This article was downloaded by: [2001:da8:d800:95::25] On: 20 November 2025, At: 19:27
Publisher: Institute for Operations Research and the Management Sciences (INFORMS)
INFORMS is located in Maryland, USA

Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Robust Actionable Prescriptive Analytics

Li Chen; , Melvyn Sim; , Xun Zhang; , Long Zhao; , Minglong Zhou

To cite this article:
Li Chen; , Melvyn Sim; , Xun Zhang; , Long Zhao; , Minglong Zhou (2025) Robust Actionable Prescriptive Analytics. Operations Research

Published online in Articles in Advance 30 May 2025

. https://doi.org/10.1287/opre.2023.0300

Full terms and conditions of use: https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2025, INFORMS

Please scroll down for article—it is on subsequent pages

With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

https://pubsonline.informs.org/journal/opre

OPERATIONS RESEARCH
Articles in Advance, pp. 1–22
ISSN 0030-364X (print), ISSN 1526-5463 (online)

Downloaded from informs.org by [2001:da8:d800:95::25] on 20 November 2025, at 19:27 . For personal use only, all rights reserved.

Methods

Robust Actionable Prescriptive Analytics

Li Chen,[a] Melvyn Sim,[b] Xun Zhang,[c,*] Long Zhao,[b] Minglong Zhou[d,*]

[a] Discipline of Business Analytics, The University of Sydney, Sydney, New South Wales 2006, Australia; [b] Department of Analytics & Operations, NUS Business School, National University of Singapore, Singapore 119245; [c] International Institute of Finance, School of Management, University of Science and Technology of China, Hefei 230036, China; [d] Department of Management Science, School of Management, Fudan University, Shanghai 200433, China

*Corresponding authors

Contact: li.chen@sydney.edu.au, https://orcid.org/0000-0002-5370-8518 (LC); dscsimm@nus.edu.sg, https://orcid.org/0000-0001-9798-2482 (MS); xunzhang2023@outlook.com, https://orcid.org/0000-0002-3721-5107 (XZ); longzhao@nus.edu.sg, https://orcid.org/0000-0002-8211-118X (LZ); minglong.zhou@outlook.com, https://orcid.org/0000-0002-3249-571X (MZ)

Received: June 7, 2023
Revised: September 11, 2024; March 5, 2025
Accepted: May 1, 2025
Published Online in Articles in Advance: May 30, 2025

Area of Review: Optimization

https://doi.org/10.1287/opre.2023.0300

Copyright: © 2025 INFORMS

Abstract. 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 safe tractable 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.

Funding: The research of L. Chen was supported by the Emerging Scholar Research Fellowships, University of Sydney Business School. The research of M. Sim and L. Zhao was supported by the Ministry of Education, Singapore under its 2019 Academic Research Fund Tier 3 [Grant MOE-2019-T3-1-010]. The research of X. Zhang was supported by the Anhui Provincial Natural Science Foundation [Grant 2408085QG222], the Fundamental Research Funds for the Central Universities [Grant BJ2040160100], and the National Natural Science Foundation of China [Grant 72025201]. The research of M. Zhou was supported by the National Natural Science Foundation of China [Grants 72301075 and 72293564/72293560].

Supplemental Material: All supplemental materials, including the code, data, and files required to reproduce the results are available at https://doi.org/10.1287/opre.2023.0300.

Keywords: robust optimization • robust satisficing • robust analytics • prescriptive analytics • side information • policy optimization

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), and Loke et al. (2021), among others, have proposed adjustments to the "predict, then optimize" approach by incorporating aspects of the

Downloaded from informs.org by [2001:da8:d800:95::25] on 20 November 2025, at 19:27 . For personal use only, all rights reserved.

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 (Hannah et al. 2010, Bertsimas and Kallus 2020, Esteban-Pérez and Morales 2021, Srivastava et al. 2021). For example, Bertsimas and Kallus (2020) employ side information to reweigh historical samples through pretrained 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) and 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, Mohajerin 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. (2024), Shafieezadeh-Abadeh et al. (2019), Blanchet et al. (2019), and Sim et al. (2021) 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 (Hao et al. 2020, Bertsimas and Van Parys 2021, Kannan et al. 2024, Nguyen et al. 2024, Sim et al. 2024). Zhang et al. (2024) solve the optimal nonparametric 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 the 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, although 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 (Lipton 2018, Bertsimas and Stellato 2021). 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.

