\text{s.t. } & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{b}(\boldsymbol{z}) \\
& \boldsymbol{y} \in \mathbb{R}^P,
\end{aligned}
\tag{25}
$$

where

$$
\boldsymbol{A}(\boldsymbol{z}) := \boldsymbol{A}_0 + \sum_{i \in [N]} \boldsymbol{A}_i z_i, \qquad \boldsymbol{b}(\boldsymbol{z}) := \boldsymbol{b}_0 + \sum_{i \in [N]} \boldsymbol{b}_i z_i, \qquad \boldsymbol{c}(\boldsymbol{z}) := \boldsymbol{c}_0 + \sum_{i \in [N]} \boldsymbol{c}_i z_i
$$

are affine mappings of $\boldsymbol{z}$. The goal is to determine the optimal here-and-now decision $\boldsymbol{x} \in \mathcal{X}$, and after the realization of the random parameters is observed, the optimal continuous wait-and-see decisions $\boldsymbol{y} \in \mathbb{R}^P$ is determined by solving Problem (25). Equivalently, we have the classical stochastic optimization problem with recourse (Dantzig 1955) as follows

$$
\begin{aligned}
Z_0 = \min \ & \frac{1}{S} \sum_{s \in [S]} \left(\boldsymbol{c}(\hat{\boldsymbol{z}}_s)^\top \boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{y}_s \right) \\
\text{s.t. } & \boldsymbol{A}(\hat{\boldsymbol{z}}_s)\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}_s \geq \boldsymbol{b}(\hat{\boldsymbol{z}}_s) \qquad \forall s \in [S] \\
& \boldsymbol{y}_s \in \mathbb{R}^P \qquad\qquad\qquad\quad \forall s \in [S] \\
& \boldsymbol{x} \in \mathcal{X}.
\end{aligned}
\tag{26}
$$

For tractability of the corresponding robust and robust satisficing model, we again focus again on the ℓ_1-norm Wasserstein metric Δ_W. The robust optimization model with recourse solves the following problem:

$$
\begin{aligned}
Z_r = \min \ \sup_{\mathbb{P} \in \mathcal{B}(r)} \ & \mathbb{E}_{\mathbb{P}} \left[\boldsymbol{c}(\tilde{\boldsymbol{z}})^\top \boldsymbol{x} + Q(\boldsymbol{x}, \tilde{\boldsymbol{z}}) \right] \\
\text{s.t. } & \boldsymbol{x} \in \mathcal{X}.
\end{aligned}
\tag{27}
$$

We can express Problem (27) as

$$
\begin{aligned}
Z_r = \min \ \sup_{\mathbb{P} \in \mathcal{B}(r)} \ & \mathbb{E}_{\mathbb{P}} \left[\boldsymbol{c}(\tilde{\boldsymbol{z}})^\top \boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{y}(\tilde{\boldsymbol{z}}) \right] \\
\text{s.t. } & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) \geq \boldsymbol{b}(\boldsymbol{z}) \qquad \forall \boldsymbol{z} \in \mathcal{Z} \\
& \boldsymbol{y} \in \mathcal{R}^{N,P} \\
& \boldsymbol{x} \in \mathcal{X},
\end{aligned}
\tag{28}
$$

where the family of recourse functions is defined as

$$
\mathcal{R}^{N,P} := \left\{ \boldsymbol{y} \ \mid \ \boldsymbol{y}(\boldsymbol{z}) : \mathbb{R}^N \mapsto \mathbb{R}^P, \boldsymbol{y} \text{ is a measurable function} \right\}.
$$

Problem (28) is generally intractable for $r > 0$ because the recourse function $\boldsymbol{y}(\boldsymbol{z})$ is unrestricted and akin to having infinite number of decision variables. A popular method to tractably solve Problem (28) is to use affine recourse adaptation, where the recourse function $\boldsymbol{y}(\boldsymbol{z})$ is restricted

to an affine function of $\boldsymbol{z}$ (see, *e.g.,* Ben-Tal et al. 2004, Kuhn et al. 2011, Bertsimas et al. 2019, Chen et al. 2020, Bertsimas et al. 2021) as follows:

$$\mathcal{L}^{N,P} := \left\{ \boldsymbol{y} \in \mathcal{R}^{N,P} \;\middle|\; \boldsymbol{y}(\boldsymbol{z}) = \boldsymbol{y}_0 + \sum_{i \in [N]} \boldsymbol{y}_i z_i \text{ for some } \boldsymbol{y}_i \in \mathbb{R}^P, i \in [N] \cup \{0\} \right\}.$$

Chen et al. (2020) propose a tractable scenario-wise lifted affine recourse adaptation to solve the problem approximately that performs almost as well as the exact model for a multi-item newsvendor problem.

For a given $\tau \geq Z_0$, we propose the adaptive linear robust satisficing model as follows:

$$\begin{aligned}
\min \; & k \\
\text{s.t. } & \mathbb{E}_{\mathbb{P}} \left[\boldsymbol{c}(\tilde{\boldsymbol{z}})^\top \boldsymbol{x} + Q(\boldsymbol{x}, \tilde{\boldsymbol{z}}) \right] - \tau \leq k \Delta_W(\mathbb{P}, \hat{\mathbb{P}}) \quad \forall \mathbb{P} \in \mathcal{P}_0(\mathcal{Z}) \\
& \boldsymbol{x} \in \mathcal{X}, \; k \geq 0.
\end{aligned}$$

THEOREM 6. *The adaptive linear robust satisficing can be equivalently written as:*

$$\begin{aligned}
\min \; & k \\
\text{s.t. } & \frac{1}{S} \sum_{s \in [S]} \sup_{(\boldsymbol{z},u) \in \bar{\mathcal{Z}}_s} \left\{ \boldsymbol{c}(\boldsymbol{z})^\top \boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{y}_s(\boldsymbol{z}, u) - ku \right\} \leq \tau & \\
& \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}_s(\boldsymbol{z}, u) \geq \boldsymbol{b}(\boldsymbol{z}) & \forall (\boldsymbol{z}, u) \in \bar{\mathcal{Z}}_s, s \in [S] \\
& \boldsymbol{y}_s \in \mathcal{R}^{N+1,P} & \forall s \in [S] \\
& \boldsymbol{x} \in \mathcal{X}, \; k \geq 0,
\end{aligned} \qquad (29)$$

where the lifted support set associated with each empirical scenario $s \in [S]$ is defined as

$$\bar{\mathcal{Z}}_s := \{ (\boldsymbol{z}, u) \in \mathcal{Z} \times \mathbb{R} \mid \|\boldsymbol{z} - \hat{\boldsymbol{z}}_s\|_1 \leq u \}. \qquad (30)$$

Similar to most adaptive optimization problems, Problem (29) is generally intractable. To obtain the optimal here-and-now solution $\boldsymbol{x} \in \mathcal{X}$ approximately, we consider a scenario-wise lifted affine recourse adaptation of Problem (29) as follows:

$$\begin{aligned}
\min \; & k \\
\text{s.t. } & \frac{1}{S} \sum_{s \in [S]} \sup_{(\boldsymbol{z},u) \in \bar{\mathcal{Z}}_s} \left\{ \boldsymbol{c}(\boldsymbol{z})^\top \boldsymbol{x} + \boldsymbol{d}^\top \left(\boldsymbol{y}_{s,0} + \sum_{i \in [N]} \boldsymbol{y}_{s,i} z_i + \boldsymbol{y}_{s,N+1} u \right) - ku \right\} \leq \tau \\
& \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B} \left(\boldsymbol{y}_{s,0} + \sum_{i \in [N]} \boldsymbol{y}_{s,i} z_i + \boldsymbol{y}_{s,N+1} u \right) \geq \boldsymbol{b}(\boldsymbol{z}) & \forall (\boldsymbol{z}, u) \in \bar{\mathcal{Z}}_s, s \in [S] \\
& \boldsymbol{y}_{s,i} \in \mathbb{R}^P & \forall i \in [N+1] \cup \{0\}, s \in [S] \\
& \boldsymbol{x} \in \mathcal{X}, \; k \geq 0.
\end{aligned}$$

$$(31)$$

Borrowing the terminology from stochastic optimization, the second stage optimization problem is said to have *relatively complete recourse* if for any $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{z} \in \mathcal{Z}$, there exists some $\boldsymbol{y} \in \mathbb{R}^P$ such

that $\boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{b}(\boldsymbol{z})$. Without the assumption of *relatively complete recourse*, the empirical optimization problem with recourse may not necessary yield a here-and-now decision, $\boldsymbol{x}$ that would ensure that the second-stage problem would always be feasible. Hence, under the true distribution, this may result in an infinite second stage expected cost under the true distribution, which is not acceptable. A strong condition called *complete recourse* is associated with the recourse matrix $\boldsymbol{B}$. Specifically, for any right hand side vector, $\boldsymbol{t} \in \mathbb{R}^M$, there exists a feasible recourse $\boldsymbol{w} \in \mathbb{R}^P$ such that $\boldsymbol{B}\boldsymbol{w} \geq \boldsymbol{t}$.

THEOREM 7. *Suppose the second stage problem* (25) *has complete recourse, then for any given* $\tau \geq Z_0$, *the scenario-wise lifted affine recourse adaptation, Model* (31), *is feasible and the objective is finite. Moreover, when* $P = 1$, *the scenario-wise lifted affine recourse adaptation would yield the optimal solution in Problem* (29).

This result is quite surprising and important. Despite the adaptive optimization model being a difficult problem to solve exactly, under complete recourse, the lifted affine recourse adaptation approximation does not limit the choice of targets $\tau \geq Z_0$ for the decision maker. The assumption of complete recourse is necessary here, and we provide a counter example in Appendix B.

We should note that in the absence of relatively complete recourse, solving the empirical optimization model or the stochastic optimization problem via Sample Average Approximation (SAA) may not even generate a feasible solution with finite risk when evaluated on the true distribution. Ben-Tal et al. (2004) observe that finding the here-and-now decision that would ensure that the second stage problem is always feasible for all realization of the uncertainty is an NP-hard problem. They propose using affine recourse adaptation as a conservative approximation to ensure feasibility of the second stage optimization problem for all $\boldsymbol{z} \in \mathcal{Z}$, which is applicable to solving the empirical optimization model as follows:

$$\bar{\bar{Z}}_0 = \min \ \mathbb{E}_{\hat{\mathbb{P}}}\left[\boldsymbol{c}(\tilde{\boldsymbol{z}})^\top \boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{y}(\tilde{\boldsymbol{z}})\right]$$
$$\text{s.t.} \ \ \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) \geq \boldsymbol{b}(\boldsymbol{z}) \ \forall \boldsymbol{z} \in \mathcal{Z}$$
$$\boldsymbol{y} \in \mathcal{L}^{N,P}$$
$$\boldsymbol{x} \in \mathcal{X}.$$

In fact, affine recourse adaptation or linear decision rule has appeared in early models of stochastic optimization, but they have been abandoned due to the lack of optimality concerns (see Garstka and Wets 1974). We can do better. Inspired by the adaptive robust satisficing, in the absence of complete recourse, we propose the following empirical optimization model with affine recourse approximation that will to ensure feasibility of the second stage optimization problem.

THEOREM 8. *The empirical linear optimization model with affine recourse approximation given by*

$$\bar{Z}_0 = \min \phi$$
$$\text{s.t.} \ \frac{1}{S} \sum_{s \in [S]} \sup_{(\boldsymbol{z},u) \in \bar{\mathcal{Z}}_s} \left\{ \boldsymbol{c}(\boldsymbol{z})^\top \boldsymbol{x} + \boldsymbol{d}^\top \left(\boldsymbol{y}_{s,0} + \sum_{i \in [N]} \boldsymbol{y}_{s,i} z_i + \boldsymbol{y}_{s,N+1} u \right) - ku \right\} \leq \phi$$
$$\boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\left(\boldsymbol{y}_{s,0} + \sum_{i \in [N]} \boldsymbol{y}_{s,i} z_i + \boldsymbol{y}_{s,N+1} u \right) \geq \boldsymbol{b}(\boldsymbol{z}) \qquad \forall (\boldsymbol{z},u) \in \bar{\mathcal{Z}}_s, s \in [S]$$
$$\boldsymbol{y}_{s,i} \in \mathbb{R}^P \qquad \forall i \in [N+1] \cup \{0\}, s \in [S]$$
$$\boldsymbol{x} \in \mathcal{X}, \ k \geq 0, \phi \in \mathbb{R},$$

(32)

has the following properties:

1. *If $\boldsymbol{x}$ is feasible to Problem (32), then $Q(\boldsymbol{x}, \boldsymbol{z}) < \infty$ for all $\boldsymbol{z} \in \mathcal{Z}$.*

2. *$\bar{\bar{Z}}_0 \geq \bar{Z}_0 \geq Z_0$.*

3. *If complete recourse holds, then $\bar{Z}_0 = Z_0$.*

4. *The adaptive linear robust satisficing is feasible for any $\tau \geq \bar{Z}_0$.*

Serendipitously, the robust satisficing model also provides a recipe for solving stochastic optimization problems approximately when the second stage problems do not have relatively complete recourse. The first property of Theorem 8 ensures that Problem (32), if feasible, also generates a here-and-now decision $\boldsymbol{x}$ that ensures that that the stage problem would be feasible for all $\boldsymbol{z} \in \mathcal{Z}$. Moreover, under complete recourse, it automatically recovers the solution of the classical two-stage stochastic optimization model with empirical distribution. To improve the approximation further, other more computationally intensive approaches can be used to improve the recourse adaptation, including piecewise affine recourse adaptation techniques (see, *e.g.*, Goh and Sim 2010, Chen et al. 2008) and Fourier-Motzkin elimination of recourse variables introduced by Zhen et al. (2018).

7. Simulation Study I: Portfolio optimization

We consider a portfolio selection problem under a data-driven setting and derive an explicit robust satisficing formulation. The decision-maker invests in N risky assets where the portfolio risk is evaluated using the empirical distribution $\hat{\mathbb{P}}$ constructed from historical data $(\hat{\boldsymbol{z}}_1, \ldots, \hat{\boldsymbol{z}}_S)$. The empirical portfolio optimization model is given by

$$Z_0 = \max \ \mathbb{E}_{\hat{\mathbb{P}}}[\boldsymbol{x}^\top \tilde{\boldsymbol{z}}]$$
$$\text{s.t.} \ \mathbb{C}_{\hat{\mathbb{P}}}^\epsilon[-\boldsymbol{x}^\top \tilde{\boldsymbol{z}}] \leq \beta$$
$$\mathbf{1}^\top \boldsymbol{x} = 1$$
$$\boldsymbol{x} \in \mathbb{R}_+^N,$$

where $\tilde{z}$ represents the daily returns of N stocks with empirical distribution $\hat{\mathbb{P}}$ and support $\mathcal{Z} = \mathbb{R}^N$. This portfolio optimization problem maximizes the expected return subject to the $1 - \epsilon$ daily CVaR being less than a β proportion of the capital.

The robust optimization (RO) solves the following optimization problem:

$$
\begin{aligned}
Z_r = \max_{} \quad & \inf_{\mathbb{P} \in \mathcal{B}(r)} \mathbb{E}_{\mathbb{P}} \left[\boldsymbol{x}^\top \tilde{\boldsymbol{z}} \right] \\
\text{s.t.} \quad & \alpha + \frac{1}{\epsilon} \mathbb{E}_{\mathbb{P}} \left[(-\boldsymbol{x}^\top \tilde{\boldsymbol{z}} - \alpha)^+ \right] \leq \beta \quad \forall \mathbb{P} \in \mathcal{B}(r) \\
& \mathbf{1}^\top \boldsymbol{x} = 1 \\
& \boldsymbol{x} \in \mathbb{R}_+^N, \ \alpha \in \mathbb{R},
\end{aligned}
$$

while the robust satisficing (RS) solves the following optimization problem:

$$
\begin{aligned}
\kappa_\tau = \min \quad & k_0 + w k_1 \\
\text{s.t.} \quad & \mathbb{E}_{\mathbb{P}} \left[\boldsymbol{x}^\top \tilde{\boldsymbol{z}} \right] \geq \tau - k_0 \Delta_W(\mathbb{P}, \hat{\mathbb{P}}) && \forall \mathbb{P} \in \mathcal{P}_0(\mathcal{Z}) \\
& \alpha + \frac{1}{\epsilon} \mathbb{E}_{\mathbb{P}} \left[(-\boldsymbol{x}^\top \tilde{\boldsymbol{z}} - \alpha)^+ \right] \leq \beta + k_1 \Delta_W(\mathbb{P}, \hat{\mathbb{P}}) && \forall \mathbb{P} \in \mathcal{P}_0(\mathcal{Z}) \\
& \mathbf{1}^\top \boldsymbol{x} = 1 \\
& \boldsymbol{x} \in \mathbb{R}_+^N,
\end{aligned}
\tag{33}
$$

for some target, $\tau \leq Z_0$, and a penalty parameter $w \geq 0$.

THEOREM 9. *Suppose the Wasserstein distance Δ_W is defined with ℓ_1-norm, and the support set is given by $\mathcal{Z} = \mathbb{R}^N$. The optimal portfolio in Problem (33) can be obtained by solving the following optimization problem:*

$$
\begin{aligned}
\min \quad & \|\boldsymbol{x}\|_\infty \\
\text{s.t.} \quad & \frac{1}{S} \sum_{s \in [S]} y_{1s} \geq \tau \\
& y_{1s} \leq \boldsymbol{x}^\top \hat{\boldsymbol{z}}_s && \forall s \in [S] \\
& \alpha + \frac{1}{\epsilon S} \sum_{s \in [S]} y_{2s} \leq \beta \\
& y_{2s} \geq -\boldsymbol{x}^\top \hat{\boldsymbol{z}}_s - \alpha && \forall s \in [S] \\
& y_{2s} \geq 0 && \forall s \in [S] \\
& \mathbf{1}^\top \boldsymbol{x} = 1 \\
& \boldsymbol{x} \in \mathbb{R}_+^N, \ \alpha \in \mathbb{R}.
\end{aligned}
\tag{34}
$$

Model (34) is a linear optimization model and has interesting insights in the context of portfolio optimization. Observe that the solution solution does not depend on the choice of penalty parameter, w. Speaking intuitively, the robust satisficing approach improves the robustness by diversifying as much as possible, while trying to attain the target level of expected utility, and safeguarding the CVaR of losses over the empirical distribution. It diversifies the portfolio more aggressively than the classical mean-variance portfolio selection of (Markowitz 1952).

PROPOSITION 7. *Whenever* $\tau \leq \mathbb{E}_{\hat{\mathbb{P}}}[\mathbf{1}^\top \tilde{\boldsymbol{z}}/N]$ *and* $\beta \geq \mathbb{C}^{\epsilon}_{\hat{\mathbb{P}}}(-\mathbf{1}^\top \tilde{\boldsymbol{z}}/N)$, *the optimal portfolio in Model* (34) *is the equal-weighted* $1/N$ *portfolio.*

As the target utility increases, the decision maker commits to less ambiguity-averse portfolio and diversify less in order to achieve a more ambitious target utility.

Despite the simplicity, it has been well-known from the empirical study of DeMiguel et al. (2009) that equal-weighted portfolio outperforms many risk models in practice. In particular, when they compare the equal-weighted portfolio against 14 models across seven empirical datasets, none of the model is consistently better than the equal-weighted portfolio in out-of-sample perfromance. When there are 25 assets, they show that mean-variance strategy and its extensions require an estimation window of over 3,000 months in order to outperform the equal-weighted portfolio, alluding to their impracticality in addressing actual portfolio selection problems. The optimality of equal weighted portfolios has also been established in distributionally robust portfolio optimization models (see, *e.g.*, Pflug et al. 2012, Mohajerin Esfahani and Kuhn 2018) as the Wasserstein ball increases in size within the ambiguity set. In the robust satisficing model, the optimal portfolio depends on the target of the decision maker, which we believe is more interpretable for the decision maker to specify than the radius of the Wasserstein ball.

