manuscript

Robust Actionable Prescriptive Analytics

Li Chen
Institute of Operations Research and Analytics, National University of Singapore
chen_l@u.nus.edu

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

Xun Zhang
Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai 200030, China
xunzhang@sjtu.edu.cn

Long Zhao
Department of Analytics & Operations (DAO), NUS Business School, National University of Singapore
longzhao@nus.edu.sg

Minglong Zhou
School of Management, Fudan University, Shanghai, China
minglong.zhou@outlook.com

We propose a new robust actionable prescriptive analytics framework that leverages past data and side information to minimize a risk-based objective function under distributional ambiguity. Our framework aims to find a policy that directly transforms the side information into implementable decisions. Specifically, we focus on developing actionable response policies that offer the benefits of interpretability and implementability. To address the potential issue of overfitting to empirical data, we adopt a data-driven robust satisficing approach that effectively handles uncertainty. We tackle the computational challenge for linear optimization models with recourse by developing a new tractable safe approximation for robust constraints, accommodating bilinear uncertainty and general norm-based uncertainty sets. Additionally, we introduce a biaffine recourse adaptation to enhance the quality of the approximation. Furthermore, we present a localized robust satisficing model that efficiently solves combinatorial optimization problems with tree-based static policies. Finally, we demonstrate the practical application of our framework through a simulation case study on risk-minimizing portfolio optimization using past returns as side information. We also provide a simulation case study on how the framework can be applied to obtain an interpretable policy for allocating taxis to different demand regions in response to weather information.

Key words: robust optimization, robust satisficing, robust analytics, prescriptive analytics, side information

1. Introduction

A business-inspired problem typically involves a decision model that incorporates relevant data from past realizations of uncertain parameters that influence the decision model, as well as side information that possesses predictive capabilities regarding those uncertainties. Predictive analytics, which focuses on determining the statistical aspects of uncertain outcomes based on the side information, is complemented by prescriptive analytics, which directly translates the side information into actionable decisions.

There has been a significant surge in research on prescriptive analytics models in recent years. Many of these models adopt a two-step approach known as "predict, then optimize," as exemplified by Ferreira et al. (2016) and Glaeser et al. (2019). In these models, decision-makers first predict the statistical properties of uncertain outcomes using the side information and subsequently employ these predictions as inputs to an optimization problem, thereby obtaining the optimal decision. However, the two-step approach could lead to inferior decisions, as elucidated by Liyanage and Shanthikumar (2005). Tulabandhula and Rudin (2013), Elmachtoub and Grigas (2022), Loke et al. (2021), among others, have proposed adjustments to the "predict, then optimize" approach by incorporating aspects of the decision process to the prediction models. Side information is also used to approximate the conditional distribution of outcome variables, which is then passed on to the stochastic optimization problem (see, *e.g.,* Hannah et al. 2010, Bertsimas and Kallus 2020, Srivastava et al. 2021, Esteban-Pérez and Morales 2021). For example, Bertsimas and Kallus (2020) employ side information to reweigh historical samples through pre-trained machine learning models, enabling the calibration of stochastic optimization problems when new observations are obtained. Kallus and Mao (2023) introduce novel approximate splitting criteria to adapt the conditional distribution for downstream decision problems by training a stochastic optimization forest.

In addition to the "predict, then optimize" approach, some prescriptive analytics models bypass the prediction step altogether and directly determine the optimal response policy based on side information. For instance, Ban and Rudin (2019) consider a newsvendor problem with an affine ordering policy that is obtained by solving an empirical optimization model. Notz and Pibernik (2022), Bertsimas and Koduri (2022) propose models that enable the search for optimal policies within a reproducing kernel Hilbert space. Bertsimas et al. (2019a) formulate an optimal prescriptive tree that accommodates discrete decision policies.

Despite the availability of historical data, the decision-maker lacks knowledge of the underlying probability distribution that generates the data. Several approaches have been proposed in the literature to mitigate the risk of overfitting when using the empirical distribution to evaluate the risk-based objective function. Notably, Esfahani and Kuhn (2018) introduce a foundational framework for data-driven distributionally robust optimization, employing an ambiguity set based on the Wasserstein metric. This framework has gained popularity and found applications in prediction and optimization models. Interestingly, Gao et al. (2022), Shafieezadeh-Abadeh et al. (2019a), Blanchet et al. (2019a) show that regularization techniques commonly used in regression and classification models can also be viewed through the lens of data-driven robust optimization. Another approach is the robust satisficing approach proposed by Long et al. (2023), which specifies a target parameter. Similar to regularization and data-driven robust optimization, the target parameter can serve as a hyper-parameter determined through cross-validation to enhance out-of-sample performance.

There are also extensions of the data-driven robust optimization models to include side information. Bertsimas et al. (2023a) estimate the conditional probability distribution of the random problem parameters using side information before incorporating it in the robust optimization model (see also, Hao et al. 2020, Kannan et al. 2020, Nguyen et al. 2021, Bertsimas and Van Parys 2021, Sim et al. 2023). Zhang et al. (2022) solve the optimal non-parametric policy for a distributionally robust newsvendor problem. Yang et al. (2022) study distributionally robust prescriptive analytics with a distributional uncertainty set based on a causal transport distance.

In addition to the challenge of handling data uncertainty, the need for interpretability of policies presents another significant barrier to the widespread adoption of prescriptive analytics in practical settings. Existing approaches often fail to provide insights into the decision-making process, leaving stakeholders hesitant to embrace decisions generated by black-box models or policies that, while technically optimal, are incomprehensible (Arrieta et al. 2020). To address this issue, our paper focuses on using tree-based static and affine policies, which are actionable policies renowned for their high interpretability (Bertsimas and Stellato 2021, Lipton 2018). By employing these approaches, we aim to provide stakeholders with a reasonable interpretation of how and why decisions are being made, facilitating their acceptance and implementation.

Summary of contributions

(i) We propose a general robust actionable prescriptive analytics framework, solving for an *actionable response policy* that offers benefits of both interpretability and implementability. This policy directly transforms side information into implementable decisions by enforcing policy feasibility. To avoid overfitting, we adopt the data-driven distributionally robust satisficing framework to mitigate data uncertainty.

(ii) To tackle the computational challenge from the proposed framework for linear optimization problems with recourse, we develop a novel tractable safe approximation for robust constraints with bilinear uncertainty and general norm-based uncertainty sets. These challenges have yet to be sufficiently addressed in the existing literature. Additionally, we introduce a biaffine recourse adaptation to enhance the quality of the approximation.

(iii) We propose a localized robust satisficing model that efficiently solves combinatorial optimization problems for a tree-based static policy.

(iv) We provide a numerical study on a portfolio optimization problem that aims to minimize the conditional Value-at-Risk (CVaR), using past returns as side information. We compare the out-of-sample performance of the tree-based affine and forest-based policies on simulated and real data. Surprisingly, despite the simplicity of the tree-based affine policy, we find that it exhibits superior performance and robustness against noise. We also provide a simulation case study on how the framework can be applied to obtain an interpretable policy for allocating taxis to different demand regions in response to weather information.

Our work contributes to the literature on optimizing policy mappings from side information to decisions. However, it distinguishes itself from existing literature in two key aspects. First, we consider general two-stage linear optimization problems with fixed recourse, unlike previous papers that often focus on specific cost functions (Ban and Rudin 2019, Zhang et al. 2022, Yang et al. 2022) or simplified decision scenarios (Bertsimas et al. 2019a). Second, while some papers address two-stage linear optimization with side information, they either do not explicitly enforce constraint feasibility, resulting in non-implementable policies potentially (Bertsimas and Koduri 2022, Notz and Pibernik 2022), or they do not cover robust optimization beyond right-hand side uncertainty (Bertsimas et al. 2023b). To the best of our knowledge, our paper is the first to provide a framework for general robust two-stage linear optimization over policies mapping from side information to implementable decisions.

Notation. We denote by $\mathbb{R}$ ($\mathbb{R}_+$) the set of real (non-negative) numbers. We use boldface lowercase letters for vectors (*e.g.*, $\boldsymbol{\theta}$), and calligraphic letters for sets (*e.g.*, $\mathcal{X}$). For a vector $\boldsymbol{\theta} \in \mathbb{R}^n$, we denote by $\text{diag}(\boldsymbol{\theta}) \in \mathbb{R}^{n\times n}$ the diagonal matrix with elements $\boldsymbol{\theta}$. We use $|\cdot|$ to denote the cardinality of a finite set. We use $[n] \triangleq \{1, 2, \ldots, n\}$ to denote the running index for a positive integer n. We denote $\mathbb{1}_{\mathcal{Z}}$ as the indicator function of the set $\mathcal{Z}$, i.e., $\mathbb{1}_{\mathcal{Z}}(\boldsymbol{x})$ equals 1 if $\boldsymbol{x} \in \mathcal{Z}$ and 0 otherwise. We denote by $\text{cl}(\mathcal{X})$ and $\text{int}(\mathcal{X})$ the closure and interior of the set $\mathcal{X}$, respectively. 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 $\tilde{v} \sim \mathbb{P}$. We use $\mathbb{E}_{\mathbb{P}}[\tilde{v}]$ to denote its expectation with respect to its distribution $\mathbb{P}$. We use $\mathcal{P}_0(\mathcal{Z})$ to represent the set of all possible distributions for a random vector that has support $\mathcal{Z} \subseteq \mathbb{R}^n$. We use $\mathcal{R}^{n_1,n_2}$ and $\mathcal{L}^{n_1,n_2}$ to denote the set of all mappings and its sub-class of affine mappings, respectively, from $\mathbb{R}^{n_1}$ to $\mathbb{R}^{n_2}$. Specifically, $\boldsymbol{x} \in \mathcal{L}^{n_1,n_2}$ implies the expression:

$$\boldsymbol{x}(\boldsymbol{u}) = \boldsymbol{x}^0 + \sum_{j\in[n_1]} \boldsymbol{x}^j u_j \qquad \forall \boldsymbol{u} \in \mathbb{R}^{n_1}$$

for some $\boldsymbol{x}^j \in \mathbb{R}^{n_2}, j \in \{0, \ldots, n_1\}$. It also applies to mapping to a matrix such as $\boldsymbol{F} \in \mathcal{R}^{n_1,n_2\times n_3}$, to represent the mapping $\boldsymbol{F} : \mathbb{R}^{n_1} \to \mathbb{R}^{n_2\times n_3}$. Finally, $\mathbf{0}$ ($\mathbf{1}$) denotes the vector of all zeros (ones) and $\boldsymbol{e}_i$ denotes the ith basis vector. The dimensions of these vectors should be clear from the context.

2. Prescriptive analytics with side information

To set up the prescriptive analytics framework, we consider a decision model evaluated on a *cost function* $g : \mathbb{R}^{n_x} \times \mathbb{R}^{n_v} \to \mathbb{R} \cup \{\infty\}$, where the input to the first argument is the *here-and-now response decision*, while the input to the second argument is the *outcome variable*, which captures the model's uncertainty. We first observe the *side information* that has some predictive power on those uncertain outcomes and then make the response decision. In other words, the output of the prescriptive analytics model is the *response policy*, $\boldsymbol{x} \in \mathcal{X} \subseteq \mathcal{R}^{n_u, n_x}$, which transforms realized side information to response decisions. Specifically, the response policy is selected over the *actionable policy set*, $\mathcal{X}$, which is designed to facilitate the interpretation of response decisions and ensure that the model's constraints are always feasible while maintaining the computational tractability of the optimization problem. Note that in the context of a prediction model such as regression, we can associate the cost function with the loss function and the response policy with the *target function* optimized over the *hypothesis set*. Hence, the hypothesis set plays a similar role as the actionable policy set in facilitating the interpretation of the target function.

We model the data stream as random variables with an unobservable data-generating probability distribution. Specifically, we use $\tilde{\boldsymbol{u}}$ to denote the random side information and $\tilde{\boldsymbol{v}}$ to denote the random outcome variables, and their support sets are represented by the sets $\mathcal{U} \subseteq \mathbb{R}^{n_u}$ and $\mathcal{V} \subseteq \mathbb{R}^{n_v}$ respectively. For convenience, we also denote the random variable $\tilde{\boldsymbol{z}} \triangleq (\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}})$ with probability distribution $\mathbb{P}^\star \in \mathcal{P}_0(\mathcal{Z})$, $\mathcal{Z} \subseteq \mathcal{U} \times \mathcal{V}$. We assume that the distribution $\mathbb{P}^\star$ is stationary and does not depend on the response decisions. In an ideal setting with full information and infinite computational resources, we would solve for optimal response policy in the following optimization problem,

$$\begin{aligned} Z^\star = \min\ & \mathbb{E}_{\mathbb{P}^\star}\left[g(\boldsymbol{x}(\tilde{\boldsymbol{u}}), \tilde{\boldsymbol{v}})\right] \\ \text{s.t.}\ & \boldsymbol{x} \in \mathcal{X} \end{aligned} \tag{1}$$

where the objective function is the expectation of the random cost function. Hence, the total costs would be minimized when the ideal optimal response policy is implemented over an infinite period under identical conditions.

Although the actual data generating distribution $\mathbb{P}^\star$ is unobservable, the decision maker has access to S historical realizations of the random variable $\tilde{\boldsymbol{z}}$, which we denote by $\hat{\boldsymbol{z}}_s = (\hat{\boldsymbol{u}}_s, \hat{\boldsymbol{v}}_s)$, $s \in [S]$. We denote the empirical distribution by $\hat{\mathbb{P}} \in \mathcal{P}_0(\mathcal{Z})$, $\tilde{\boldsymbol{z}} \sim \hat{\mathbb{P}}$, such that $\hat{\mathbb{P}}\left[\tilde{\boldsymbol{z}} = \hat{\boldsymbol{z}}_s\right] = 1/S$, $s \in [S]$. As an approximation to Problem (1), it is reasonable to use the empirical distribution $\hat{\mathbb{P}}$ to evaluate the objective function by solving the following *empirical optimization problem*,

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

Assumption 1. *We assume that the empirical optimization problem* (2) *is solvable,* i.e.*, there exists* $\hat{\boldsymbol{x}} \in \mathcal{X}$ *such that*

$$Z_0 = \frac{1}{S}\sum_{s\in[S]} g(\hat{\boldsymbol{x}}(\hat{\boldsymbol{u}}_s), \hat{\boldsymbol{v}}_s).$$

Actionable policy set

The response policy is optimized over the actionable policy set, which takes the general form,

$$\mathcal{X} = \{\boldsymbol{x} \in \mathcal{A} \,|\, \boldsymbol{A}(\boldsymbol{u})\boldsymbol{x}(\boldsymbol{u}) \leq \boldsymbol{b}(\boldsymbol{u}), \boldsymbol{x}(\boldsymbol{u}) \in \mathcal{D},\ \forall \boldsymbol{u} \in \mathcal{U}\},$$

for some feasible set, $\mathcal{D} \subseteq \mathbb{R}^{n_x}$, mappings $\boldsymbol{A}$ and $\boldsymbol{b}$ to appropriate dimensions of matrix and vector respectively, and a sub-class of mappings $\mathcal{A} \subseteq \mathcal{R}^{n_u, n_x}$ that would facilitate interpretation of the policy. Note that apart from the linear constraints that must be satisfied for all possible realization of the side information, the feasible set $\mathcal{D}$ can also enforce nonlinear constraints, such as discrete constraints commonly used in practice. We posit that the actionable policy set should endow with the following desirable properties:

- **Interpretability:** The response policy should be intuitive and easy to understand and interpret.
- **Implementability:** Every response decision induced by the policy must satisfy the constraints mandated by the prescriptive analytics model. It should also be computationally tractable to optimize policies over the actionable policy set.

The property of interpretability is crucial, as it enhances the degree to which a human can discern the rationale behind a data-driven decision and understand its implications (Bertsimas and Stellato 2021). A policy endowed with greater transparency also facilitates the model calibration for improving the quality of various decisions or predictions (Lipton 2018).

Additionally, it is important to equip the decision-maker with the ability to elucidate why specific choices are made, especially when responding to side information (Arrieta et al. 2020). An unconventional correlation might be an intriguing new discovery, but more often than not, it can be traced back to issues with the data. Correcting such data-related problems often nullifies the correlation. Certain policy mappings, such as affine mappings, which are prevalent in statistical and econometric analyses, and tree-based mappings, typically employed in decision analysis, boast high interpretability. One noteworthy example is the regression tree, which synthesizes these two highly interpretable mappings. While there is an array of machine learning models like neural networks and ensemble models, their interpretability is not given. Therefore, these models are not within our purview for consideration.

Successful implementation of prescriptive analytics models often requires communicating the response policies to stakeholders. The implementable property ensures that the response decisions can be implemented in practice. For instance, they can include capacity constraints and restrict that ordering quantities must satisfy non-negative. In contrast, a hypothesis set in a predictive analytics model does not need to adhere to this property because target functions are typically not the prescribed solutions. Finally, the implementability property should also ensure that the optimal response policy can be obtained with reasonable computational efforts.

Tree-based policies

We focus on tree-based policies; specifically, we consider a decision tree on the side information with L leaf nodes that is constructed by splitting the support $\mathcal{U}$ into L non-overlapping and bounded subsets, $\mathcal{U}_\ell \subseteq \mathcal{U}$, $\ell \in [L]$. Specifically, each subset is a nonempty bounded hyper-rectangular polyhedron, which may not be closed such that

$$\mathrm{cl}(\mathcal{U}_\ell) = \{\boldsymbol{u} \in \mathbb{R}^{n_u} \mid \underline{\boldsymbol{u}}_\ell \le \boldsymbol{u} \le \bar{\boldsymbol{u}}_\ell\}.$$

Each node $\ell \in [L]$ can be associated with a set of historical samples $\mathcal{S}_\ell \subseteq [S]$ such that $\mathcal{S}_\ell = \{s \in [S] \mid \hat{\boldsymbol{u}}_s \in \mathcal{U}_\ell\}$, and the hyper-rectangular support $\mathcal{V}_\ell \subseteq \mathcal{V}$ is given by

$$\mathcal{V}_\ell = \{\boldsymbol{v} \in \mathbb{R}^{n_v} \mid \underline{\boldsymbol{v}}_\ell \le \boldsymbol{v} \le \bar{\boldsymbol{v}}_\ell\},$$

such that $\hat{\boldsymbol{v}}_s \in \mathcal{V}_\ell$ for all $s \in \mathcal{S}_\ell$, $\ell \in [L]$. Therefore, we also can express $\mathcal{Z} = \cup_{\ell \in [L]} \mathcal{Z}_\ell$, where $\mathcal{Z}_\ell = \mathcal{U}_\ell \times \mathcal{V}_\ell$. To some degree, this also captures a relation between side information and the outcome variables. We next define the sub-class of tree-based static mappings as follows,

$$\mathcal{T}^{n_1,n_2} \triangleq \left\{ \boldsymbol{x} \in \mathcal{R}^{n_1,n_2} \;\middle|\; \begin{array}{l} \exists \boldsymbol{x}_\ell^0 \in \mathbb{R}^{n_2}, \ell \in [L]: \\ \boldsymbol{x}(\boldsymbol{u}) = \boldsymbol{x}_\ell^0 \text{ if } \boldsymbol{u} \in \mathcal{U}_\ell \text{ for some } \ell \in [L] \end{array} \right\},$$

and the sub-class of tree-based affine mappings as follows,

$$\bar{\mathcal{T}}^{n_1,n_2} \triangleq \left\{ \boldsymbol{x} \in \mathcal{R}^{n_1,n_2} \;\middle|\; \begin{array}{l} \exists \boldsymbol{x}_\ell \in \mathcal{L}^{n_1,n_2}, \ell \in [L]: \\ \boldsymbol{x}(\boldsymbol{u}) = \boldsymbol{x}_\ell(\boldsymbol{u}) \text{ if } \boldsymbol{u} \in \mathcal{U}_\ell \text{ for some } \ell \in [L] \end{array} \right\}.$$

The choice of the sub-class of mappings $\mathcal{A}$ has ramifications on the computational tractability of the overall optimization problem in ensuring the feasibility of the model's constraints. In particular, the sub-class of tree-based static mappings allows us to include more sophisticated constraints within the actionable policy set. When $\mathcal{A} = \mathcal{T}^{n_u,n_x}$, we can let $\boldsymbol{A}$ and $\boldsymbol{b}$ be affine mappings, and we express the empirical optimization problem of Problem (2), as the following robust optimization problem,

$$\begin{aligned} Z_0 = \min\ & \frac{1}{S} \sum_{\ell \in [L]} \sum_{s \in \mathcal{S}_\ell} g(\boldsymbol{x}_\ell^0, \hat{\boldsymbol{v}}_s) \\ \text{s.t.}\ & \boldsymbol{A}(\boldsymbol{u}) \boldsymbol{x}_\ell^0 \le \boldsymbol{b}(\boldsymbol{u}) && \forall \boldsymbol{u} \in \mathcal{U}_\ell, \ell \in [L] \\ & \boldsymbol{x}_\ell^0 \in \mathcal{D} && \forall \ell \in [L]. \end{aligned} \tag{3}$$

The sub-class of tree-based affine mappings, *i.e.*, $\mathcal{A} = \bar{\mathcal{T}}^{n_u,n_x}$ would maintain a computationally tractable format when $\mathcal{D} = \mathbb{R}^{n_x}$, $\boldsymbol{A}$ is a static mapping and $\boldsymbol{b}$ is an affine mapping. Note that the ubiquitous affine policy analogous to the affine target function in linear regression is a special case of a tree-based affine mapping with one leaf node, $L = 1$. Accordingly, we can express the empirical optimization problem of Problem (2), as the following robust optimization problem,

$$\begin{aligned} Z_0 = \min\ & \frac{1}{S} \sum_{\ell \in [L]} \sum_{s \in \mathcal{S}_\ell} g(\boldsymbol{x}_\ell(\hat{\boldsymbol{u}}_s), \hat{\boldsymbol{v}}_s) \\ \text{s.t.}\ & \boldsymbol{A}\boldsymbol{x}_\ell(\boldsymbol{u}) \leq \boldsymbol{b}(\boldsymbol{u}) && \forall \boldsymbol{u} \in \mathcal{U}_\ell, \ell \in [L] \\ & \boldsymbol{x}_\ell \in \mathcal{L}^{n_u,n_x} && \forall \ell \in [L]. \end{aligned} \tag{4}$$

Since the support sets $\mathcal{U}_\ell$, $\ell \in [L]$ are nonempty and bounded, we can replace $\mathcal{U}_\ell$ with its closure when solving the robust constraints. It is important to note that the set of robust constraints is an important feature in prescriptive analytics, which could ensure that the coefficients associated with affine mappings are also bounded. Hence, issues such as identifiability, which could arise in a regression model, may not pose a problem.