In this paper, we propose a general robust actionable prescriptive analytics framework. This framework provides tree-based static and affine mappings from side information to feasible operational decisions and mitigates data uncertainty via robust satisficing and optimization. From a modeling perspective, the proposed model addresses data uncertainty, policy interpretability, and implementability simultaneously in prescriptive analytics. We provide a numerical study on a portfolio optimization problem. We compare the out-of-sample performance of the tree-based affine and forest-based policies on simulated and real data. Our model 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.

1.1. Summary of Contributions to the Literature

- Contributions to data-driven policy optimization:

  i. Our work advances data-driven policy optimization by addressing general linear optimization problems with recourse, distinguishing it from existing literature that often focuses on specific cost functions (Ban and Rudin 2019, Yang et al. 2022, Zhang et al. 2024) or simplified decision scenarios (Bertsimas et al. 2019a). Although some studies address linear optimization with recourse, they either overlook constraint feasibility, leading to nonimplementable policies (Bertsimas and Koduri 2022, Notz and Pibernik 2022), or limit their scope to right-hand-side uncertainty (Bertsimas et al. 2023b), which is not applicable to the portfolio optimization problems we consider. To our knowledge, this is the first framework to learn policies

Downloaded from informs.org by [2001:da8:d800:95::25] on 20 November 2025, at 19:27 . For personal use only, all rights reserved.

that map side information to implementable decisions in such a general two-stage linear optimization context.

ii. Our policy optimization approach not only yields interpretable and implementable policies but also offers a unique advantage in risk-aware optimization. Unlike two-step approaches that optimize risk over the conditional distribution of outcome variables given the realization of side information—potentially leading to suboptimal results because of the exclusion of the side information distribution—our method optimizes risk over the joint distribution of side information and outcome variables. This distinction is highlighted in our portfolio optimization example in Section 6.

- Contributions to robust optimization and satisficing:

i. Our work significantly extends the existing robust satisficing framework (Long et al. 2023) to the context of prescriptive analytics to incorporate side information. We focus on a general robust actionable prescriptive analytics framework, solving for an actionable response policy that maps side information directly to feasible operational decisions.

ii. The proposed model presents new computational challenges that have not been sufficiently addressed in the literature. We develop a novel, tractable, safe approximation for robust constraints involving bilinear uncertainty and general norm-based uncertainty sets (Theorem 2) and discuss its relation to existing techniques (Remark 1). We establish new feasibility and optimality results associated with the approximation (Theorem 3), generalizing existing findings (Remark 2), and generating new insights (Remark 3). Additionally, we introduce a biaffine recourse adaptation to enhance approximation quality.

iii. We propose a localized robust satisficing model as a more computationally accessible alternative to our primary model. This localized model, when implemented with a tree-based static policy, maintains computational efficiency even in combinatorial optimization scenarios. We also establish new feasibility results (Theorem 4) and finite sample guarantees (Theorem 5) for the localized model and test its numerical performance.

1.2. Notation

We denote by $\mathbb{R}$ ($\mathbb{R}_+$) the set of real (nonnegative) 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 $\mathrm{diag}(\boldsymbol{\theta}) \in \mathbb{R}^{n\times n}$ 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 $\mathcal{Z}$, that is, $\mathbb{1}_{\mathcal{Z}}(\boldsymbol{x})$ equals one if $\boldsymbol{x} \in \mathcal{Z}$ and zero otherwise. We denote by $\mathrm{cl}(\mathcal{X})$ and $\mathrm{int}(\mathcal{X})$ 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 subclass 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 establish the prescriptive analytics framework, we model the data stream as random variables governed by 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*, with support sets $\mathcal{U} \subseteq \mathbb{R}^{n_u}$ and $\mathcal{V} \subseteq \mathbb{R}^{n_v}$, respectively. The side information has predictive power over outcome variables. 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 independent of the decision-making process.

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 first argument represents the *here-and-now response decision*, whereas the second argument represents the outcome variable, which captures the model's uncertainty. The decision-making process follows a sequential structure: We first observe the side information and then determine the response decision. In this framework, the output of the prescriptive analytics model is the *response policy*, $\boldsymbol{x} \in \mathcal{X} \subseteq \mathcal{R}^{n_u,n_x}$, which maps observed side information to response decisions. The response policy is selected over the *actionable policy set*, $\mathcal{X}$, which is designed to ensure interpretability, maintain feasibility under model constraints, and preserve the computational tractability of the optimization problem. Note that this framework parallels prediction models such as regression: The cost function aligns with the loss function, whereas the response policy corresponds to the *target function*, optimized over the *hypothesis set*. In this analogy, the hypothesis set plays

Downloaded from informs.org by [2001:da8:d800:95::25] on 20 November 2025, at 19:27 . For personal use only, all rights reserved.

a similar role as the actionable policy set in facilitating the interpretation of the target function.