In the following, we illustrate the model performance. Following a similar setting as in Mohajerin Esfahani and Kuhn (2018), we consider $N = 10$ stocks. We assume the return $\tilde{z}_n$, for $n \in [N]$, can be decomposed into a systematic risk $\varepsilon_n \sim \mathcal{N}(0, 5\%)$ and an idiosyncratic risk $\delta_n \sim \mathcal{N}(n \times 2\%, n \times 3\%)$. By construction, stocks with higher indices provide higher mean returns but at the same time result in higher risks. The support of the return is $\mathbb{R}^N$. We set $\epsilon = 0.05$, and β varies in different instances. All optimization models are solved based on a sample of size $S \in \{30, 300\}$, and the out-of-sample performance of solutions is evaluated on a set of verification samples of size 10,000. For each instance, we solve the robust satisficing model (34) over a sequence of expected utility targets τ and solve the robust optimization model over a sequence of radius r. Then, we compare the efficient frontier on both out-of-sample CVaR and average return. Here, out-of-sample CVaR and average return represent CVaR and average return evaluated under the set of verification samples. Besides robust satisficing and robust optimization models, we also present performance of the equal-weighted portfolio and the empirical optimization model.

First, we fix sample size $S = 30$. In this case, the historical sample is not large enough to accurately reflect the true distribution of returns; therefore, one should expect that the in-sample and out-of-sample metrics would vary by a lot, indicating the need of robustness. To begin with, we evaluate the in-sample performance of the equal-weighted $1/N$ portfolio—the in-sample average return is 0.119 and the in-sample CVaR is -0.008. When $\beta < -0.008$, the equal-weighted portfolio

is infeasible for both robust satisficing and robust optimization models. When $\beta \geq -0.008$, the equal-weighted portfolio is always feasible and optimal for the robust satisficing model when $\tau \leq 0.119$. However, an equal-weighted portfolio may not be attained as an optimal solution to the robust optimization model for some radius r. In our experiments, the empirical optimization leads to a higher in-sample return than the equal-weighted portfolio. For the robust satisficing, we vary $\tau \in [0.119, Z_0]$, and we can guarantee that the robust satisficing has a non-empty feasible set for $\tau \in [0.119, Z_0]$. In practice, we can vary τ as a percentage of Z_0, which is intuitive to specify. For the robust optimization, we vary the radius $r \in [0, \bar{r}]$, where $\bar{r}$ is the largest radius such that the robust optimization model has a non-empty feasible set. Note that $\bar{r}$ varies for different instances, and this value has to be tuned manually.

In Figure 1, we compare the efficient frontiers of robust satisficing and robust optimization when $\beta = 0.1$. Robust satisficing approach achieves a dominant efficient frontier. For the same out-of-sample average utility, the robust satisficing solutions often achieves a lower out-of-sample CVaR. In this figure, we also mark the performance of the empirical optimization model and the equal-weighted portfolio. The empirically optimized solution coincides with the robust solution when $r = 0$ and the robust satisficing solution when $\tau = Z_0$. As we can see, the empirically optimized solution overfit to the sample and produce a high risk. The out-of-sample CVaR is 0.25, which significantly exceeds the imposed acceptable level of $\beta = 0.1$. This indicates the importance of obtaining robustness to improve the empirical optimization model and control the out-of-sample CVaR, because of the huge discrepancy between in-sample and out-of-sample metrics.

From Figure 1, we also notice that the efficient frontier of the robust satisficing model is smoother than that of the robust optimization model. This is because the solutions to the robust optimization model is more sensitive to the parameter r. In Figure 2, we show the out-of-sample metrics with respect to model parameters, *i.e.,* τ in robust satisficing and r in robust optimization. As we can see, the out-of-sample metrics of robust optimization model can experience some sudden jumps within a small neighborhood of r, or they can stay constant for a large range of r. On the contrary, the out-of-sample metrics of the robust satisficing model is smooth in parameter τ.

We provide two more instances in Figure 3, where we keep $S = 30$ and vary $\beta \in \{-0.05, 0.0\}$. In both comparisons, we tune the upper bound of radius, $\bar{r}$, until the robust optimization model is no longer feasible. We observe that the robust satisficing models again lead to better efficient frontiers than those of the corresponding robust optimization models, further illustrating the potentiality of our approach.

Then, we consider the case with a large sample, $S = 300$. In this example, the equal-weighted portfolio has an in-sample average return of 0.118 and an in-sample CVaR of 0.016. In Figure 4, we compare the efficient frontiers of robust satisficing and robust optimization when $\beta = 0.05$.
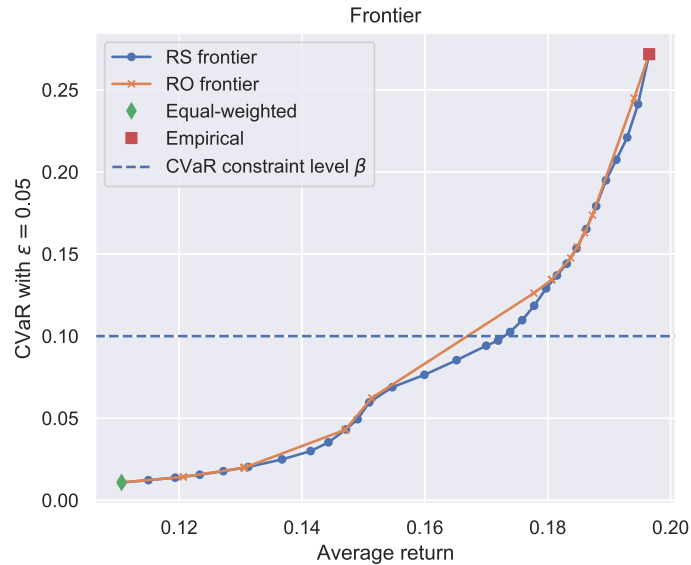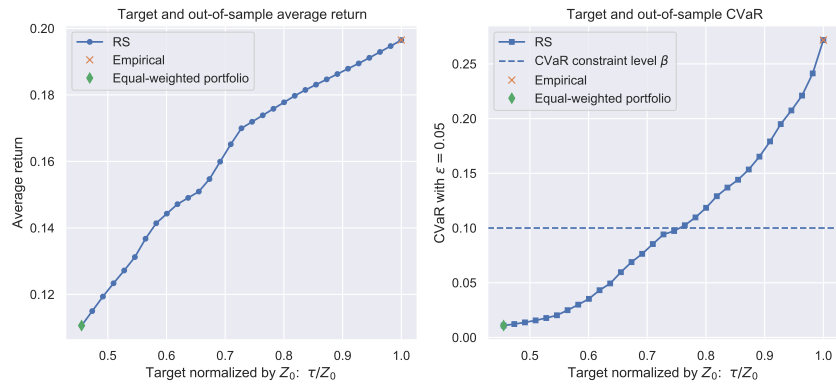
Figure 1 **Comparison of efficient frontiers by RO and RS:** $S = 30$ **and** $\beta = 0.1$.

We observe that the efficient frontier of the robust satisficing model dominates that of the robust optimization model. Although the differences between in-sample and out-of-sample metrics are smaller than those when $S = 30$, there is still a non-negligible gap, *e.g.,* the out-of-sample CVaR of the empirically optimized solution exceeds that acceptable level of $\beta = 0.05$. This indicates the need of robustness to reduce the out-of-sample risk of the empirical optimization model even when we have a large sample. We provide two more instances in Figure 5, where we keep $S = 300$ and vary $\beta \in \{0.01, 0.1\}$. We observe that the robust satisficing model again leads to better efficient frontiers than that of the robust optimization model when $\beta = 0.01$. When $\beta = 0.1$, the empirical CVaR constraint is rather loose in both models, reducing the differences in solutions and leading to similar efficient frontiers.

Finally, we investigate the effect of β. In Figure 6, we plot the efficient frontiers of the robust satisficing models under different β values. As β decreases, the set of feasible solutions diminishes. In addition, the solution will tend to over-fit to the sample as β decreases. This is because we face an ambitious constraint on in-sample CVaR, and one must over-fit to the sample to achieve it. It is important to note that β is a parameter that the decision maker prescribes to reflect his risk attitude; hence, this should be treated differently as τ, which is a model parameter that we tune.

8. Simulation Study II: Adaptive network lot-sizing

We consider a similar network lot-sizing context as in Bertsimas and de Ruiter (2016). The decision maker sells a single item at N different stores. We must determine the initial stock allocation
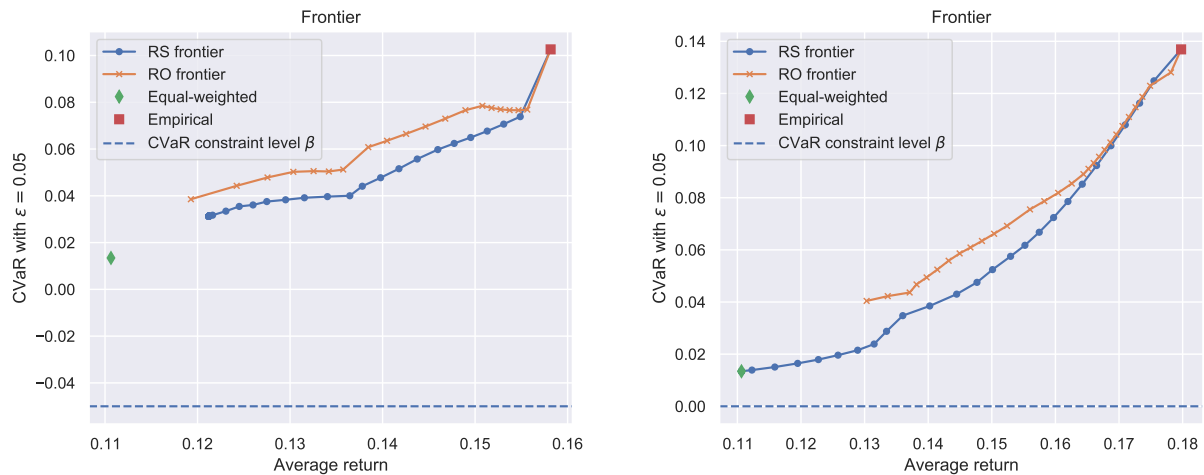
(a) RS out-of-sample metrics w.r.t. normalized target τ/Z_0.
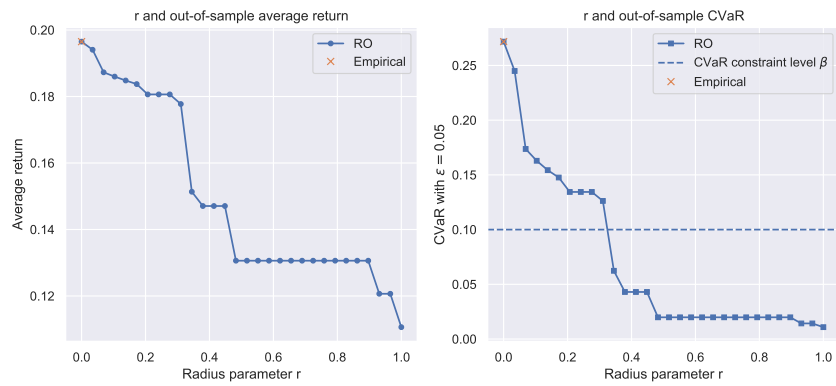
(b) RO out-of-sample metrics w.r.t. Wasserstein radius parameter r.

Figure 2   **Out-of-sample metrics with respect to model parameters:** $S = 30$ **and** $\beta = 0.1$**.**

(a) $\beta = -0.05$ (b) $\beta = 0.0$

Figure 3   **Comparison of efficient frontiers by RO and RS:** $S = 30$; $\beta = -0.05$ **(left), and** $\beta = 0$ **(right).**
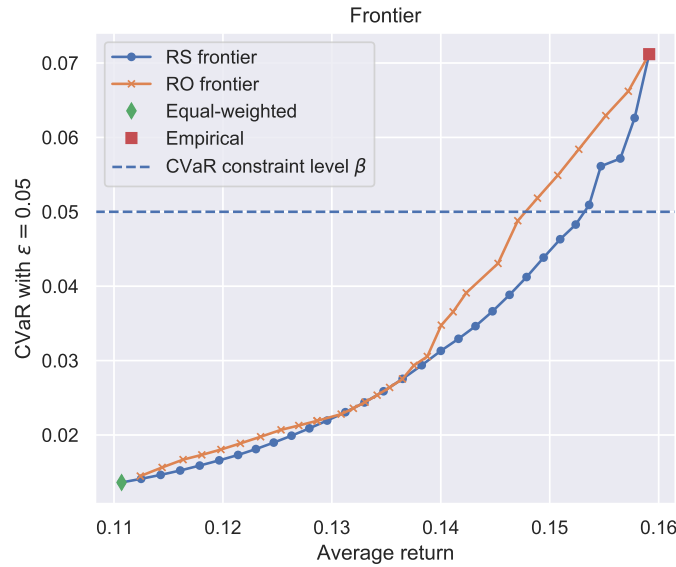
Figure 4 **Comparison of efficient frontiers by RO and RS:** $S = 300$ **and** $\beta = 0.05$.

(a) $\beta = 0.01$ (b) $\beta = 0.1$
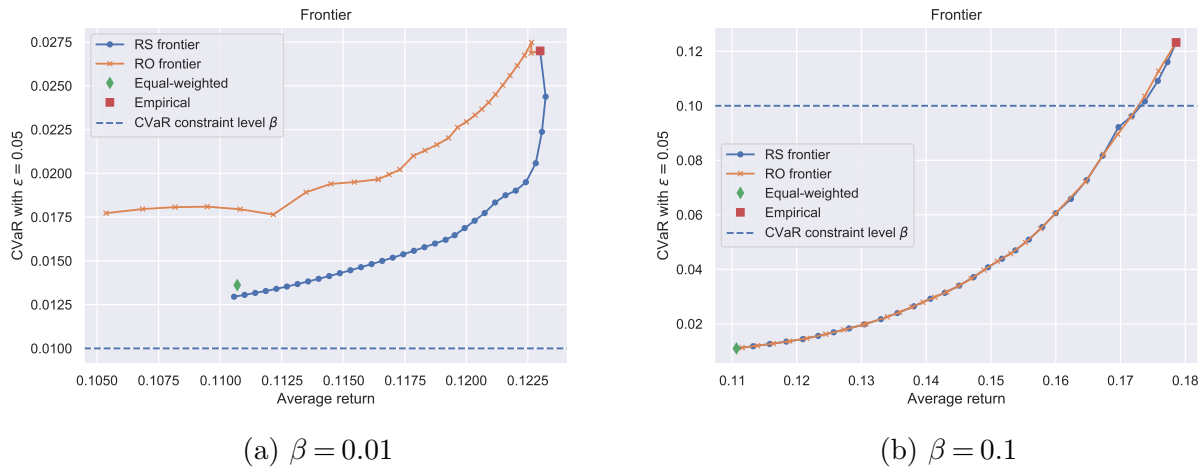
Figure 5 **Comparison of efficient frontiers by RO and RS:** $S = 300$; $\beta = 0.01$ **(left) and** $\beta = 0.1$ **(right).**

$x_i \in [0, \bar{\delta}_i]$ at a unit ordering cost c_i, for different stores $i \in [N]$, prior to the realization of random demands $\tilde{z}_i$ for $i \in [N]$. After observing the demands, we can transport stock y_{ij} from store i to store j at a unit transportation cost d_{ij} to better satisfy demands. We use w_i to represents the emergency orders at store i, at a unit cost $l_i > c_i$. We randomly generate the N stores in a 10×10 grid. The transportation cost d_{ij} is set to be proportional to the euclidean distance, D_{ij}, between stores i and j.
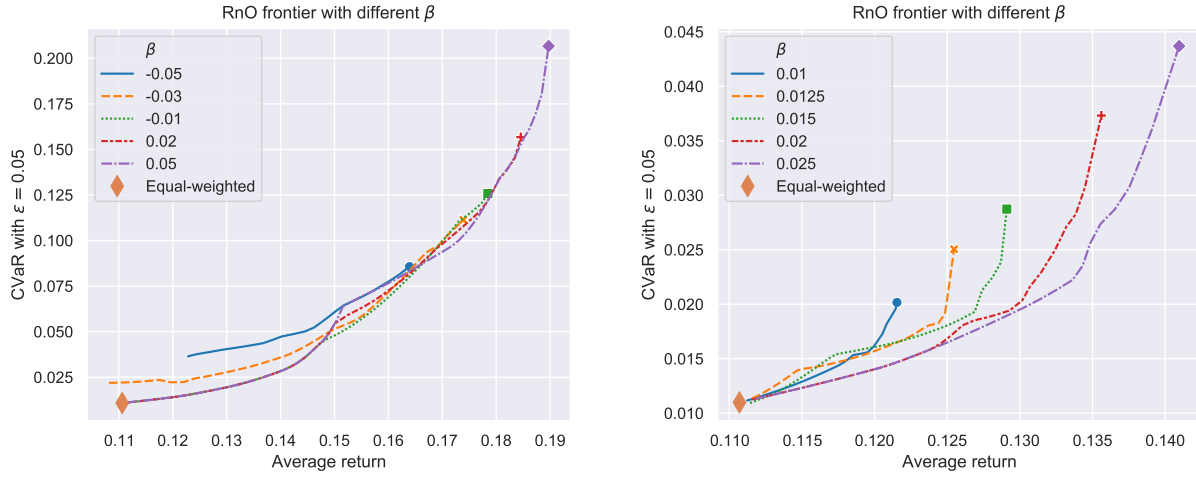
Figure 6 Effect of β on the solution frontier of RS: $S = 30$ **(left)** and $S = 300$ **(right)**.

Following Theorem 6, the adaptive robust satisficing with scenario-wise affine recourse adaptation can be written as follows:

$$
\min k
$$
$$
\text{s.t. } \boldsymbol{c}^\top \boldsymbol{x} - \tau + \frac{1}{S} \sum_{s \in [S]} v_s \leq 0
$$
$$
\sum_{i \in [N]} \boldsymbol{d}_i^\top \boldsymbol{y}_i^{(s)}(\boldsymbol{z}, u) + \boldsymbol{l}^\top \boldsymbol{w}^{(s)}(\boldsymbol{z}, u) - ku \leq v_s \qquad \forall (\boldsymbol{z}, u) \in \bar{\mathcal{Z}}_s, \ \forall s \in [S]
$$
$$
x_i + w_i^{(s)}(\boldsymbol{z}, u) + \sum_{j \in [N]} y_{ji}^{(s)}(\boldsymbol{z}, u) - \sum_{j \in [N]} y_{ij}^{(s)}(\boldsymbol{z}, u) - z_i \geq 0 \quad \forall (\boldsymbol{z}, u) \in \bar{\mathcal{Z}}_s, \ \forall s \in [S], \ \forall i \in [N]
$$
$$
\boldsymbol{y}^{(s)}(\boldsymbol{z}, u) \geq \boldsymbol{0}, \ \boldsymbol{w}^{(s)}(\boldsymbol{z}, u) \geq \boldsymbol{0} \qquad \forall (\boldsymbol{z}, u) \in \bar{\mathcal{Z}}, s \in [S]
$$
$$
\boldsymbol{0} \leq \boldsymbol{x} \leq \bar{\boldsymbol{\delta}}
$$
$$
\boldsymbol{y}^{(s)} \in \mathcal{L}^{N+1, N \times N}, \ \boldsymbol{w}^{(s)} \in \mathcal{L}^{N+1, N} \qquad \forall s \in [S],
$$

where the lifted uncertainty set associated with each scenario $s \in [S]$ is as defined in Equation (30).