Constructing a tree from data

Similar to the classical classification and regression tree (CART) procedure, we can construct the policy tree using the *binary recursive partitioning* heuristics, which is an iterative procedure that splits the data into partitions or branches. At each split, we need to determine a side information index and a threshold such that the samples at one node are divided into two groups. The split should be determined based on the best performance on the empirical optimization models, *i.e.*, Problem (3) for the tree-based static policy and Problem (4) for the tree-based affine policy. For instance, for the tree-based affine policy, at each node $\ell \in [L]$ and side information index $i \in [n_u]$, we solve for $\omega^*_{\ell i} = \arg\min_{\omega \in \Omega_{\ell i}} \zeta_{\ell i}(\omega)$, where $\Omega_{\ell i}$ is a discrete subset within the interior of $[\underline{u}_{\ell i}, \bar{u}_{\ell i}]$, and

$$\begin{aligned} \zeta_{\ell i}(\omega) \triangleq \min\ & \left(\sum_{s \in \mathcal{S}_{\ell i}(\omega)} g(\boldsymbol{x}_1(\hat{\boldsymbol{u}}_s), \hat{\boldsymbol{v}}_s) + \sum_{s \in \mathcal{S}_\ell \setminus \mathcal{S}_{\ell i}(\omega)} g(\boldsymbol{x}_2(\hat{\boldsymbol{u}}_s), \hat{\boldsymbol{v}}_s) \right) \\ \text{s.t.}\ & \boldsymbol{A}\boldsymbol{x}_1(\boldsymbol{u}) \leq \boldsymbol{b}(\boldsymbol{u}) && \forall \boldsymbol{u} \in \mathcal{U}_{\ell i}(\omega) \\ & \boldsymbol{A}\boldsymbol{x}_2(\boldsymbol{u}) \leq \boldsymbol{b}(\boldsymbol{u}) && \forall \boldsymbol{u} \in \mathcal{U}_\ell \setminus \mathcal{U}_{\ell i}(\omega) \\ & \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{L}^{n_u,n_x}, \end{aligned} \tag{5}$$

with $\mathcal{U}_{\ell i}(\omega) \triangleq \{\boldsymbol{u} \in \mathcal{U}_\ell \mid u_i \leq \omega\}$ and $\mathcal{S}_{\ell i}(\omega) \triangleq \{s \in \mathcal{S}_\ell \mid \hat{\boldsymbol{u}}_s \in \mathcal{U}_{\ell,i}(\omega)\}$. Note that Problem (5) is typically a tractable optimization problem, which allows us to determine $\omega^*_{\ell i}$ efficiently by solving all instances of the problem for $\omega \in \Omega_{\ell i}$. Observe that $\mathcal{U}_{\ell i}(\infty) = \mathcal{U}_\ell$ and $\zeta_{\ell i}(\omega) \leq \zeta_{\ell i}(\infty)$ for all $\omega \in [\underline{u}_{\ell i}, \bar{u}_{\ell i}]$. Hence, the next node and side information index to split is determined by

$$(\ell^*, i^*) = \arg\max_{\ell \in [L], i \in [n_u]} \{\zeta_{\ell i}(\infty) - \zeta_{\ell i}(\omega^*_{\ell i})\},$$

which corresponds to the greatest reduction in the objective value in the empirical optimization model after the split. After that, we continue splitting each partition into smaller groups as the heuristics move up each branch. We summarize the binary recursive partitioning algorithm below:

1. Initialize $\Lambda \leftarrow 1, \mathcal{U}_1 \leftarrow \mathcal{U}$, and the desired number of leaf nodes $L > 1$.
2. For every $\ell \in [\Lambda]$ and $i \in [n_u]$, determine $\omega^*_{\ell i} = \arg\min_{\omega \in \Omega_{\ell i}} \zeta_{\ell i}(\omega)$.
3. Determine $(\ell^*, i^*) = \arg\max_{\ell \in [\Lambda], i \in [n_u]} \{\zeta_{\ell i}(\infty) - \zeta_{\ell i}(\omega^*_{\ell i})\}$.
4. If $\Lambda < L$, then:
    - Create $\mathcal{U}_{\Lambda+1} \leftarrow \mathcal{U}_{\ell^*} \backslash \mathcal{U}_{\ell^* i^*}(\omega^*_{\ell^* i^*})$
    - Replace $\mathcal{U}_{\ell^*} \leftarrow \mathcal{U}_{\ell^* i^*}(\omega^*_{\ell^* i^*})$.
    - Set $\Lambda \leftarrow \Lambda + 1$
    - Proceed to Step 2
5. Algorithm terminates with $\mathcal{U}_\ell$ for $\ell \in [L]$.

We can adopt a cross-validation procedure to determine the desired number of leaf nodes. Let $\bar{L}$ be the maximum number of leaf nodes to consider. Next, we apply the binary recursive partitioning algorithm for each $L \in [\bar{L}]$ to determine the tree structure with L leaf nodes. Among the $\bar{L}$ different tree structures, we use a K-fold cross-validation procedure to determine the desired tree configuration. For this purpose, we split the historical data into K non-overlapping testing data sets. For each $k \in [K]$, the k-th training data set comprises the historical data without the k-th testing data set. Subsequently, for a given tree configuration, we solve the empirical optimization problems on the training sets and evaluate the performance of the response policies on their corresponding testing data sets. For each $L \in [\bar{L}]$, we record the average out-of-sample performance over the K iterations, and we select the tree configuration that gives the best average out-of-sample performance.

3. Robust optimization and satisficing

It has been well-known that solutions from empirical optimization models would have inferior out-of-sample results (Smith and Winkler 2006). To address the issue of overfitting, Esfahani and Kuhn (2018) propose a framework for data-driven robust optimization by incorporating an ambiguity set of probability distributions that are proximal to the empirical distribution with respect to the Wasserstein metric. Likewise, we consider the following data-driven robust optimization problem

$$\begin{aligned} \min \ & \sup_{\mathbb{P} \in \mathcal{F}(\Gamma)} \mathbb{E}_{\mathbb{P}}\left[g(\boldsymbol{x}(\tilde{\boldsymbol{u}}), \tilde{\boldsymbol{v}})\right] \\ \text{s.t.} \ & \boldsymbol{x} \in \mathcal{X}, \end{aligned} \tag{6}$$

where the ambiguity set is defined by the type-I Wasserstein distance as follows

$$\mathcal{F}(\Gamma) \triangleq \left\{ \mathbb{P} \in \mathcal{P}_0(\mathcal{Z}) \ \middle|\ \begin{array}{l} (\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}}) \sim \mathbb{P} \\ \Delta(\mathbb{P}, \hat{\mathbb{P}}) \leq \Gamma \end{array} \right\},$$

and

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

The confidence guarantees that the true distribution, $\mathbb{P}^\star$, resides within the Wassterstein-based ambiguity set, $\mathcal{F}(\Gamma)$ has been established in Fournier and Guillin (2015). In this context, if the true data-generating distribution $\mathbb{P}^\star$ (where $(\tilde{\boldsymbol{u}},\tilde{\boldsymbol{v}})\sim\mathbb{P}^\star$) is a light-tailed distribution, and $\mathbb{P}^S$ is the distribution guiding the dispersion of independent samples $(\hat{\boldsymbol{u}}_s,\hat{\boldsymbol{v}}_s)$, with $s\in[S]$, drawn from $\mathbb{P}^\star$. Then for any $\Gamma>0$, the probability that the divergence $\Delta(\mathbb{P}^\star,\hat{\mathbb{P}})$ exceeds Γ, as governed by $\mathbb{P}^S$, will decrease exponentially towards zero as the sample size, S, increases. Esfahani and Kuhn (2018) adopt the concentration results in Fournier and Guillin (2015) to derive a finite-sample guarantee. Although tighter results have also been established in Blanchet et al. (2019b), Shafieezadeh-Abadeh et al. (2019b), Si et al. (2020), and Gao (2022), as these bounds are implicitly derived, they are typically not used in practice to determine the size of the Wasserstein ambiguity set, Γ. Moreover, it is impossible to know the parameters for the bounds or whether the data is independently generated. Instead, like the regularization term in machine learning models, the parameter Γ is a hyper-parameter that should be determined via cross-validation techniques.

From the results of Esfahani and Kuhn (2018), we can reformulate Problem (6) as a robust optimization problem,

$$\begin{aligned} Z_\Gamma = \min\ & \kappa\Gamma + \frac{1}{S}\sum_{s\in[S]} t_s \\ \text{s.t.}\ & \sup_{(\boldsymbol{u},\boldsymbol{v})\in\mathcal{Z}} \{g(\boldsymbol{x}(\boldsymbol{u}),\boldsymbol{v}) - \kappa(\|\boldsymbol{v}-\hat{\boldsymbol{v}}_s\|+\|\boldsymbol{u}-\hat{\boldsymbol{u}}_s\|)\} \le t_s \ \forall s\in[S] \\ & \boldsymbol{x}\in\mathcal{X}, \kappa\ge 0. \end{aligned} \tag{7}$$

Long et al. (2023) propose a robust satisficing model that can also be used to address the issue of overfitting in data-driven optimization problems. Instead of sizing the ambiguity set with Γ, the robust satisficing model is specified by a target $\tau\ge Z_0$, which is easier to interpret than Γ. In this context, we can formulate the robust satisficing model as the following optimization problem,

$$\begin{aligned} \min\ & \kappa \\ \text{s.t.}\ & \sup_{\mathbb{P}\in\mathcal{P}_0(\mathcal{Z})} \left\{ \mathbb{E}_{\mathbb{P}}\left[g(\boldsymbol{x}(\tilde{\boldsymbol{u}}),\tilde{\boldsymbol{v}})\right] - \kappa\Delta(\mathbb{P},\hat{\mathbb{P}}) \right\} \le \tau \\ & \boldsymbol{x}\in\mathcal{X},\ \kappa\ge 0, \end{aligned} \tag{8}$$

or equivalently as the following robust optimization problem,

$$\begin{aligned} \kappa_\tau = \min\ & \kappa \\ \text{s.t.}\ & \frac{1}{S}\sum_{s\in[S]} t_s \le \tau \\ & t_s \ge \sup_{(\boldsymbol{u},\boldsymbol{v})\in\mathcal{Z}} \{g(\boldsymbol{x}(\boldsymbol{u}),\boldsymbol{v}) - \kappa(\|\boldsymbol{v}-\hat{\boldsymbol{v}}_s\|+\|\boldsymbol{u}-\hat{\boldsymbol{u}}_s\|)\} \ \forall s\in[S] \\ & \boldsymbol{x}\in\mathcal{X}, \kappa\ge 0. \end{aligned} \tag{9}$$

We establish the feasibility of the robust satisficing model.

Theorem 1. *Suppose Assumption 1 holds and the function $g(\boldsymbol{x}(\boldsymbol{u}),\boldsymbol{v})$ is Lipschitz continuous with respect to $\boldsymbol{u}$ and $\boldsymbol{v}$ such that*

$$g(\boldsymbol{x}(\boldsymbol{u}_1),\boldsymbol{v}_1) - g(\boldsymbol{x}(\boldsymbol{u}_2),\boldsymbol{v}_2) \le \bar{L}(\|\boldsymbol{u}_1-\boldsymbol{u}_2\| + \|\boldsymbol{v}_1-\boldsymbol{v}_2\|) \qquad \forall \boldsymbol{x}\in\mathcal{X}, \boldsymbol{u}_1,\boldsymbol{u}_2\in\mathcal{U}, \boldsymbol{v}_1,\boldsymbol{v}_2\in\mathcal{V}.$$

Then, the robust satisficing model, Problem (9)*, is feasible for all $\tau \ge Z_0$ for some $\kappa_\tau \le \bar{L}$, where Z_0 is the optimal value to Problem* (2)*.*

Proof. The proof is relegated to Appendix A.

We next note that the robust satisficing model naturally results in an out-of-sample performance guarantee via the concentration of empirical measures.

Proposition 1. *Under the feasibility condition of Theorem 1, for any $\tau \ge Z_0$, the optimal solution of Problem* (9) *satisfies the following:*

$$\mathbb{P}^S\left[\mathbb{E}_{\mathbb{P}^\star}\left[g(\boldsymbol{x}(\tilde{\boldsymbol{u}}),\tilde{\boldsymbol{v}})\right] > \tau + \kappa_\tau\Gamma\right] \le \mathbb{P}^S\left[\Delta(\mathbb{P}^\star,\hat{\mathbb{P}}) > \Gamma\right] \qquad \forall \Gamma \ge 0.$$

Proposition 1 provides the statistical justification of the robust satisficing model from the target attainment perspective. Reducing this violation probability is consistent with obtaining the lowest possible κ_τ, which the robust satisficing model minimizes. We also refer interested readers to Sim et al. (2023), regarding the connections between robust optimization and robust satisficing regarding their consistency analysis.

Apart from being more intuitive to interpret, the performance benefits of robust satisficing over its robust optimization counterpart have also been demonstrated in Long et al. (2023), Ramachandra et al. (2021). When applied to regression and classification problems, Sim et al. (2021) demonstrate with well-known data sets that the robust satisficing approach would generally outperform the robust optimization approach when the respective target and size parameters are determined via cross-validation. The process of tuning the target parameter has been also discussed in Sim et al. (2023), where the target parameter is determined by a K-fold cross-validation procedure to improve out-of-sample performance.

The pending question now is whether we can efficiently solve Problem (9), which can be equivalently written as the following:

$$\begin{aligned}
\kappa_\tau = \min\ & \kappa \\
\text{s.t.}\ & \frac{1}{S}\sum_{s\in[S]} t_s \le \tau \\
& t_s \ge \sup_{(\boldsymbol{u},\boldsymbol{v})\in\mathcal{Z}_\ell} \left\{ g(\boldsymbol{x}_\ell(\boldsymbol{u}),\boldsymbol{v}) - \kappa(\|\boldsymbol{v}-\hat{\boldsymbol{v}}_s\| + \|\boldsymbol{u}-\hat{\boldsymbol{u}}_s\|)\right\} \ \forall s\in[S], \ell\in[L] \\
& \boldsymbol{x}\in\mathcal{X}, \kappa \ge 0.
\end{aligned} \tag{10}$$

To obtain a tractable exact formulation, we consider the case where g is a saddle function, *i.e.*, $g(\boldsymbol{x},\boldsymbol{v})$ is convex in $\boldsymbol{x}$ for any given $\boldsymbol{v}\in\mathcal{V}$. It is upper-semicontinuous and concave in $\boldsymbol{v}$ for any given $\boldsymbol{x}\in\mathcal{D}$. When we focus on tree-based static mapping, *i.e.*, $\mathcal{A}=\mathcal{T}^{n_x,n_u}$, the above can be reformulated as a modest-sized convex optimization problem if g is a saddle function. However, when we focus on tree-based affine mapping, *i.e.*, $\mathcal{A}=\bar{\mathcal{T}}^{n_x,n_u}$, we would further require that $g(\boldsymbol{x}(\boldsymbol{u}),\boldsymbol{v})$ should be jointly concave in $(\boldsymbol{u},\boldsymbol{v})$ for any affine mapping $\boldsymbol{x}$. In the following, we consider the case where the cost function g is represented as a linear optimization with recourse, where we illustrate suitable approximations under a tree-based affine policy and tractable exact reformulations under special cases.

4. Linear optimization with recourse

Linear optimization is arguably the most important optimization format used in practice. As in the stochastic optimization literature, we consider the cost function g of the form of a linear optimization problem with fixed recourse as follows

$$
\begin{aligned}
g(\boldsymbol{x},\boldsymbol{v}) = \min\ & \boldsymbol{d}^\top \boldsymbol{y} \\
\text{s.t.}\ & \boldsymbol{F}(\boldsymbol{v})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \ge \boldsymbol{f}(\boldsymbol{v}) \\
& \boldsymbol{y}\in\mathbb{R}^{n_y},
\end{aligned}
\tag{11}
$$

where $\boldsymbol{F}\in\mathcal{L}^{n_v,n_f\times n_x}$ and $\boldsymbol{f}\in\mathcal{L}^{n_v,n_f}$. Since the decision $\boldsymbol{y}$ in Problem (11) is made after observing the response decision $\boldsymbol{x}$ and outcome variables $\boldsymbol{v}$, we call $\boldsymbol{y}$ the *recourse* decision. We assume *complete recourse*, *i.e.*, for any $\boldsymbol{w}\in\mathbb{R}^{n_f}$, there exists a $\boldsymbol{y}\in\mathbb{R}^{n_y}$ such that $\boldsymbol{B}\boldsymbol{y}\ge\boldsymbol{w}$.

Similar to the approach of Long et al. (2023), we can express the exact robust satisficing problem as the following adaptive robust optimization problem

$$
\begin{aligned}
\min\ & \kappa \\
\text{s.t.}\ & \frac{1}{S}\sum_{s\in[S]} t_s \le \tau \\
& \boldsymbol{d}^\top \boldsymbol{y}_{\ell s}(\boldsymbol{u},\boldsymbol{v},\sigma,\nu) - \kappa(\sigma+\nu) \le t_s && \forall(\boldsymbol{u},\boldsymbol{v},\sigma,\nu)\in\bar{\mathcal{Z}}_{\ell s}, \ell\in[L], s\in[S] \\
& \boldsymbol{F}(\boldsymbol{v})\boldsymbol{x}(\boldsymbol{u}) + \boldsymbol{B}\boldsymbol{y}_{\ell s}(\boldsymbol{u},\boldsymbol{v},\sigma,\nu) \ge \boldsymbol{f}(\boldsymbol{v}) && \forall(\boldsymbol{u},\boldsymbol{v},\sigma,\nu)\in\bar{\mathcal{Z}}_{\ell s}, \ell\in[L], s\in[S] \\
& \boldsymbol{x}\in\mathcal{X}, \kappa\ge 0 \\
& \boldsymbol{y}_{\ell s}\in\mathcal{R}^{n_u+n_v+2,n_y} && \forall \ell\in[L], s\in[S]
\end{aligned}
\tag{12}
$$

where the lifted uncertainty sets are defined as

$$
\bar{\mathcal{Z}}_{\ell s} \triangleq \{(\boldsymbol{u},\boldsymbol{v},\sigma,\nu)\in\mathcal{U}_\ell\times\mathcal{V}_\ell\times\mathbb{R}\times\mathbb{R} \mid \sigma\ge\|\boldsymbol{u}-\hat{\boldsymbol{u}}_s\|, \nu\ge\|\boldsymbol{v}-\hat{\boldsymbol{v}}_s\|\}
$$

for all $\ell\in[L], s\in[S]$.

When Problem (11) has only "right-hand-side" uncertainty, the matrix $\boldsymbol{F}$ does not depend on $\boldsymbol{v}$, *i.e.*, $\boldsymbol{F}(\boldsymbol{v})=\boldsymbol{F}^0$. This simplification ensures that the function $g(\boldsymbol{x}_\ell(\boldsymbol{u}),\boldsymbol{v})$ is jointly convex with respect to $(\boldsymbol{u},\boldsymbol{v})$, for any affine mapping $\boldsymbol{x}_\ell$. This has been well-studied in the literature. In particular, we can directly apply Long et al. (2023), Bertsimas et al. (2023b) to obtain a safe tractable approximation to the robust satisficing problem.