In an ideal setting with full information and infinite computational resources, we would solve for optimal response policy in the following optimization problem:

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

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 $\hat{\mathbb{P}}[\tilde{\boldsymbol{z}} = \hat{\boldsymbol{z}}_s] = 1/S$, $s \in [S]$. 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*:

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

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

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

2.1. Actionable Policy Set

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

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

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 subclass 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 realizations 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 because 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). 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. Although 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 nonnegative. 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.

2.2. 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 nonoverlapping 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

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

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

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

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 subclass of tree-based static mappings as follows:

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

Downloaded from informs.org by [2001:da8:d800:95::25] on 20 November 2025, at 19:27 . For personal use only, all rights reserved.

and the subclass 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 subclass 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 subclass 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 optimization problem:

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

The subclass of tree-based affine mappings, that is, $\mathcal{A} = \bar{\mathcal{T}}^{n_u,n_x}$ 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 optimization problem:

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

Because 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.

2.3. Constructing a Tree from Data

To properly define the tree-based policy, we must prescribe a tree structure. The tree structure then serves as the input to a downstream robust satisficing model, where we optimize for a robust tree-based response policy. Therefore, constructing a tree is a preliminary step of the robust prescriptive analytics framework. 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, that is, 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 $[\underline{u}_{\ell i}, \overline{u}_{\ell i}]$, and

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

with $\mathcal{U}_{\ell i}(\omega) \triangleq \{\boldsymbol{u} \in \mathcal{U}_\ell \,|\, u_i \le \omega\}$ and $\mathcal{S}_{\ell i}(\omega) \triangleq \{s \in \mathcal{S}_\ell \,|\, \hat{\boldsymbol{u}}_s \in \mathcal{U}_{\ell,i}(\omega)\}$. 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 $\zeta_{\ell i}(\omega) \le \zeta_{\ell i}(\infty)$ for all $\omega \in [\underline{u}_{\ell i}, \overline{u}_{\ell i}]$. 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 $\mathcal{U}_{\Lambda+1} \leftarrow \mathcal{U}_{\ell^*}\setminus\mathcal{U}_{\ell^* i^*}(\omega_{\ell^* i^*}^*)$
   - 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 $\ell \in [L]$.

Similar to CART, we can impose certain requirements for splitting. For instance, we might require that the number of observations in a leaf exceed a specific threshold before it can be split. These requirements only impact the search space of (ℓ^*, i^*), and we omit them here to maintain the readability of the algorithm.

We can adopt a cross-validation procedure to determine the desired number of leaf nodes. Let $\bar{L}$ be the

Downloaded from informs.org by [2001:da8:d800:95::25] on 20 November 2025, at 19:27 . For personal use only, all rights reserved.

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 nonoverlapping testing data sets. For each $k \in [K]$, the kth training data set comprises the historical data without the kth 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. For interpretability and practicality, we do not wish to grow a complex tree guiding the operational policy while having too few observations in any leaf node and overfitting to the samples in this step. Therefore, the maximum number of leaf nodes, $\bar{L}$, should not be large. In our numerical experiment, we set $\bar{L} = 4$ and observe that this number is sufficiently large for obtaining satisfactory out-of-sample performance.

The heuristic approach we propose here is one of many ways one can grow a tree. For example, one may adapt the optimal classification tree approach (Bertsimas and Dunn 2017, Aghaei et al. 2024) to construct a tree that achieves the best in-sample average (regularized) objective. However, we remark that in-sample optimality of the tree does not necessarily imply better performance of the downstream robust optimization models. In our numerical study, we also test how varying the granularity of the set $\Omega_{\ell i}$ and constructing the tree using mixed-integer optimization would affect the model performance and total computation time (see Online Appendix B.2). The results reassure that our heuristic approach is efficient and effective.