Similarly, we can formulate the benchmark robust optimization model under the Wasserstein metric and lifted affine recourse adaptation as:

$$
Z_r = \min kr + \boldsymbol{c}^\top \boldsymbol{x} + \frac{1}{S} \sum_{s \in [S]} v_s
$$
$$
\text{s.t. } \sum_{i \in [N]} \boldsymbol{d}_i^\top \boldsymbol{y}_i^{(s)}(\boldsymbol{z}, u) + \boldsymbol{l}^\top \boldsymbol{w}^{(s)}(\boldsymbol{z}, u) - ku \leq v_s \qquad \forall (\boldsymbol{z}, u) \in \bar{\mathcal{Z}}_s, \ \forall s \in [S]
$$
$$
x_i + w_i^{(s)}(\boldsymbol{z}, u) + \sum_{j \in [N]} y_{ji}^{(s)}(\boldsymbol{z}, u) - \sum_{j \in [N]} y_{ij}^{(s)}(\boldsymbol{z}, u) - z_i \geq 0 \quad \forall (\boldsymbol{z}, u) \in \bar{\mathcal{Z}}_s, \ \forall s \in [S], \ \forall i \in [N]
$$
$$
\boldsymbol{y}^{(s)}(\boldsymbol{z}, u) \geq \boldsymbol{0}, \ \boldsymbol{w}^{(s)}(\boldsymbol{z}, u) \geq \boldsymbol{0} \qquad \forall (\boldsymbol{z}, u) \in \bar{\mathcal{Z}}_s, s \in [S]
$$
$$
\boldsymbol{0} \leq \boldsymbol{x} \leq \bar{\boldsymbol{\delta}}
$$
$$
\boldsymbol{y}^{(s)} \in \mathcal{L}^{N+1, N \times N}, \ \boldsymbol{w}^{(s)} \in \mathcal{L}^{N+1, N} \qquad \forall s \in [S].
$$

As r approaches zero, the robust model becomes the empirical optimization model:

$$Z_0 = \min \; \boldsymbol{c}^\top \boldsymbol{x} + \frac{1}{S} \sum_{s \in [S]} v_s$$

$$\text{s.t.} \; \sum_{i \in [N]} \boldsymbol{d}_i^\top \boldsymbol{y}_i^{(s)} + \boldsymbol{l}^\top \boldsymbol{w}^{(s)} \leq v_s \qquad \forall s \in [S]$$

$$x_i + w_i^{(s)} + \sum_{j \in [N]} y_{ji}^{(s)} - \sum_{j \in [N]} y_{ij}^{(s)} - z_i \geq 0 \quad \forall s \in [S], \; \forall i \in [N]$$

$$\boldsymbol{y}^{(s)} \geq \boldsymbol{0}, \; \boldsymbol{w}^{(s)} \geq \boldsymbol{0} \qquad \forall s \in [S]$$

$$\boldsymbol{0} \leq \boldsymbol{x} \leq \bar{\boldsymbol{\delta}}.$$

Note that with the provision of the emergency orders in the model would make this a complete recourse problem. In this study, we set $N = 20$ and $\mathcal{Z} = [0, 40]^N$. For all $i \in [N], j \in [N]$, we set $c_i = 10$, $l_i = 30$, $\delta_i = 40$, $d_{ij} = 2D_{ij}$. We let $S=5$, *i.e.*, the empirical distribution consists of five historical samples. To test the solutions, we generate $10,000$ samples as the testing data. In each sample, demands at different locations are independently generated. We use a normal distribution, $\mathcal{N}(20, 10^2)$, to generate these demands, and the demands are truncated to ensure they lie in the support set $\mathcal{Z}$.

We solve the robust optimization (RO) model with different values of Wasserstein radius r and the robust satisficing (RS) model with different values of target τ. Then, we compare the out-of-sample costs with respect to the first-stage ordering cost. We also benchmark the baseline empirical optimization model. We present the performance comparison in Figure 7. The data-driven optimization only gives a single solution, and it performs poorly because it over-fits to the historical data. In many practical settings, the historical data is not rich enough to depict the true distribution; therefore, the performance of data-driven optimization is inferior in these cases. With different input parameters, both RO and RS models can improve cost metrics significantly, especially cost at higher quantiles. Furthermore, RS model outperforms RO model in terms of both average cost and costs at different quantiles. For the average total cost, the standard deviations of sample mean of the robust satisficing solutions under different targets range from one to five, which are small enough to conclude the improvement is statistically significant. We have tried varying the standard deviation of the underlying normal distribution and generating demands from some uniform distributions and Poisson distributions, and we observe similar results. We present results for an example where demands are generated from a normal distribution $\mathcal{N}(20, 12^2)$ in Appendix C. In Appendix D, we present a sample code on how to model RO and RS in this problem using RSOME (Chen et al. 2020).

Another benefit of robust satisficing is that it is much easier to select a target parameter, as opposed to selecting a radius parameter for the Wasserstein ball. The target is a management-related parameter while the Wasserstein radius is an abstract tuning parameter. We present how

the out-of-sample cost metrics changes with respect to the two model parameters in Figure 8. We plot both the average cost and the cost at the 90th percentile. The curve for the 95th percentile is similar to the 90th percentile (see Figure 7) and is therefore omitted. The parameter r is not interpretable operationally, and the out-of-sample performance is sensitive to the choice of this parameter. As we can see, the average cost and 90th percentile change rapidly within a small range of radius r. In other words, a small change in r can lead to a very different solution. On the other hand, the change in out-of-sample performance is much more smooth with respect to the change in target in RS. Hence, a small change in the target parameter will not lead to a significant change in solution.

In addition, the "nice" range of target τ/Z_0 is relatively stable with respect to changes in other model parameters, *e.g.,* support set. However, the "nice" range of radius r can change significantly as other parameters change. We provide an illustration in Figure 9, where we expand the capacity of store $i \in [N]$ to $200 + 10i$ and expand the support set of demand at this store from $[0, 40]$ to $[0, 200 + 10i]$, for $i \in [N]$. We plot the out-of-sample cost metrics with respect to model parameters and overlay with the curves in Figure 8 to present the difference. The cost metrics of each instance are normalized by the respective baseline out-of-sample average cost so that they are on a similar scale. Clearly, the robust satisficing approach is far less sensitive to the specification of support set compared to the robust optimization model. As we observe, the "nice" range of radius r changes significantly while the performance with respect to the normalized target τ/Z_0 is very stable to this change. Given the interpretation of the model parameter $\tau \geq Z_0$ in robust satisficing, decision maker can always choose the target as a percentage of Z_0. However, the radius parameter r is usually a hyper-parameter that does not have a physical meaning. Hence, in addressing data-driven optimization problems, there are remarkable challenges faced by robust optimization models to determine to right hyper-parameter for the ambiguity set, which can be highly sensitive to how the support set is being specified. In contrast, besides the target parameter being more interpretable and intuitive to specify in management decision problems, it is also far less sensitive to the specification of the support set of the random parameters.
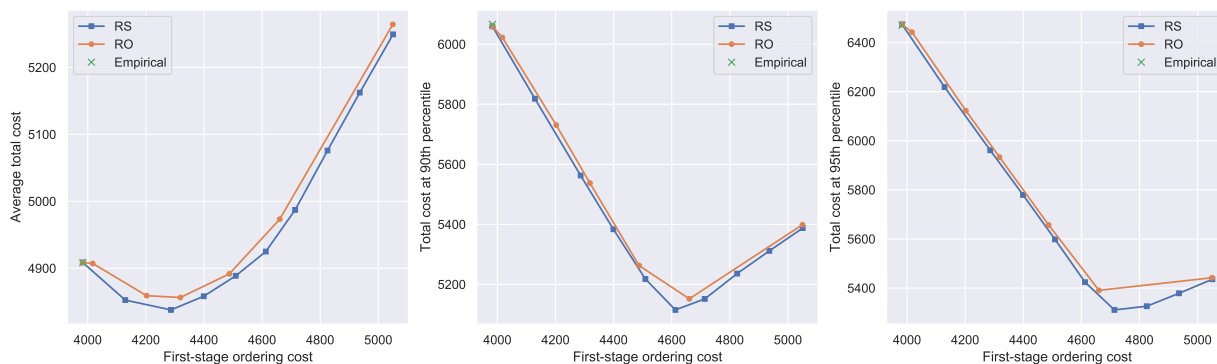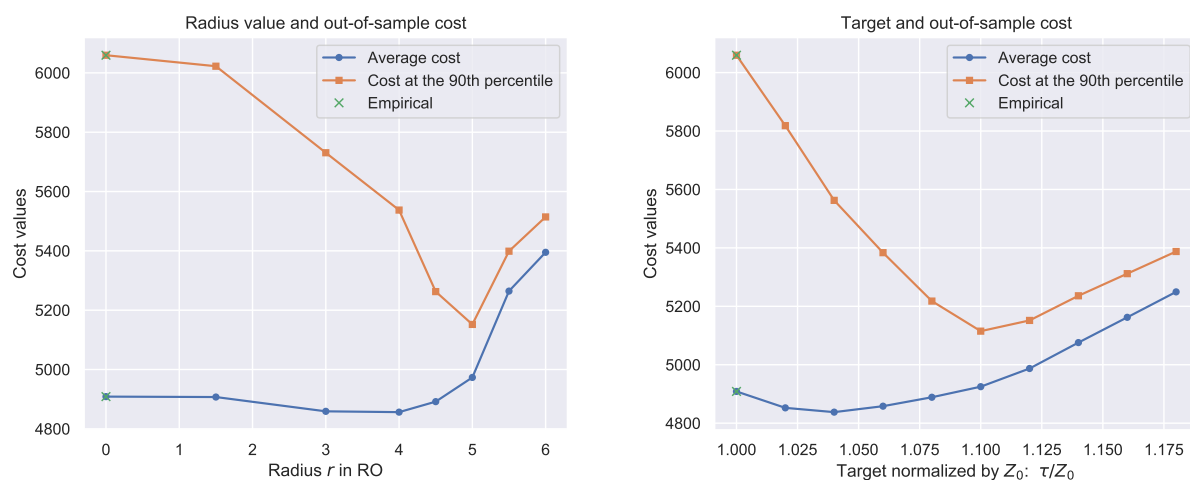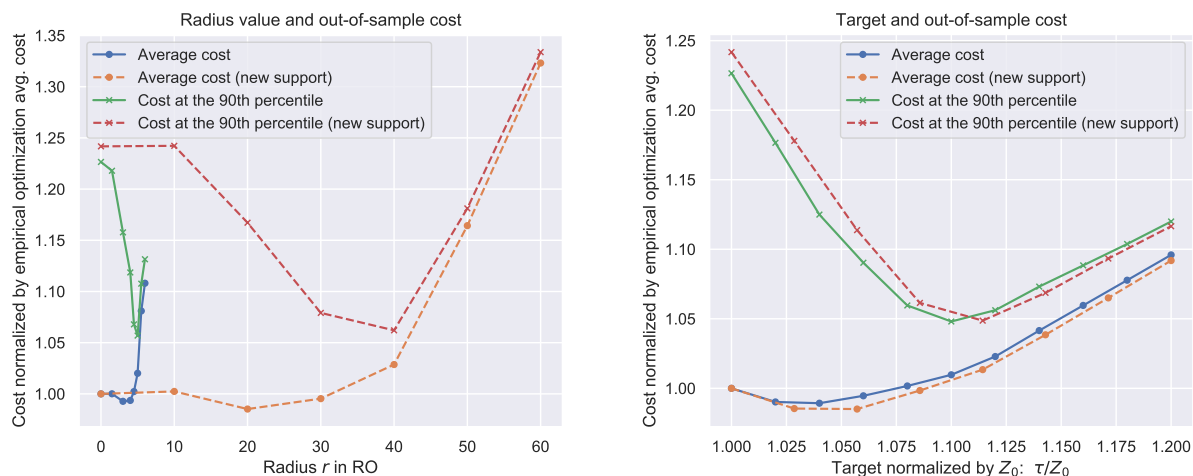
Figure 7   Summary of performance: RS and RO models at different first-stage costs

(a) Cost metrics w.r.t. Wasserstein radius r (b) Cost metrics w.r.t. target τ

Figure 8   Out-of-sample cost metrics w.r.t. Wasserstein radius r in RO (left) and target τ in RS (right).

(a) Cost metrics w.r.t. Wasserstein radius r (b) Cost metrics w.r.t. target τ

Figure 9   Change of the "nice" range of radius in RO (left) and normalized target in RS (right).

9. Conclusion

In this paper, we introduce the data-driven optimization robust satisficing framework that minimizes the model's fragility to uncertainty in achieving its prescribed target. We establish tractable robust satisficing models for risk-based linear optimization, combinatorial optimization, and linear optimization problems with recourse. In these problem contexts, we show that the robust satisficing model is at least as easy to solve as the corresponding data-driven robust optimization models. We perform numerical studies on portfolio optimization and network lot-sizing problems to elucidate that robust satisficing model is more effective than the data-driven robust optimization model in improving the out-of-sample performance. While we have illustrated the rationale of using our robust satisficing model and its potential benefits over several numerical experiments, we have not ruled out problem instances where robust optimization could attain better performance than robust satisficing. We hope our work could inspire others to explore robust satisficing as one of the candidates approaches, alongside robust optimization and stochastic programming, for addressing real world optimization problems under uncertainty.

Acknowledgement

References

Artzner, P., F. Delbaen, J.M. Eber, D. Heath. 1999. Coherent measures of risk. *Mathematical Finance* **9**(3) 203–228.

Aumann, R.J., R. Serrano. 2008. An economic index of riskiness. *Journal of Political Economy* **116**(5).

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

Ben-Tal, A., R. Brekelmans, D. den Hertog, J.P. Vial. 2017. Globalized robust optimization for nonlinear uncertain inequalities. *INFORMS Journal on Computing* **29**(2) 350–366.

Ben-Tal, A., D. den Hertog, A. De Waegenaere, B. Melenberg, G. Rennen. 2013. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science* **59**(2) 341–357.

Ben-Tal, A., D. den Hertog, J.P. Vial. 2015. Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical Programming* **149**(1-2) 265–299.

Ben-Tal, A., A. Goryashko, E. Guslitzer, A. Nemirovski. 2004. Adjustable robust solutions of uncertain linear programs. *Mathematical programming* **99**(2) 351–376.

Ben-Tal, A., A. 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, D., F.J. de Ruiter. 2016. Duality in two-stage adaptive linear optimization: faster computation and stronger bounds. *INFORMS Journal on Computing* **28**(3) 500–511.

Bertsimas, D., S. Shtern, B. Sturt. 2021. Two-stage sample robust optimization. *Operations Research* URL `https://arxiv.org/abs/1907.07142`.

Bertsimas, D., M. Sim. 2003. Robust discrete optimization and network flows. *Mathematical Programming* **98**(1) 49–71.

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

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

Bertsimas, Dimitris, Vishal Gupta, Nathan Kallus. 2018a. Data-driven robust optimization. *Mathematical Programming* **167**(2) 235–292.

Bertsimas, Dimitris, Vishal Gupta, Nathan Kallus. 2018b. Robust sample average approximation. *Mathematical Programming* **171**(1) 217–282.

Box, George EP. 1976. Science and statistics. *Journal of the American Statistical Association* **71**(356) 791–799.

Brown, D.B., E.D. Giorgi, M. Sim. 2012. Aspirational preferences and their representation by risk measures. *Management Science* **58**(11) 2095–2113.

Brown, D.B., M. Sim. 2009. Satisficing measures for analysis of risky positions. *Management Science* **55**(1) 71–84.

Charnes, Abraham, William W Cooper. 1959. Chance-constrained programming. *Management science* **6**(1) 73–79.

Chen, L.G., D.Z. Long, M. Sim. 2015. On dynamic decision making to meet consumption targets. *Operations Research* **63**(5) 1117–1130.

Chen, L.G., Q. Tang. 2019. Supply chain performance with a target-oriented retailer. *NUS Working Paper* .

Chen, X., M. Sim, P. Sun, J. Zhang. 2008. A linear decision-based approximation approach to stochastic programming. *Operations Research* **56**(2) 344–357.

Chen, Z., M. Sim, P. Xiong. 2020. Robust stochastic optimization made easy with RSOME. *Management science* URL `https://doi.org/10.1287/mnsc.2020.3603`.

Chen, Zhi, Daniel Kuhn, Wolfram Wiesemann. 2018. Data-driven chance constrained programs over wasserstein balls. *arXiv preprint arXiv:1809.00210* .

Dantzig, George B. 1955. Linear programming under uncertainty. *Management science* **1**(3-4) 197–206.

DeMiguel, V., L. Garlappi, R. Uppal. 2009. Optimal versus naive diversification: how inefficient is the 1/N portfolio strategy? *The Review of Financial Studies* **22**(5) 1915–1953.

Follmer, H., A. Schied. 2002. Convex measures of risk and trading constraints. *Finance and Stochastics* **6**(4) 429–447.

Follmer, H., A. Schied. 2004. *Stochastic finance: an introduction in discrete time*. Walter de Gruyter.

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, R., X. Chen, A.J. Kleywegt. 2017. Wasserstein distributional robustness and regularization in statistical learning URL `https://arxiv.org/abs/1712.06050`.

Gao, R., A.J. Kleywegt. 2016. Distributionally robust stochastic optimization with Wasserstein distance URL `https://arxiv.org/abs/1604.02199`.

Garstka, Stanley J, Roger J-B Wets. 1974. On decision rules in stochastic programming. *Mathematical Programming* **7**(1) 117–143.

Goh, J., M. Sim. 2010. Distributionally robust optimization and its tractable approximations. *Operations research* **58**(4) 902–917.

Gotoh, J.Y., M.J. Kim, A.E.B. Lim. 2018. Robust empirical optimization is almost the same as mean–variance optimization. *Operations research letters* **46**(4) 448–452.

Gotoh, J.Y., M.J. Kim, A.E.B. Lim. 2020. Calibration of distributionally robust empirical optimization models. *arXiv preprint arXiv:1711.06565v2* .

Hall, N.G., D.Z. Long, J. Qi, M. Sim. 2015. Managing underperformance risk in project portfolio selection. *Operations Research* **63**(3) 660–675.

Hanasusanto, Grani A, Daniel Kuhn, Wolfram Wiesemann. 2016. A comment on "computational complexity of stochastic programming problems". *Mathematical Programming* **159**(1) 557–569.

Iancu, D.A., N. Trichakis. 2014. Pareto efficiency in robust optimization. *Management Science* **60**(1) 131–147.

Jaillet, P., J. Qi, M. Sim. 2016. Routing optimization under uncertainty. *Operations Research* **64**(1) 186–200.

Kleywegt, A.J., A. Shapiro, T. Homem-de Mello. 2002. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization* **12**(2) 479–502.

Kuhn, D., W. Wiesemann, A. Georghiou. 2011. Primal and dual linear decision rules in stochastic and robust optimization. *Mathematical Programming* **130**(1) 177–209.

Luedtke, J., S. Ahmed, G.L. Nemhauser. 2010. An integer programming approach for linear programs with probabilistic constraints. *Mathematical programming* **122**(2) 247–272.

Mao, J.C.T. 1970. Survey of capital budgeting: Theory and practice. *The Journal of Finance* **25**(2) 349–360.

Markowitz, H.M. 1952. Portfolio selection. *Journal of Finance* **7**(1) 77–91.

Mohajerin Esfahani, P., D. Kuhn. 2018. Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. *Mathematical Programming* **171**(1-2) 1–52.

Nemirovski, Arkadi, Alexander Shapiro. 2007. Convex approximations of chance constrained programs. *SIAM Journal on Optimization* **17**(4) 969–996.

Pardo, L. 2006. *Statistical inference based on divergence measures*. Chapman and Hall/CRC.