EXAMPLE 1 (JOINT PRODUCTION AND PROCUREMENT PROBLEM). We consider a joint production and procurement problem with n_x products and n_r resources. The actionable policy set that characterizes the production and procurement constraints is given by

$$\mathcal{X}=\left\{(\boldsymbol{x},\boldsymbol{r})\in\bar{\mathcal{T}}^{n_u,n_x+n_r}\ \mid \boldsymbol{A}\boldsymbol{x}(\boldsymbol{u})\le\boldsymbol{r}(\boldsymbol{u}),\boldsymbol{x}(\boldsymbol{u})\ge\boldsymbol{0},\boldsymbol{r}(\boldsymbol{u})\ge\boldsymbol{0}\ \forall\boldsymbol{u}\in\mathcal{U}\right\},$$

where $\boldsymbol{x}$ and $\boldsymbol{r}$ represent the respective production and procurement decisions. The historical samples of the side information and demands are denoted by $(\hat{\boldsymbol{u}}_s,\hat{\boldsymbol{v}}_s)$, $s\in[S]$. The total cost associated with procuring $\boldsymbol{r}$, and subtracting the sales revenue for demand $\boldsymbol{v}$, assuming zero salvage values for unmet demands, is given by

$$\begin{aligned}g(\boldsymbol{x},\boldsymbol{r},\boldsymbol{v})&=\boldsymbol{c}^\top\boldsymbol{r}-\sum_{i\in[n_x]}p_i\min\{x_i,v_i\}\\&=\sum_{i\in[n_x]}\max\left\{\frac{1}{n_x}\boldsymbol{c}^\top\boldsymbol{r}-p_ix_i,\frac{1}{n_x}\boldsymbol{c}^\top\boldsymbol{r}-p_iv_i\right\},\end{aligned}$$

or equivalently

$$\begin{aligned}g(\boldsymbol{x},\boldsymbol{r},\boldsymbol{v})=\min\ &\boldsymbol{1}^\top\boldsymbol{y}\\\text{s.t. }&\boldsymbol{P}\boldsymbol{x}-\boldsymbol{C}\boldsymbol{r}+\boldsymbol{y}\ge\boldsymbol{0}\\&-\boldsymbol{C}\boldsymbol{r}+\boldsymbol{y}\ge-\boldsymbol{P}\boldsymbol{v},\end{aligned}$$

where $\boldsymbol{P}=\mathrm{diag}(\boldsymbol{p})$ and $\boldsymbol{C}=\frac{1}{n_x}\boldsymbol{1}\boldsymbol{c}^\top$. This is a complete recourse problem with only right-hand-side uncertainty. Hence, we have the flexibility to use the tree-based affine response policy and propose the following safe tractable approximation for the robust satisficing problem,

$$\begin{aligned}\min\ &\kappa\\\text{s.t. }&\frac{1}{S}\sum_{s\in[S]}t_s\le\tau\\&\boldsymbol{1}^\top\boldsymbol{y}_{\ell s}(\boldsymbol{u},\boldsymbol{v},\sigma,\nu)-\kappa\sigma-\kappa\nu\le t_s&&\forall(\boldsymbol{u},\boldsymbol{v},\sigma,\nu)\in\bar{\mathcal{Z}}_{\ell s},\ell\in[L],s\in[S]\\&\boldsymbol{P}\boldsymbol{x}_\ell(\boldsymbol{u})+\boldsymbol{y}_{\ell s}(\boldsymbol{u},\boldsymbol{v},\sigma,\nu)\ge\boldsymbol{C}\boldsymbol{r}_\ell(\boldsymbol{u})&&\forall(\boldsymbol{u},\boldsymbol{v},\sigma,\nu)\in\bar{\mathcal{Z}}_{\ell s},\ell\in[L],s\in[S]\\&\boldsymbol{y}_{\ell s}(\boldsymbol{u},\boldsymbol{v},\sigma,\nu)\ge\boldsymbol{C}\boldsymbol{r}_\ell(\boldsymbol{u})-\boldsymbol{P}\boldsymbol{v}&&\forall(\boldsymbol{u},\boldsymbol{v},\sigma,\nu)\in\bar{\mathcal{Z}}_{\ell s},\ell\in[L],s\in[S]\\&\boldsymbol{A}\boldsymbol{x}_\ell(\boldsymbol{u})\le\boldsymbol{r}(\boldsymbol{u})&&\forall\boldsymbol{u}\in\mathcal{U}_\ell,\ell\in[L]\\&\boldsymbol{x}_\ell(\boldsymbol{u})\ge\boldsymbol{0},\boldsymbol{r}_\ell(\boldsymbol{u})\ge\boldsymbol{0}&&\forall\boldsymbol{u}\in\mathcal{U}_\ell,\ell\in[L]\\&\boldsymbol{y}_{\ell s}\in\mathcal{L}^{n_u+n_v+2,n_y}&&\forall\ell\in[L],s\in[S]\\&\boldsymbol{x}_\ell\in\mathcal{L}^{n_u,n_x},\boldsymbol{r}_\ell\in\mathcal{L}^{n_u,n_r}&&\forall\ell\in[L]\\&\kappa\ge0,\end{aligned}\tag{13}$$

which can be easily implemented in RSOME developed by Chen et al. (2020). Since this problem has complete recourse, following the analysis of Long et al. (2023), Problem (13) is feasible for any chosen target τ greater than the optimum objective value of the corresponding empirical optimization problem. Moreover, the approximation is exact if $n_x = 1$.

Similarly, under the tree-based static policy, $\mathcal{A} = \mathcal{T}^{n_u,n_x}$, we can also use the same approach to provide a safe tractable approximation for the robust satisficing problem. Hence, we consider the general model of Problem (11) with $\boldsymbol{F} \in \mathcal{L}^{n_v, n_f \times n_x}$ and under the tree-based affine response policy, $\mathcal{A} = \bar{\mathcal{T}}^{n_u,n_x}$. To obtain a tractable safe approximation, we first consider the following lifted affine recourse adaptation (Bertsimas et al. 2019b, Chen et al. 2020) as follows,

$$
\begin{aligned}
\min\ & \kappa \\
\text{s.t.}\ & \frac{1}{S}\sum_{s\in[S]} t_s \le \tau \\
& \boldsymbol{d}^\top \boldsymbol{y}_{\ell s}(\boldsymbol{u},\boldsymbol{v},\sigma,\nu) - \kappa(\sigma+\nu) \le t_s && \forall(\boldsymbol{u},\boldsymbol{v},\sigma,\nu) \in \bar{\mathcal{Z}}_{\ell s}, \ell\in[L], s\in[S] \\
& \boldsymbol{F}(\boldsymbol{v})\boldsymbol{x}_\ell(\boldsymbol{u}) + \boldsymbol{B}\boldsymbol{y}_{\ell s}(\boldsymbol{u},\boldsymbol{v},\sigma,\nu) \ge \boldsymbol{f}(\boldsymbol{v}) && \forall(\boldsymbol{u},\boldsymbol{v},\sigma,\nu) \in \bar{\mathcal{Z}}_{\ell s}, \ell\in[L], s\in[S] \\
& \boldsymbol{x}\in\mathcal{X}, \kappa \ge 0 \\
& \boldsymbol{y}_{\ell s} \in \mathcal{L}^{n_u+n_v+2, n_y} && \forall \ell\in[L], s\in[S].
\end{aligned}
\tag{14}
$$

However, Problem (14) remains intractable due to the bilinear uncertainty in $\boldsymbol{v}$ and $\boldsymbol{u}$ in some of the robust constraints. Although there are tractable safe approximations for robust constraints with bilinear uncertainty under the assumptions of at least one of the uncertainty sets being polyhedral (de Ruiter et al. 2022, Zhen et al. 2022a,b), they do not apply to the robust constraints in Problem (14) involving non-polyhedral uncertainty sets for both uncertain parameters because the norms used in $\bar{\mathcal{Z}}_{\ell s}$ can be general norms. To our knowledge, existing literature on multi-stage robust linear optimization such as Bertsimas et al. (2023b) does not address such robust constraints either. To tackle this challenge, we extend the existing dualization technique with affine recourse approximation (Zhen et al. 2022b) to derive a tractable safe approximation for Problem (14) in Theorem 2 by leveraging the structure of $\bar{\mathcal{Z}}_{\ell s}$ and the positive homogeneity of the norms.

THEOREM 2. *The robust constraint*

$$
\boldsymbol{q}_1^\top \boldsymbol{u} + \boldsymbol{q}_2^\top \boldsymbol{v} + \boldsymbol{v}^\top \boldsymbol{Q}_3 \boldsymbol{u} \le q_4\sigma + q_5\nu + q_6, \quad \forall(\boldsymbol{u},\boldsymbol{v},\sigma,\nu)\in\bar{\mathcal{Z}}_{\ell s} \tag{15}
$$

has a tractable safe approximation $(\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{Q}_3, q_4, q_5, q_6) \in \mathcal{Q}_{\ell s}$ *where*

$$\mathcal{Q}_{\ell s} \triangleq \left\{ \begin{array}{c|l} & (\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{Q}_3, q_4, q_5, q_6) \in \mathbb{R}^{n_u} \times \mathbb{R}^{n_v} \times \mathbb{R}^{n_v \times n_u} \times \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R} \\ & \exists \boldsymbol{\theta}, \overline{\boldsymbol{\lambda}}, \underline{\boldsymbol{\lambda}} \in \mathbb{R}^{n_u}, \overline{\boldsymbol{p}}, \underline{\boldsymbol{p}} \in \mathbb{R}^{n_v}, \overline{\boldsymbol{P}}, \underline{\boldsymbol{P}}, \overline{\boldsymbol{\Gamma}}, \underline{\boldsymbol{\Gamma}}, \boldsymbol{\Theta} \in \mathbb{R}^{n_v \times n_u} : \\ & \boldsymbol{q}_1^\top \hat{\boldsymbol{u}}_s + \boldsymbol{q}_2^\top \hat{\boldsymbol{v}}_s + \hat{\boldsymbol{v}}_s^\top \boldsymbol{Q}_3 \hat{\boldsymbol{u}}_s + (\overline{\boldsymbol{p}} + \overline{\boldsymbol{P}} \hat{\boldsymbol{u}}_s)^\top (\overline{\boldsymbol{v}}_\ell - \hat{\boldsymbol{v}}_s) \\ & \qquad + (\underline{\boldsymbol{p}} + \underline{\boldsymbol{P}} \hat{\boldsymbol{u}}_s)^\top (\hat{\boldsymbol{v}}_s - \underline{\boldsymbol{v}}_\ell) + \overline{\boldsymbol{\lambda}}^\top (\overline{\boldsymbol{u}}_\ell - \hat{\boldsymbol{u}}_s) + \underline{\boldsymbol{\lambda}}^\top (\hat{\boldsymbol{u}}_s - \underline{\boldsymbol{u}}_\ell) \le q_6 \\ & \|\boldsymbol{q}_1 + \overline{\boldsymbol{P}}^\top (\overline{\boldsymbol{v}}_\ell - \hat{\boldsymbol{v}}_s) + \underline{\boldsymbol{P}}^\top (\hat{\boldsymbol{v}}_s - \underline{\boldsymbol{v}}_\ell) + \boldsymbol{Q}_3^\top \hat{\boldsymbol{v}}_s - \overline{\boldsymbol{\lambda}} + \underline{\boldsymbol{\lambda}}\|_* \le q_4 \\ & \|\boldsymbol{q}_2 - \overline{\boldsymbol{p}} + \underline{\boldsymbol{p}} + \boldsymbol{\Theta} \overline{\boldsymbol{u}}_\ell - (\boldsymbol{\Theta} - \boldsymbol{Q}_3 + \overline{\boldsymbol{P}} - \underline{\boldsymbol{P}}) \underline{\boldsymbol{u}}_\ell\|_* + (\overline{\boldsymbol{u}}_\ell - \underline{\boldsymbol{u}}_\ell)^\top \boldsymbol{\theta} \le q_5 \\ (\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{Q}_3, & \sum_{i \in [n_u]} \|\boldsymbol{\Theta} \boldsymbol{e}_i\|_* \boldsymbol{e}_i \le \boldsymbol{\theta} \\ q_4, q_5, q_6) & \sum_{i \in [n_u]} \|(\boldsymbol{\Theta} - \boldsymbol{Q}_3 + \overline{\boldsymbol{P}} - \underline{\boldsymbol{P}}) \boldsymbol{e}_i\|_* \boldsymbol{e}_i \le \boldsymbol{\theta} \\ & \overline{\boldsymbol{p}} - \overline{\boldsymbol{\Gamma}} \overline{\boldsymbol{u}}_\ell + (\overline{\boldsymbol{P}} + \overline{\boldsymbol{\Gamma}}) \underline{\boldsymbol{u}}_\ell \ge \boldsymbol{0} \\ & \overline{\boldsymbol{P}} + \overline{\boldsymbol{\Gamma}} \ge \boldsymbol{0} \\ & \underline{\boldsymbol{p}} - \underline{\boldsymbol{\Gamma}} \overline{\boldsymbol{u}}_\ell + (\underline{\boldsymbol{P}} + \underline{\boldsymbol{\Gamma}}) \underline{\boldsymbol{u}}_\ell \ge \boldsymbol{0} \\ & \underline{\boldsymbol{P}} + \underline{\boldsymbol{\Gamma}} \ge \boldsymbol{0} \\ & \overline{\boldsymbol{\Gamma}} \ge \boldsymbol{0}, \underline{\boldsymbol{\Gamma}} \ge \boldsymbol{0}, \overline{\boldsymbol{\lambda}} \ge \boldsymbol{0}, \underline{\boldsymbol{\lambda}} \ge \boldsymbol{0} \end{array} \right\}. \quad (16)$$

Moreover, the following properties of $\mathcal{Q}_{\ell s}$ *hold:*

(a) If $(\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{Q}_3, q_4, q_5, q_6) \in \mathcal{Q}_{\ell s}$ *then it is also feasible in Constraint* (15)*. The converse is true if* $\boldsymbol{Q}_3 = \boldsymbol{0}$.

(b) If $\boldsymbol{q}_1^\top \hat{\boldsymbol{u}}_s + \boldsymbol{q}_2^\top \hat{\boldsymbol{v}}_s + \hat{\boldsymbol{v}}_s^\top \boldsymbol{Q}_3 \hat{\boldsymbol{u}}_s \le q_6$*, there exists* $q_4, q_5 \in \mathbb{R}_+$ *such that* $(\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{Q}_3, q_4, q_5, q_6) \in \mathcal{Q}_{\ell s}$.

Proof. The proof is relegated to Appendix A.

Based on Theorem 2, we propose the following tractable safe approximation of Problem (14).

$$
\begin{aligned}
\min\ & \kappa \\
\text{s.t.}\ & \frac{1}{S}\sum_{s\in[S]} t_s \le \tau \\
& \boldsymbol{d}^\top \boldsymbol{y}_{\ell s}(\boldsymbol{u},\boldsymbol{v},\sigma,\nu) - \kappa(\sigma+\nu) \le t_s && \forall (\boldsymbol{u},\boldsymbol{v},\sigma,\nu)\in \bar{\mathcal{Z}}_{\ell s}, \ell\in[L], s\in[S] \\
& \left.\begin{aligned}
& (\boldsymbol{q}^{k1}_{\ell s}, \boldsymbol{q}^{k2}_{\ell s}, \boldsymbol{Q}^{k3}_{\ell s}, q^{k4}_{\ell s}, q^{k5}_{\ell s}, q^{k6}_{\ell s}) \in \mathcal{Q}_{\ell s} \\
& \boldsymbol{q}^{k1}_{\ell s} = -\sum_{i\in[n_u]} \boldsymbol{e}_i \boldsymbol{e}_k^\top (\boldsymbol{F}^0 \boldsymbol{x}^i_\ell + \boldsymbol{B}\boldsymbol{y}^i_{\ell s}) \\
& \boldsymbol{q}^{k2}_{\ell s} = \sum_{j\in[n_v]} \boldsymbol{e}_j \boldsymbol{e}_k^\top (\boldsymbol{f}^j - \boldsymbol{F}^j \boldsymbol{x}^0_\ell - \boldsymbol{B}\boldsymbol{y}^{n_u+j}_{\ell s}) \\
& \boldsymbol{Q}^{k3}_{\ell s} = -\sum_{i\in[n_u]}\sum_{j\in[n_v]} \boldsymbol{e}_j \boldsymbol{e}_k^\top \left(\boldsymbol{F}^j \boldsymbol{x}^i_\ell\right) \boldsymbol{e}_i^\top \\
& q^{k4}_{\ell s} = \boldsymbol{e}_k^\top \boldsymbol{B}\boldsymbol{y}^{n_u+n_v+1}_{\ell s} \\
& q^{k5}_{\ell s} = \boldsymbol{e}_k^\top \boldsymbol{B}\boldsymbol{y}^{n_u+n_v+2}_{\ell s} \\
& q^{k6}_{\ell s} = \boldsymbol{e}_k^\top (\boldsymbol{F}^0 \boldsymbol{x}^0_\ell + \boldsymbol{B}\boldsymbol{y}^0_{\ell s} - \boldsymbol{f}^0)
\end{aligned}\right\} && \forall \ell\in[L], s\in[S], k\in[n_f] \\
& \boldsymbol{x}\in\mathcal{X}, \kappa\ge 0 \\
& \boldsymbol{y}_{\ell s}\in \mathcal{L}^{n_u+n_v+2, n_y} && \forall \ell\in[L], s\in[S].
\end{aligned}
\tag{17}
$$

Observe that Property (a) of $\mathcal{Q}_{\ell s}$ in Theorem 2 implies that the approximation (16) is exact without bilinear uncertainty. Hence, Problem (17) is a generalization of the special cases without the bilinear uncertainty in the robust constraints. It is important to note that if the norm is a polyhedral norm such as the ℓ_1 and ℓ_∞ norm, then Problem (17) would retain the same computationally attractive format as a linear optimization problem. Property (b) in Theorem 2 has ramifications on the feasibility of the approximate robust satisficing problem when a reasonable target $\tau > Z_0$ is chosen, as we present in the following result.

Theorem 3. *Suppose Assumption 1 holds, Problem* (11) *has complete recourse, and* $\hat{\boldsymbol{u}}_s \in \operatorname{int}(\mathcal{U}_\ell)$ *for each* $\ell\in[L]$, $s\in\mathcal{S}_\ell$. *Then for any* $\tau > Z_0$, *there exist reduced affine mappings* $\boldsymbol{y}_{\ell s}\in\mathcal{L}^{n_u+n_v+2,n_y}$, $s\in[S], \ell\in[L]$ *such that*

$$
\boldsymbol{y}_{\ell s}(\boldsymbol{u},\boldsymbol{v},\sigma,\nu) = \boldsymbol{y}_{\ell s}(\boldsymbol{0},\boldsymbol{0},\sigma,\nu) \qquad \forall (\boldsymbol{u},\boldsymbol{v},\sigma,\nu)\in\mathbb{R}^{n_u+n_v+2},
$$

that are feasible in Problem (17). *Moreover, When* $n_y = 1$, *and* $d\neq 0$, *then the reduced affine mappings of the form*

$$
y_{\ell s}(\boldsymbol{u},\boldsymbol{v},\sigma,\nu) = (t_s + \kappa\sigma + \kappa\nu)/d \qquad \forall s\in[S], \ell\in[L]
$$

is optimal in Problem (17).

Proof. The proof is relegated to Appendix A.

Remark 1. Theorem 3 extends Theorem 7 in Long et al. (2023) to two-stage linear optimization with side information. The feasibility result implies that the reduced affine mappings do not limit the choice of targets $\tau > Z_0$ for the decision maker, which is important for adjusting robustness.

Remark 2. Note the existence of the reduced affine mappings that ensure feasibility is a new insight that has not been observed in Bertsimas et al. (2019b), Long et al. (2023) because their corresponding recourse function has the following two-stage representation:

$$g(\boldsymbol{x},\boldsymbol{v}) = \boldsymbol{c}(\boldsymbol{v})^\top \boldsymbol{x} + \min_{\boldsymbol{y}\in\mathbb{R}^{n_y}} \left\{ \boldsymbol{d}^\top \boldsymbol{y} \mid \boldsymbol{F}(\boldsymbol{v})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{f}(\boldsymbol{v}) \right\}.$$

If we can subsume the first stage cost to the recourse optimization problem without increasing the number of recourse variables, we would also improve the approximation quality via affine recourse adaptation. Specifically, if $\boldsymbol{d} \neq \boldsymbol{0}$ then we can formulate

$$g(\boldsymbol{x},\boldsymbol{v}) = \min_{\boldsymbol{y}\in\mathbb{R}^{n_y}} \left\{ \boldsymbol{d}^\top \boldsymbol{y} \mid \bar{\boldsymbol{F}}(\boldsymbol{v})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{f}(\boldsymbol{v}) \right\},$$

where $\bar{\boldsymbol{F}}(\boldsymbol{v}) = \boldsymbol{F}(\boldsymbol{v}) - \frac{1}{d_r}\boldsymbol{B}\boldsymbol{e}_r\boldsymbol{c}(\boldsymbol{v})^\top$, for some $r \in [n_y]$ such that $d_r \neq 0$ (see also Example 1).