3. Robust Optimization and Satisficing

It is 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, Mohajerin 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:

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

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

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

and

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

The confidence guarantees that the true distribution, $\mathbb{P}^\star$, resides within the Wasserstein-based ambiguity set, $\mathcal{F}(\Gamma)$, 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 toward zero as the sample size, S, increases. Mohajerin 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. (2019), Shafieezadeh-Abadeh et al. (2019), Si et al. (2020), and Gao (2023), because 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 are 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 Mohajerin Esfahani and Kuhn (2018), we can reformulate Problem (6) as a robust optimization problem:

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

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:

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

Downloaded from informs.org by [2001:da8:d800:95::25] on 20 November 2025, at 19:27 . For personal use only, all rights reserved.

or equivalently as the following robust optimization problem:

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

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*

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

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 Online 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 $\boldsymbol{x}$ of Problem (9), which is inherently random as it depends on the independent samples $\{(\hat{\boldsymbol{u}}_s,\hat{\boldsymbol{v}}_s), s\in[S]\}$, satisfies the following:*

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

Here, $\mathbb{P}^S$ is the distribution that governs the independent samples $(\hat{\boldsymbol{u}}_s,\hat{\boldsymbol{v}}_s)$, $s\in[S]$.

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.

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). When applied to regression and classification problems, Sim et al. (2021) demonstrate with well-known data sets that the robust satisficing approach performs well when the target parameter is determined via cross-validation. The process of tuning the target parameter has been also discussed in Sim et al. (2024), where the target parameter is determined by a K-fold cross-validation procedure to improve out-of-sample performance. Henceforth, in the rest of the paper, we focus on the robust satisficing model (9) as in Chen et al. (2025), although all the development also applies to the distributionally robust optimization model (7).

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

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

To obtain a tractable exact formulation, we consider the case where g is a saddle function, that is, $g(\boldsymbol{x},\boldsymbol{v})$ 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, that is, $\mathcal{A} = \mathcal{T}^{n_x,n_u}$, 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, that is, $\mathcal{A} = \bar{\mathcal{T}}^{n_x,n_u}$, 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. We summarize in Table 1 the conditions in robust prescriptive analytics models.

Table 1. Conditions in Robust Prescriptive Analytics Models

	Elements in constraints			
Policy $\mathcal{A}$	Map $\boldsymbol{A}$	Map $\boldsymbol{b}$	Domain $\mathcal{D}$	Cost function $g(\boldsymbol{x},\boldsymbol{v})$
$\mathcal{T}^{n_u,n_x}$	$\mathcal{L}^{n_u,n_b\times n_x}$	$\mathcal{L}^{n_u,n_b}$	Conic Representable	Saddle function Linear optimization with complete recourse (11)
$\bar{\mathcal{T}}^{n_u,n_x}$	$\mathbb{R}^{n_b\times n_x}$	$\mathcal{L}^{n_u,n_b}$	$\mathbb{R}^{n_x}$	Linear optimization with complete recourse (11)

Downloaded from informs.org by [2001:da8:d800:95::25] on 20 November 2025, at 19:27 . For personal use only, all rights reserved.

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 in the form of a linear optimization problem with fixed recourse as follows:

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

where $\boldsymbol{F} \in \mathcal{L}^{n_v, n_f \times n_x}$ and $\boldsymbol{f} \in \mathcal{L}^{n_v, n_f}$. Because 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*; that is, for any $\boldsymbol{w} \in \mathbb{R}^{n_f}$, there exists a $\boldsymbol{y} \in \mathbb{R}^{n_y}$ such that $\boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{w}$.

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

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

where the lifted uncertainty sets are defined as

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

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 $\boldsymbol{v}$, that is, $\boldsymbol{F}(\boldsymbol{v}) = \boldsymbol{F}^0$. 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) and 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

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

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

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

or equivalently

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

where $\boldsymbol{P} = \operatorname{diag}(\boldsymbol{p})$ and $\boldsymbol{C} = \frac{1}{n_x}\mathbf{1}\boldsymbol{c}^\top$. 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:

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

Downloaded from informs.org by [2001:da8:d800:95::25] on 20 November 2025, at 19:27 . For personal use only, all rights reserved.

which can be easily implemented in RSOME developed by Chen et al. (2020). Because 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 safe tractable approximation, we first consider the following lifted affine recourse adaptation (Bertsimas et al. 2019b, Chen et al. 2020) as follows:

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

However, Problem (14) remains intractable because of the bilinear uncertainty in $\boldsymbol{v}$ and $\boldsymbol{u}$ in some of the robust constraints. Although there are safe tractable approximations for robust constraints with bilinear uncertainty under the assumptions of at least one of the uncertainty sets being polyhedral (Zhen et al. 2022a, b; de Ruiter et al. 2023), 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 $\bar{Z}_{\ell s}$ can be general norms. To our knowledge, existing literature on multistage 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 safe approximation for Problem (14) in Theorem 2 by leveraging the structure of $\bar{Z}_{\ell s}$ and the positive homogeneity of the norms.