Pflug, G.C., A. Pichler, D. Wozabal. 2012. The 1/N investment strategy is optimal under high model ambiguity. *Journal of Banking and Finance* **36**(2) 410–417.

Rockafellar, R.T., S. Uryasev. 2002. Conditional value-at-risk for general loss distributions. *Journal of banking & finance* **26**(7) 1443–1471.

Schwartz, Barry, YAKOV Ben-Haim, Cliff Dacso. 2011. What makes a good decision? Robust satisficing as a normative standard of rational decision making. *Journal for the Theory of Social Behaviour* **41**(2) 209–227.

Shapiro, Alexander, Arkadi Nemirovski. 2005. On complexity of stochastic programming problems. *Continuous optimization*. Springer, 111–146.

Simon, H.A. 1959. Theories of decision-making in economics and behavioral science. *The American Economic Review* **49**(3) 253–283.

Sion, Maurice, et al. 1958. On general minimax theorems. *Pacific Journal of mathematics* **8**(1) 171–176.

Smith, J.E., R.L. Winkler. 2006. The optimizer's curse: skepticism and postdecision surprise in decision analysis. *Management Science* **52**(3) 311–322.

Sturt, B. 2021. A nonparametric algorithm for optimal stopping based on robust optimization. *working paper* .

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

Zhang, Y., R. Baldacci, M. Sim, J. Tang. 2019. Routing optimization with time windows under uncertainty. *Mathematical Programming* **175**(1) 263–305.

Zhen, J., D. den Hertog, M. Sim. 2018. Adjustable robust optimization via fourier–motzkin elimination. *Operations Research* **66**(4) 1086–1100.

A. Proof of Results

Proof of Proposition 2.   The robust satisficing model under this setting is given by

$$\kappa_\tau = \min k$$
$$\text{s.t. } \frac{1}{S} \sum_{s \in [S]} y_s \leq \tau$$
$$y_s \geq \sup_{\boldsymbol{z}_s \in \mathcal{Z}} \{\max_{i \in [I]} \{a_i \boldsymbol{x}^\top \boldsymbol{z} + b_i\} - k\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}_s\|\} \quad \forall s \in [S]$$
$$\boldsymbol{x} \in \mathcal{X}, k \geq 0.$$

First, for any $s \in [s]$, we focus on

$$y_s \geq \sup_{\boldsymbol{z}_s \in \mathcal{Z}} \{\max_{i \in [I]} \{a_i \boldsymbol{x}^\top \boldsymbol{z} + b_i\} - k\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}_s\|\},$$

which is then equivalent to:

$$y_s \geq \sup_{\boldsymbol{z}_s \in \mathcal{Z}} \{a_i \boldsymbol{x}^\top \boldsymbol{z} + b_i - k\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}_s\|\} \quad \forall i \in [I].$$

Now, for any $i \in [I], s \in [S]$, we focus on the inner maximization:

$$\sup_{\boldsymbol{z}_s \in \mathcal{Z}} \{a_i \boldsymbol{x}^\top \boldsymbol{z} + b_i - k\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}_s\|\}$$
$$= \sup_{\boldsymbol{z} \in \mathcal{Z}} \left\{ a_i \boldsymbol{x}^\top \boldsymbol{z} + b_i - \sup_{\|\boldsymbol{\gamma}_{is}\|_* \leq k} \boldsymbol{\gamma}_{is}^\top (\boldsymbol{z} - \hat{\boldsymbol{z}}_s) \right\}$$
$$= \sup_{\boldsymbol{z} \in \mathcal{Z}} \inf_{\|\boldsymbol{\gamma}_{is}\|_* \leq k} \left\{ a_i \boldsymbol{x}^\top \boldsymbol{z} + b_i - \boldsymbol{\gamma}_{is}^\top (\boldsymbol{z} - \hat{\boldsymbol{z}}_s) \right\}$$
$$= \inf_{\|\boldsymbol{\gamma}_{is}\|_* \leq k} \sup_{\boldsymbol{z} \in \mathcal{Z}} \left\{ a_i \boldsymbol{x}^\top \boldsymbol{z} + b_i - \boldsymbol{\gamma}_{is}^\top (\boldsymbol{z} - \hat{\boldsymbol{z}}_s) \right\},$$

where the interchange of maximization and minimization follows because the function is affine in both variables, the set $\{\boldsymbol{\gamma}_{is} \in \mathbb{R}^N \mid \|\boldsymbol{\gamma}_{is}\|_* \leq k\}$ is compact, and the set $\mathcal{Z}$ is convex and closed. Now, we focus on the inner maximization. For any $i \in [I], s \in [S]$, let $Z_{is} = \sup_{\boldsymbol{z} \in \mathcal{Z}} \left\{ a_i \boldsymbol{x}^\top \boldsymbol{z} - \boldsymbol{\gamma}_{is}^\top \boldsymbol{z} \right\}$, where $\mathcal{Z} = \{\boldsymbol{z} \in \mathbb{R}^N \mid \boldsymbol{C}\boldsymbol{z} \leq \boldsymbol{h}\}$. By strong duality, we have:

$$Z_{is} = \inf \boldsymbol{\eta}_{is}^\top \boldsymbol{h}$$
$$\text{s.t. } \boldsymbol{C}^\top \boldsymbol{\eta}_{is} = a_i \boldsymbol{x} - \boldsymbol{\gamma}_{is}$$
$$\boldsymbol{\eta}_{is} \geq \boldsymbol{0}.$$

Finally, by above reformulations, the robust satisficing model becomes:

$$\min k$$
$$\text{s.t. } \frac{1}{S} \sum_{s \in [S]} y_s \leq \tau$$
$$y_s \geq a_i \boldsymbol{x}^\top \hat{\boldsymbol{z}}_s + b_i + \boldsymbol{\eta}_{is}^\top \boldsymbol{h} - \boldsymbol{\eta}_{is}^\top \boldsymbol{C} \hat{\boldsymbol{z}}_s \quad \forall s \in [S], \ i \in [I]$$
$$k \geq \|a_i \boldsymbol{x} - \boldsymbol{C}^\top \boldsymbol{\eta}_{is}\|_* \qquad\qquad \forall s \in [S], \ i \in [I]$$
$$\boldsymbol{\eta}_{is} \geq \boldsymbol{0} \qquad\qquad\qquad\qquad \forall s \in [S], \ i \in [I]$$
$$\boldsymbol{x} \in \mathcal{X},$$

which gives the final formulation in Proposition 2. $\square$

Proof of Proposition 3. By the definition of Problem (6), we know

$$\mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right] \leq \tau + k\Delta_W(\mathbb{P}, \hat{\mathbb{P}}) \quad \forall \mathbb{P} \in \mathcal{P}_0(\mathcal{Z}). \tag{35}$$

Therefore,

$$\mathbb{P}^S\left[\mathbb{E}_{\mathbb{P}^*}\left[f\left(\boldsymbol{x}, \tilde{\boldsymbol{z}}\right)\right] > \tau + kr\right] \leq \mathbb{P}^S\left[k\Delta_W(\mathbb{P}^*, \hat{\mathbb{P}}) > kr\right] \leq \mathbb{P}^S\left[\Delta_W(\mathbb{P}^*, \hat{\mathbb{P}}) > r\right],$$

where the inequality and equality are due to (35). Hence, our result follows from Theorem 1. $\qquad\square$

Proof of Proposition 4. Note that $\sum_{i \in [N]} 2\max\{z_i - x_i, 0\}$ is convex in $\boldsymbol{z}$; hence, the worst-case scenario must occur at the extreme points of the uncertainty set. Specifically, the worst-case scenario must satisfy that $\|\boldsymbol{z}\|_1 = r$ and at least $N-1$ components of $\boldsymbol{z}$ are either 0 or d. In addition, demand could only occur at $\lceil r/d \rceil$ locations. Because $r \leq d\lceil N/2 - 1\rceil$, we know $\lceil r/d \rceil \leq \lceil N/2 - 1\rceil$.

Consider any here-and-now ordering decision $\bar{\boldsymbol{x}}$. Without loss of generality, we order the components of $\bar{\boldsymbol{x}}$ with respect to the ordering quantity in a descending order, *i.e.*, $\bar{x}_1 \geq \bar{x}_2 \geq \cdots \geq \bar{x}_N \geq 0$. In this case, the worst-case demand would occur only at the last $\lceil r/d \rceil$ locations such that $z_{N-\lceil r/d \rceil+1} = r - d\lceil r/d \rceil$ and $z_{N-\lceil r/d \rceil+2} = \cdots = z_N = d$.

Suppose $\bar{\boldsymbol{x}}$ only orders inventory at less than or equal to $N - \lceil r/d \rceil$ locations, *i.e.*, $\bar{x}_{N-\lceil r/d \rceil+1} = \cdots = \bar{x}_N = 0$. One could see that such an ordering decision is strictly dominated by a no-ordering decision. Suppose $\bar{\boldsymbol{x}}$ only orders inventory at more than $N - \lceil r/d \rceil$ locations, *i.e.*, $\bar{x}_{N-\lceil r/d \rceil+1} > 0$. By reverting to a no-ordering decision, we would save a here-and-now ordering cost of

$$\sum_{i \in [N]} x_i,$$

while incurring an additional emergency fulfillment cost of at most

$$2\sum_{i \in [N]\setminus[N-\lceil r/d \rceil]} x_i \leq \sum_{i \in [N]} x_i.$$

The last inequality is due to that $N \geq 2\lceil r/d \rceil$ and the ordering of $\bar{\boldsymbol{x}}$. Hence, any non-zero ordering decision is dominated by a no-ordering decision. $\qquad\square$

Proof of Proposition 5. To prove this proposition, we consider two candidate solution $\boldsymbol{x}_1 = \boldsymbol{0}$ and $\boldsymbol{x}_2 = \tau\boldsymbol{1}/N$ and show that $\boldsymbol{x}_2$ would always achieve a better objective value than $\boldsymbol{x}_1$.

When we fix the here-and-now solution as $\boldsymbol{x}_1 = \boldsymbol{0}$, we have

$$\kappa_\tau^{(1)} = \max_{\boldsymbol{z} \in [0,d]^N}\left\{2 - \frac{\tau}{\|\boldsymbol{z}\|_1}\right\} = 2 - \frac{\tau}{Nd}.$$

When we fix the here-and-now solution as $\boldsymbol{x}_2 = \tau \boldsymbol{1}/N$, we have

$$\kappa_\tau^{(2)} = \max_{\boldsymbol{z} \in [0,d]^N} \left\{ \frac{2 \sum\limits_{i \in [N]} \max\{z_i - \tau/N, 0\}}{\|\boldsymbol{z}\|_1} \right\} = 2 - \frac{2\tau}{Nd}.$$

Because $\kappa_\tau^{(2)} < \kappa_\tau^{(1)}$, we know that $\boldsymbol{x}_2$ is strictly better than $\boldsymbol{x}_1$. Hence, the optimal here-and-now ordering quantity in the above robust satisficing network lot-sizing problem is non-zero. $\qquad\square$

Proof of Theorem 2. Given a fragility measure ρ defined by Equation (11), we first show that it has all the five properties in this theorem and is lower semi-continuous, *i.e.*, $\{\tilde{v} \mid \rho(\tilde{v}) \leq a\}$ is a closed set for any $a \geq 0$. For convenience, we define

$$\mathcal{K}(\tilde{v}) := \left\{ k \geq 0 \ \mid \ \mathbb{E}_\mathbb{P}[\tilde{v}] \leq k\Delta(\mathbb{P}, \hat{\mathbb{P}}), \ \ \forall \mathbb{P} \in \mathcal{P}_0 \right\}, \tag{36}$$

and hence $\rho(\tilde{v}) = \inf \mathcal{K}(\tilde{v})$.

1. **Monotonicity**. If $\tilde{v}_1 \geq \tilde{v}_2$, then $\mathbb{E}_\mathbb{P}[\tilde{v}_1] \geq \mathbb{E}_\mathbb{P}[\tilde{v}_2]$ for any $\mathbb{P} \in \mathcal{P}_0$. That is, for any $k \in \mathcal{K}(\tilde{v}_1)$, we must have $k \in \mathcal{K}(\tilde{v}_2)$. Therefore, $\mathcal{K}(\tilde{v}_1) \subseteq \mathcal{K}(\tilde{v}_2)$. Taking the infimum gives $\rho(\tilde{v}_1) \geq \rho(\tilde{v}_2)$.

2. **Positive homogeneity**. The case of $\lambda = 0$ is trivial. Consider any $\lambda > 0$. Notice that

$$\begin{aligned} \rho(\lambda\tilde{v}) &= \inf\left\{ k > 0 \ \mid \ \mathbb{E}_\mathbb{P}[\lambda\tilde{v}] \leq k\Delta(\mathbb{P}, \hat{\mathbb{P}}), \ \ \forall \mathbb{P} \in \mathcal{P}_0 \right\} \\ &= \inf\left\{ k > 0 \ \mid \ \mathbb{E}_\mathbb{P}[\tilde{v}] \leq \tfrac{k}{\lambda}\Delta(\mathbb{P}, \hat{\mathbb{P}}), \ \ \forall \mathbb{P} \in \mathcal{P}_0 \right\} \\ &= \lambda\inf\left\{ \beta > 0 \ \mid \ \mathbb{E}_\mathbb{P}[\tilde{v}] \leq \beta\Delta(\mathbb{P}, \hat{\mathbb{P}}), \ \ \forall \mathbb{P} \in \mathcal{P}_0 \right\} \\ &= \lambda\rho(\tilde{v}). \end{aligned}$$

3. **Subadditivity**. Suppose $k_1 \in \mathcal{K}(\tilde{v}_1)$ and $k_2 \in \mathcal{K}(\tilde{v}_2)$. It is not hard to see that

$$\mathbb{E}_\mathbb{P}[\tilde{v}_1 + \tilde{v}_2] \leq (k_1 + k_2)\Delta(\mathbb{P}, \hat{\mathbb{P}}), \ \ \forall \mathbb{P} \in \mathcal{P}_0,$$

which indicates $(k_1 + k_2) \in \mathcal{K}(\tilde{v}_1 + \tilde{v}_2)$. The subadditivity then follows by taking the infimum.

4. **Pro-robustness**. If $\tilde{v} \leq 0$, then for all $\mathbb{P} \in \mathcal{P}_0$ and $k > 0$ we have $\mathbb{E}_\mathbb{P}[\tilde{v}] \leq 0 \leq k\Delta(\mathbb{P}, \hat{\mathbb{P}})$. That implies $\rho(\tilde{v}) = 0$.

5. **Anti-fragility**. Suppose $\mathbb{E}_{\hat{\mathbb{P}}}[\tilde{v}] > 0$, then $\mathcal{K}(\tilde{v}) = \emptyset$ because $\Delta(\hat{\mathbb{P}}, \hat{\mathbb{P}}) = 0$. Therefore, $\rho(\tilde{v}) = \inf \mathcal{K}(\tilde{v}) = \inf \emptyset = \infty$.

The $\sigma(L^\infty(\hat{\mathbb{P}}), L^1(\hat{\mathbb{P}}))$-lower semi-continuity of ρ can be shown as follows. Consider any converging sequence of random variable $\tilde{v}_1, \ldots, \tilde{v}_n$ such that $\tilde{v}_n \to \tilde{v}$ in probability as $n \to +\infty$. For any fixed value $a \geq 0$, we need to show that $\rho(\tilde{v}) \leq a$ if $\rho(\tilde{v}_n) \leq a$ for all $n > 0$. Note that we assume that $\mathcal{L}$ is the space of bounded real-valued function, and $|\tilde{v}| \leq M$ for some positive constant M for all $\tilde{v} \in \mathcal{L}$. Under this assumption, the above statement is true because $\lim\limits_{n \to +\infty} \mathbb{E}_{\hat{\mathbb{P}}}[\tilde{v}_n] = \mathbb{E}_{\hat{\mathbb{P}}}[\tilde{v}]$.

We now prove that the function Δ defined by Equation (14) is exactly the probability distribution function associated with ρ. To this end, we first show that the function Δ defined by Equation (14) is a probability distance function. Then, we show with such Δ, we have $\inf \mathcal{K}(\tilde{v}) = \rho(\tilde{v})$, *i.e.*, the representation (14) is valid.

We note that since $\rho(0) = 0 < 1$ by pro-robustness property, $\Delta(\mathbb{P}, \hat{\mathbb{P}}) \geq \mathbb{E}_{\mathbb{P}}[0] = 0$ for all $\mathbb{P} \in \mathcal{P}_0$. Moreover, due to the property of anti-fragility, $\rho(\tilde{v}) = \infty > 1$ for all $\tilde{v}$ with $\mathbb{E}_{\hat{\mathbb{P}}}[\tilde{v}] > 0$. Therefore, by Equation (14),

$$\Delta(\hat{\mathbb{P}}, \hat{\mathbb{P}}) = \sup_{\tilde{v} \in \mathcal{L}} \{ \mathbb{E}_{\hat{\mathbb{P}}}[\tilde{v}] \mid \rho(\tilde{v}) \leq 1, \mathbb{E}_{\hat{\mathbb{P}}}[\tilde{v}] \leq 0 \} \leq 0;$$

together with $\Delta(\mathbb{P}, \hat{\mathbb{P}}) \geq 0$ for all $\mathbb{P} \in \mathcal{P}_0$ we know $\Delta(\hat{\mathbb{P}}, \hat{\mathbb{P}}) = 0$. Hence, Δ is a probability distance function.

Now, given any $\tilde{v} \in \mathcal{L}$ and with Δ defined as in Equation (14), we define the set $\mathcal{K}(\tilde{v})$ as in Equation (36). It remains to show $\inf \mathcal{K}(\tilde{v}) = \rho(\tilde{v})$. To prove the result, we start from the case where $\rho(\tilde{v}) \in (0, \infty)$.

We first show $\inf \mathcal{K}(\tilde{v}) \leq \rho(\tilde{v})$. Consider any $k \geq \rho(\tilde{v})$. By Positive homogeneity, $\rho(\tilde{v}/k) = \rho(\tilde{v})/k \leq 1$. Given any $\mathbb{P} \in \mathcal{P}_0$, $\Delta(\mathbb{P}, \hat{\mathbb{P}}) = \sup_{\tilde{w} \in \mathcal{L}} \{ \mathbb{E}_{\mathbb{P}}[\tilde{w}] \mid \rho(\tilde{w}) \leq 1 \} \geq \mathbb{E}_{\mathbb{P}}[\tilde{v}/k]$, which implies $\mathbb{E}_{\mathbb{P}}[\tilde{v}] \leq k\Delta(\mathbb{P}, \hat{\mathbb{P}})$. Hence, $k \in \mathcal{K}(\tilde{v})$. This indicates $\inf \mathcal{K}(\tilde{v}) \leq \rho(\tilde{v})$.