Example 2 (Portfolio optimization). We consider a data-driven portfolio optimization problem with n_x stocks. The historical samples of the side information and random returns are denoted by $(\hat{\boldsymbol{u}}_s, \hat{\boldsymbol{v}}_s)$, $s \in [S]$. We minimize the conditional value-at-risk (CVaR) of the portfolio,

$$\mathbb{C}^{\epsilon}_{\mathbb{P}}\left[\tilde{\boldsymbol{v}}^\top \boldsymbol{x}(\tilde{\boldsymbol{u}})\right]$$

where CVaR is defined as

$$\mathbb{C}^{\epsilon}_{\mathbb{P}}\left[\tilde{z}\right] \triangleq \inf_{\eta\in\mathbb{R}} \left\{ \eta + \frac{1}{\epsilon}\mathbb{E}_{\mathbb{P}}\left[(-\tilde{z}-\eta)^+\right] \right\}$$

with $\epsilon \in (0,1)$. Equivalently, we can write the problem as

$$g((\boldsymbol{x},\eta),\boldsymbol{v}) = \min_{y\in\mathbb{R}} \left\{ y \mid y \geq \eta - (\eta + \boldsymbol{v}^\top \boldsymbol{x})/\epsilon,\ y \geq \eta \right\}.$$

We observe that the function $g((\boldsymbol{x}(\boldsymbol{u}),\eta),\boldsymbol{v})$ is no longer jointly convex in $(\boldsymbol{u},\boldsymbol{v})$ for $\mathcal{A} = \bar{\mathcal{T}}^{n_u,n_x}$,

$$\mathcal{X} = \left\{ (\boldsymbol{x},\eta) \in \bar{\mathcal{T}}^{n_u,n_x} \times \mathbb{R} \;\middle|\; \boldsymbol{1}^\top \boldsymbol{x}(\boldsymbol{u}) = 1, \boldsymbol{x}(\boldsymbol{u}) \geq \boldsymbol{0},\ \forall \boldsymbol{u} \in \mathcal{U} \right\}.$$

Observe that this is a complete recourse problem with $n_y = 1$; hence, from Theorem 3, we can use the reduced tree-based affine mapping to formulate the robust satisficing problem as follows,

$$
\begin{aligned}
\min\ & \kappa \\
\text{s.t. } & \frac{1}{S}\sum_{s\in[S]} t_s \le \tau \\
& t_s + \kappa(\sigma+\nu) \ge \eta - (\eta + \boldsymbol{v}^\top \boldsymbol{x}(\boldsymbol{u}))/\epsilon && \forall(\boldsymbol{u},\boldsymbol{v},\sigma,\nu)\in\bar{\mathcal{Z}}_{\ell s}, \ell\in[L], s\in[S] \\
& t_s + \kappa(\sigma+\nu) \ge \eta && \forall(\boldsymbol{u},\boldsymbol{v},\sigma,\nu)\in\bar{\mathcal{Z}}_{\ell s}, \ell\in[L], s\in[S] \\
& \mathbf{1}^\top \boldsymbol{x}_\ell(\boldsymbol{u}) = 1 && \forall \boldsymbol{u}\in\mathcal{U}_\ell, \ell\in[L] \\
& \boldsymbol{x}_\ell(\boldsymbol{u}) \ge \mathbf{0} && \forall \boldsymbol{u}\in\mathcal{U}_\ell, \ell\in[L] \\
& \boldsymbol{x}_\ell \in \mathcal{L}^{n_u,n_x} && \forall \ell\in[L] \\
& \boldsymbol{t}\in\mathbb{R}^S, \eta\in\mathbb{R}, \kappa\ge 0.
\end{aligned}
$$

Observe that, since $\kappa \ge 0$,

$$
\begin{aligned}
& t_s + \kappa(\sigma+\nu) \ge \eta && \forall(\boldsymbol{u},\boldsymbol{v},\sigma,\nu)\in\bar{\mathcal{Z}}_{\ell s}, \ell\in[L], s\in[S] \\
\Longrightarrow\ & t_s + \kappa(\sigma+\nu) \ge \eta && \forall(\boldsymbol{u},\boldsymbol{v},\sigma,\nu)\in\bar{\mathcal{Z}}_{\ell s}, \ell\in[L], s\in\mathcal{S}_\ell \\
\Longrightarrow\ & t_s \ge \eta && \forall s\in[S] \\
\Longrightarrow\ & t_s + \kappa(\sigma+\nu) \ge \eta && \forall(\boldsymbol{u},\boldsymbol{v},\sigma,\nu)\in\bar{\mathcal{Z}}_{\ell s}, \ell\in[L], s\in[S]
\end{aligned}
$$

We can apply Theorem 2 to obtain a safe tractable approximation that guarantees a solution for any reasonably chosen target, $\tau > Z_0$, as follows,

$$
\begin{aligned}
\min\ & \kappa \\
\text{s.t. } & \frac{1}{S}\sum_{s\in[S]} t_s \le \tau \\
& \left(\mathbf{0}, -\boldsymbol{x}_\ell^0/\epsilon, -\boldsymbol{X}_\ell/\epsilon, \kappa, \kappa, t_s - \eta + \eta/\epsilon\right) \in \mathcal{Q}_{\ell s} && \forall \ell\in[L], s\in[S] \\
& t_s \ge \eta && \forall s\in[S] \\
& \mathbf{1}^\top(\boldsymbol{x}_\ell^0 + \boldsymbol{X}_\ell \boldsymbol{u}) = 1 && \forall \boldsymbol{u}\in\mathcal{U}_\ell, \ell\in[L] \\
& \boldsymbol{x}_\ell^0 + \boldsymbol{X}_\ell\boldsymbol{u} \ge \mathbf{0} && \forall \boldsymbol{u}\in\mathcal{U}_\ell, \ell\in[L] \\
& \boldsymbol{x}_\ell^0 \in \mathbb{R}^{n_x}, \boldsymbol{X}_\ell \in \mathbb{R}^{n_x\times n_u} && \forall \ell\in[L] \\
& \boldsymbol{t}\in\mathbb{R}^S, \eta\in\mathbb{R}, \kappa\ge 0.
\end{aligned} \tag{18}
$$

where we write $\boldsymbol{x}_\ell(\boldsymbol{u}) = \boldsymbol{x}_\ell^0 + \boldsymbol{X}_\ell\boldsymbol{u}$ for each $\ell\in[L]$.

We note that there is a distinction between the recourse adaptation mapping and the actionable response policy. Specifically, the actionable response policy is optimized over the actionable policy set to determine the actual response decisions directly and intuitively from the side information. In contrast, the recourse adaptation mapping aims to provide a tractable approximation to the adaptive robust optimization models. After the outcome variables $\boldsymbol{v}$ have been realized, instead of using the recourse adaptation, we should solve for the optimal recourse $\boldsymbol{y}$ directly from Problem (11) (see discussions in Bertsimas et al. 2019b). Since the purpose of the recourse adaptation mapping is purely computational, it would not be necessary to consider its interpretability.

Biaffine recourse adaptation

To further improve the solutions to the adaptive robust optimization problems, various extensions of the affine adaptations have been proposed in the literature (see, *e.g.*, Ben-Tal et al. 2004, Chen et al. 2008, Goh and Sim 2010, Georghiou et al. 2015, Kuhn et al. 2011, Bertsimas et al. 2019b, Zhen et al. 2018), which can also be applied here. Inspired by the biaffine perturbation of our robust models, we propose a new *biaffine recourse adaptation* tailored to improve the solution to Problem (12). We first introduced the following sub-class of bilinear mappings:

$$\mathcal{B}^{n_u,n_v,n_y} \triangleq \left\{ \boldsymbol{y} \in \mathcal{R}^{n_u+n_v,n_y} \;\middle|\; \begin{array}{l} \exists \boldsymbol{y}^{ij} \in \mathbb{R}^{n_y} \;\; \forall i \in [n_u], j \in [n_v]: \\ \boldsymbol{y}(\boldsymbol{u},\boldsymbol{v}) = \displaystyle\sum_{i\in[n_u]} \sum_{j\in[n_v]} \boldsymbol{y}^{ij} u_i v_j \quad \forall \boldsymbol{u} \in \mathbb{R}^{n_u}, \boldsymbol{v} \in \mathbb{R}^{n_v} \end{array} \right\}.$$

Accordingly, we consider lifted biaffine recourse adaptation as follows,

$$\begin{array}{lll} \min & \kappa & \\ \text{s.t.} & \dfrac{1}{S}\displaystyle\sum_{s\in[S]} t_s \le \tau & \\ & \boldsymbol{d}^\top \left(\boldsymbol{y}_{\ell s}(\boldsymbol{u},\boldsymbol{v},\sigma,\nu) + \bar{\boldsymbol{y}}_{\ell s}(\boldsymbol{u},\boldsymbol{v})\right) - \kappa(\sigma+\nu) \le t_s & \forall (\boldsymbol{u},\boldsymbol{v},\sigma,\nu) \in \bar{\mathcal{Z}}_{\ell s}, \ell \in [L], s \in [S] \\ & \boldsymbol{F}(\boldsymbol{v})\boldsymbol{x}_\ell(\boldsymbol{u}) + \boldsymbol{B}\left(\boldsymbol{y}_{\ell s}(\boldsymbol{u},\boldsymbol{v},\sigma,\nu) + \bar{\boldsymbol{y}}_{\ell s}(\boldsymbol{u},\boldsymbol{v})\right) \ge \boldsymbol{f}(\boldsymbol{v}) & \forall (\boldsymbol{u},\boldsymbol{v},\sigma,\nu) \in \bar{\mathcal{Z}}_{\ell s}, \ell \in [L], s \in [S] \\ & \boldsymbol{x} \in \mathcal{X}, \kappa \ge 0 & \\ & \boldsymbol{y}_{\ell s} \in \mathcal{L}^{n_u+n_v+2,n_y}, \bar{\boldsymbol{y}}_{\ell s} \in \mathcal{B}^{n_u,n_v,n_y} & \forall \ell \in [L], s \in [S]. \end{array} \tag{19}$$

Although Theorem 3 shows that there exists a reduced affine recourse adaptation that is optimal in complete recourse and $n_y = 1$, it is still essential to consider the more complex model when $n_y \ge 2$. For instance, suppose the cost function is given by

$$g(\boldsymbol{x},\boldsymbol{v}) = \min_{\boldsymbol{y}\in\mathbb{R}^2} \left\{ y_1 + y_2 \mid y_1 \ge \boldsymbol{f}^\top(\boldsymbol{v})\boldsymbol{x},\; y_2 \ge -\boldsymbol{f}^\top(\boldsymbol{v})\boldsymbol{x} \right\}$$

which is a complete recourse problem with $n_y = 2$. Observe that for any $\boldsymbol{x} \in \bar{\mathcal{T}}^{n_u,n_x}$, we have $g(\boldsymbol{x}(\boldsymbol{u}),\boldsymbol{v}) = 0$ for all $(\boldsymbol{u},\boldsymbol{v}) \in \mathcal{Z}$. Indeed, the optimal recourse adaptation can be replicated in Problem (19), when

$$\left.\begin{array}{l} y_{\ell 1}(\boldsymbol{u},\boldsymbol{v},\sigma,\nu) + \bar{y}_{\ell 1}(\boldsymbol{u},\boldsymbol{v},\sigma) = \boldsymbol{f}^\top(\boldsymbol{v})\boldsymbol{x}_\ell(\boldsymbol{u}) \\ y_{\ell 2}(\boldsymbol{u},\boldsymbol{v},\sigma,\nu) + \bar{y}_{\ell 2}(\boldsymbol{u},\boldsymbol{v},\sigma) = -\boldsymbol{f}^\top(\boldsymbol{v})\boldsymbol{x}_\ell(\boldsymbol{u}) \end{array}\right\} \forall (\boldsymbol{u},\boldsymbol{v},\sigma,\nu) \in \bar{\mathcal{Z}}_{\ell s}, \ell \in [L], s \in [S]$$

but not in the simpler model of Problem (14).

Based on Theorem 2, the tractable safe approximation of Problem (19) is given by

$$
\begin{aligned}
&\min\ \kappa \\
&\text{s.t. } \frac{1}{S}\sum_{s\in[S]} t_s \le \tau \\
&\quad (\boldsymbol{q}_{\ell s}^{k1}, \boldsymbol{q}_{\ell s}^{k2}, \boldsymbol{Q}_{\ell s}^{k3}, q_{\ell s}^{k4}, q_{\ell s}^{k5}, q_{\ell s}^{k6}) \in \mathcal{Q}_{\ell s} && \forall \ell\in[L], s\in[S], k\in\{0\}\cup[n_f] \\
&\left.\begin{aligned}
&\boldsymbol{q}_{\ell s}^{01} = \sum_{i\in[n_u]} \boldsymbol{d}^\top \boldsymbol{y}_{s,\ell}^{i} \boldsymbol{e}_i \\
&\boldsymbol{q}_{\ell s}^{02} = \sum_{j\in[n_v]} \boldsymbol{d}^\top \boldsymbol{y}_{s,\ell}^{n_u+j} \boldsymbol{e}_j \\
&\boldsymbol{Q}_{\ell s}^{03} = \sum_{i\in[n_u]}\sum_{j\in[n_v]} \boldsymbol{d}^\top \bar{\boldsymbol{y}}_{s,\ell}^{ij} \boldsymbol{e}_j \boldsymbol{e}_i^\top \\
&q_{\ell s}^{04} = \kappa - \boldsymbol{d}^\top \boldsymbol{y}_{s,\ell}^{n_u+n_v+1} \\
&q_{\ell s}^{05} = \kappa - \boldsymbol{d}^\top \boldsymbol{y}_{s,\ell}^{n_u+n_v+2} \\
&q_{\ell s}^{06} = t_s - \boldsymbol{d}^\top \boldsymbol{y}_{\ell s}^{0}
\end{aligned}\right\} && \forall \ell\in[L], s\in[S] \\
&\left.\begin{aligned}
&\boldsymbol{q}_{\ell s}^{k1} = -\sum_{i\in[n_u]} \boldsymbol{e}_i \boldsymbol{e}_k^\top (\boldsymbol{F}^0 \boldsymbol{x}_\ell^i + \boldsymbol{B}\boldsymbol{y}_{\ell s}^i) \\
&\boldsymbol{q}_{\ell s}^{k2} = \sum_{j\in[n_v]} \boldsymbol{e}_j \boldsymbol{e}_k^\top (\boldsymbol{f}^j - \boldsymbol{F}^j \boldsymbol{x}_\ell^0 - \boldsymbol{B}\boldsymbol{y}_{\ell s}^{n_u+j}) \\
&\boldsymbol{Q}_{\ell s}^{k3} = -\sum_{i\in[n_u]}\sum_{j\in[n_v]} \boldsymbol{e}_j \boldsymbol{e}_k^\top \left(\boldsymbol{F}^j \boldsymbol{x}_\ell^i + \boldsymbol{B}\bar{\boldsymbol{y}}_{\ell s}^{ij}\right) \boldsymbol{e}_i^\top \\
&q_{\ell s}^{k4} = \boldsymbol{e}_k^\top \boldsymbol{B}\boldsymbol{y}_{\ell s}^{n_u+n_v+1} \\
&q_{\ell s}^{k5} = \boldsymbol{e}_k^\top \boldsymbol{B}\boldsymbol{y}_{\ell s}^{n_u+n_v+2} \\
&q_{\ell s}^{k6} = \boldsymbol{e}_k^\top (\boldsymbol{F}^0 \boldsymbol{x}_\ell^0 + \boldsymbol{B}\boldsymbol{y}_{\ell s}^0 - \boldsymbol{f}^0)
\end{aligned}\right\} && \forall \ell\in[L], s\in[S], k\in[n_f] \\
&\quad \boldsymbol{x}\in\mathcal{X}, \kappa\ge 0 \\
&\quad \boldsymbol{y}_{\ell s}\in\mathcal{L}^{n_u+n_v+2,n_y}, \bar{\boldsymbol{y}}_{\ell s}\in\mathcal{B}^{n_u,n_v,n_y} && \forall \ell\in[L], s\in[S].
\end{aligned}
\tag{20}
$$

5. A localized tree-based model

Under the tree-based policy, the size of Problem (10) grows linearly with the product $S\times L$. This can pose computational challenges, particularly when dealing with large data size S. To alleviate this issue, we propose a localized tree-based robust satisficing model associated with the hyper-parameter, $\theta > 0$, as follows:

$$
\begin{aligned}
\kappa_\tau = &\min\ \kappa \\
&\text{s.t. } \sup_{\boldsymbol{w}\in\mathcal{W}} \left\{ \sum_{\ell\in[L]} w_\ell \left(\sup_{\mathbb{P}\in\mathcal{P}_0(\mathcal{Z}_\ell)} \left\{ \mathbb{E}_{\mathbb{P}}\left[g(\boldsymbol{x}(\tilde{\boldsymbol{u}}), \tilde{\boldsymbol{v}})\right] - \kappa\Delta(\mathbb{P}, \hat{\mathbb{P}}_\ell) \right\} \right) - \kappa\theta\|\boldsymbol{w}-\hat{\boldsymbol{w}}\|_1 \right\} \le \tau \\
&\quad \boldsymbol{x}\in\mathcal{X},\ \kappa\ge 0,
\end{aligned}
\tag{21}
$$

where $\hat{w}_\ell = |\mathcal{S}_\ell|/S,\ \ell \in [L], \mathcal{W} = \{\boldsymbol{w} \in \mathbb{R}^L_+ | \mathbf{1}^\top \boldsymbol{w} = 1\}$, and $\hat{\mathbb{P}}_\ell$ denotes the empirical distribution of $(\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}})$, conditional on $\tilde{\boldsymbol{u}} \in \mathcal{U}_\ell$. We have the following equivalent formulation of Problem (21),

$$\begin{aligned}
\kappa_\tau = \min\ & \kappa \\
\text{s.t.}\ & \sup_{\boldsymbol{w} \in \mathcal{W}} \left\{ \sum_{\ell \in [L]} w_\ell r_\ell - \kappa\theta \|\boldsymbol{w} - \hat{\boldsymbol{w}}\|_1 \right\} \le \tau \\
& \frac{1}{|\mathcal{S}_\ell|} \sum_{s \in \mathcal{S}_\ell} t_s \le r_\ell && \forall \ell \in [L] \\
& \sup_{(\boldsymbol{u},\boldsymbol{v}) \in \mathcal{Z}_\ell} \{ g(\boldsymbol{x}(\boldsymbol{u}), \boldsymbol{v}) - \kappa(\|\boldsymbol{u} - \hat{\boldsymbol{u}}_s\| + \|\boldsymbol{v} - \hat{\boldsymbol{v}}_s\|) \} \le t_s && \forall \ell \in [L], s \in \mathcal{S}_\ell \\
& \boldsymbol{x} \in \mathcal{X},\ \kappa \ge 0.
\end{aligned} \tag{22}$$

Note that the number of constraints of Problem (22) is reduced to the magnitude of $S = \sum_{\ell \in [L]} |\mathcal{S}_\ell|$. To establish the feasibility of Problem (22), we focus on Lipschitz continuous cost function g and derive a counterpart of Theorem 1 under the tree-based policy set. We first define

$$Z_{0,\ell} = \min_{\boldsymbol{x}_\ell \in \mathcal{X}} \frac{1}{|\mathcal{S}_\ell|} \sum_{s \in \mathcal{S}_\ell} g(\boldsymbol{x}_\ell(\hat{\boldsymbol{u}}_s), \hat{\boldsymbol{v}}_s),$$

and let $Z_0 = \sum_{\ell \in [L]} \hat{w}_\ell Z_{0,\ell}$ denote the optimal value of the empirical optimization model.

THEOREM 4. *Suppose $g(\boldsymbol{x}(\boldsymbol{u}), \boldsymbol{v})$ is Lipschitz continuous with respect to $\boldsymbol{u}$ and $\boldsymbol{v}$ with a maximum Lipschitz constant of $\bar{L}$ for any mapping $\boldsymbol{x} \in \mathcal{X}$. Then, the robust satisficing model, Problem (22), is feasible for all $\tau \ge Z_0$ for some $\kappa_\tau \le \max\{\bar{L}, \max_{\ell \in [L]}\{|Z_{0,\ell}|\}/\theta\}$.*

Proof. The proof is relegated to Appendix A.

Analogous to Proposition 1, we can derive the following result on target shortfall avoidance and target attainment asymptotic guarantees.

THEOREM 5. *Suppose the feasibility condition of Theorem 4 holds. For any $\tau \ge Z_0$ and $\Gamma \ge 0$, the optimal solution of Problem (21) satisfies*

$$\mathbb{P}^S \left[\mathbb{E}_{\mathbb{P}^\star} \left[g(\boldsymbol{x}(\tilde{\boldsymbol{u}}), \tilde{\boldsymbol{v}}) \right] > \tau + \kappa_\tau \Gamma \right] \le \min_{\gamma_1 + \gamma_2 = \Gamma} \left\{ \sum_{\ell \in [L]} \mathbb{P}^{\mathcal{S}_\ell} \left[\Delta(\mathbb{P}^\star_\ell, \hat{\mathbb{P}}_\ell) > \gamma_1 \right] + 2L \exp\left(\frac{-2S\gamma_2^2}{L^2\theta^2} \right) \right\},$$

where $\mathbb{P}^{\mathcal{S}_\ell}$ is the distribution that governs the distribution of independent samples $(\hat{\boldsymbol{u}}_s, \hat{\boldsymbol{v}}_s)$, $s \in \mathcal{S}_\ell$.

Proof. The proof is relegated to Appendix A.

Theorem 5 demonstrates that the theoretical likelihood limit diminishes exponentially towards zero as the quantity of samples increases. This suggests a robust performance by the model when dealing with large datasets. Furthermore, the most significant degree of robustness in our localized robust satisficing model aligns with optimizing for the smallest conceivable value of κ. Despite these findings, the theorem does not offer direct instructions on selecting the optimal values for the hyper-parameters, θ and τ. As in previous cases, these parameters should be determined through cross-validation, a robust method for estimating the performance of a model on an independent dataset.

The computational benefits of the localized tree-based model cannot be understated, as we demonstrate an important application in combinatorial optimization problems. We first observe that, under the tree-based static response policy,

$$\sup_{\boldsymbol{u}\in\mathcal{U}_\ell,\boldsymbol{v}\in\mathcal{V}_\ell}\left\{g(\boldsymbol{x}_\ell^0,\boldsymbol{v})-\kappa\|\boldsymbol{u}-\hat{\boldsymbol{u}}_s\|-\kappa\|\boldsymbol{v}-\hat{\boldsymbol{v}}_s\|\right\}=\sup_{\boldsymbol{v}\in\mathcal{V}_\ell}\left\{g(\boldsymbol{x}_\ell^0,\boldsymbol{v})-\kappa\|\boldsymbol{v}-\hat{\boldsymbol{v}}_s\|\right\},$$

because $\hat{\boldsymbol{u}}_s\in\mathcal{U}_\ell$ for all $s\in\mathcal{S}_\ell$.

Example 3 (Combinatorial optimization). Consider the following localized empirical combinatorial optimization problem,

$$\begin{aligned} Z_0=\min\ & \sum_{\ell\in[L]}\hat{w}_\ell\left(\frac{1}{|\mathcal{S}_\ell|}\sum_{s\in\mathcal{S}_\ell}\sum_{n\in[n_x]}c_n\hat{v}_{sn}x_{\ell n}\right) \\ \text{s.t. }& \boldsymbol{x}_\ell\in\mathcal{D} \qquad\qquad \forall\ell\in[L], \end{aligned}$$

where the cost function is $g(\boldsymbol{x},\boldsymbol{v})=\sum_{n\in[n_x]}c_nv_nx_n$, for some $\boldsymbol{c}\geq\boldsymbol{0}$ and $\mathcal{D}\subseteq\{0,1\}^{n_x}$. Note that the objective function is separable in $\boldsymbol{x}_\ell$ for $\ell\in[L]$; one could solve L combinatorial problems,

$$\min_{\boldsymbol{x}_\ell\in\mathcal{D}_\ell}\left\{\frac{1}{|\mathcal{S}_\ell|}\sum_{s\in\mathcal{S}_\ell}\sum_{n\in[n_x]}c_n\hat{v}_{sn}x_{\ell n}\right\},$$

to determine the optimal decisions $\boldsymbol{x}_\ell$, $\ell\in[L]$. Hence, if the underlying combinatorial optimization problem is computationally tractable, the localized empirical combinatorial optimization problem is also tractable.

Accordingly, the localized robust satisficing combinatorial model with side information can be written as

$$\begin{aligned} \min\ & \kappa \\ \text{s.t. }& \sup_{\boldsymbol{w}\in\mathcal{W}}\left\{\sum_{\ell\in[L]}w_\ell r_\ell-\kappa\theta\|\boldsymbol{w}-\hat{\boldsymbol{w}}\|\right\}\leq\tau \\ & \frac{1}{|\mathcal{S}_\ell|}\sum_{s\in\mathcal{S}_\ell}\left(\sup_{\boldsymbol{v}\in\mathcal{V}_\ell}\left\{\sum_{n\in[n_x]}c_nv_nx_{\ell n}-\kappa\|\boldsymbol{v}-\hat{\boldsymbol{v}}_s\|\right\}\right)\leq r_\ell \quad \forall\ell\in[L],s\in\mathcal{S}_\ell \\ & \boldsymbol{x}_\ell\in\mathcal{D} \qquad\qquad \forall\ell\in[L] \\ & \kappa\geq 0. \end{aligned} \tag{23}$$

Hence, following the same analysis as Long et al. (2023), suppose the norm is given by ℓ_1-norm and recall that $\mathcal{V}_\ell = [\underline{\boldsymbol{v}}_\ell, \bar{\boldsymbol{v}}_\ell]$ for all $\ell \in [L]$, then Problem (23) admits the following explicit formulation:

$$\begin{aligned} \min\ & \kappa \\ \text{s.t.}\ & \sup_{\boldsymbol{w} \in \mathcal{W}} \left\{ \sum_{\ell \in [L]} w_\ell \left(\boldsymbol{d}_\ell^\top(\kappa) \boldsymbol{x}_\ell \right) - \kappa\theta \|\boldsymbol{w} - \hat{\boldsymbol{w}}\| \right\} \leq \tau \\ & \boldsymbol{x}_\ell \in \mathcal{D} \qquad\qquad \forall \ell \in [L] \\ & \kappa \geq 0, \end{aligned}$$

where

$$d_{\ell n}(\kappa) = \frac{1}{|\mathcal{S}_\ell|} \sum_{s \in \mathcal{S}_\ell} (c_n \hat{v}_{sn} + (c_n - \kappa)^+ (\bar{v}_{\ell n} - \hat{v}_{sn})) \qquad \forall n \in [n_x].$$

Because $\boldsymbol{d}_\ell(\kappa)$ is non-increasing in κ, the above problem can be solved via a bisection search where each subproblem involves solving L individual combinatorial optimization problems with linear cost functions.

6. Numerical study on portfolio optimization

In this section, we present the computational results of the CVaR-based portfolio optimization problem, as introduced in Example 2. Our numerical study focuses on the robust satisficing model with side information under a tree-based affine response policy, i.e., Model (18). We denote the solution to this model for a tree with L leaf nodes as the "RobustTreeAffine" policy. It is worth noting that when the tree has only one leaf node, we refer to the solution as the "RobustAffine" policy. We set the target τ in Model (18) as $Z_0^{Affine} + \alpha$, where $\alpha > 0$ denotes a relative target margin, and Z_0^{Affine} is the optimal value of the following empirical optimization problem,

$$\begin{aligned} Z_0^{Affine} = \min\ & \frac{1}{S} \sum_{\ell \in [L]} \sum_{s \in \mathcal{S}_\ell} t_{\ell s} \\ \text{s.t.}\ & t_{\ell s} \geq \eta - \frac{\eta + \hat{\boldsymbol{v}}_s^\top \boldsymbol{x}_\ell(\hat{\boldsymbol{u}}_s)}{\epsilon} && \forall \ell \in [L], s \in \mathcal{S}_\ell \\ & t_{\ell s} \geq \eta && \forall \ell \in [L], s \in \mathcal{S}_\ell \\ & \boldsymbol{1}^\top \boldsymbol{x}_\ell(\hat{\boldsymbol{u}}_s) = 1 && \forall \ell \in [L], s \in \mathcal{S}_\ell \\ & \boldsymbol{x}_\ell(\hat{\boldsymbol{u}}_s) \geq \boldsymbol{0} && \forall \ell \in [L], s \in \mathcal{S}_\ell \\ & t_{\ell s} \in \mathbb{R} && \forall \ell \in [L], s \in \mathcal{S}_\ell \\ & \boldsymbol{x}_\ell \in \mathcal{L}^{n_u, n_x} && \forall \ell \in [L]. \end{aligned} \tag{24}$$

We compare our proposed model, Problem (18), with two benchmarks. The first one is the "Forest" policy, which is obtained from the StochOptForest algorithm via apx-soln criteria (Kallus and Mao 2023). The second one is the "RobustStatic" policy, which is the solution to the following robust satisficing model without side information,

$$\begin{aligned} \min\ & \kappa \\ \text{s.t.}\ & \frac{1}{S}\sum_{s\in[S]} t_s \le \tau \\ & t_s + \kappa\nu \ge \eta - (\eta + \boldsymbol{v}^\top \boldsymbol{x})/\epsilon \qquad \forall (\boldsymbol{v},\nu)\in \bar{\mathcal{V}}_s \\ & t_s + \kappa\nu \ge \eta \qquad \forall (\boldsymbol{v},\nu)\in \bar{\mathcal{V}}_s \\ & \mathbf{1}^\top \boldsymbol{x} = 1 \\ & \boldsymbol{x}\in\mathbb{R}_+^{n_x}, \boldsymbol{t}\in\mathbb{R}^S, \eta\in\mathbb{R}, \kappa\ge 0. \end{aligned} \tag{25}$$

We set the target τ in Model (25) as $Z_0^{Static}+\alpha$ for some target margin $\alpha>0$, and Z_0^{Static} is the optimal value of the empirical optimization model ignoring side information,

$$\begin{aligned} Z_0^{Static} = \min\ & \frac{1}{S}\sum_{s\in[S]} t_s \\ \text{s.t.}\ & t_s \ge \eta - (\eta + \hat{\boldsymbol{v}}_s^\top \boldsymbol{x})/\epsilon \\ & t_s \ge \eta \\ & \mathbf{1}^\top \boldsymbol{x} = 1 \\ & \boldsymbol{x}\in\mathbb{R}_+^{n_x}, \boldsymbol{t}\in\mathbb{R}^S, \eta\in\mathbb{R}. \end{aligned} \tag{26}$$

The solution to Model (26) is called the "EOStatic" policy.

Simulated data

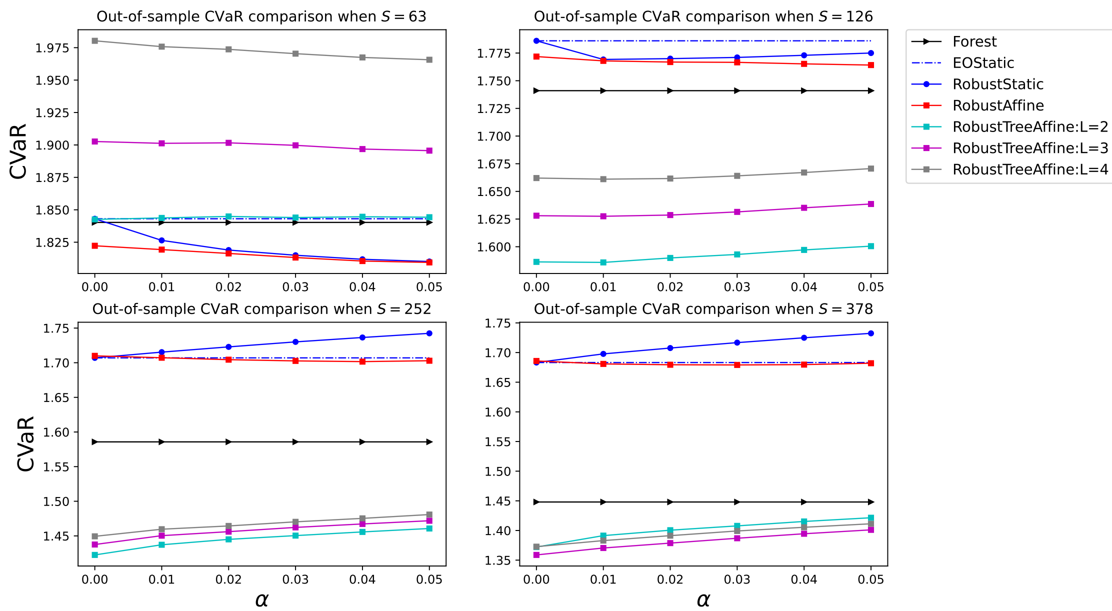
We next present the results of numerical experiments based on simulated data using the same setup as Kallus and Mao (2023). Our simulation includes $n_x=3$ stocks and $n_u=10$ covariates, with a fixed risk level $\epsilon=0.2$. We independently generate each component of the side information $\hat{\boldsymbol{u}}$ from a standard normal distribution and generate the returns in the following way:

$$\begin{aligned} \tilde{v}_1 &= 1+0.2\exp(\tilde{u}_1) - LN(0, 1-0.5\mathbb{1}_{[-3,-1]}(\tilde{u}_2)) \\ \tilde{v}_2 &= 1-0.2\tilde{u}_1 - LN(0, 1-0.5\mathbb{1}_{[-1,1]}(\tilde{u}_2)) \\ \tilde{v}_3 &= 1+0.2|\tilde{u}_1| - LN(0, 1-0.5\mathbb{1}_{[1,3]}(\tilde{u}_2)) \end{aligned}$$

where $LN(\mu,\sigma^2)$ is a log-normal distribution with parameters μ and σ.

To mimic daily returns over the past three months, six months, one year, or one year and a half, we generate a training set $\{\hat{\boldsymbol{u}}_s, \hat{\boldsymbol{v}}_s\}_{s\in[S]}$ consists of $S\in\{63,126,252,378\}$ independent identically distributed (i.i.d.) samples. Then we implement different approaches and obtain the corresponding policies. For the robust satisficing models, we select $\alpha\in\{0,0.01,0.02,0.03,0.04,0.05\}$ and limit the number of leaf nodes to 4. We train the tree in our model based on the algorithm described in

Figure 1 **CVaR comparison under Kallus and Mao (2023)'s setting**



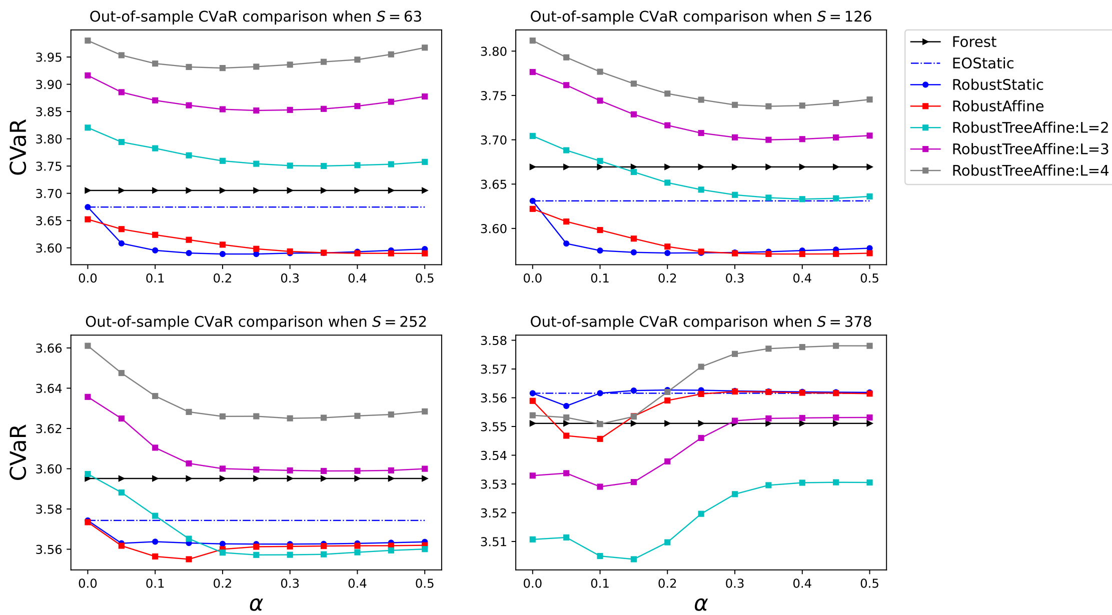
Section 2. We compute the out-of-sample CVaR for each policy using the test set $\{\hat{\hat{\boldsymbol{u}}}_s, \hat{\hat{\boldsymbol{v}}}_s\}_{s\in[10000]}$. We repeat this procedure 100 times and report the average results in Figure 1.

We observe that simple policies like RobustStatic and RobustAffine perform the best when the training size is small ($S = 63$), as they avoid overfitting the limited data. However, as the training size increases, there is little improvement in the performance of these simple policies, suggesting underfitting when the training size is relatively large ($S \geq 126$). In contrast, the performance of more complex policies improves significantly as the number of training samples increases. When the training size is relatively large ($S \geq 126$), the RobustTreeAffine policy outperforms the RobustStatic policy, demonstrating the benefits of using side information. Perhaps surprisingly, the RobustTreeAffine policies also outperform the Forest policy when $S \geq 126$, indicating that a simple tree with fewer than four leaf nodes is sufficient to provide high-quality solutions. This phenomenon can be partially explained by the modeling philosophy, as the Forest policy attempts to evaluate out-of-sample CVaR based on a large set of return realizations corresponding to the same observed side information, optimizing the *conditional* CVaR over possible portfolios. However, we can only collect one out-of-sample daily return realization for the given side information, as the side information changes every day. Therefore, we assess the out-of-sample *unconditional* CVaR based on pairs of different side information and their corresponding returns. As a result, the Forest policy may not necessarily be optimal, as we illustrate in our numerical comparisons.

A low signal-noise ratio setting

According to Gu et al. (2020), the out-of-sample R^2 for monthly return prediction is below 1% using well-tuned random forest, gradient boost, or neural networks, indicating a much lower signal-noise

Figure 2 CVaR comparison in a low signal-noise ratio setting



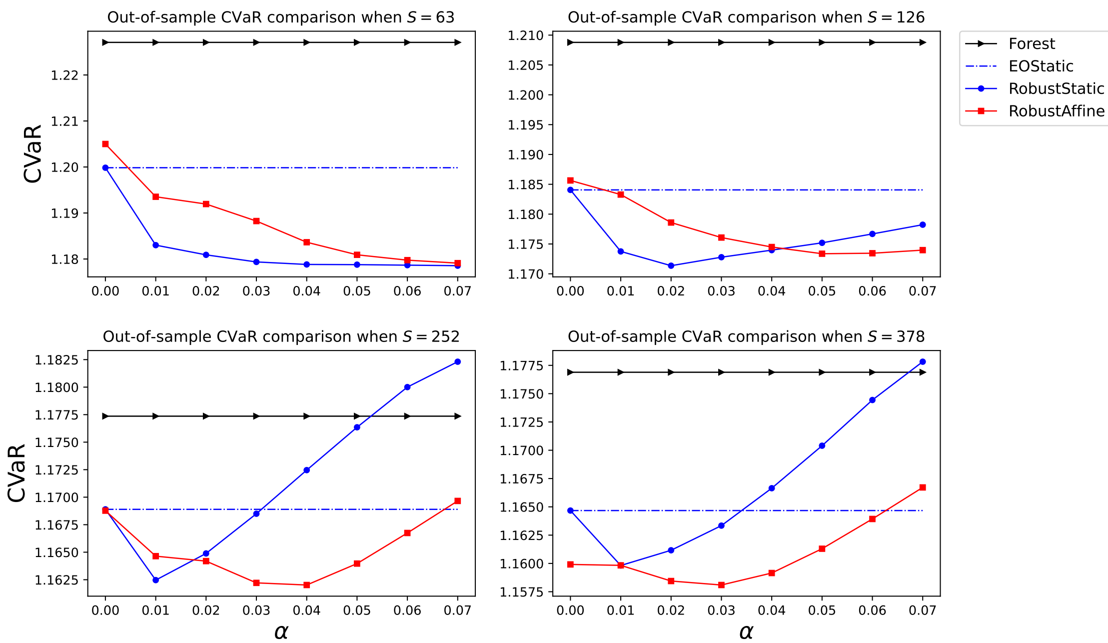
ratio for the real-world setting than in the simulation considered above. Thus, we are motivated to explore a low signal-noise simulation setting, which we achieve by introducing an independent noise term $LN(0,0.8)$. Specifically, the data-generating process becomes:

$$\tilde{v}_1 = 1 + 0.2\exp(\tilde{u}_1) - LN(0, 1 - 0.5\mathbb{1}_{[-3,-1]}(\tilde{u}_2)) - LN(0, 0.8)$$
$$\tilde{v}_2 = 1 - 0.2\tilde{u}_1 - LN(0, 1 - 0.5\mathbb{1}_{[-1,1]}(\tilde{u}_2)) - LN(0, 0.8)$$
$$\tilde{v}_3 = 1 + 0.2|\tilde{u}_1| - LN(0, 1 - 0.5\mathbb{1}_{[1,3]}(\tilde{u}_2)) - LN(0, 0.8).$$

We conduct the same experiment as before and vary $\alpha \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5\}$ to increase robustness given the low signal-noise ratio. We report the numerical results in Figure 2.

We observe that all robust satisficing policies, including RobustStatic, RobustAffine, and RobustTreeAffine policies, can achieve a lower CVaR by increasing the target margin α to some extent. This finding highlights the importance of incorporating robustness into the optimization problem, which is attained in our model by setting a less ambitious CVaR target. It suggests it can provide a better solution to the portfolio problem in a low signal-noise ratio setting. Moreover, we observe similar results to those in Figure 1. Specifically, simple policies, such as RobustStatic, and RobustAffine, outperform complex policies when the training size is relatively small ($S \leq 252$). We require more observations to learn complex policies effectively as the signal-noise ratio decreases. When the training size is $S = 378$, complex policies can achieve a lower CVaR than more straightforward policies. This observation indicates that the Forest and RobustTreeAffine policies can better utilize side information as the training size increases.

Figure 3 CVaR comparison based on the S&P data



Real-world dataset

In the next set of numerical experiments, we consider using real data based on five stocks in the S&P500 from January 01, 2010, to January 01, 2020. We utilize the short-term (last 21 days or one month) average return and standard deviation as side information, as empirical studies by Medhat and Schmeling (2022) and Moreira and Muir (2017) have demonstrated that leveraging such information can lead to high risk-adjusted returns.

We conduct the experiments in a rolling horizon manner with a time window of 21 periods. At the beginning of a time window (time t), we use the dataset $\mathcal{H}_t \triangleq \{\hat{\boldsymbol{u}}_s, \hat{\boldsymbol{v}}_s\}_{s=t-S}^{t-1}$ as the training set. We consider various training sizes, $S \in \{63, 126, 252, 378\}$. We obtain different policies using the training set and apply them to the next 21 periods $(t, t+1, \ldots, t+20)$. After obtaining the returns, we set $t \leftarrow t+21$ and repeat the above procedure 30 times. This process yields 630 out-of-sample returns for each policy, and we calculate the corresponding CVaR. To mitigate the impact of the selected stocks, we randomly sample five stocks 20 times and report the average of 20 out-of-sample CVaRs in Figure 3. Hence, we fix $L = 1$ and focus on the benefits of the RobustAffine policy.

Similar to the findings in Figure 2, we observe that when the sample size is relatively small $(S \leq 126)$, the RobustStatic policy outperforms the RobustAffine policy, indicating the difficulty in leveraging side information in a low signal-noise ratio setting with limited data. However, as the sample size increases, the performance of the RobustAffine policy improves. For the largest training size $(S = 378)$, the RobustAffine policy outperforms the RobustStatic policy for a wide range of target parameters α, highlighting the benefits of incorporating side information even in a low signal-noise ratio setting. Additionally, we note that the RobustAffine policy outperforms the Forest policy across all experiments, which is consistent with previous simulation results.

7. A case study on interpretable taxi allocation

To illustrate the interpretability of the proposed framework, we examine a taxi allocation problem based on the work of Hao et al. (2020), where a taxi operator manages a fleet of taxis and must allocate them across various service regions. Some regions experience a surplus of taxis, while others face a shortage. In this context, we define a region with an abundance of taxi supply as a "supply region" and a region with an insufficient number of taxis to meet passenger demands as a "demand region". The taxi operator can address this supply-demand imbalance by deploying taxis from the supply regions to fulfill uncertain demands in the demand regions.

We represent the maximum number of taxis that can be deployed from a supply region $i \in [I]$ to the demand regions as q_i. Additionally, we denote the uncertain demand for taxis at each demand region $j \in [J]$ as $\tilde{v}_j$. Using real data from a taxi operator provider in Singapore, Hao et al. (2020) elucidate that weather information, specifically the level of precipitation (measured in millimeters), has some predictive power on the demands of taxis in the different demand regions of the city. Therefore, we use the level of precipitation as the side information and denote it by a univariate random variable $\tilde{u}$.