We then show $\inf \mathcal{K}(\tilde{v}) \geq \rho(\tilde{v})$. Consider any $0 < k < \rho(\tilde{v})$, and hence $\rho(\tilde{v}/k) = \rho(\tilde{v})/k > 1$. Denote a set $\mathcal{W} = \{ \tilde{w} \in \mathcal{L} \mid \rho(\tilde{w}) \leq 1 \}$. We next apply Hahn-Banach separation theorem similarly to the proof for Theorem 6 in Follmer and Schied (2002). Specifically, by the convexity and $\sigma(L^\infty(\hat{\mathbb{P}}), L^1(\hat{\mathbb{P}}))$-lower semi-continuity of ρ, $\mathcal{W}$ is a weak*-closed, convex set and $\tilde{v}/k \notin \mathcal{W}$. Therefore, by Hahn-Banach separation theorem in the locally convex space $(L^\infty(\hat{\mathbb{P}}), \sigma(L^\infty(\hat{\mathbb{P}}), L^1(\hat{\mathbb{P}})))$, there exists a linear functional l with

$$\infty > l(\tilde{v}/k) > \beta > l(\tilde{w}), \quad \forall \tilde{w} \in \mathcal{W}$$

for some $\beta \in \mathbb{R}$. Consider any $\tilde{w} \leq \epsilon$ with a $\epsilon < 0$, then for all $\lambda > 0$, $\lambda \tilde{w} \leq \lambda \epsilon < 0$ and hence $\lambda \tilde{w} \in \mathcal{W}$ by pro-robustness property. Therefore, $\beta > l(\lambda \tilde{w}) = \lambda l(\tilde{w})$, where the equality holds since l is a linear functional. As it is true for all $\lambda > 0$, we know $l(\tilde{w}) \leq 0$. It further implies that l is a positive linear functional. WLOG, we can normalize l such that $l(1) = 1$. In this case there exists $\mathbb{P} \in \mathcal{P}_0$ such that $l(\tilde{w}) = \mathbb{E}_{\mathbb{P}}[\tilde{w}]$ for all $\tilde{w} \in \mathcal{L}$. With this particular $\mathbb{P}$, $\mathbb{E}_{\mathbb{P}}[\tilde{v}/k] > \beta \geq \sup_{\tilde{w} \in \mathcal{W}} \mathbb{E}_{\mathbb{P}}[\tilde{w}] = \Delta(\mathbb{P}, \hat{\mathbb{P}})$. This indicates $\mathbb{E}_{\mathbb{P}}[\tilde{v}] > k\Delta(\mathbb{P}, \hat{\mathbb{P}})$ and $k \notin \mathcal{K}(\tilde{v})$. Therefore, $\inf \mathcal{K}(\tilde{v}) \geq \rho(\tilde{v})$.

We hence conclude $\inf \mathcal{K}(\tilde{v}) = \rho(\tilde{v})$ whenever $\rho(\tilde{v}) \in (0, \infty)$. For the case of $\rho(\tilde{v}) = 0$, we just need the above proof of $\inf \mathcal{K}(\tilde{v}) \leq \rho(\tilde{v})$ to conclude $\inf \mathcal{K}(\tilde{v}) = 0$. For the case of $\rho(\tilde{v}) = \infty$, we just need the above proof of $\inf \mathcal{K}(\tilde{v}) \geq \rho(\tilde{v})$ to conclude $\inf \mathcal{K}(\tilde{v}) = \infty$. $\qquad \square$

Proof of Proposition 6. The proof follows largely from existing proofs, see, *e.g.,* Hall et al. (2015). For the purpose of completeness, we include the proof here nevertheless.

First, we prove the "if" direction. Suppose there exists some normalized convex risk measure μ as stated in the proposition, and let $\rho(\tilde{v}) = \inf\{k > 0 \mid k\mu(\tilde{v}/k) \leq 0\}$. For convenience, define $\mathcal{K}_\mu(\tilde{v}) := \{k > 0 \mid k\mu(\tilde{v}/k) \leq 0\}$. We show that ρ has the desired properties as follows.

1. Monotonicity. If $\tilde{v}_1 \geq \tilde{v}_2$, then $\forall k > 0$, we have $\tilde{v}_1/k \geq \tilde{v}_2/k$, $k\mu(\tilde{v}_1/k) \geq k\mu(\tilde{v}_2/k)$. Hence, for any $k \in \mathcal{K}_\mu(\tilde{v}_1)$, we must have $k \in \mathcal{K}_\mu(\tilde{v}_2)$. Taking the infimum gives us $\rho(\tilde{v}_1) \geq \rho(\tilde{v}_2)$.

2. Positive homogeneity. The case of $\lambda = 0$ is trivial and now we consider any $\lambda > 0$. $\rho(\lambda\tilde{v}) = \inf\{k > 0|k\mu(\lambda\tilde{v}/k) \leq 0\} = \inf\{\lambda\hat{k} > 0|\lambda\hat{k}\mu(\lambda\tilde{v}/\lambda\hat{k}) \leq 0\} = \lambda\inf\{\hat{k} > 0|\hat{k}\mu(\tilde{v}/\hat{k}) \leq 0\} = \lambda\rho(\tilde{v})$.

3. Subadditivity. Consider any $\tilde{v}_1, \tilde{v}_2 \in \mathcal{L}$ and $k_1 \in \mathcal{K}_\mu(\tilde{v}_2), k_2 \in \mathcal{K}_\mu(\tilde{v}_2)$. Then, by the convexity of μ, we have $(k_1 + k_2)\mu((k_1\tilde{v}_1/k_1 + k_2\tilde{v}_2/k_2)/(k_1 + k_2)) \leq k_1\mu(\tilde{v}_1/k_1) + k_2\mu(\tilde{v}_2/k_2) \leq 0$. Therefore, we have $(k_1 + k_2) \in \mathcal{K}_\mu(\tilde{v}_2 + \tilde{v}_2)$. Taking the infimum gives us $\rho(\tilde{v}_2 + \tilde{v}_2) \leq \rho(\tilde{v}_1) + \rho(\tilde{v}_2)$.

4. Pro-robustness. Because $\tilde{v} \leq 0$, we have $\mu(\tilde{v}/k) \leq \mu(0) = 0$ for any $k > 0$. Therefore, $k\mu(\tilde{v}/k) \leq 0$ for all $k > 0$, which indicates $\rho(\tilde{v}) = 0$.

5. Anti-fragility. Suppose $\mathbb{E}_{\hat{\mathbb{P}}}[\tilde{v}] > 0$ and consider any $k > 0$. Then $\mathbb{E}_{\hat{\mathbb{P}}}[\tilde{v}/k] > 0$, and hence $\mu(\tilde{v}/k) \geq \mathbb{E}_{\hat{\mathbb{P}}}[\tilde{v}/k] > 0$. Therefore, $\mathcal{K}_\mu(\tilde{v})$ is an empty set. Taking the infimum gives us $\rho(\tilde{v}) = +\infty$.

The lower semi-continuity of ρ can be shown as follows. Consider any converging sequence of random variable $\tilde{v}_1, \ldots, \tilde{v}_n$ such that $\tilde{v}_n \to \tilde{v}$ as $n \to +\infty$. For any fixed value $a \geq 0$, we need to show that $\rho(\tilde{v}) \leq a$ if $\rho(\tilde{v}_n) \leq a$ for all $n > 0$. This is true because μ is a lower semi-continuous function, *i.e.,* $\lim_{n \to +\infty} \mu(\tilde{v}_n) = \mu(\tilde{v})$. Specifically, if $\rho(\tilde{v}_n) \leq a$ for all $n > 0$, then $\mu(\tilde{v}_n/a) \leq 0$ for all $n > 0$, indicating $\mu(\tilde{v}/a) \leq 0$. Therefore, we have $\rho(\tilde{v}) \leq a$.

We now prove the "only if" direction. Consider any fragility measure ρ. Then we need to prove that there exists a normalized convex risk measure μ satisfying $\mu(\tilde{v}) \geq \mathbb{E}_{\hat{\mathbb{P}}}[\tilde{v}]$ and $\rho(\tilde{v}) = \inf\{k > 0 \mid k\mu(\tilde{v}/k) \leq 0\} \; \forall \tilde{v} \in \mathcal{L}$. To this end, we consider the μ defined as follows,

$$\mu(\tilde{v}) = \inf\{a \mid \rho(\tilde{v} - a) \leq 1\}.$$

We now show that such μ satisfies the requirement. First, we prove that this μ is a convex risk measure. Define $\mathcal{K}_\rho(\tilde{v}) := \{a \mid \rho(\tilde{v} - a) \leq 1\}$.

1. Monotonicity. For any $\tilde{v}_1 \geq \tilde{v}_2$, we have $\rho(\tilde{v}_1 - a) \geq \rho(\tilde{v}_2 - a)$ for all $a \in \mathbb{R}$. Then, for any $a \in \mathcal{K}_\rho(\tilde{v}_1)$, we must have $a \in \mathcal{K}_\rho(\tilde{v}_2)$. Taking the infimum gives us $\mu(\tilde{v}_1) \geq \mu(\tilde{v}_2)$.

2. Translation invariance. For any $a' \in \mathbb{R}$, we have $\mu(\tilde{v} + a') = \inf\{a \mid \rho(\tilde{v} - (a - a')) \leq 1\} = a' + \inf\{a - a' \mid \rho(\tilde{v} - (a - a')) \leq 1\} = a' + \mu(\tilde{v})$.

3. Convexity. Consider any $\tilde{v}_1, \tilde{v}_2 \in \mathcal{L}$, $\lambda \in [0,1]$, $a_1 \in \mathcal{K}_\rho(\tilde{v}_1)$ and $a_2 \in \mathcal{K}_\rho(\tilde{v}_2)$. Then, $\rho(\tilde{v}_1 - a_1) \leq 1$ and $\rho(\tilde{v}_2 - a_2) \leq 1$. Therefore, $\rho(\lambda \tilde{v}_1 + (1-\lambda)\tilde{v}_2 - (\lambda a_1 + (1-\lambda)a_2)) = \rho(\lambda(\tilde{v}_1 - a_1) + (1-\lambda)(\tilde{v}_2 - a_2)) \leq \lambda \rho(\tilde{v}_1 - a_1) + (1-\lambda)\rho(\tilde{v}_2 - a_2) \leq 1$, where the first inequality is due to Subadditivity and Positive homogeneity of ρ. It indicates $\lambda a_1 + (1-\lambda)a_2 \in \mathcal{K}_\rho(\lambda \tilde{v}_1 + (1-\lambda)\tilde{v}_2)$. Taking the infimum gives us the convexity.

4. Normalization. Because of the property of anti-fragility, we have $\rho(a') = +\infty$ for all $a' > 0$. Therefore, $\mu(0) = 0$.

To show $\mu(\tilde{v}) \geq \mathbb{E}_{\hat{\mathbb{P}}}[\tilde{v}]$, notice that $\forall a < \mathbb{E}_{\hat{\mathbb{P}}}[\tilde{v}]$, $\mathbb{E}_{\hat{\mathbb{P}}}[\tilde{v} - a] > 0$ and hence $\rho(\tilde{v} - a) = \infty$, $a \notin \mathcal{K}_\rho(\tilde{v})$. Therefore, $\mu(\tilde{v}) = \inf \mathcal{K}_\rho(\tilde{v}) \geq \mathbb{E}_{\hat{\mathbb{P}}}[\tilde{v}]$.

To show $\rho(\tilde{v}) = \inf \{k > 0 \mid k\mu(\tilde{v}/k) \leq 0\}$, we observe $\inf \{k > 0 \mid k\mu(\tilde{v}/k) \leq 0\} = \inf\{k > 0 | \exists a \leq 0 : \rho(\tilde{v}/k - a) \leq 1\} = \inf\{k > 0 | \rho(\tilde{v}/k) \leq 1\} = \inf\{k > 0 | \rho(\tilde{v}) \leq k\} = \rho(\tilde{v})$. $\qquad\square$

Proof of Theorem 3. For generality, we consider a support set $\mathcal{Z} = [\underline{z}, \bar{z}]$ in this proof. For any $i \in [I]$, the robust satisficing constraint in Problem (18) can be written as:

$$\mathbb{E}_{\mathbb{Q}}\left[\alpha_i + \frac{1}{\epsilon}(\boldsymbol{a}_i(\tilde{\boldsymbol{z}})'\boldsymbol{x} - b_i(\tilde{\boldsymbol{z}}) - \alpha_i)^+ - k_i\|\tilde{\boldsymbol{z}} - \tilde{\boldsymbol{u}}\|\right] \leq 0 \quad \forall \mathbb{Q} \in \mathcal{Q},$$

where

$$\mathcal{Q} := \left\{\mathbb{Q} \in \mathcal{P}_0(\mathcal{Z}^2) \ \middle| \ (\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{v}}) \sim \mathbb{Q}, \ \tilde{\boldsymbol{z}} \sim \mathbb{P}, \ \tilde{\boldsymbol{v}} \sim \hat{\mathbb{P}} \right\}.$$

The above constraint can be equivalently written as

$$\sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}\left[\alpha_i + \frac{1}{\epsilon}(\boldsymbol{a}_i(\tilde{\boldsymbol{z}})'\boldsymbol{x} - b_i(\tilde{\boldsymbol{z}}) - \alpha_i)^+ - k_i\|\tilde{\boldsymbol{z}} - \tilde{\boldsymbol{u}}\|\right] \leq 0$$
$$\Longleftrightarrow \alpha_i + \frac{1}{S\epsilon} \sum_{s \in [S]} \sup_{\boldsymbol{z}_s \in \mathcal{Z}} \{(\boldsymbol{a}_i(\boldsymbol{z}_s)'\boldsymbol{x} - b_i(\boldsymbol{z}_s) - \alpha_i)^+ - k_i\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}_s\|\} \leq 0.$$

Similarly, for $i = 0$, we have

$$\mathbb{E}_{\mathbb{Q}}\left[\alpha_0 + \frac{1}{\epsilon}(-\boldsymbol{c}(\tilde{\boldsymbol{z}})'\boldsymbol{x} - \alpha_0)^+ - k_0\|\tilde{\boldsymbol{z}} - \tilde{\boldsymbol{u}}\|\right] \leq 0 \quad \forall \mathbb{Q} \in \mathcal{Q}$$
$$\Longleftrightarrow \alpha_0 + \frac{1}{S\epsilon} \sum_{s \in [S]} \sup_{\boldsymbol{z}_s \in \mathcal{Z}} \{(-\boldsymbol{c}(\boldsymbol{z}_s)'\boldsymbol{x} - \alpha_0)^+ - k_0\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}_s\|\} \leq 0.$$

Hence, the model can be written as:

$$
\begin{aligned}
\min \quad & \sum_{i \in [I] \cup \{0\}} w_i k_i \\
\text{s.t.} \quad & \alpha_i + \frac{1}{S\epsilon} \sum_{s \in [S]} y_{is} \leq 0 && \forall i \in [I] \\
& \alpha_0 + \frac{1}{S\epsilon} \sum_{s \in [S]} y_{0s} + \tau \leq 0 \\
& y_{is} \geq \sup_{\boldsymbol{z}_s \in \mathcal{Z}} -k_i \|\boldsymbol{z}_s - \hat{\boldsymbol{z}}_s\| && \forall i \in [I] \cup \{0\}, \ s \in [S] \\
& y_{is} \geq \sup_{\boldsymbol{z}_s \in \mathcal{Z}} \{\boldsymbol{a}_i(\boldsymbol{z}_s)'\boldsymbol{x} - b_i(\boldsymbol{z}_s) - \alpha_i - k_i \|\boldsymbol{z}_s - \hat{\boldsymbol{z}}_s\|\} && \forall i \in [I], \ s \in [S] \\
& y_{0s} \geq \sup_{\boldsymbol{z}_s \in \mathcal{Z}} \{-\boldsymbol{c}(\boldsymbol{z}_s)'\boldsymbol{x} - \alpha_0 - k_0 \|\boldsymbol{z}_s\} - \hat{\boldsymbol{z}}_s\| && \forall s \in [S] \\
& \boldsymbol{x} \geq \boldsymbol{0} \\
& \alpha_i \in \mathbb{R}, \ k_i \geq 0, \ y_{is} \geq 0 && \forall i \in [I] \cup \{0\}, s \in [S].
\end{aligned}
$$

First, we focus on $\sup_{\boldsymbol{z}_s \in \mathcal{Z}} -k_i \|\boldsymbol{z}_s - \hat{\boldsymbol{z}}_s\|$. For any $i \in [I] \cup \{0\}, s \in [S]$, we have

$$
\sup_{\boldsymbol{z}_s \in \mathcal{Z}} -k_i \|\boldsymbol{z}_s - \hat{\boldsymbol{z}}_s\| = 0.
$$

Then, we focus on $\sup_{\boldsymbol{z}_s \in \mathcal{Z}} \{\boldsymbol{a}_i(\boldsymbol{z}_s)'\boldsymbol{x} - b_i(\boldsymbol{z}_s) - \alpha_i - k_i \|\boldsymbol{z}_s - \hat{\boldsymbol{z}}_s\|\}$. For any $i \in [I], s \in [S]$, we have

$$
\begin{aligned}
& \sup_{\boldsymbol{z}_s \in \mathcal{Z}} \boldsymbol{a}_i(\boldsymbol{z}_s)'\boldsymbol{x} - b_i(\boldsymbol{z}_s) - \alpha_i - k_i \|\boldsymbol{z}_s - \hat{\boldsymbol{z}}_s\| \\
={} & \sup_{\boldsymbol{z}_s \in \mathcal{Z}} \boldsymbol{a}_{i,0}^\top \boldsymbol{x} + \sum_{n \in [N]} \boldsymbol{a}_{i,n}^\top \boldsymbol{x} z_{sn} - b_{i,0} - \sum_{n \in [N]} b_{i,n} z_{sn} - \alpha_i - k_i \|\boldsymbol{z}_s - \hat{\boldsymbol{z}}_s\| \\
={} & \sup_{\boldsymbol{z}_s \in \mathcal{Z}} (\boldsymbol{A}_i^\top \boldsymbol{x} - \boldsymbol{b}_i)^\top \boldsymbol{z}_s + \boldsymbol{a}_{i,0}^\top \boldsymbol{x} - b_{i,0} - \alpha_i - k_i \|\boldsymbol{z}_s - \hat{\boldsymbol{z}}_s\| \\
={} & \sup_{\boldsymbol{z}_s \in \mathcal{Z}} \inf_{\|\boldsymbol{\gamma}_{is}\|_* \leq k_i} (\boldsymbol{A}_i^\top \boldsymbol{x} - \boldsymbol{b}_i - \boldsymbol{\gamma}_{is})^\top \boldsymbol{z}_s + \boldsymbol{a}_{i,0}^\top \boldsymbol{x} - b_{i,0} - \alpha_i + \boldsymbol{\gamma}_{is}^\top \hat{\boldsymbol{z}}_s \\
={} & \inf_{\|\boldsymbol{\gamma}_{is}\|_* \leq k_i} \{\sup_{\boldsymbol{z}_s \in \mathcal{Z}} \{(\boldsymbol{A}_i^\top \boldsymbol{x} - \boldsymbol{b}_i - \boldsymbol{\gamma}_{is})^\top \boldsymbol{z}_s\} + \boldsymbol{a}_{i,0}^\top \boldsymbol{x} - b_{i,0} - \alpha_i + \boldsymbol{\gamma}_{is}^\top \hat{\boldsymbol{z}}_s\},
\end{aligned}
$$

where the interchange of sup and inf is valid because the function is biaffine, $\mathcal{Z}$ is convex and closed, and the set $\{\boldsymbol{\gamma}_{is} \in \mathbb{R}^N \mid \|\boldsymbol{\gamma}_{is}\|_* \leq k_i\}$ is convex and compact. Now, we focus on the inner maximization problem:

$$
\begin{aligned}
\sup \quad & (\boldsymbol{A}_i^\top \boldsymbol{x} - \boldsymbol{b}_i - \boldsymbol{\gamma}_{is})^\top \boldsymbol{z}_s \\
\text{s.t.} \quad & \boldsymbol{z}_s \leq \bar{\boldsymbol{z}} \\
& -\boldsymbol{z}_s \leq -\underline{\boldsymbol{z}}.
\end{aligned}
$$

By strong duality in linear optimization, the above is equivalent to:

$$
\begin{aligned}
\inf \quad & \boldsymbol{\eta}_{1is}^\top \bar{\boldsymbol{z}} - \boldsymbol{\eta}_{2is}^\top \underline{\boldsymbol{z}} \\
\text{s.t.} \quad & \boldsymbol{\eta}_{1is} - \boldsymbol{\eta}_{2is} = \boldsymbol{A}_i^\top \boldsymbol{x} - \boldsymbol{b}_i - \boldsymbol{\gamma}_{is} \\
& \boldsymbol{\eta}_{1is} \geq \boldsymbol{0}, \ \boldsymbol{\eta}_{2is} \geq \boldsymbol{0}.
\end{aligned}
$$