Let x_{ij} represent the number of taxis allocated from supply region $i \in [I]$ to demand region $j \in [J]$, and let c_{ij} denote the associated unit allocation cost. The allocation decision $\boldsymbol{x}$ can adapt to the side information $\tilde{u}$. It is important to note that demands can only be satisfied when taxis are available from the supply region. Consequently, the number of taxi demands satisfied at demand region $j \in [J]$ is given by the minimum between the sum of $x_{ij}(\tilde{u})$ over all $i \in [I]$ and $\tilde{v}_j$. We assume that the decision-maker does not incur any cost for unmet demands and earns a revenue of r_j for fulfilling each demand at demand region $j \in [J]$. We formulate the taxi allocation problem as follows,

$$\begin{aligned} \min\ & \mathbb{E}_{\mathbb{P}}\left[g(\boldsymbol{x}(\tilde{u}), \tilde{\boldsymbol{v}})\right] \\ \text{s.t}\ & \sum_{j \in [J]} x_{ij}(u) \leq q_i \ \ \forall u \in \mathcal{U}, i \in [I] \\ & \boldsymbol{x}(u) \geq \mathbf{0} \qquad\quad \forall u \in \mathcal{U} \end{aligned} \tag{27}$$

where the cost function g can be written as the optimal value of a linear optimization problem,

$$\begin{aligned} g(\boldsymbol{x}, \boldsymbol{v}) = \min\ & \sum_{j \in [J]} y_j \\ \text{s.t}\ & y_j \geq \sum_{i \in [I]} (c_{ij} - r_j) x_{ij} \ \ \forall j \in [J] \\ & y_j \geq \sum_{i \in [I]} c_{ij} x_{ij} - r_j v_j \ \ \forall j \in [J]. \end{aligned}$$

Given the historical samples of the side information and demand $(\hat{u}_s, \hat{\boldsymbol{v}}_s)$, $s \in [S]$, the corresponding robust satisficing taxi allocation model with side information is as follows,

$$
\begin{aligned}
\min\ & \kappa \\
\text{s.t. } & \frac{1}{S}\sum_{s\in[S]} t_s \le \tau \\
& \mathbf{1}^\top \boldsymbol{y}_{\ell s}(u,\boldsymbol{v},\sigma,\nu) - \kappa\sigma - \kappa\nu \le t_s && \forall (u,\boldsymbol{v},\sigma,\nu) \in \bar{\mathcal{Z}}_{\ell s}, \ell\in[L], s\in[S] \\
& y_{\ell s,j}(u,\boldsymbol{v},\sigma,\nu) \ge \sum_{i\in[I]} (c_{ij} - r_j) x_{\ell,ij}(u) && \forall (u,\boldsymbol{v},\sigma,\nu) \in \bar{\mathcal{Z}}_{\ell s}, \ell\in[L], s\in[S], j\in[J] \\
& y_{\ell s,j}(u,\boldsymbol{v},\sigma,\nu) \ge \sum_{i\in[I]} c_{ij} x_{\ell,ij}(u) - r_j v_j && \forall (u,\boldsymbol{v},\sigma,\nu) \in \bar{\mathcal{Z}}_{\ell s}, \ell\in[L], s\in[S], j\in[J] \\
& \sum_{j\in[J]} x_{\ell,ij}(u) \le q_i && \forall u \in \mathcal{U}_\ell, \ell\in[L], i\in[I] \\
& \boldsymbol{x}_\ell(u) \ge \mathbf{0} && \forall u \in \mathcal{U}_\ell, \ell\in[L] \\
& \boldsymbol{y}_{\ell s} \in \mathcal{L}^{n_u+n_v+2,n_y} && \forall \ell\in[L], s\in[S] \\
& \boldsymbol{x}_\ell \in \mathcal{L}^{n_u,n_x} && \forall \ell\in[L] \\
& \kappa \ge 0.
\end{aligned}
\tag{28}
$$

Note that the recourse problem in this case study has only right-hand-side uncertainty. As we have discussed at the beginning of Section 4, the reformulation of Problem (28) directly follows from Long et al. (2023); hence, we omit the reformulation here. As a benchmark model, we consider the robust satisficing model without side information, *i.e.*, the response policy is a tree-based static policy with $L = 1$.

Data generation. Similar to the simulation experiment conducted in Hao et al. (2020), we make the following assumptions: $I = 1$ and $J = 5$. The cost and revenue parameters are set as $c_j = 3$ and $r_j = 0.05(12.5 - 0.5j) + c_j$ for each $j \in [J]$. The mean and standard deviation of u are denoted as $\mu_u = 10$ and $\sigma_u = 0.3\mu_u$, respectively. We assume that the support set of u falls within the interval $\mathcal{I}_u = [1, 19]$. We divide the interval $\mathcal{I}_u$ into $n_d = 4$ non-overlapping subsets of the same length and denote the i-th subset as $\mathcal{I}_u^i = [\underline{u}_i, \overline{u}_i]$ for each $i \in [n_d]$.

In our model, we consider n_d linear demand functions, $\boldsymbol{v}^i(u, \tilde{\boldsymbol{\epsilon}}^i) = \boldsymbol{w}_0^i + \boldsymbol{w}_1^i u + \tilde{\boldsymbol{\epsilon}}^i$, where $i \in [n_d]$. We randomly generate $\bar{\boldsymbol{w}}$ from the uniform distribution on $[0,1]^J$ and set $\boldsymbol{w}_1^i = (1 + 0.2i)\bar{\boldsymbol{w}}$ for each $i \in [n_d]$. For the intercept $\boldsymbol{w}_0^i$, we set $\boldsymbol{w}_0^1 = 10 \times \mathbf{1}$ and $\boldsymbol{w}_0^i = \boldsymbol{w}_0^{i-1} + \underline{u}_i \boldsymbol{w}_1^{i-1} - \underline{u}_i \boldsymbol{w}_1^i$ for $i \in \{2, \ldots, n_d\}$. The error term $\tilde{\boldsymbol{\epsilon}}^i$ has a zero mean, and its variance is $\boldsymbol{\delta}_e^i = 0.1\boldsymbol{\mu}_v^i$, where $\boldsymbol{\mu}_v^i = \boldsymbol{w}_0^i + \boldsymbol{w}_1^i \mu_u$. The support set of the demand is a box, $\mathcal{I}_v \triangleq [\boldsymbol{w}^1 + \boldsymbol{w}_1^1 \underline{u}_1, \boldsymbol{w}^{n_d} + \boldsymbol{w}_1^{n_d} \overline{u}_{n_d}]$. Next, we set the supply capacity as $q = 0.5 \times \mathbf{1}^\top(\boldsymbol{w}_0^{n_d} + \boldsymbol{w}_1^{n_d} \overline{u}_{n_d})$. Note that such a supply capacity q is only half of the total mean demand when $u = \overline{u}_{n_d}$; hence, we may not satisfy the total expected demand when u is large.

To generate samples for training and evaluation, we randomly generate $S = 60$ samples of $\hat{u}_s$. For each $s \in [S]$, we randomly generate samples of the error term, $\hat{\boldsymbol{\epsilon}}_s$, from the multivariate normal

Table 1 Revenue improvement in ten random sample instances

Sample instance	1	2	3	4	5	6	7	8	9	10	Average
R^S	34.16	34.54	34.48	34.20	34.34	34.54	34.59	34.56	34.08	34.64	34.41
R^N	32.52	31.90	32.46	32.40	32.49	32.15	32.44	32.24	32.44	32.53	32.36
$\frac{R^S-R^N}{R^N}$	5.04%	8.28%	6.23%	5.53%	5.67%	7.42%	6.62%	7.20%	5.05%	6.49%	6.35%

distribution $\mathcal{N}(\mathbf{0}, \operatorname{diag}(\boldsymbol{\delta}_e^{j^*}))$ if $\hat{u}_s \in \mathcal{I}_u^{j^*}$ for some $j^* \in [n_d]$. The demand sample with observation $\hat{u}_s$ is then $\hat{\boldsymbol{v}}_s = \boldsymbol{v}^{j^*}(\hat{u}_s, \hat{\boldsymbol{\epsilon}}_s) = \boldsymbol{w}_0^{j^*} + \boldsymbol{w}_1^{j^*}\hat{u}_s + \hat{\boldsymbol{\epsilon}}_s$. We truncate demand samples so that they fall into the support set $\mathcal{I}_v$. Then, $\{(\hat{u}_s, \hat{\boldsymbol{v}}_s), s \in [S]\}$ constitutes our training samples. We generate $10,000$ test samples with the same procedure to evaluate the out-of-sample performance.

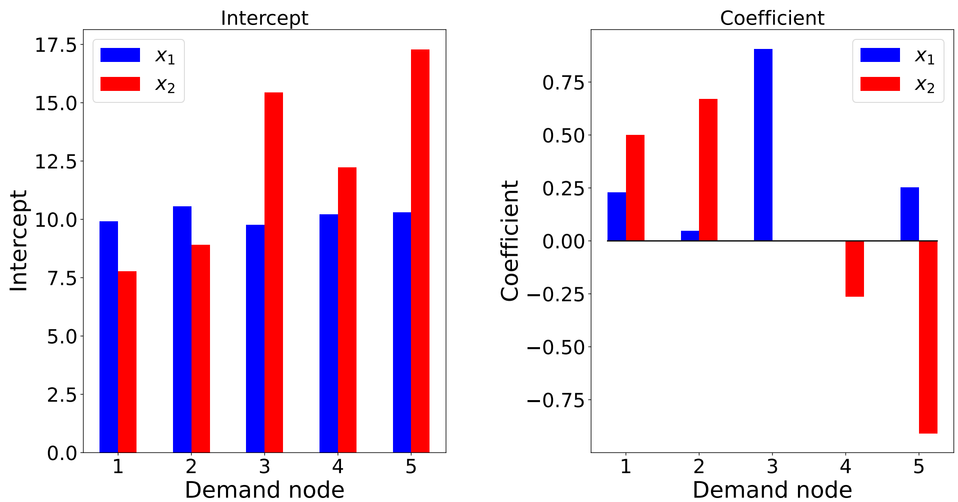
Evaluation of revenue improvement. We first generate the tree structures with $L \in [4]$ using the binary recursive partitioning algorithm in Section 2. Based on cross-validation, we fix the number of leaf nodes as $L = 2$. Then, we calibrate the target parameter in our robust satisficing model (28) by selecting the target margin $\alpha \in [0, 4]$ via a 5-fold cross-validation combined with a Golden search procedure. We solve the benchmark robust satisficing model without side information via the same cross-validation procedure. In our numerical experiments, we solve the two models with ten different training data sets and test them on the same set of test data of size $10,000$.

Let R^S and R^N denote the out-of-sample revenue of the robust satisficing models with tree-based affine policy and the benchmark robust satisficing model (with static policy ignoring side information), respectively. We compute the revenue ratio between the two models, $\frac{R^S}{R^N}$, and summarize the results in Table 1. We note that the average revenue ratio over the ten random instances is 1.064, indicating an improvement in revenue of 6.4%. A significant reason for the improvement is that the solutions to the prescriptive analytics model with side information can adapt to the precipitation data to better align with the demands. This is similarly observed in the portfolio optimization example in the previous section.

Interpretable taxi allocation policy. The impact of precipitation on our optimal taxi allocation policy is substantial, as illustrated by the first random instance's optimal response policy shown in Figure 4. One immediate advantage of the tree-based affine allocation policy is its real-time adaptability to realized precipitation information. The policy, consisting of intercepts and coefficients, is easily understandable and interpretable. The non-negligible magnitudes of the coefficients indicate the strong influence of precipitation information on the allocation quantities.

Note that the mean demands at all demand regions increase with precipitation u. When u falls within the first leaf node, the overall demands do not overwhelm the supply capacity q. Hence, it is reasonable to see positive coefficients, *i.e.,* we should deploy more taxis to all demand regions

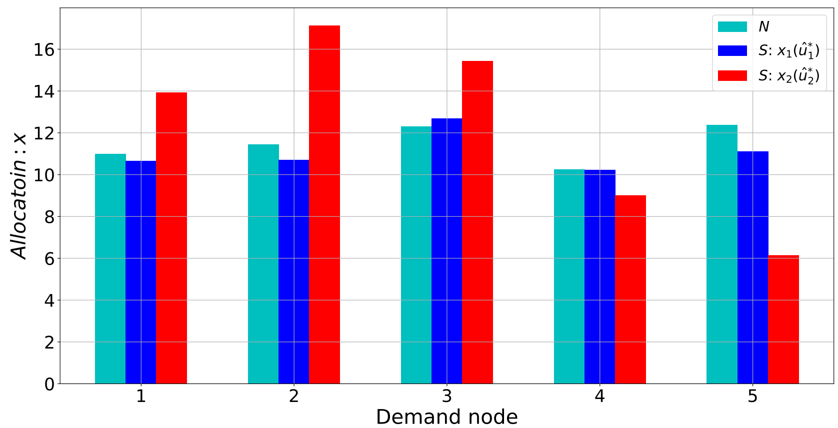
Figure 4 **Illustration of the tree-based affine allocation policy**



as u increases. An interesting observation arises when u lies within the second leaf node: the coefficients of the affine allocation policy for demand regions 4 and 5 become negative, seemingly contradicting the general trend of increasing demands with u. However, this behavior can be explained by considering the limited supply capacity q. In this scenario, the overall mean demand becomes overwhelming, requiring the system to prioritize which demands to fulfill. Given that the unit revenue earned from satisfying demand in a particular demand region $j \in [J]$ decreases as j increases, it becomes optimal to sacrifice some demands from demand regions 4 and 5 in order to meet as many demands as possible in the earlier demand regions. This intuitive and interpretable allocation policy highlights the practical usefulness of the robust actionable prescriptive analytics framework for managers and practitioners.

To elucidate the difference between our adaptive allocation policy and the static policy of the benchmark robust satisficing model ignoring side information, we visualize them in Figure 5. Specifically, we plot the allocation quantities for two instances: $\hat{u}_1^*$ and $\hat{u}_2^*$, which lie in leaf nodes one and two, respectively. We set the covariate values as $\hat{u}_1^* = \frac{1+u^*}{2}$ and $\hat{u}_2^* = \frac{19+u^*}{2}$, where u^* is the cutoff value that divides the samples into two leaf nodes. The corresponding allocation quantities $\boldsymbol{x}_1(\hat{u}_1^*)$ and $\boldsymbol{x}_2(\hat{u}_2^*)$ are presented in Figure 5. The most notable difference is that our policy prioritizes demands at the first three demand regions when u lies in the second leaf node due to the overwhelming total demand. This demonstrates the benefits of having a response policy that could leverage the side information with some predictive powers for uncertain demands.

Figure 5 **Illustration of allocation quantities of robust satisficing with (S) and without (N) side information.**



References

Arrieta AB, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, García S, Gil-López S, Molina D, Benjamins R, et al. (2020) Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* 58:82–115.

Ban GY, Rudin C (2019) The big data newsvendor: Practical insights from machine learning. *Operations Research* 67(1):90–108.

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

Bertsimas D, Dunn J, Mundru N (2019a) Optimal prescriptive trees. *INFORMS Journal on Optimization* 1(2):164–183.

Bertsimas D, Kallus N (2020) From predictive to prescriptive analytics. *Management Science* 66(3):1025–1044.

Bertsimas D, Koduri N (2022) Data-driven optimization: A reproducing kernel Hilbert space approach. *Operations Research* 70(1):454–471.

Bertsimas D, McCord C, Sturt B (2023a) Dynamic optimization with side information. *European Journal of Operational Research* 304(2):634–651.

Bertsimas D, Shtern S, Sturt B (2023b) A data-driven approach to multistage stochastic linear optimization. *Management Science* 69(1):51–74.

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

Bertsimas D, Stellato B (2021) The voice of optimization. *Machine Learning* 110(2):249–277.

Bertsimas D, Van Parys B (2021) Bootstrap robust prescriptive analytics. *Mathematical Programming* 1–40.

Blanchet J, Kang Y, Murthy K (2019a) Robust Wasserstein profile inference and applications to machine learning. *Journal of Applied Probability* 56(3):830–857.

Blanchet J, Kang Y, Murthy K (2019b) Robust wasserstein profile inference and applications to machine learning. *Journal of Applied Probability* 56(3):830–857.

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

Chen Z, Sim M, Xiong P (2020) Robust stochastic optimization made easy with RSOME. *Management Science* 66(8):3329–3339.

de Ruiter FJ, Zhen J, den Hertog D (2022) Dual approach for two-stage robust nonlinear optimization. *Operations Research* .

Elmachtoub AN, Grigas P (2022) Smart "predict, then optimize". *Management Science* 68(1):9–26.

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

Esteban-Pérez A, Morales JM (2021) Distributionally robust stochastic programs with side information based on trimmings. *Mathematical Programming* 1–37.

Ferreira K, Lee B, Simchi-Levi D (2016) Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & Service Operations Management* 18(1):69–88.

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

Gao R (2022) Finite-sample guarantees for Wasserstein distributionally robust optimization: Breaking the curse of dimensionality. *Operations Research* .

Gao R, Chen X, Kleywegt AJ (2022) Wasserstein distributionally robust optimization and variation regularization. *Operations Research* .

Georghiou A, Wiesemann W, Kuhn D (2015) Generalized decision rule approximations for stochastic programming via liftings. *Mathematical Programming* 152(1):301–338.

Glaeser C, Fisher M, Su X (2019) Optimal retail location: Empirical methodology and application to practice. *Manufacturing & Service Operations Management* 21(1):86–102.

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

Gu S, Kelly B, Xiu D (2020) Empirical asset pricing via machine learning. *The Review of Financial Studies* 33(5):2223–2273.

Hannah L, Powell W, Blei D (2010) Nonparametric density estimation for stochastic optimization with an observable state variable. *Advances in Neural Information Processing Systems* 23.

Hao Z, He L, Hu Z, Jiang J (2020) Robust vehicle pre-allocation with uncertain covariates. *Production and Operations Management* 29(4):955–972.

Kallus N, Mao X (2023) Stochastic optimization forests. *Management Science* 69(4):1975–1994.

Kannan R, Bayraksan G, Luedtke JR (2020) Residuals-based distributionally robust optimization with covariate information. *arXiv preprint arXiv:2012.01088* .

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

Lipton ZC (2018) The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue* 16(3):31–57.

Liyanage LH, Shanthikumar JG (2005) A practical inventory control policy using operational statistics. *Operations Research Letters* 33(4):341–348.

Loke GG, Tang Q, Xiao Y (2021) Decision-driven regularization: A blended model for predict-then-optimize. *Available at SSRN 3623006* .

Long DZ, Sim M, Zhou M (2023) Robust satisficing. *Operations Research* 71(1):61–82.

Medhat M, Schmeling M (2022) Short-term momentum. *The Review of Financial Studies* 35(3):1480–1526.

Moreira A, Muir T (2017) Volatility-managed portfolios. *The Journal of Finance* 72(4):1611–1644.

Nguyen VA, Zhang F, Blanchet J, Delage E, Ye Y (2021) Robustifying conditional portfolio decisions via optimal transport. *arXiv preprint arXiv:2103.16451* .

Notz PM, Pibernik R (2022) Prescriptive analytics for flexible capacity management. *Management Science* 68(3):1756–1775.

Ramachandra A, Rujeerapaiboon N, Sim M (2021) Robust conic satisficing. *Available at SSRN 3842446* .

Shafieezadeh-Abadeh S, Kuhn D, Esfahani PM (2019a) Regularization via mass transportation. *Journal of Machine Learning Research* 20(103):1–68.

Shafieezadeh-Abadeh S, Kuhn D, Esfahani Mohajerin P (2019b) Regularization via mass transportation. *Journal of Machine Learning Research* 20(103):1–68.

Si N, Blanchet J, Ghosh S, Squillante M (2020) Quantifying the empirical Wasserstein distance to a set of measures: Beating the curse of dimensionality. *Advances in Neural Information Processing Systems* 33:21260–21270.

Sim M, Tang Q, Zhou M, Zhu T (2023) The analytics of robust satisficing: Predict, optimize, satisfice, then fortify. *Available at SSRN 3981205* .

Sim M, Zhao L, Zhou M (2021) Tractable robust supervised learning models. *Available at SSRN 3981205* .

Smith JE, Winkler RL (2006) The optimizer's curse: Skepticism and postdecision surprise in decision analysis. *Management Science* 52(3):311–322.

Srivastava PR, Wang Y, Hanasusanto GA, Ho CP (2021) On data-driven prescriptive analytics with side information: A regularized Nadaraya-Watson approach. *arXiv preprint arXiv:2110.04855* .

Tulabandhula T, Rudin C (2013) Machine learning with operational costs. *Journal of Machine Learning Research* 14(25):1989–2028.

Yang J, Zhang L, Chen N, Gao R, Hu M (2022) Decision-making with side information: A causal transport robust approach. *Available at Optimization Online* .

Zhang L, Yang J, Gao R (2022) Optimal robust policy for feature-based newsvendor. *Management Science* .

Zhen J, de Ruiter FJ, Roos E, den Hertog D (2022a) Robust optimization for models with uncertain second-order cone and semidefinite programming constraints. *INFORMS Journal on Computing* 34(1):196–210.

Zhen J, Den Hertog D, Sim M (2018) Adjustable robust optimization via Fourier–Motzkin elimination. *Operations Research* 66(4):1086–1100.

Zhen J, Marandi A, de Moor D, den Hertog D, Vandenberghe L (2022b) Disjoint bilinear optimization: A two-stage robust optimization perspective. *INFORMS Journal on Computing* 34(5):2410–2427.

A. Proofs

Proof of Theorem 1. Let $\hat{\boldsymbol{x}}$ be an optimal solution to Problem (2). For any $(\boldsymbol{u}_s, \boldsymbol{v}_s) \in \mathcal{Z}$, $s \in [S]$, we have

$$
\begin{aligned}
& g(\hat{\boldsymbol{x}}(\boldsymbol{u}_s), \boldsymbol{v}_s) - \kappa(\|\boldsymbol{v}_s - \hat{\boldsymbol{v}}_s\| + \|\boldsymbol{u}_s - \hat{\boldsymbol{u}}_s\|) \\
= & \, g(\hat{\boldsymbol{x}}(\hat{\boldsymbol{u}}_s), \hat{\boldsymbol{v}}_s) + (g(\hat{\boldsymbol{x}}(\boldsymbol{u}_s), \boldsymbol{v}_s) - g(\hat{\boldsymbol{x}}(\hat{\boldsymbol{u}}_s), \hat{\boldsymbol{v}}_s) - \kappa(\|\boldsymbol{v}_s - \hat{\boldsymbol{v}}_s\| + \|\boldsymbol{u}_s - \hat{\boldsymbol{u}}_s\|)) \\
\leq & \, g(\hat{\boldsymbol{x}}(\hat{\boldsymbol{u}}_s), \hat{\boldsymbol{v}}_s) + (\bar{L} - \kappa)(\|\boldsymbol{u}_s - \hat{\boldsymbol{u}}_s\| + \|\boldsymbol{v}_s - \hat{\boldsymbol{v}}_s\|).
\end{aligned}
$$

For any $\kappa \geq \bar{L}$ and $(\boldsymbol{u}_s, \boldsymbol{v}_s) \in \mathcal{Z}$, $s \in [S]$, we have

$$
\frac{1}{S} \sum_{s \in [S]} (g(\hat{\boldsymbol{x}}(\boldsymbol{u}_s), \boldsymbol{v}_s) - \kappa(\|\boldsymbol{v}_s - \hat{\boldsymbol{v}}_s\| + \|\boldsymbol{u}_s - \hat{\boldsymbol{u}}_s\|)) \leq \frac{1}{S} \sum_{s \in [S]} g(\hat{\boldsymbol{x}}(\hat{\boldsymbol{u}}_s), \hat{\boldsymbol{v}}_s) = Z_0 \leq \tau.
$$

Hence, it follows that Problem (9) is feasible for some $\kappa_\tau \leq \bar{L}$. □

Proof of Theorem 2. We first observe q_4 and q_5 are necessarily non-negative; otherwise, Constraint (15) is never feasible. Then we dualize $\boldsymbol{v}$ to obtain the following equivalence

$$
\begin{aligned}
& \sup_{(\boldsymbol{u}, \boldsymbol{v}, \sigma, \nu) \in \bar{\mathcal{Z}}_{\ell s}} \left\{ \boldsymbol{q}_1^\top \boldsymbol{u} + \boldsymbol{q}_2^\top \boldsymbol{v} + \boldsymbol{v}^\top \boldsymbol{Q}_3 \boldsymbol{u} - q_4 \sigma - q_5 \nu \right\} \\
= & \sup_{\boldsymbol{u} \in \mathcal{U}_\ell, \boldsymbol{v} \in \mathcal{V}_\ell} \left\{ \boldsymbol{q}_1^\top \boldsymbol{u} + \boldsymbol{q}_2^\top \boldsymbol{v} + \boldsymbol{v}^\top \boldsymbol{Q}_3 \boldsymbol{u} - q_4 \|\boldsymbol{u} - \hat{\boldsymbol{u}}_s\| - q_5 \|\boldsymbol{v} - \hat{\boldsymbol{v}}_s\| \right\} \\
= & \sup_{\boldsymbol{u} \in \mathcal{U}_\ell} \inf_{\overline{\boldsymbol{\rho}} \geq \boldsymbol{0}, \underline{\boldsymbol{\rho}} \geq \boldsymbol{0}} \left\{ \boldsymbol{q}_1^\top \boldsymbol{u} + \overline{\boldsymbol{\rho}}^\top \overline{\boldsymbol{v}}_\ell - \underline{\boldsymbol{\rho}}^\top \underline{\boldsymbol{v}}_\ell + (\boldsymbol{q}_2 + \boldsymbol{Q}_3 \boldsymbol{u} + \underline{\boldsymbol{\rho}} - \overline{\boldsymbol{\rho}})^\top \hat{\boldsymbol{v}}_s - q_4 \|\boldsymbol{u} - \hat{\boldsymbol{u}}_s\| : \|\boldsymbol{q}_2 + \boldsymbol{Q}_3 \boldsymbol{u} - \overline{\boldsymbol{\rho}} + \underline{\boldsymbol{\rho}}\|_* \leq q_5 \right\}.
\end{aligned}
$$

The left-hand side of the above constraint can be viewed as an adaptive robust optimization problem with uncertain parameter $\boldsymbol{u}$ and recourse variable $\overline{\boldsymbol{\rho}}, \underline{\boldsymbol{\rho}}$:

$$
\begin{aligned}
\inf \; & \sup_{\boldsymbol{u} \in \mathcal{U}_\ell} \left\{ \boldsymbol{q}_1^\top \boldsymbol{u} + \overline{\boldsymbol{\rho}}(\boldsymbol{u})^\top \overline{\boldsymbol{v}}_\ell - \underline{\boldsymbol{\rho}}(\boldsymbol{u})^\top \underline{\boldsymbol{v}}_\ell + (\boldsymbol{q}_2 + \boldsymbol{Q}_3 \boldsymbol{u} + \underline{\boldsymbol{\rho}}(\boldsymbol{u}) - \overline{\boldsymbol{\rho}}(\boldsymbol{u}))^\top \hat{\boldsymbol{v}}_s - q_4 \|\boldsymbol{u} - \hat{\boldsymbol{u}}_s\| \right\} \\
\text{s.t.} \; & \|\boldsymbol{q}_2 + \boldsymbol{Q}_3 \boldsymbol{u} - \overline{\boldsymbol{\rho}}(\boldsymbol{u}) + \underline{\boldsymbol{\rho}}(\boldsymbol{u})\|_* \leq q_5 \\
& \overline{\boldsymbol{\rho}}(\boldsymbol{u}) \geq \boldsymbol{0}, \underline{\boldsymbol{\rho}}(\boldsymbol{u}) \geq \boldsymbol{0} \\
& \overline{\boldsymbol{\rho}}, \underline{\boldsymbol{\rho}} \in \mathcal{R}^{n_u, n_v}.
\end{aligned}
$$

We apply affine recourse adaptation $\overline{\boldsymbol{\rho}}(\boldsymbol{u}) = \overline{\boldsymbol{p}} + \overline{\boldsymbol{P}} \boldsymbol{u}$, $\underline{\boldsymbol{\rho}}(\boldsymbol{u}) = \underline{\boldsymbol{p}} + \underline{\boldsymbol{P}} \boldsymbol{u}$ to obtain its safe approximation of the robust constraint (15):

$$
\begin{aligned}
& \sup_{\boldsymbol{u} \in \mathcal{U}_\ell} \left\{ \boldsymbol{q}_1^\top \boldsymbol{u} + (\overline{\boldsymbol{p}} + \overline{\boldsymbol{P}} \boldsymbol{u})^\top \overline{\boldsymbol{v}}_\ell - (\underline{\boldsymbol{p}} + \underline{\boldsymbol{P}} \boldsymbol{u})^\top \underline{\boldsymbol{v}}_\ell + (\boldsymbol{q}_2 + \boldsymbol{Q}_3 \boldsymbol{u} + \underline{\boldsymbol{p}} + \underline{\boldsymbol{P}} \boldsymbol{u} - \overline{\boldsymbol{p}} - \overline{\boldsymbol{P}} \boldsymbol{u})^\top \hat{\boldsymbol{v}}_s - q_4 \|\boldsymbol{u} - \hat{\boldsymbol{u}}_s\| \right\} \leq q_6 \\
& \sup_{\boldsymbol{u} \in \mathcal{U}_\ell} \left\{ \|\boldsymbol{q}_2 + \boldsymbol{Q}_3 \boldsymbol{u} - \overline{\boldsymbol{p}} - \overline{\boldsymbol{P}} \boldsymbol{u} + \underline{\boldsymbol{p}} + \underline{\boldsymbol{P}} \boldsymbol{u}\|_* \right\} \leq q_5 \\
& \inf_{\boldsymbol{u} \in \mathcal{U}_\ell} \left\{ \boldsymbol{e}_i^\top (\overline{\boldsymbol{p}} + \overline{\boldsymbol{P}} \boldsymbol{u}) \right\} \geq 0 \qquad \forall i \in [n_v] \\
& \inf_{\boldsymbol{u} \in \mathcal{U}_\ell} \left\{ \boldsymbol{e}_i^\top (\underline{\boldsymbol{p}} + \underline{\boldsymbol{P}} \boldsymbol{u}) \right\} \geq 0 \qquad \forall i \in [n_v]
\end{aligned}
\tag{29}
$$

for some $\overline{\boldsymbol{p}}, \underline{\boldsymbol{p}} \in \mathbb{R}^{n_v}, \overline{\boldsymbol{P}}, \underline{\boldsymbol{P}} \in \mathbb{R}^{n_v \times n_u}$. By dualizing variable $\boldsymbol{u}$, the first robust constraint of (29) has the following equivalence

$$\begin{aligned}
&\boldsymbol{q}_1^\top \hat{\boldsymbol{u}}_s + \boldsymbol{q}_2^\top \hat{\boldsymbol{v}}_s + \hat{\boldsymbol{v}}_s^\top \boldsymbol{Q}_3 \hat{\boldsymbol{u}}_s + (\overline{\boldsymbol{p}} + \overline{\boldsymbol{P}}\hat{\boldsymbol{u}}_s)^\top (\overline{\boldsymbol{v}}_\ell - \hat{\boldsymbol{v}}_s) \\
&\qquad + (\underline{\boldsymbol{p}} + \underline{\boldsymbol{P}}\hat{\boldsymbol{u}}_s)^\top (\hat{\boldsymbol{v}}_s - \underline{\boldsymbol{v}}_\ell) + \overline{\boldsymbol{\lambda}}^\top (\overline{\boldsymbol{u}}_\ell - \hat{\boldsymbol{u}}_s) + \underline{\boldsymbol{\lambda}}^\top (\hat{\boldsymbol{u}}_s - \underline{\boldsymbol{u}}_\ell) \le q_6 \\
&\|\boldsymbol{q}_1 + \overline{\boldsymbol{P}}^\top (\overline{\boldsymbol{v}}_\ell - \hat{\boldsymbol{v}}_s) + \underline{\boldsymbol{P}}^\top (\hat{\boldsymbol{v}}_s - \underline{\boldsymbol{v}}_\ell) + \boldsymbol{Q}_3^\top \hat{\boldsymbol{v}}_s - \overline{\boldsymbol{\lambda}} + \underline{\boldsymbol{\lambda}}\|_* \le q_4 \\
&\overline{\boldsymbol{\lambda}} \ge \boldsymbol{0}, \underline{\boldsymbol{\lambda}} \ge \boldsymbol{0}
\end{aligned} \tag{30}$$

by strong duality. Similarly, the last two sets of robust constraints of (29) are equivalent to the existence of $\overline{\boldsymbol{\Gamma}}, \underline{\boldsymbol{\Gamma}} \in \mathbb{R}^{n_v \times n_u}$ such that

$$\begin{aligned}
&\overline{\boldsymbol{p}} - \overline{\boldsymbol{\Gamma}}\overline{\boldsymbol{u}}_\ell + (\overline{\boldsymbol{P}} + \overline{\boldsymbol{\Gamma}})\underline{\boldsymbol{u}}_\ell \ge \boldsymbol{0} \\
&\overline{\boldsymbol{\Gamma}} \ge \boldsymbol{0}, \overline{\boldsymbol{P}} + \overline{\boldsymbol{\Gamma}} \ge \boldsymbol{0} \\
&\underline{\boldsymbol{p}} - \underline{\boldsymbol{\Gamma}}\overline{\boldsymbol{u}}_\ell + (\underline{\boldsymbol{P}} + \underline{\boldsymbol{\Gamma}})\underline{\boldsymbol{u}}_\ell \ge \boldsymbol{0} \\
&\underline{\boldsymbol{\Gamma}} \ge \boldsymbol{0}, \underline{\boldsymbol{P}} + \underline{\boldsymbol{\Gamma}} \ge \boldsymbol{0}.
\end{aligned} \tag{31}$$

For the second robust constraint of (29), we have the following equivalence

$$\begin{aligned}
&\sup_{\boldsymbol{u} \in \mathcal{U}_\ell} \left\{ \|\boldsymbol{q}_2 + \boldsymbol{Q}_3 \boldsymbol{u} - \overline{\boldsymbol{p}} - \overline{\boldsymbol{P}}\boldsymbol{u} + \underline{\boldsymbol{p}} + \underline{\boldsymbol{P}}\boldsymbol{u}\|_* \right\} \le q_5 \\
\iff &\sup_{\boldsymbol{u} \in \mathcal{U}_\ell, \|\boldsymbol{\gamma}\| \le 1} \left\{ \boldsymbol{\gamma}^\top (\boldsymbol{q}_2 + \boldsymbol{Q}_3 \boldsymbol{u} - \overline{\boldsymbol{p}} - \overline{\boldsymbol{P}}\boldsymbol{u} + \underline{\boldsymbol{p}} + \underline{\boldsymbol{P}}\boldsymbol{u}) \right\} \le q_5 \\
\iff &\sup_{\|\boldsymbol{\gamma}\| \le 1} \inf_{\overline{\boldsymbol{\theta}} \ge \boldsymbol{0}, \underline{\boldsymbol{\theta}} \ge \boldsymbol{0}} \left\{ \boldsymbol{\gamma}^\top (\boldsymbol{q}_2 - \overline{\boldsymbol{p}} + \underline{\boldsymbol{p}}) + \overline{\boldsymbol{u}}_\ell^\top \overline{\boldsymbol{\theta}} - \underline{\boldsymbol{u}}_\ell^\top \underline{\boldsymbol{\theta}} : (\boldsymbol{Q}_3 - \overline{\boldsymbol{P}} + \underline{\boldsymbol{P}})^\top \boldsymbol{\gamma} = \overline{\boldsymbol{\theta}} - \underline{\boldsymbol{\theta}} \right\} \le q_5
\end{aligned}$$

where the first equivalence is by the definition of the dual norm, and the second is by strong duality. The problem can be regarded as an adaptive robust optimization with uncertainty $\boldsymbol{\gamma}$ and recourse variable $\overline{\boldsymbol{\theta}}, \underline{\boldsymbol{\theta}}$. However, we can still apply affine recourse adaptation

$$\overline{\boldsymbol{\theta}}(\gamma) = \boldsymbol{\theta} + \boldsymbol{\Theta}^\top \gamma, \quad \underline{\boldsymbol{\theta}}(\gamma) = \boldsymbol{\theta} + \boldsymbol{\Theta}^\top \boldsymbol{\gamma} - (\boldsymbol{Q}_3 - \overline{\boldsymbol{P}} + \underline{\boldsymbol{P}})^\top \boldsymbol{\gamma}$$

for some $\boldsymbol{\theta} \in \mathbb{R}^{n_u}, \boldsymbol{\Theta} \in \mathbb{R}^{n_v \times n_u}$ to obtain a conservative approximation as follows,

$$\begin{aligned}
&\boldsymbol{\gamma}^\top (\boldsymbol{q}_2 - \overline{\boldsymbol{p}} + \underline{\boldsymbol{p}}) + (\boldsymbol{\theta} + \boldsymbol{\Theta}^\top \gamma)^\top \overline{\boldsymbol{u}}_\ell - (\boldsymbol{\theta} + (\boldsymbol{\Theta} - \boldsymbol{Q}_3 + \overline{\boldsymbol{P}} - \underline{\boldsymbol{P}})^\top \gamma)^\top \underline{\boldsymbol{u}}_\ell \le q_5 && \forall \|\boldsymbol{\gamma}\| \le 1 \\
&\boldsymbol{\theta} + \boldsymbol{\Theta}^\top \gamma \ge \boldsymbol{0} && \forall \|\boldsymbol{\gamma}\| \le 1 \\
&\boldsymbol{\theta} + (\boldsymbol{\Theta} - \boldsymbol{Q}_3 + \overline{\boldsymbol{P}} - \underline{\boldsymbol{P}})^\top \boldsymbol{\gamma} \ge \boldsymbol{0} && \forall \|\boldsymbol{\gamma}\| \le 1,
\end{aligned}$$

which is equivalent to

$$\begin{aligned}
&\|\boldsymbol{q}_2 - \overline{\boldsymbol{p}} + \underline{\boldsymbol{p}} + \boldsymbol{\Theta}\overline{\boldsymbol{u}}_\ell - (\boldsymbol{\Theta} - \boldsymbol{Q}_3 + \overline{\boldsymbol{P}} - \underline{\boldsymbol{P}})\underline{\boldsymbol{u}}_\ell\|_* + (\overline{\boldsymbol{u}}_\ell - \underline{\boldsymbol{u}}_\ell)^\top \boldsymbol{\theta} \le q_5 \\
&\boldsymbol{\theta} \ge \sum_{i \in [n_u]} \|\boldsymbol{\Theta}\boldsymbol{e}_i\|_* \boldsymbol{e}_i \\
&\boldsymbol{\theta} \ge \sum_{i \in [n_u]} \|(\boldsymbol{\Theta} - \boldsymbol{Q}_3 + \overline{\boldsymbol{P}} - \underline{\boldsymbol{P}})\boldsymbol{e}_i\|_* \boldsymbol{e}_i.
\end{aligned} \tag{32}$$

Combining the equations (30), (31), (32) together and noting they hold for each $s\in[S]$, we obtain the tractable formulation (16).

We have already established the conservative approximation of Constraint (15). When $\boldsymbol{Q}_3=\boldsymbol{0}$, the converse is true by noticing that the Constraint (15) is affine in the uncertain parameters so that we have its exact reformulation:

$$
\begin{aligned}
&\exists\overline{\boldsymbol{\lambda}},\underline{\boldsymbol{\lambda}}\in\mathbb{R}_+^{n_u},\overline{\boldsymbol{p}},\underline{\boldsymbol{p}}\in\mathbb{R}_+^{n_v}:\\
&\boldsymbol{q}_1^\top\hat{\boldsymbol{u}}_s+\boldsymbol{q}_2^\top\hat{\boldsymbol{v}}_s+\overline{\boldsymbol{p}}^\top(\overline{\boldsymbol{v}}_\ell-\hat{\boldsymbol{v}}_s)+\underline{\boldsymbol{p}}^\top(\hat{\boldsymbol{v}}_s-\underline{\boldsymbol{v}}_\ell)+\overline{\boldsymbol{\lambda}}^\top(\overline{\boldsymbol{u}}_\ell-\hat{\boldsymbol{u}}_s)+\underline{\boldsymbol{\lambda}}^\top(\hat{\boldsymbol{u}}_s-\underline{\boldsymbol{u}}_\ell)\le q_6\\
&\|\boldsymbol{q}_1-\overline{\boldsymbol{\lambda}}+\underline{\boldsymbol{\lambda}}\|_*\le q_5\\
&\|\boldsymbol{q}_2-\overline{\boldsymbol{p}}+\underline{\boldsymbol{p}}\|_*\le q_4
\end{aligned}
$$

based on the standard duality approach. Clearly $(\boldsymbol{q}_1,\boldsymbol{q}_2,\boldsymbol{0},q_4,q_5,q_6)\in\mathcal{Q}_{\ell s}$ by setting $\overline{\boldsymbol{P}}=\underline{\boldsymbol{P}}=\overline{\boldsymbol{\Gamma}}=\underline{\boldsymbol{\Gamma}}=\boldsymbol{\Theta}=\boldsymbol{0}$ and $\boldsymbol{\theta}=\boldsymbol{0}$ in Equation (16).

To show the second property, we can take $\overline{\boldsymbol{p}}=\underline{\boldsymbol{p}}=\boldsymbol{0},\overline{\boldsymbol{\lambda}}=\underline{\boldsymbol{\lambda}}=\boldsymbol{0},\overline{\boldsymbol{P}}=\underline{\boldsymbol{P}}=\overline{\boldsymbol{\Gamma}}=\underline{\boldsymbol{\Gamma}}=\boldsymbol{\Theta}=\boldsymbol{0}$, $q_4=\|\boldsymbol{q}_1+\boldsymbol{Q}_3^\top\hat{\boldsymbol{v}}_s\|_*$, $\boldsymbol{\theta}=\sum_{i\in[n_u]}\|\boldsymbol{Q}_3\boldsymbol{e}_i\|_*\boldsymbol{e}_i$ and $q_5=\|\boldsymbol{q}_2+\boldsymbol{Q}_3\underline{\boldsymbol{u}}_\ell\|_*+(\overline{\boldsymbol{u}}_\ell-\underline{\boldsymbol{u}}_\ell)^\top\boldsymbol{\theta}$ to ensure $(\boldsymbol{q}_1,\boldsymbol{q}_2,\boldsymbol{Q}_3,q_4,q_5,q_6)\in\mathcal{Q}_{\ell s}$.