Hence, we know

$$\inf_{\|\boldsymbol{\gamma}_{is}\|_* \le k_i} \sup_{\boldsymbol{z}_s \in \mathcal{Z}} \{(\boldsymbol{A}_i^\top \boldsymbol{x} - \boldsymbol{b}_i - \boldsymbol{\gamma}_{is})^\top \boldsymbol{z}_s\} + \boldsymbol{a}_{i,0}^\top \boldsymbol{x} - b_{i,0} - \alpha_i + \boldsymbol{\gamma}_{is}^\top \hat{\boldsymbol{z}}_s$$

is equivalent to

$$\inf \ (\boldsymbol{\eta}_{1is}^\top \bar{\boldsymbol{z}} - \boldsymbol{\eta}_{2is}^\top \underline{\boldsymbol{z}} + \boldsymbol{a}_{i,0}^\top \boldsymbol{x} - b_{i,0} - \alpha_i) + (\boldsymbol{A}_i^\top \boldsymbol{x} - \boldsymbol{b}_i - \boldsymbol{\eta}_{1is} + \boldsymbol{\eta}_{2is})^\top \hat{\boldsymbol{z}}_s$$
$$\text{s.t.} \ \|\boldsymbol{A}_i^\top \boldsymbol{x} - \boldsymbol{b}_i - \boldsymbol{\eta}_{1is} - \boldsymbol{\eta}_{2is}\|_* \le k_i$$
$$\boldsymbol{\eta}_{1is} \ge \boldsymbol{0}, \ \boldsymbol{\eta}_{2is} \ge \boldsymbol{0}.$$

After re-arranging the terms, the above is

$$\inf \ \boldsymbol{\eta}_{1is}^\top \bar{\boldsymbol{z}} - \boldsymbol{\eta}_{2is}^\top \underline{\boldsymbol{z}} - (\boldsymbol{\eta}_{1is} - \boldsymbol{\eta}_{2is})^\top \hat{\boldsymbol{z}}_s + (\boldsymbol{A}_i^\top \boldsymbol{x} - \boldsymbol{b}_i)^\top \hat{\boldsymbol{z}}_s + \boldsymbol{a}_{i,0}^\top \boldsymbol{x} - b_{i,0} - \alpha_i$$
$$\text{s.t.} \ \|\boldsymbol{A}_i^\top \boldsymbol{x} - \boldsymbol{b}_i - \boldsymbol{\eta}_{1is} + \boldsymbol{\eta}_{2is}\|_* \le k_i$$
$$\boldsymbol{\eta}_{1is} \ge \boldsymbol{0}, \ \boldsymbol{\eta}_{2is} \ge \boldsymbol{0}.$$

Similarly, for $s \in [S]$, we have

$$\sup_{\boldsymbol{z}_s \in \mathcal{Z}} \{(-\boldsymbol{c}(\boldsymbol{z}_s)'\boldsymbol{x} - \alpha_0)^+ - k_0\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}_s\|\}$$
$$= \inf_{\|\boldsymbol{\gamma}_{0s}\|_* \le k_0} \{\sup_{\boldsymbol{z}_s \in \mathcal{Z}} \{(-\boldsymbol{C}^\top \boldsymbol{x} - \boldsymbol{\gamma}_{0s})^\top \boldsymbol{z}_s\} - \boldsymbol{c}_0^\top \boldsymbol{x} - \alpha_0 + \boldsymbol{\gamma}_{0s}^\top \hat{\boldsymbol{z}}_s\}$$

By strong duality of the inner maximization problem, the above is equivalent to

$$\inf \ \boldsymbol{\eta}_{10s}^\top \bar{\boldsymbol{z}} - \boldsymbol{\eta}_{20s}^\top \underline{\boldsymbol{z}} - (\boldsymbol{\eta}_{10s} - \boldsymbol{\eta}_{20s})^\top \hat{\boldsymbol{z}}_s - (\boldsymbol{C}^\top \boldsymbol{x})^\top \hat{\boldsymbol{z}}_s - \boldsymbol{c}_0^\top \boldsymbol{x} - \alpha_0$$
$$\text{s.t.} \ \| -\boldsymbol{C}^\top \boldsymbol{x} - \boldsymbol{\eta}_{1is} + \boldsymbol{\eta}_{2is}\|_* \le k_0$$
$$\boldsymbol{\eta}_{1is} \ge \boldsymbol{0}, \ \boldsymbol{\eta}_{2is} \ge \boldsymbol{0}.$$

Hence, after replacing $\bar{\boldsymbol{z}} = \boldsymbol{1}$ and $\underline{\boldsymbol{z}} = -\boldsymbol{1}$, the final reformulation for the robust satisficing problem (18) becomes:

$$\min \ \sum_{i \in [I] \cup \{0\}} w_i k_i$$

$$\text{s.t.} \ \alpha_i + \frac{1}{S\epsilon_i} \sum_{s \in [S]} y_{is} \le 0 \qquad\qquad\qquad \forall i \in [I]$$

$$y_{is} \ge \boldsymbol{a}_i(\hat{\boldsymbol{z}}_s)^\top \boldsymbol{x} - \boldsymbol{b}_i(\hat{\boldsymbol{z}}_s) + (\boldsymbol{\eta}_{1is} + \boldsymbol{\eta}_{2is})^\top \boldsymbol{1} - (\boldsymbol{\eta}_{1is} - \boldsymbol{\eta}_{2is})^\top \hat{\boldsymbol{z}}_s - \alpha_i \quad \forall i \in [I], \ s \in [S]$$

$$\alpha_0 + \frac{1}{S\epsilon_0} \sum_{s \in [S]} y_{0s} + \tau \le 0$$

$$y_{0s} \ge -\boldsymbol{c}(\hat{\boldsymbol{z}}_s)^\top \boldsymbol{x} + (\boldsymbol{\eta}_{10s} + \boldsymbol{\eta}_{20s})^\top \boldsymbol{1} - (\boldsymbol{\eta}_{10s} - \boldsymbol{\eta}_{20s})^\top \hat{\boldsymbol{z}}_s - \alpha_0 \qquad \forall s \in [S]$$

$$\|\boldsymbol{A}_i^\top \boldsymbol{x} - \boldsymbol{b}_i - \boldsymbol{\eta}_{1is} + \boldsymbol{\eta}_{2is}\|_* \le k_i \qquad\qquad\qquad \forall i \in [I], \ s \in [S]$$

$$\|\boldsymbol{C}^\top \boldsymbol{x} + \boldsymbol{\eta}_{10s} - \boldsymbol{\eta}_{20s}\|_* \le k_0 \qquad\qquad\qquad \forall s \in [S]$$

$$\boldsymbol{\eta}_{1is} \ge \boldsymbol{0}, \ \boldsymbol{\eta}_{2is} \ge \boldsymbol{0} \qquad\qquad\qquad \forall i \in [I] \cup \{0\}, \ s \in [S]$$

$$\boldsymbol{x} \ge \boldsymbol{0}$$

$$\alpha_i \in \mathbb{R}, \ k_i \ge 0, \ y_{is} \ge 0 \qquad\qquad\qquad \forall i \in [I] \cup \{0\}, s \in [S].$$

$\square$

Proof of Theorem 4. We first focus on the maximization problem:

$$\sup_{\mathcal{B}(r)} \mathbb{E}_{\hat{\mathbb{P}}}\left[\boldsymbol{c}^\top\boldsymbol{x} + \sum_{n\in[N]} d_n\tilde{z}_n x_n\right].$$

By classic results (Mohajerin Esfahani and Kuhn 2018), the above can be written as:

$$\inf_{k\geq 0}\left\{kr + \frac{1}{S}\sum_{s\in[S]}\sup_{\boldsymbol{z}_s\in\mathcal{Z}}\left\{\boldsymbol{c}^\top\boldsymbol{x} + \sum_{n\in[N]} d_n z_{sn}x_n - k\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}_s\|_1\right\}\right\},$$

which can be further reduced to

$$\boldsymbol{c}^\top\boldsymbol{x} + \inf_{k\geq 0}\left\{kr + \frac{1}{S}\sum_{s\in[S]}\sum_{n\in[N]}\sup_{z_{sn}\in[-\underline{z}_n,1]}\left\{d_n z_{sn}x_n - k|z_{sn} - \hat{z}_{sn}|\right\}\right\}.$$

We can now focus on the inner maximization problem, for $s\in[S], n\in[N]$:

$$\sup_{z_{sn}\in[-\underline{z}_n,1]}\left\{d_n z_{sn}x_n - k|z_{sn} - \hat{z}_{sn}|\right\}$$

$$= x_n\sup_{z_{sn}\in[-\underline{z}_n,1]}\left\{d_n z_{sn} - k|z_{sn} - \hat{z}_{sn}|\right\}$$

$$= x_n\max\{d_n\hat{z}_{sn}, d_n - k(1 - \hat{z}_{sn})\}$$

$$= x_n(d_n\hat{z}_{sn} + (d_n - k)^+(1 - \hat{z}_{sn})),$$

where the first equality follows because $x_n\in\{0,1\}$ and $k\geq 0$, and the second equality follows by noticing that $\boldsymbol{d}\geq\boldsymbol{0}$ and the maximum can only occur at $z_{sn}\in\{\hat{z}_{sn},1\}$. Hence, we have

$$\frac{1}{S}\sum_{s\in[S]}\sum_{n\in[N]}\sup_{z_{sn}\in[-\underline{z}_n,1]}\left\{d_n z_{sn}x_n - k|z_{sn} - \hat{z}_{sn}|\right\}$$

$$= \sum_{n\in[N]}\frac{1}{S}\sum_{s\in[S]}(d_n\hat{z}_{sn} + (d_n - k)^+(1 - \hat{z}_{sn}))x_n$$

$$= \sum_{n\in[N]}(d_n\mathbb{E}_{\hat{\mathbb{P}}}[\tilde{z}_n] + (d_n - k)^+(1 - \mathbb{E}_{\hat{\mathbb{P}}}[\tilde{z}_n]))x_n$$

$$= \sum_{n\in[N]}(d_n - k)^+ x_n,$$

where the last equality is due to $\mathbb{E}_{\hat{\mathbb{P}}}[\tilde{\boldsymbol{z}}] = \boldsymbol{0}$. Hence, the reformulation in the theorem follows from the results of Bertsimas and Sim (2003) that it suffices to restrict $k\in\{0,d_1,\ldots,d_N\}$. $\square$

Proof of Theorem 5. First, the robust satisficing constraint

$$\mathbb{E}_{\mathbb{P}}\left[\boldsymbol{c}^\top\boldsymbol{x} + \sum_{n\in[N]} d_n\tilde{z}_n x_n\right] \leq \tau + k\Delta_W(\mathbb{P},\hat{\mathbb{P}}) \quad \forall\mathbb{P}\in\mathcal{P}_0(\mathcal{Z})$$

can be equivalently written as

$$\frac{1}{S} \sum_{s \in [S]} \sup_{\boldsymbol{z}_s \in \mathcal{Z}} \{ \boldsymbol{c}^\top \boldsymbol{x} + \sum_{n \in [N]} d_n z_{sn} x_n - k \|\boldsymbol{z}_s - \hat{\boldsymbol{z}}_s\|_1 \} \leq \tau$$

$$\Longleftrightarrow \boldsymbol{c}^\top \boldsymbol{x} + \frac{1}{S} \sum_{s \in [S]} \sup_{\boldsymbol{z}_s \in \mathcal{Z}} \left\{ \sum_{n \in [N]} (d_n z_{sn} x_n - k|z_{sn} - \hat{z}_{sn}|) \right\} \leq \tau$$

$$\Longleftrightarrow \boldsymbol{c}^\top \boldsymbol{x} + \frac{1}{S} \sum_{s \in [S]} \sum_{n \in [N]} \sup_{z_{sn} \in [-\underline{z}_n, 1]} \{ d_n z_{sn} x_n - k|z_{sn} - \hat{z}_{sn}| \} \leq \tau.$$

By the proof of Theorem 4, the inner maximization can be written as

$$\sup_{z_{sn} \in [-\underline{z}_n, 1]} \{ d_n z_{sn} x_n - k|z_{sn} - \hat{z}_{sn}| \} = x_n (d_n \hat{z}_{sn} + (d_n - k)^+ (1 - \hat{z}_{sn})).$$

Hence, we can write the robust satisficing constraints as:

$$\boldsymbol{c}^\top \boldsymbol{x} + \frac{1}{S} \sum_{s \in [S]} \sum_{n \in [N]} x_n (d_n \hat{z}_{sn} + (d_n - k)^+ (1 - \hat{z}_{sn})) \leq \tau$$

$$\Longleftrightarrow \boldsymbol{c}^\top \boldsymbol{x} + \sum_{n \in [N]} \frac{1}{S} \sum_{s \in [S]} (d_n \hat{z}_{sn} + (d_n - k)^+ (1 - \hat{z}_{sn})) x_n \leq \tau$$

$$\Longleftrightarrow \boldsymbol{c}^\top \boldsymbol{x} + \sum_{n \in [N]} (d_n \mathbb{E}_{\hat{\mathbb{P}}}[\tilde{z}_n] + (d_n - k)^+ (1 - \mathbb{E}_{\hat{\mathbb{P}}}[\tilde{z}_n])) x_n \leq \tau$$

$$\Longleftrightarrow \boldsymbol{c}^\top \boldsymbol{x} + \sum_{n \in [N]} (d_n - k)^+ x_n \leq \tau,$$

where the last equality is due to $\mathbb{E}_{\hat{\mathbb{P}}}[\tilde{\boldsymbol{z}}] = \boldsymbol{0}$. Hence, we arrive at Problem (23).

Because we minimize k in Problem (23), we can use a bisection search algorithm. Whenever $k \geq \|\boldsymbol{d}\|_\infty$, Problem (23) is feaisible because $\tau \geq Z_0$. Hence, we can safely set the upper bound of the bisection search to be $\|\boldsymbol{d}\|_\infty$. Finally, for each fixed k, the complexity of the resulting feasibility subproblem has the same conplexity as the empirical combinatorial optimization problem. $\qquad \square$

Proof of Theorem 6. Based on the definition of Wasserstein metric, we can rewrite the stochastic linear robust satisficing model as

$$\min \; k$$
$$\text{s.t.} \;\; \mathbb{E}_{\mathbb{Q}} \left[\boldsymbol{c}(\tilde{\boldsymbol{z}})^\top \boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{y}(\tilde{\boldsymbol{z}}) - k\|\tilde{\boldsymbol{z}} - \tilde{\boldsymbol{v}}\|_1 \right] - \tau \leq 0, \quad \forall \mathbb{Q} \in \mathcal{Q},$$
$$\boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) \geq \boldsymbol{b}(\boldsymbol{z}), \qquad\qquad\qquad \forall \boldsymbol{z} \in \mathcal{Z},$$
$$\boldsymbol{y} \in \mathcal{R}^{N,P}, \; \boldsymbol{x} \in \mathcal{X}, \; k \geq 0$$

where the ambiguity set $\mathcal{Q}$ is defined as

$$\mathcal{Q} := \left\{ \mathbb{Q} \in \mathcal{P}_0(\mathcal{Z}^2) \;\middle|\; (\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{v}}) \sim \mathbb{Q}, \; \tilde{\boldsymbol{z}} \sim \mathbb{P}, \; \tilde{\boldsymbol{v}} \sim \hat{\mathbb{P}} \right\}.$$

Given the structure of $\hat{\mathbb{P}}$, we can rewrite the above problem as

$$
\begin{aligned}
&\min \ k \\
&\text{s.t. } \frac{1}{S} \sum_{s \in [S]} \mathbb{E}_{\mathbb{Q}_s}[\boldsymbol{c}(\tilde{\boldsymbol{z}})^\top \boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{y}(\tilde{\boldsymbol{z}}) - k\|\tilde{\boldsymbol{z}} - \hat{\boldsymbol{z}}_s\|_1] \leq \tau, \quad \forall \mathbb{Q}_s \in \mathcal{P}_0(\mathcal{Z}), s \in [S], \\
&\qquad \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) \geq \boldsymbol{b}(\boldsymbol{z}), \qquad\qquad\qquad \forall \boldsymbol{z} \in \mathcal{Z}, \\
&\qquad \boldsymbol{y} \in \mathcal{R}^{N,P}, \ \boldsymbol{x} \in \mathcal{X}, \ k \geq 0,
\end{aligned}
$$

where $\mathbb{Q}_s$ can be seen as the conditional probability distribution of $\tilde{\boldsymbol{z}}$ given $\tilde{\boldsymbol{v}} = \hat{\boldsymbol{z}}_s$. The worst case distribution $\mathbb{Q}_s^* \in \mathcal{P}_0(\mathcal{Z})$, $s \in [S]$, is a one-point distribution. Therefore, the above problem is equivalent to

$$
\begin{aligned}
&\min \ k \\
&\text{s.t. } \frac{1}{S} \sum_{s \in [S]} \sup_{\boldsymbol{z} \in \mathcal{Z}} \left\{ \boldsymbol{c}(\boldsymbol{z})^\top \boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{y}(\boldsymbol{z}) - k\|\boldsymbol{z} - \hat{\boldsymbol{z}}_s\|_1 \right\} \leq \tau, \\
&\qquad \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) \geq \boldsymbol{b}(\boldsymbol{z}), \qquad\qquad\qquad \forall \boldsymbol{z} \in \mathcal{Z}, \\
&\qquad \boldsymbol{y} \in \mathcal{R}^{N,P}, \ \boldsymbol{x} \in \mathcal{X}, \ k \geq 0,
\end{aligned}
\tag{37}
$$

We will show that for any optimal solution in (37) we can find a feasible solution in (29) with the same objective value and *vice versa*.

Consider any optimal solution $(\bar{k}, \bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})$ in Problem (37). We define $\hat{\boldsymbol{y}}_s(\boldsymbol{z}, u) := \bar{\boldsymbol{y}}(\boldsymbol{z})$ for all $(\boldsymbol{z}, u) \in \bar{\mathcal{Z}}_s, s \in [S]$. Because $\boldsymbol{A}(\boldsymbol{z})\bar{\boldsymbol{x}} + \boldsymbol{B}\bar{\boldsymbol{y}}(\boldsymbol{z}) \geq \boldsymbol{b}(\boldsymbol{z})$ for all $\boldsymbol{z} \in \mathcal{Z}$, we also have $\boldsymbol{A}(\boldsymbol{z})\bar{\boldsymbol{x}} + \boldsymbol{B}\hat{\boldsymbol{y}}_s(\boldsymbol{z}, u) \geq \boldsymbol{b}(\boldsymbol{z})$ for all $(\boldsymbol{z}, u) \in \bar{\mathcal{Z}}_s, s \in [S]$.

In addition,

$$
\frac{1}{S} \sum_{s \in [S]} \sup_{\boldsymbol{z} \in \mathcal{Z}} \left\{ \boldsymbol{c}(\boldsymbol{z})^\top \bar{\boldsymbol{x}} + \boldsymbol{d}^\top \bar{\boldsymbol{y}}(\boldsymbol{z}) - \bar{k}\|\boldsymbol{z} - \hat{\boldsymbol{z}}_s\|_1 \right\} \leq \tau
$$

indicates

$$
\frac{1}{S} \sum_{s \in [S]} \sup_{(\boldsymbol{z}, u) \in \bar{\mathcal{Z}}_s} \left\{ \boldsymbol{c}(\boldsymbol{z})^\top \bar{\boldsymbol{x}} + \boldsymbol{d}^\top \hat{\boldsymbol{y}}_s(\boldsymbol{z}, u) - \bar{k}u \right\} \leq \tau
$$

because of the definition of $\hat{\boldsymbol{y}}$ and $\bar{\mathcal{Z}}_s, s \in [S]$. Therefore, $(\bar{k}, \bar{\boldsymbol{x}}, \hat{\boldsymbol{y}}_1, \ldots, \hat{\boldsymbol{y}}_S)$ is feasible in Problem (29) and leads to the same objective value.