□

Proof of Theorem 3. For any $\boldsymbol{x}\in\bar{\mathcal{T}}^{n_u,n_x}$, let $\ell^*(s)\in[L]$ such that $\hat{\boldsymbol{u}}_s\in\operatorname{int}(\mathcal{U}_{\ell^*(s)})$ for each $s\in[S]$, we have

$$
\begin{aligned}
&\frac{1}{S}\sum_{s\in[S]}t_s\le\tau\\
&\boldsymbol{d}^\top\hat{\boldsymbol{y}}_s\le t_s && \forall s\in[S]\\
&\boldsymbol{F}(\hat{\boldsymbol{v}}_s)\boldsymbol{x}_{\ell^*(s)}(\hat{\boldsymbol{u}}_s)+\boldsymbol{B}\hat{\boldsymbol{y}}_s\ge\boldsymbol{f}(\hat{\boldsymbol{v}}_s) && \forall s\in[S]
\end{aligned}
$$

for some $\hat{\boldsymbol{y}}_s\in\mathbb{R}^{n_y}$, $s\in[S]$. We take $\boldsymbol{y}^0_{\ell^*(s)s}=\hat{\boldsymbol{y}}_s$ and $\boldsymbol{y}^i_{\ell^*(s)s}=\boldsymbol{0}$ for each $i\in[n_u+n_v]$ so that

$$
\begin{aligned}
&\boldsymbol{d}^\top\boldsymbol{y}^0_{\ell^*(s)s}=\boldsymbol{d}^\top\hat{\boldsymbol{y}}_s\le t_s\\
&\boldsymbol{q}^{k1\top}_{\ell^*(s)s}\hat{\boldsymbol{u}}_s+\boldsymbol{q}^{k2\top}_{\ell^*(s)s}\hat{\boldsymbol{v}}_s+\hat{\boldsymbol{v}}_s^\top\boldsymbol{Q}^{k3}_{\ell^*(s)s}\hat{\boldsymbol{u}}_s-q^{k6}_{\ell^*(s)s}=\boldsymbol{e}_k^\top\left(\boldsymbol{f}(\hat{\boldsymbol{v}}_s)-\boldsymbol{F}(\hat{\boldsymbol{v}}_s)\boldsymbol{x}_{\ell^*(s)}(\hat{\boldsymbol{u}}_s)-\boldsymbol{B}\hat{\boldsymbol{y}}_s\right)\le 0\ \forall k\in[n_f].
\end{aligned}
$$

Therefore, Property (b) in Theorem 2 implies the existence of $\lambda_{\ell^*(s)s},\mu_{\ell^*(s)s}\in\mathbb{R}_+$ and $\boldsymbol{\zeta}_{\ell^*(s)s},\boldsymbol{\xi}_{\ell^*(s)s}\in\mathbb{R}_+^{n_f}$ such that

$$
\begin{aligned}
&(\boldsymbol{0},\boldsymbol{0},\boldsymbol{0},\lambda_{\ell^*(s)s},\mu_{\ell^*(s)s},t_s-\boldsymbol{d}^\top\boldsymbol{y}^0_{\ell^*(s)s})\in\mathcal{Q}_{\ell^*(s)s}\\
&\left(\boldsymbol{q}^{k1}_{\ell^*(s)s},\boldsymbol{q}^{k2}_{\ell^*(s)s},\boldsymbol{Q}^{k3}_{\ell^*(s)s},\boldsymbol{e}_k^\top\boldsymbol{\zeta}_{\ell^*(s)s},\boldsymbol{e}_k^\top\boldsymbol{\xi}_{\ell^*(s)s},q^{k6}_{\ell^*(s)s}\right)\in\mathcal{Q}_{\ell^*(s)s}\ \forall k\in[n_f].
\end{aligned}
$$

By the complete recourse assumption, there exists $\boldsymbol{\alpha}_{\ell^*(s)s},\boldsymbol{\beta}_{\ell^*(s)s}\in\mathbb{R}^y$ such that $\boldsymbol{B}\boldsymbol{\alpha}_{\ell^*(s)s}\ge\boldsymbol{\zeta}_{\ell^*(s)s}$ and $\boldsymbol{B}\boldsymbol{\beta}_{\ell^*(s)s}\ge\boldsymbol{\xi}_{\ell^*(s)s}$. Then we have

$$
\left(\boldsymbol{q}^{k1}_{\ell^*(s)s},\boldsymbol{q}^{k2}_{\ell^*(s)s},\boldsymbol{Q}^{k3}_{\ell^*(s)s},\boldsymbol{e}_k^\top\boldsymbol{B}\boldsymbol{\alpha}_{\ell^*(s)s},\boldsymbol{e}_k^\top\boldsymbol{B}\boldsymbol{\beta}_{\ell^*(s)s},q^{k6}_{\ell^*(s)s}\right)\in\mathcal{Q}_{\ell^*(s)s}\ \forall k\in[n_f].
$$

Then we take $\boldsymbol{y}_{\ell^*(s)s}^{n_u+n_v+1}=\boldsymbol{\alpha}_{\ell^*(s)s}$, $\boldsymbol{y}_{\ell^*(s)s}^{n_u+n_v+2}=\boldsymbol{\beta}_{\ell^*(s)s}$ and

$$\kappa \geq \bar{\kappa} \triangleq \max_{s\in[S]}\left\{\max\left\{\lambda_{\ell^*(s)s}+\boldsymbol{d}^\top \boldsymbol{y}_{\ell^*(s)s}^{n_u+n_v+1}, \mu_{\ell^*(s)s}+\boldsymbol{d}^\top \boldsymbol{y}_{\ell^*(s)s}^{n_u+n_v+2}\right\}\right\}$$

to ensure $(\boldsymbol{0},\boldsymbol{0},\boldsymbol{0},\kappa-\boldsymbol{d}^\top \boldsymbol{y}_{\ell^*(s)s}^{n_u+n_v+1},\kappa-\boldsymbol{d}^\top \boldsymbol{y}_{\ell^*(s)s}^{n_u+n_v+2},t_s-\boldsymbol{d}^\top \boldsymbol{y}_{\ell^*(s)s}^{0})\in\mathcal{Q}_{\ell^*(s)s}$, which is equivalent to

$$\boldsymbol{d}^\top \boldsymbol{y}_{\ell^*(s)s}(\boldsymbol{u},\boldsymbol{v},\sigma,\nu)-\kappa(\sigma+\nu)\leq t_s, \quad \forall(\boldsymbol{u},\boldsymbol{v},\sigma,\nu)\in\bar{\mathcal{Z}}_{\ell^*(s)s}$$

by Property (a) in Theorem 2.

For $\ell\in[L]\backslash\{\ell^*(s)\}$, by complete recourse assumption there exists $\hat{\boldsymbol{y}}_{\ell s}$ such that

$$\boldsymbol{B}\hat{\boldsymbol{y}}_{\ell s}\geq \boldsymbol{f}(\hat{\boldsymbol{v}}_s)-\boldsymbol{F}(\hat{\boldsymbol{v}}_s)\boldsymbol{x}_\ell(\hat{\boldsymbol{u}}_s)$$

Let $\boldsymbol{y}_{\ell s}^0=\hat{\boldsymbol{y}}_{\ell s}$ and $\boldsymbol{y}_{\ell s}^i=\boldsymbol{0}$ for each $i\in[n_u+n_v]$. By a similar argument as above, there exist $\boldsymbol{y}_{\ell s}^{n_u+n_v+1}$ and $\boldsymbol{y}_{\ell s}^{n_u+n_v+2}$ such that

$$\left(\boldsymbol{q}_{\ell s}^{k1},\boldsymbol{q}_{\ell s}^{k2},\boldsymbol{Q}_{\ell s}^{k3},\boldsymbol{e}_k^\top \boldsymbol{B}\boldsymbol{y}_{\ell s}^{n_u+n_v+1},\boldsymbol{e}_k^\top \boldsymbol{B}\boldsymbol{y}_{\ell s}^{n_u+n_v+2},q_{\ell s}^{k6}\right)\in\mathcal{Q}_{\ell s}, \quad \forall k\in[n_f].$$

It remains to show there exists $\kappa\geq\bar{\kappa}$ such that

$$\boldsymbol{d}^\top \boldsymbol{y}_{\ell s}^0-t_s\leq\left(\kappa-\boldsymbol{d}^\top \boldsymbol{y}_{\ell s}^{n_u+n_v+1}\right)\sigma+\left(\kappa-\boldsymbol{d}^\top \boldsymbol{y}_{\ell s}^{n_u+n_v+2}\right)\nu \;\forall(\boldsymbol{u},\boldsymbol{v},\sigma,\nu)\in\bar{\mathcal{Z}}_{\ell s}.$$

Indeed, it suffices to take

$$\kappa=|\bar{\kappa}|+\max_{s\in[S]}\max_{\ell\in[L]\backslash\{\ell^*(s)\}}\left\{|\boldsymbol{d}^\top \boldsymbol{y}_{\ell s}^{n_u+n_v+1}|+\frac{|\boldsymbol{d}^\top \boldsymbol{y}_{\ell s}^0-t_s|}{\inf_{\boldsymbol{u}\in\mathcal{U}_\ell}\|\boldsymbol{u}-\hat{\boldsymbol{u}}_s\|}+|\boldsymbol{d}^\top \boldsymbol{y}_{\ell s}^{n_u+n_v+2}|\right\}$$

where $\inf_{\boldsymbol{u}\in\mathcal{U}_\ell}\|\boldsymbol{u}-\hat{\boldsymbol{u}}_s\|>0$ for any $s\in[S],\ell\in[L]\backslash\{\ell^*(s)\}$ because $\hat{\boldsymbol{u}}_s\notin\mathrm{cl}(\mathcal{U}_\ell)$. It follows that

$$\begin{aligned}
&\inf_{(\boldsymbol{u},\boldsymbol{v},\sigma,\nu)\in\bar{\mathcal{Z}}_{\ell s}}\left(\kappa-\boldsymbol{d}^\top \boldsymbol{y}_{\ell s}^{n_u+n_v+1}\right)\sigma+\left(\kappa-\boldsymbol{d}^\top \boldsymbol{y}_{\ell s}^{n_u+n_v+2}\right)\nu\\
=&\inf_{\boldsymbol{u}\in\mathcal{U}_\ell,\boldsymbol{v}\in\mathcal{V}_\ell}\left(\kappa-\boldsymbol{d}^\top \boldsymbol{y}_{\ell s}^{n_u+n_v+1}\right)\|\boldsymbol{u}-\hat{\boldsymbol{u}}_s\|+\left(\kappa-\boldsymbol{d}^\top \boldsymbol{y}_{\ell s}^{n_u+n_v+2}\right)\|\boldsymbol{v}-\hat{\boldsymbol{v}}_s\|\\
\geq&\inf_{\boldsymbol{u}\in\mathcal{U}_\ell}\left(\kappa-\boldsymbol{d}^\top \boldsymbol{y}_{\ell s}^{n_u+n_v+1}\right)\|\boldsymbol{u}-\hat{\boldsymbol{u}}_s\|\\
\geq&\inf_{\boldsymbol{u}\in\mathcal{U}_\ell}\frac{|\boldsymbol{d}^\top \boldsymbol{y}_{\ell s}^0-t_s|\cdot\|\boldsymbol{u}-\hat{\boldsymbol{u}}_s\|}{\inf_{\boldsymbol{u}\in\mathcal{U}_\ell}\|\boldsymbol{u}-\hat{\boldsymbol{u}}_s\|}\\
\geq&\,\boldsymbol{d}^\top \boldsymbol{y}_{\ell s}^0-t_s.
\end{aligned}$$

When $n_y=1$, we must have $\boldsymbol{B}>\boldsymbol{0}$ or $\boldsymbol{B}<\boldsymbol{0}$ because of the complete recourse assumption. Without loss of generality, we assume $\boldsymbol{B}=[b_1,...,b_{n_f}]^\top>\boldsymbol{0}$ and $d>0$ to avoid trivial cases. For Problem (12), the recourse function $y_{\ell s}(\boldsymbol{u},\boldsymbol{v},\sigma,\nu)$ must satisfy the constraints

$$\begin{aligned}
&dy_{\ell s}(\boldsymbol{u},\boldsymbol{v},\sigma,\nu)\leq t_s+\kappa\sigma+\kappa\nu && \forall(\boldsymbol{u},\boldsymbol{v},\sigma,\nu)\in\bar{\mathcal{Z}}_{\ell s},\ell\in[L],s\in[S]\\
&y_{\ell s}(\boldsymbol{u},\boldsymbol{v},\sigma,\nu)\geq\max_{k\in[n_f]}\left\{(\boldsymbol{e}_k^\top \boldsymbol{f}(\boldsymbol{v})-\boldsymbol{e}_k^\top \boldsymbol{F}(\boldsymbol{v})\boldsymbol{x}_\ell(\boldsymbol{u}))/b_k\right\} && \forall(\boldsymbol{u},\boldsymbol{v},\sigma,\nu)\in\bar{\mathcal{Z}}_{\ell s},\ell\in[L],s\in[S],
\end{aligned}$$

which indicates the optimal recourse function for any $s \in [S], \ell \in [L]$ would be

$$y_{\ell s}(\boldsymbol{u}, \boldsymbol{v}, \sigma, \nu) = (t_s + \kappa\sigma + \kappa\nu)/d,$$

which is affine in σ and ν and does not depend on $\boldsymbol{u}$ or $\boldsymbol{v}$. Therefore, there exist reduced affine mappings $y_{\ell s}$ that are optimal for Problem (12) and hence solve Problem (17). □

Proof of Theorem 4. Let $\boldsymbol{x}_\ell^\dagger$ be an optimal solution that achieves $Z_{0,\ell}$, for $\ell \in [L]$. For any $(\boldsymbol{u}_s, \boldsymbol{v}_s) \in \mathcal{Z}_\ell$, $s \in [S]$, $\ell \in [L]$, we have

$$\begin{aligned} & g(\boldsymbol{x}_\ell^\dagger(\boldsymbol{u}_s), \boldsymbol{v}_s) - \kappa(\|\boldsymbol{v}_s - \hat{\boldsymbol{v}}_s\| + \|\boldsymbol{u}_s - \hat{\boldsymbol{u}}_s\|) \\ \leq\ & g(\boldsymbol{x}_\ell^\dagger(\hat{\boldsymbol{u}}_s), \hat{\boldsymbol{v}}_s) + (\bar{L} - \kappa)(\|\boldsymbol{u}_s - \hat{\boldsymbol{u}}_s\| + \|\boldsymbol{v}_s - \hat{\boldsymbol{v}}_s\|). \end{aligned}$$

Then, for any $\kappa \geq \bar{L}$, we have

$$\frac{1}{|\mathcal{S}_\ell|} \sum_{s \in \mathcal{S}_\ell} \sup_{(\boldsymbol{u}_s, \boldsymbol{v}_s) \in \mathcal{Z}_\ell} \left\{ g(\boldsymbol{x}_\ell^\dagger(\boldsymbol{u}_s), \boldsymbol{v}_s) - \kappa(\|\boldsymbol{v}_s - \hat{\boldsymbol{v}}_s\| + \|\boldsymbol{u}_s - \hat{\boldsymbol{u}}_s\|) \right\} \leq \frac{1}{|\mathcal{S}_\ell|} \sum_{s \in \mathcal{S}_\ell} g(\boldsymbol{x}_\ell^\dagger(\hat{\boldsymbol{u}}_s), \hat{\boldsymbol{v}}_s) = Z_{0,\ell}.$$

Now, consider any $\boldsymbol{w} \in \mathcal{W}$ and $r_\ell = Z_{0,\ell}$ for $\ell \in [L]$, we have

$$\boldsymbol{w}^\top \boldsymbol{r} - \kappa\theta \|\boldsymbol{w} - \hat{\boldsymbol{w}}\|_1 \leq \hat{\boldsymbol{w}}^\top \boldsymbol{r} + (\|\boldsymbol{r}\|_\infty - \kappa\theta)\|\boldsymbol{w} - \hat{\boldsymbol{w}}\|_1.$$

If κ further satisfies $\kappa \geq \max_{\ell \in [L]} \{|r_\ell|\}/\theta$, we have

$$\sup_{\boldsymbol{w} \in \mathcal{W}} \left\{ \sum_{\ell \in [L]} w_\ell r_\ell - \kappa\theta \|\boldsymbol{w} - \hat{\boldsymbol{w}}\|_1 \right\} \leq \hat{\boldsymbol{w}}^\top \boldsymbol{r} = Z_0 \leq \tau.$$

Hence, it follows that Problem (22) is feasible for some $\kappa_\tau \leq \max\{\bar{L}, \max_{\ell \in [L]} \{|Z_{0,\ell}|\}/\theta\}$. □

Proof of Theorem 5. For any feasible solutions to Problem (21), we have

$$\sum_{\ell \in [L]} w_\ell^\star \left(\mathbb{E}_{\mathbb{P}_\ell^\star}[g(\boldsymbol{x}(\tilde{\boldsymbol{u}}), \tilde{\boldsymbol{v}})] - \kappa_\tau \Delta(\mathbb{P}_\ell^\star, \hat{\mathbb{P}}_\ell) \right) - \kappa_\tau \theta \|\boldsymbol{w}^\star - \hat{\boldsymbol{w}}\|_1 \leq \tau,$$

where $w^\star_\ell = \mathbb{P}^\star[\tilde{\boldsymbol{u}} \in \mathcal{U}_\ell]$ and $\mathbb{P}^\star_\ell$ is the true conditional distribution of $\tilde{\boldsymbol{v}}$ given $\tilde{\boldsymbol{u}} \in \mathcal{U}_\ell$. Hence, for all $\Gamma \geq 0$, we have

$$\begin{aligned}
&\mathbb{P}^S\left[\mathbb{E}_{\mathbb{P}^\star}\left[g(\boldsymbol{x}(\tilde{\boldsymbol{u}}), \tilde{\boldsymbol{v}})\right] > \tau + \kappa_\tau \Gamma\right] \\
&= \mathbb{P}^S\left[\sum_{\ell \in [L]} w^\star_\ell \mathbb{E}_{\mathbb{P}^\star_\ell}\left[g(\boldsymbol{x}(\tilde{\boldsymbol{u}}), \tilde{\boldsymbol{v}})\right] > \tau + \kappa_\tau \Gamma\right] \\
&\leq \mathbb{P}^S\left[\sum_{\ell \in [L]} w^\star_\ell \Delta(\mathbb{P}^\star_\ell, \hat{\mathbb{P}}_\ell) + \theta \|\boldsymbol{w}^\star - \hat{\boldsymbol{w}}\|_1 > \Gamma\right] \\
&\leq \min_{\gamma_1 + \gamma_2 = \Gamma} \left\{\mathbb{P}^S\left[\sum_{\ell \in [L]} w^\star_\ell \Delta(\mathbb{P}^\star_\ell, \hat{\mathbb{P}}_\ell) > \gamma_1\right] + \mathbb{P}^S\left[\theta \|\boldsymbol{w}^\star - \hat{\boldsymbol{w}}\|_1 > \gamma_2\right]\right\} \\
&\leq \min_{\gamma_1 + \gamma_2 = \Gamma} \left\{\mathbb{P}^S\left[\max_{\ell \in [L]} \Delta(\mathbb{P}^\star_\ell, \hat{\mathbb{P}}_\ell) > \gamma_1\right] + \mathbb{P}^S\left[\theta \|\boldsymbol{w}^\star - \hat{\boldsymbol{w}}\|_1 > \gamma_2\right]\right\} \\
&\leq \min_{\gamma_1 + \gamma_2 = \Gamma} \left\{\sum_{\ell \in [L]} \mathbb{P}^{\mathcal{S}_\ell}\left[\Delta(\mathbb{P}^\star_\ell, \hat{\mathbb{P}}_\ell) > \gamma_1\right] + \mathbb{P}^S\left[\theta \|\boldsymbol{w}^\star - \hat{\boldsymbol{w}}\|_1 > \gamma_2\right]\right\} \\
&\leq \min_{\gamma_1 + \gamma_2 = \Gamma} \left\{\sum_{\ell \in [L]} \mathbb{P}^{\mathcal{S}_\ell}\left[\Delta(\mathbb{P}^\star_\ell, \hat{\mathbb{P}}_\ell) > \gamma_1\right] + 2L \exp\left(\frac{-2S\gamma_2^2}{L^2\theta^2}\right)\right\}.
\end{aligned}$$

The last inequality follows from

$$\begin{aligned}
&\mathbb{P}^S\left[\theta \|\boldsymbol{w}^\star - \hat{\boldsymbol{w}}\|_1 > \gamma_2\right] \\
&\leq \mathbb{P}^S\left[|w^\star_\ell - \hat{w}_\ell| > \gamma_2/(\ell\theta) \text{ for some } \ell \in [L]\right] \\
&\leq \sum_{\ell \in [L]} \mathbb{P}^S\left[\left|\frac{\sum_{s \in [S]} \mathbb{1}_{\mathcal{U}_\ell}(\hat{\boldsymbol{u}}_s)}{S} - \mathbb{E}_{\mathbb{P}^S}\left[\frac{\sum_{s \in [S]} \mathbb{1}_{\mathcal{U}_\ell}(\hat{\boldsymbol{u}}_s)}{S}\right]\right| > \frac{\gamma_2}{L\theta}\right] && \text{(Union bound)} \\
&\leq \sum_{\ell \in [L]} 2 \exp\left(-\frac{2S\gamma_2^2}{L^2\theta^2}\right) && \text{(Hoeffding's inequality)} \\
&= 2L \exp\left(-\frac{2S\gamma_2^2}{L^2\theta^2}\right).
\end{aligned}$$

Therefore, for any $\tau \geq Z_0$, $\boldsymbol{x}$ and κ_τ feasible in Problem (21) and $\Gamma \geq 0$, we have

$$\mathbb{P}^S\left[\mathbb{E}_{\mathbb{P}^\star}\left[g(\boldsymbol{x}(\tilde{\boldsymbol{u}}), \tilde{\boldsymbol{v}})\right] > \tau + \kappa_\tau \Gamma\right] \leq \min_{\gamma_1 + \gamma_2 = \Gamma} \left\{\sum_{\ell \in [L]} \mathbb{P}^{\mathcal{S}_\ell}\left[\Delta(\mathbb{P}^\star_\ell, \hat{\mathbb{P}}_\ell) > \gamma_1\right] + 2L \exp\left(\frac{-2S\gamma_2^2}{L^2\theta^2}\right)\right\},$$

for all $\Gamma \geq 0$. □