Now, consider any optimal solution $(\bar{k}, \bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}_1, \ldots, \bar{\boldsymbol{y}}_S)$ in Problem (29). We define $\hat{\boldsymbol{y}}(\boldsymbol{z}) := \bar{\boldsymbol{y}}_{s^*(\boldsymbol{z})}(\boldsymbol{z}, \|\boldsymbol{z} - \tilde{\boldsymbol{z}}_{s^*(\boldsymbol{z})})\|_1)$ for all $\boldsymbol{z} \in \mathcal{Z}$, where

$$
s^*(\boldsymbol{z}) := \underset{s \in [S]}{\arg\min}\{\boldsymbol{d}^\top \bar{\boldsymbol{y}}_s(\boldsymbol{z}, \|\boldsymbol{z} - \hat{\boldsymbol{z}}_s\|_1)\}.
$$

Because $\boldsymbol{A}(\boldsymbol{z})\bar{\boldsymbol{x}} + \boldsymbol{B}\bar{\boldsymbol{y}}_s(\boldsymbol{z}, u) \geq \boldsymbol{b}(\boldsymbol{z})$ for all $(\boldsymbol{z}, u) \in \bar{\mathcal{Z}}_s, s \in [S]$, we must have

$$
\boldsymbol{A}(\boldsymbol{z})\bar{\boldsymbol{x}} + \boldsymbol{B}\bar{\boldsymbol{y}}_s(\boldsymbol{z}, \|\boldsymbol{z} - \hat{\boldsymbol{z}}_s\|_1) \geq \boldsymbol{b}(\boldsymbol{z}), \quad \forall \boldsymbol{z} \in \mathcal{Z}, s \in [S].
$$

By the definition of $\hat{\boldsymbol{y}}$, we have

$$\boldsymbol{A}(\boldsymbol{z})\bar{\boldsymbol{x}} + \boldsymbol{B}\hat{\boldsymbol{y}}(\boldsymbol{z}) \geq \boldsymbol{b}(\boldsymbol{z}), \quad \forall \boldsymbol{z} \in \mathcal{Z}.$$

Notice that

$$\frac{1}{S} \sum_{s \in [S]} \sup_{(\boldsymbol{z},u) \in \bar{\mathcal{Z}}_s} \left\{ \boldsymbol{c}(\boldsymbol{z})^\top \bar{\boldsymbol{x}} + \boldsymbol{d}^\top \bar{\boldsymbol{y}}_s(\boldsymbol{z},u) - \bar{k}u \right\} \leq \tau,$$

which indicates

$$\frac{1}{S} \sum_{s \in [S]} \sup_{\boldsymbol{z} \in \mathcal{Z}} \left\{ \boldsymbol{c}(\boldsymbol{z})^\top \bar{\boldsymbol{x}} + \boldsymbol{d}^\top \bar{\boldsymbol{y}}_s(\boldsymbol{z}, \|\boldsymbol{z} - \hat{\boldsymbol{z}}_s\|_1) - \bar{k}\|\boldsymbol{z} - \hat{\boldsymbol{z}}_s\|_1 \right\} \leq \tau.$$

Then, by the definition of $\hat{\boldsymbol{y}}$, we have

$$\frac{1}{S} \sum_{s \in [S]} \sup_{\boldsymbol{z} \in \mathcal{Z}} \left\{ \boldsymbol{c}(\boldsymbol{z})^\top \bar{\boldsymbol{x}} + \boldsymbol{d}^\top \hat{\boldsymbol{y}}(\boldsymbol{z}) - \bar{k}\|\boldsymbol{z} - \hat{\boldsymbol{z}}_s\|_1 \right\} \leq \tau.$$

The above follows from the definition of $\hat{\boldsymbol{y}}$; specifically,

$$\boldsymbol{d}^\top \hat{\boldsymbol{y}}(\boldsymbol{z}) \leq \boldsymbol{d}^\top \bar{\boldsymbol{y}}_s(\boldsymbol{z}, \|\boldsymbol{z} - \hat{\boldsymbol{z}}_s\|_1), \quad \forall \boldsymbol{z} \in \mathcal{Z}, \ \forall s \in [S].$$

Therefore, from $(\bar{k}, \bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})$, we can get a feasible solution $(\bar{k}, \bar{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ for Problem (37) that leads to the same objective value. $\qquad\square$

Proof of Theorem 7. Because $\tau \geq Z_0$, there exists some $\bar{\boldsymbol{x}} \in \mathcal{X}$ and $\bar{\boldsymbol{y}}_s \in \mathbb{R}^P, s \in [S]$ such that

$$\frac{1}{S} \sum_{s \in [S]} (\boldsymbol{c}(\hat{\boldsymbol{z}}_s)^\top \bar{\boldsymbol{x}} + \boldsymbol{d}^\top \bar{\boldsymbol{y}}_s) \leq \tau,$$

$$\boldsymbol{A}(\hat{\boldsymbol{z}}_s)\bar{\boldsymbol{x}} + \boldsymbol{B}\bar{\boldsymbol{y}}_s \geq \boldsymbol{b}(\hat{\boldsymbol{z}}_s), \quad \forall s \in [S].$$

Because matrix B has complete recourse, we can find a $\hat{\boldsymbol{y}}_{sN+1}$ such that:

$$\boldsymbol{B}\hat{\boldsymbol{y}}_{s,N+1} \geq \begin{pmatrix} \max\limits_{i \in [N]} \{|[\boldsymbol{b}_i - \boldsymbol{A}_i\bar{\boldsymbol{x}}]_1|\} \\ \vdots \\ \max\limits_{i \in [N]} \{|[\boldsymbol{b}_i - \boldsymbol{A}_i\bar{\boldsymbol{x}}]_M|\} \end{pmatrix}.$$

Therefore, for any $(\boldsymbol{z}, u) \in \bar{\mathcal{Z}}_s, s \in [S]$ and $m \in [M]$, we have

$$[\boldsymbol{B}\hat{\boldsymbol{y}}_{N+1}]_m u \geq \sum_{i \in [N]} [\boldsymbol{B}\hat{\boldsymbol{y}}_{N+1}]_m |z_i - \hat{z}_{si}| \geq \sum_{i \in [N]} |[\boldsymbol{b}_i - \boldsymbol{A}_i\bar{\boldsymbol{x}}]_m| |z_i - \hat{z}_{si}| \geq \sum_{i \in [N]} [\boldsymbol{b}_i - \boldsymbol{A}_i\bar{\boldsymbol{x}}]_m (z_i - \hat{z}_{si}),$$

which indicates

$$\boldsymbol{B}\hat{\boldsymbol{y}}_{N+1} u \geq \sum_{i \in [N]} (\boldsymbol{b}_i - \boldsymbol{A}_i\bar{\boldsymbol{x}})(z_i - \hat{z}_{si}), \quad \forall (\boldsymbol{z}, u) \in \bar{\mathcal{Z}}_s, \ \forall s \in [S].$$

Notice that we can rewrite $\boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}$ and $\boldsymbol{b}(\boldsymbol{z})$ as:

$$\boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} = \boldsymbol{A}(\boldsymbol{z} - \hat{\boldsymbol{z}}_s + \hat{\boldsymbol{z}}_s)\boldsymbol{x} = \boldsymbol{A}(\hat{\boldsymbol{z}}_s)\boldsymbol{x} + \sum_{i \in [N]} \boldsymbol{A}_i \boldsymbol{x}(z_i - \hat{z}_{si}), \quad \forall s \in [S], \ \forall \boldsymbol{z} \in \mathcal{Z},$$

$$\boldsymbol{b}(\boldsymbol{z}) = \boldsymbol{b}(\boldsymbol{z} - \hat{\boldsymbol{z}}_s + \hat{\boldsymbol{z}}_s) = \boldsymbol{b}(\hat{\boldsymbol{z}}_s) + \sum_{i \in [N]} \boldsymbol{b}_i (z_i - \hat{z}_{si}), \qquad \forall s \in [S], \ \forall \boldsymbol{z} \in \mathcal{Z}.$$

Let us define $\hat{\boldsymbol{y}}_{s,0} := \bar{\boldsymbol{y}}_s$ and $\hat{\boldsymbol{y}}_{s,i} := \boldsymbol{0}$ for $i \in [N]$. By above inequality and the alternative form of $\boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}$ and $\boldsymbol{b}(\boldsymbol{z})$, we have

$$\boldsymbol{A}(\boldsymbol{z})\bar{\boldsymbol{x}} + \boldsymbol{B}(\hat{\boldsymbol{y}}_{s,0} + \hat{\boldsymbol{y}}_{s,N+1}u + \sum_{i \in [N]} \hat{\boldsymbol{y}}_{s,i}z_i) \geq \boldsymbol{b}(\boldsymbol{z}), \quad \forall (\boldsymbol{z}, u) \in \bar{\mathcal{Z}}_s, \ \forall s \in [S].$$

Next, we define

$$\hat{k} := \max\left\{0, \max_{i \in [N], s \in [S]}\{|\boldsymbol{c}_i^\top \bar{\boldsymbol{x}}| + \boldsymbol{d}^\top \hat{\boldsymbol{y}}_{s,N+1}\}\right\},$$

which satisfies

$$\sum_{i \in [N]} \boldsymbol{c}_i^\top \bar{\boldsymbol{x}}(z_i - \hat{z}_{si}) + \boldsymbol{d}^\top \hat{\boldsymbol{y}}_{s,N+1}u - \hat{k}u \leq 0, \quad \forall (\boldsymbol{z}, u) \in \bar{\mathcal{Z}}_s, \ \forall s \in [S].$$

Notice that we can rewrite $\boldsymbol{c}(\boldsymbol{z})^\top \boldsymbol{x}$ as

$$\boldsymbol{c}(\boldsymbol{z})^\top \boldsymbol{x} = \boldsymbol{c}(\boldsymbol{z} - \hat{\boldsymbol{z}}_s + \hat{\boldsymbol{z}}_s)^\top \boldsymbol{x} = \boldsymbol{c}(\hat{\boldsymbol{z}}_s)^\top \boldsymbol{x} + \sum_{i \in [N]} \boldsymbol{c}_i^\top \boldsymbol{x}(z_i - \hat{z}_{si}), \quad \forall s \in [S], \ \forall \boldsymbol{z} \in \mathcal{Z}.$$

By above inequality and alternative form of $\boldsymbol{c}(\boldsymbol{z})^\top \boldsymbol{x}$, we have

$$\frac{1}{S} \sum_{s \in [S]} \sup_{(\boldsymbol{z}, u) \in \bar{\mathcal{Z}}_s} \left\{ \boldsymbol{c}(\boldsymbol{z})^\top \bar{\boldsymbol{x}} + \boldsymbol{d}^\top \left(\hat{\boldsymbol{y}}_{s,0} + \sum_{i \in [N]} \hat{\boldsymbol{y}}_{s,i}z_i + \hat{\boldsymbol{y}}_{s,N+1}u \right) - \hat{k}u \right\}$$

$$\leq \frac{1}{S} \sum_{s \in [S]} (\boldsymbol{c}(\hat{\boldsymbol{z}}_s)^\top \bar{\boldsymbol{x}} + \boldsymbol{d}^\top \bar{\boldsymbol{y}}_s)$$

$$\leq \tau.$$

Therefore, $(\hat{k}, \bar{\boldsymbol{x}}, \hat{\boldsymbol{y}}_{s,0}, \ldots, \hat{\boldsymbol{y}}_{s,N+1})$ constitutes a feasible solution to Problem (31).

Now, we show that the scenario-wise lifted affine recourse adaptation would yield the exact objective value as Problem (29) when $P = 1$. In this case, matrix $\boldsymbol{B}$ has dimension $M \times 1$, *i.e.*, $\boldsymbol{B} \in \mathbb{R}^{M,1}$ and it is either strictly positive or negative. When $d\boldsymbol{B} \leq \boldsymbol{0}$, the solutions to both Problem (31) and Problem (29) are trivial and their objective values coincide. For non-trivial cases, we can, without loss of generality, focus on $d > 0$ and $\boldsymbol{B} > \boldsymbol{0}$.

In Problem (29), the recourse function $y_s(\boldsymbol{z})$ must satisfy

$$y_s(\boldsymbol{z}, u) \geq \max_{m \in [M]} \left\{ \frac{[\boldsymbol{b}(\boldsymbol{z}) - \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}]_m}{[\boldsymbol{B}]_m} \right\}, \quad \forall (\boldsymbol{z}, u) \in \bar{\mathcal{Z}}_s, \ \forall s \in [S].$$

In addition,

$$\frac{1}{S}\sum_{s\in[S]}\sup_{(\boldsymbol{z},u)\in\bar{\mathcal{Z}}_s}\left\{\boldsymbol{c}(\boldsymbol{z})^\top\boldsymbol{x}+dy_s(\boldsymbol{z},u)-ku\right\}\leq\tau$$

can be equivalently written as

$$\frac{1}{S}\sum_{s\in[S]}v_s\leq\tau$$

$$v_s\geq\boldsymbol{c}(\boldsymbol{z})^\top\boldsymbol{x}+dy_s(\boldsymbol{z},u)-ku,\quad\forall(\boldsymbol{z},u)\in\bar{\mathcal{Z}}_s,\ \forall s\in[S].$$

The above indicates that the recourse function $y_s(\boldsymbol{z})$ must satisfy

$$y_s(\boldsymbol{z},u)\leq\frac{v_s-\boldsymbol{c}(\boldsymbol{z})^\top\boldsymbol{x}+ku}{d},\quad\forall(\boldsymbol{z},u)\in\bar{\mathcal{Z}}_s,\ \forall s\in[S].$$

Therefore, an optimal recourse function for any $s\in[S]$ would be

$$y_s(\boldsymbol{z},u):=\frac{v_s-\boldsymbol{c}(\boldsymbol{z})^\top\boldsymbol{x}+ku}{d},$$

which is an affine function of $\boldsymbol{z}$ and u for a given scenario s. Therefore, there exists an optimal scenario-wise lifted affine function for Problem (31) that achieves the same optimal solution as Problem (29). $\qquad\square$

Proof of Theorem 8. To see the first property, consider any $\boldsymbol{x},\boldsymbol{y}_{s,i},k,\phi$ which are feasible to Problem (32). For any $\boldsymbol{z}\in\mathcal{Z}$, there exists $s\in[S]$ and $u\in\mathbb{R}$ such that $(\boldsymbol{z},u)\in\bar{\mathcal{Z}}_s$. Correspondingly, choose $\boldsymbol{y}=\boldsymbol{y}_{s,0}+\sum_{i\in[N]}\boldsymbol{y}_{s,i}z_i+\boldsymbol{y}_{s,N+1}u$, we have $\boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}+\boldsymbol{B}\boldsymbol{y}=\boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}+\boldsymbol{B}\left(\boldsymbol{y}_{s,0}+\sum_{i\in[N]}\boldsymbol{y}_{s,i}z_i+\boldsymbol{y}_{s,N+1}u\right)\geq\boldsymbol{b}(\boldsymbol{z})$, where the inequality is due to the feasibility of $\boldsymbol{x},\boldsymbol{y}_{s,i}$ for Problem (32). Therefore, $Q(\boldsymbol{x},\boldsymbol{z})<\infty$.

To see $\bar{\bar{Z}}_0\geq\bar{Z}_0$, we consider an optimal $\boldsymbol{x},\boldsymbol{y}(\boldsymbol{z})=\boldsymbol{y}_0+\sum_{i\in[N]}\boldsymbol{y}_iz_i\in\mathcal{L}^{N,P}$ for the problem in defining $\bar{Z}_0$. We then choose $\boldsymbol{y}_{s,i}=\boldsymbol{y}_i,\forall i\in\{0,1,\dots,N\}$, $\boldsymbol{y}_{s,N+1}=\boldsymbol{0},\forall s\in[S]$, $\phi=\bar{\bar{Z}}_0,k=\max_{i\in[N]}\{|\boldsymbol{x}^\top\boldsymbol{c}_i+\boldsymbol{d}^\top\boldsymbol{y}_i|\}$, and show that they are feasible to Problem (32). Consider any $s\in[S]$ and $(\boldsymbol{z},u)\in\bar{\mathcal{Z}}_S$,

$$\left(\boldsymbol{c}(\boldsymbol{z})^\top\boldsymbol{x}+\boldsymbol{d}^\top\left(\boldsymbol{y}_{s,0}+\sum_{i\in[N]}\boldsymbol{y}_{s,i}z_i+\boldsymbol{y}_{s,N+1}u\right)-ku\right)-\left(\boldsymbol{c}(\hat{\boldsymbol{z}}_s)^\top\boldsymbol{x}+\boldsymbol{d}^\top\left(\boldsymbol{y}_0+\sum_{i\in[N]}\boldsymbol{y}_i\hat{z}_{si}\right)\right)$$

$$=\sum_{i\in[N]}\left(\boldsymbol{c}_i\left(z_i-\hat{z}_{si}\right)\right)^\top\boldsymbol{x}+\boldsymbol{d}^\top\left(\sum_{i\in[N]}\boldsymbol{y}_i\left(z_i-\hat{z}_{si}\right)\right)-ku$$

$$=\sum_{i\in[N]}\left(\boldsymbol{x}^\top\boldsymbol{c}_i+\boldsymbol{d}^\top\boldsymbol{y}_i\right)\left(z_i-\hat{z}_{si}\right)-ku$$

$$\leq\sum_{i\in[N]}\left(\boldsymbol{x}^\top\boldsymbol{c}_i+\boldsymbol{d}^\top\boldsymbol{y}_i\right)\left(z_i-\hat{z}_{si}\right)-k\sum_{i\in[N]}|z_i-\hat{z}_{si}|$$

$$\leq\sum_{i\in[N]}k|z_i-\hat{z}_{si}|-k\sum_{i\in[N]}|z_i-\hat{z}_{si}|$$

$$=0,$$

where the first two inequalities are due to the definition of $\bar{\mathcal{Z}}_S$ and k, respectively. Therefore, we always have $\left(c(z)^\top x + d^\top \left(y_{s,0} + \sum_{i\in[N]} y_{s,i} z_i + y_{s,N+1} u\right) - ku\right) \leq \left(c(\hat{z}_s)^\top x + d^\top \left(y_0 + \sum_{i\in[N]} y_i \hat{z}_{si}\right)\right)$, and hence

$$\frac{1}{S} \sum_{s\in[S]} \sup_{(z,u)\in\bar{\mathcal{Z}}_s} \left\{ c(z)^\top x + d^\top \left(y_{s,0} + \sum_{i\in[N]} y_{s,i} z_i + y_{s,N+1} u\right) - ku \right\}$$
$$\leq \frac{1}{S} \sum_{s\in[S]} \left(c(\hat{z}_s)^\top x + d^\top \left(y_0 + \sum_{i\in[N]} y_i \hat{z}_{si}\right) \right)$$
$$= \bar{\bar{Z}}_0$$
$$= \phi.$$

Hence, the first constraint in Problem (32) is satisfied. The second constraint in Problem (32) follows directly from the feasibility of $x, y(z) = y_0 + \sum_{i\in[N]} y_i z_i \in \mathcal{L}^{N,P}$ for the problem in defining $\bar{\bar{Z}}_0$. Therefore, we conclude that such $x, y_{s,i}, \phi, k$ are feasible to Problem (32), and hence $\bar{Z}_0 \leq \phi = \bar{\bar{Z}}_0$.

To see $\bar{Z}_0 \geq Z_0$, consider any $x, y_{s,i}, \phi, k$ which are feasible to Problem (32). For any $s \in [S]$, choose $y_s = y_{s,0} + \sum_{i\in[N]} y_{s,i} \hat{z}_{si}$. Noticing that $x, y_{s,i}$ are feasible to the second constraint in Problem (32) and choosing $(z, u) = (\hat{z}_s, 0)$, we have $A(\hat{z}_s)x + B\left(y_{s,0} + \sum_{i\in[N]} y_{s,i} \hat{z}_{si} + y_{s,N+1} \cdot 0\right) \geq b(\hat{z}_s)$, the left-hand-side of which is indeed $A(\hat{z}_s)x + By_s$. Therefore, x, y_s are feasible to Problem (26). Consequently,

$$Z_0 \leq \frac{1}{S} \sum_{s\in[S]} \left(c(\hat{z}_s)^\top x + d^\top y_s \right)$$
$$= \frac{1}{S} \sum_{s\in[S]} \left(c(\hat{z}_s)^\top x + d^\top \left(y_{s,0} + \sum_{i\in[N]} y_{s,i} \hat{z}_{si} + y_{s,N+1} \cdot 0 \right) - k \cdot 0 \right)$$
$$\leq \frac{1}{S} \sum_{s\in[S]} \sup_{(z,u)\in\bar{\mathcal{Z}}_s} \left\{ c(z)^\top x + d^\top \left(y_{s,0} + \sum_{i\in[N]} y_{s,i} z_i + y_{s,N+1} u \right) - ku \right\}$$
$$\leq \phi,$$

where the first inequality is due to the feasibility of the x, y_s to Problem(26), the second inequality holds since $(\hat{z}_s, 0) \in \bar{\mathcal{Z}}_s$, the last inequality follows from the feasibility of $x, y_{s,i}, \phi, k$ to Problem (32). Since it holds for all feasible ϕ, we conclude $Z_0 \leq \bar{Z}_0$.

In the case of complete recourse, by Theorem 7, $\phi = Z_0$ is feasible to Problem (32), which implies $\bar{Z}_0 \leq Z_0$; together with $Z_0 \leq \bar{Z}_0$, we conclude $Z_0 = \bar{Z}_0$.

The last property is straightforward. Specifically, with the optimal $x, y_{s,i}, k$ for Problem (32), we can construct $y_s(z, u) = y_{s,0} + \sum_{i\in[N]} y_{s,i} z_i + y_{s,N+1} u$. Then $x, y_s(z, u), k$ are feasible to Problem (29) with any $\tau \geq \bar{Z}_0$. $\qquad\square$

Proof of Theorem 9.   We first look at the first set of constraints:

$$\mathbb{E}_{\mathbb{P}}\left[\boldsymbol{x}^{\top}\tilde{\boldsymbol{z}}\right] \geq \tau - k_0 \Delta_W(\mathbb{P},\hat{\mathbb{P}}) \quad \forall \mathbb{P} \in \mathcal{P}_0(\mathcal{Z}),$$

which is equivalent to

$$\mathbb{E}_{\mathbb{P}}\left[-\boldsymbol{x}^{\top}\tilde{\boldsymbol{z}}\right] \leq -\tau + k_0 \Delta_W(\mathbb{P},\hat{\mathbb{P}}) \quad \forall \mathbb{P} \in \mathcal{P}_0(\mathcal{Z}).$$

Then, the reformulation of this constraint follows from the proof of Proposition 2 by noticing $I = 1$, $a_1 = -1$, $b_1 = 0$, $\boldsymbol{C} = \boldsymbol{0}$, $\boldsymbol{h} = \boldsymbol{0}$. Specifically, this constraint is equivalent to that $\exists \boldsymbol{y}_1$ such that

$$\frac{1}{S}\sum_{s\in[S]} y_{1s} \leq -\tau$$
$$y_{1s} \geq -\boldsymbol{x}^{\top}\hat{\boldsymbol{z}}_s \quad \forall s \in [S]$$
$$k_0 \geq \|\boldsymbol{x}\|_{\infty}.$$

By replacing the variable $\boldsymbol{y}_1$ with $-\boldsymbol{y}_1$, the above is equivalent to

$$\frac{1}{S}\sum_{s\in[S]} y_{1s} \geq \tau$$
$$y_{1s} \leq \boldsymbol{x}^{\top}\hat{\boldsymbol{z}}_s \quad \forall s \in [S]$$
$$k_0 \geq \|\boldsymbol{x}\|_{\infty}.$$

Now, we focus on the second set of constraints

$$\alpha + \frac{1}{\epsilon}\mathbb{E}_{\mathbb{P}}\left[(-\boldsymbol{x}^{\top}\tilde{\boldsymbol{z}} - \alpha)^+\right] \leq \beta + k_1\Delta_W(\mathbb{P},\hat{\mathbb{P}}) \quad \forall \mathbb{P} \in \mathcal{P}_0(\mathcal{Z}),$$

which is equivalent to

$$\mathbb{E}_{\mathbb{P}}\left[\max\left\{\alpha, -\frac{1}{\epsilon}\boldsymbol{x}\tilde{\boldsymbol{z}} - \frac{\epsilon-1}{\epsilon}\alpha\right\}\right] \leq \beta + k_1\Delta_W(\mathbb{P},\hat{\mathbb{P}}) \quad \forall \mathbb{P} \in \mathcal{P}_0(\mathcal{Z}).$$

Then, we can reformulate this by the proof of Proposition 2, noticing that $I = 2$, $a_1 = 0$, $b_1 = \alpha$, $a_2 = -1/\epsilon$, $b_2 = -\alpha(\epsilon-1)/\epsilon$, $\boldsymbol{C} = \boldsymbol{0}$, $\boldsymbol{h} = \boldsymbol{0}$. Specifically, after rearranging the term, this constraint is equivalent to that $\exists \boldsymbol{y}_2$ such that

$$\alpha + \frac{1}{\epsilon S}\sum_{s\in[S]} y_{2s} \leq \beta$$
$$y_{2s} \geq -\boldsymbol{x}^{\top}\hat{\boldsymbol{z}} - \alpha \quad \forall s \in [S]$$
$$y_{2s} \geq 0 \quad \forall s \in [S]$$
$$k_1 \geq \|\boldsymbol{x}\|_{\infty}.$$

From these reformulations, we see that the objective function is only subject to one constraint: $k_0 + wk_1 \leq (1+w)\|\boldsymbol{x}\|_{\infty}$, for which the ranking only depends on $\|\boldsymbol{x}\|_{\infty}$. Hence, the final formulation (34) follows. $\qquad\square$

Proof of Proposition 7. This follows because the objective is essentially $\|\boldsymbol{x}\|_\infty$ and $\mathbf{1}^\top\boldsymbol{x} = 1$. Hence, the best possible objective is achieved by a equal-weighted portfolio $\mathbf{1}/N$. Since $\tau \leq \mathbb{E}_{\hat{\mathbb{P}}}[\mathbf{1}^\top\tilde{\boldsymbol{z}}/N]$ and $\beta \geq \mathbb{C}_{\hat{\mathbb{P}}}^\epsilon(-\mathbf{1}^\top\tilde{\boldsymbol{z}}/N)$, the equal-weighted portfolio is feasible in Model (34) and it is also optimal. $\qquad\square$

B.   Infeasibility of Lifted Affine Recourse Adaptation under Relatively Complete Recourse

We adopt the example used in Bertsimas et al. (2019). Consider the case of unbounded support $\mathcal{Z} = [-1,1]^2$ and

$$
\begin{aligned}
\min\ & k \\
\text{s.t.}\ & 0 \times y(\boldsymbol{z}) - 0 \leq k\|\boldsymbol{z}\|_1, && \forall \boldsymbol{z} \in \mathcal{Z} \\
& y(\boldsymbol{z}) \geq z_1 - z_2 && \forall z \in \mathcal{Z} \\
& y(\boldsymbol{z}) \geq z_2 - z_1 && \forall z \in \mathcal{Z} \\
& y(\boldsymbol{z}) \leq z_1 + z_2 + 2 && \forall z \in \mathcal{Z} \\
& y(\boldsymbol{z}) \leq -z_1 - z_2 + 2 && \forall z \in \mathcal{Z} \\
& y \in \mathcal{R}^{3,1}, k \geq 0,
\end{aligned}
$$

for which a feasible recourse function would be $y(\boldsymbol{z}) = |z_1 - z_2|$. Hence, this is a relatively complete recourse problem. Under the lifted affine recourse adaptation, we solve the following problem:

$$
\begin{aligned}
\min\ & k \\
\text{s.t.}\ & 0 \leq ku, && \forall (\boldsymbol{z}, u) : u \geq \|\boldsymbol{z}\|_1, \boldsymbol{z} \in \mathcal{Z} \\
& y_0 + y_1 z_1 + y_2 z_2 + y_3 u \geq z_1 - z_2 && \forall (\boldsymbol{z}, u) : u \geq \|\boldsymbol{z}\|_1, \boldsymbol{z} \in \mathcal{Z} \\
& y_0 + y_1 z_1 + y_2 z_2 + y_3 u \geq z_2 - z_1 && \forall (\boldsymbol{z}, u) : u \geq \|\boldsymbol{z}\|_1, \boldsymbol{z} \in \mathcal{Z} \\
& y_0 + y_1 z_1 + y_2 z_2 + y_3 u \leq z_1 + z_2 + 2 && \forall (\boldsymbol{z}, u) : u \geq \|\boldsymbol{z}\|_1, \boldsymbol{z} \in \mathcal{Z} \\
& y_0 + y_1 z_1 + y_2 z_2 + y_3 u \leq -z_1 - z_2 + 2 && \forall (\boldsymbol{z}, u) : u \geq \|\boldsymbol{z}\|_1, \boldsymbol{z} \in \mathcal{Z} \\
& y_0, y_1, y_2, y_3 \in \mathbb{R}, k \geq 0.
\end{aligned}
$$

For $\boldsymbol{z} = \mathbf{0}$, the pair of semi-infinite constraints $y_0 + y_3 u \geq 0$ and $y_0 + y_3 u \leq 2$ for all $u \geq 0$, implies that $y_3 = 0$. With $z_1 = z_2 = 1$, we have $y_0 + y_1 + y_2 = 0$, and with $z_1 = z_2 = -1$, we have $y_0 - y_1 - y_2 = 0$, both implying $y_0 = 0$. Moreover, with $z_1 = 1, z_2 = -1$, we would require $y_1 - y_2 = 2$. However, infeasibility occurs when $z_1 = -1, z_2 = 1$, which mandates $-y_1 + y_2 = 2$. Hence, no lifted affine recourse function would be feasible in the above problem.

C. Additional Simulation Results

The adaptive network lot-sizing problem In the main text, we used a normal distribution, $\mathcal{N}(20, 10^2)$, to generate the test data. Here, we use a normal distribution $\mathcal{N}(20, 12^2)$. The results are summarized in Figure 10, Figure 11, and Figure 12.
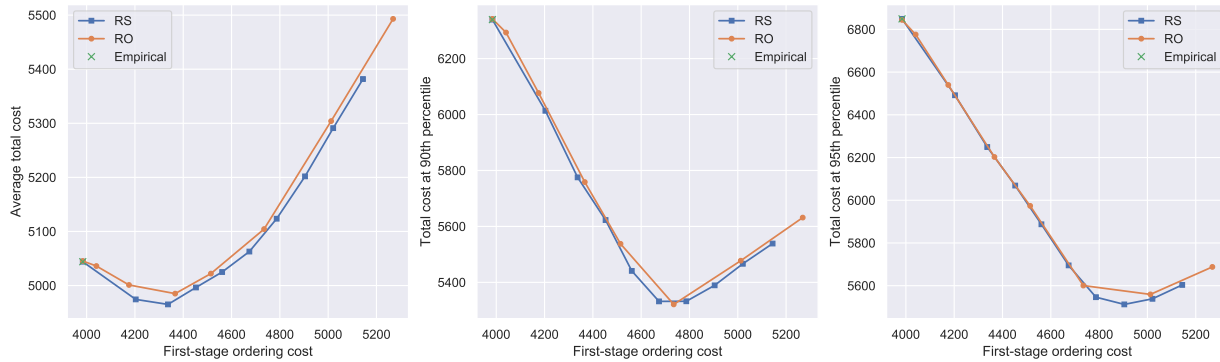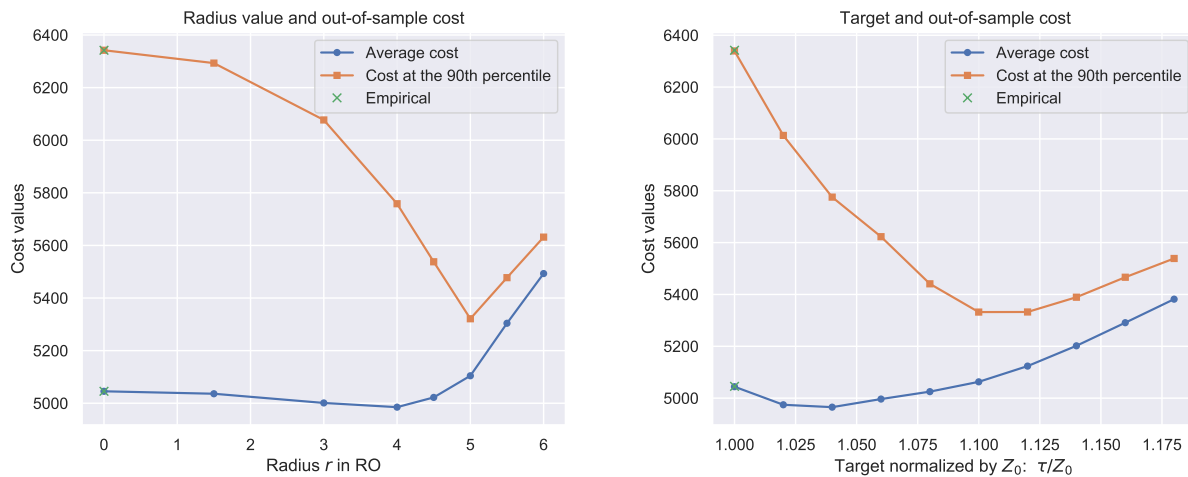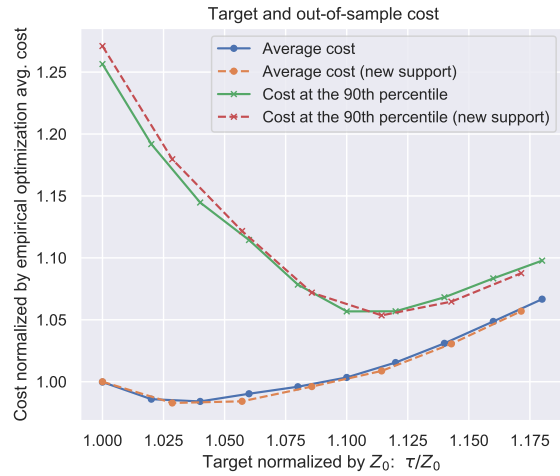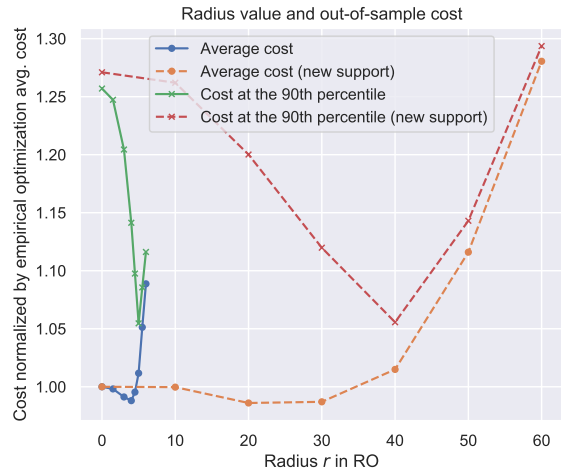

Figure 10   **Summary of performance: RS and RO models at different first-stage costs**

(a) Cost metrics w.r.t. Wasserstein radius r (b) Cost metrics w.r.t. target τ

Figure 11   **Out-of-sample cost metrics w.r.t. Wasserstein radius r in RO (left) and target τ in RS (right).**

60

(a) Cost metrics w.r.t. Wasserstein radius r

(b) Cost metrics w.r.t. target τ

Figure 12   Change of the "nice" range of radius in RO (left) and normalized target in RS (right).

D. A Sample Code for RS and RO using RSOME

RSOME is a Matlab modeling package for distributionally robust optimization (Chen et al. 2020), which can be used to solve robust satisficing and robust optimization. Here, we provide a sample code of the adaptive network lot-sizing problem. The sample code for the RO model is as follows:

```matlab
  %%%%Parameters%%%%
% S: number of samples
% N: number of stores
% D: maximum possible demand
% c: cost of ordering
% l: cost of emergency ordering
% t: transportation cost
% d: historical samples of demands
% r: radius of the Wasserstein ball

%%%%Variables%%%%
% x: initial stock allocation
% w: second-stage emergency ordering
% y: second-stage transhipment


%%%%Model%%%%
model = rsome('RO');

% Define random variables
z = model.random(N,1);        %demand
u = model.random;             %lifted variable

% Define scenarios and the lifted joint ambiguity set
P = model.ambiguity(S);
for s = 1:S
    P(s).suppset(0 <= z, z <= D, norm(z - d(s,:)') <= u);
end
pr = P.prob;
P.probset(pr == 1/S);

% Define event-wise expectation
P.exptset(expect(u) <= r);

% Declare Warsserstein ambiguity set
model.with(P);

% Define decision variables
x = model.decision(N,1);
y = model.decision(N,N);
w = model.decision(N,1);

% Define scenario-wise adaptation
for s = 1:S
    y.evtadapt(s);
```

```
    w.evtadapt(s);
end
% Define affine adaptation
y.affadapt(z);
y.affadapt(u);
w.affadapt(z);
w.affadapt(u);
% Define objective function
model.min(c'*x+expect(sum(sum(t.*y)) + c2'*w));

% Define constraints
for i = 1:N
    model.append(z(i)-x(i)-w(i)+sum(y(i,:))-sum(y(:,i)) <= 0);
end

model.append(y >= 0);
model.append(w >= 0);
model.append(x >= 0);
model.append(x <= D);

% Solution
model.solve;
```

The sample code for the RS model is as follows:

```
  %%%%Parameters%%%%
% S: number of samples
% N: number of stores
% D: maximum possible demand
% c: cost of ordering
% l: cost of emergency ordering
% t: transportation cost
% d: historical samples of demands
% T: target


%%%%Variables%%%%
% x: initial stock allocation
% w: second-stage emergency ordering
% y: second-stage transhipment


%%%%Model%%%%
model = rsome('RS');

% Define random variables
z = model.random(N,1);        %demand
u = model.random;             %lifted variable

% Define scenarios and the lifted joint ambiguity set
P = model.ambiguity(S);
```

```
for s = 1:S
    P(s).suppset(0 <= z, z <= D, norm(z - d(s,:)',1) <= u);
end

pr = P.prob;
P.probset(pr == 1/S);
% Declare ambiguity set
model.with(P);

% Define decision variables
k = model.decision();
x = model.decision(N,1);
y = model.decision(N,N);
w = model.decision(N,1);

% Define scenario-wise adaptation
for s = 1:S
    y.evtadapt(s);
    w.evtadapt(s);
end
% Define affine adaptation
y.affadapt(z);
y.affadapt(u);
w.affadapt(z);
w.affadapt(u);
% Define objective function
model.min(k);

% Define constraints
model.append(c'*x+expect(sum(sum(t.*y)) + c2'*w - k*u) <= T)
for i = 1:N
    model.append(z(i)-x(i)-w(i)+sum(y(i,:))-sum(y(:,i)) <= 0);
end

model.append(y >= 0);
model.append(w >= 0);
model.append(x >= 0);
model.append(x <= D);

% Solution
model.solve;
```