Theorem 2. *The robust constraint*

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

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

$$
\begin{aligned}
\mathcal{Q}_{\ell s} \triangleq \Bigg\{ (\boldsymbol{q}_1,\boldsymbol{q}_2,\boldsymbol{Q}_3, & q_4,q_5,q_6) \;\Bigg|\; (\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 (\bar{\boldsymbol{v}}_\ell - \hat{\boldsymbol{v}}_s) \\
& \quad + (\underline{\boldsymbol{p}} + \underline{\boldsymbol{P}}\hat{\boldsymbol{u}}_s)^\top (\hat{\boldsymbol{v}}_s - \underline{\boldsymbol{v}}_\ell) + \overline{\boldsymbol{\lambda}}^\top (\bar{\boldsymbol{u}}_\ell - \hat{\boldsymbol{u}}_s) \\
& \quad + \underline{\boldsymbol{\lambda}}^\top (\hat{\boldsymbol{u}}_s - \underline{\boldsymbol{u}}_\ell) \le q_6 \\
& \|\boldsymbol{q}_1 + \overline{\boldsymbol{P}}^\top (\bar{\boldsymbol{v}}_\ell - \hat{\boldsymbol{v}}_s) + \underline{\boldsymbol{P}}^\top (\hat{\boldsymbol{v}}_s - \underline{\boldsymbol{v}}_\ell) \\
& \quad + \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}\bar{\boldsymbol{u}}_\ell - (\boldsymbol{\Theta} - \boldsymbol{Q}_3 + \overline{\boldsymbol{P}} - \underline{\boldsymbol{P}})\underline{\boldsymbol{u}}_\ell\|_* \\
& \quad + (\bar{\boldsymbol{u}}_\ell - \underline{\boldsymbol{u}}_\ell)^\top \boldsymbol{\theta} \le q_5 \\
& \sum_{i\in[n_u]} \|\boldsymbol{\Theta}\boldsymbol{e}_i\|_* \boldsymbol{e}_i \le \boldsymbol{\theta} \\
& \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}}\bar{\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}}\bar{\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{aligned}
\Bigg\}. \tag{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* $(\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{Q}_3, q_4, q_5, q_6) \in \mathcal{Q}_{\ell s}$.

c. *The set* $\mathcal{Q}_{\ell s}$ *is a convex cone, that is, for any* $(\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{Q}_3, q_4, q_5, q_6) \in \mathcal{Q}_{\ell s}$, $(\boldsymbol{q}_1', \boldsymbol{q}_2', \boldsymbol{Q}_3', q_4', q_5', q_6') \in \mathcal{Q}_{\ell s}$, *and* $\alpha, \beta \in \mathbb{R}_+$, *we have* $\alpha \cdot (\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{Q}_3, q_4, q_5, q_6) + \beta \cdot (\boldsymbol{q}_1', \boldsymbol{q}_2', \boldsymbol{Q}_3', q_4', q_5', q_6') \in \mathcal{Q}_{\ell s}$.

Proof. The proof is relegated to Online Appendix A.

Remark 1. We remark that when one of the uncertainty sets is polyhedral, one can apply techniques from existing literature (Zhen et al. 2022b, de Ruiter et al. 2023) to derive safe tractable approximations that are no worse than the result in Theorem 2. This is because the proposed approximation in Theorem 2 is a general result that does not leverage the explicit polyhedral structure of the uncertainty set. Nonetheless, if the polyhedral structure is known, we can easily modify the proposed approximation in Theorem 2 to match the existing approximation (Bertsimas and de Ruiter 2016, Zhen et al. 2022b).

Downloaded from informs.org by [2001:da8:d800:95::25] on 20 November 2025, at 19:27 . For personal use only, all rights reserved.

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

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

Observe that Property (a) of $\mathcal{Q}_{\ell s}$ in Theorem 2 implies that 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 implications on the feasibility of the approximate robust satisficing problem when a reasonable target $\tau \ge Z_0$ is chosen, as we present in the following theorem. Property (c) in Theorem 2 is a natural characteristic of the set $\mathcal{Q}_{\ell s}$, as any conical combination of robust constraints in (15) must also remain valid.

Theorem 3. *Suppose Assumption 1 holds, Problem (11) has complete recourse, and $\hat{\boldsymbol{u}}_s \in \mathrm{int}(\mathcal{U}_\ell)$ for each $\ell\in[L]$, $s\in\mathcal{S}_\ell$. Then for any $\tau \ge Z_0$, 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 $d \ne 0$, then the reduced affine mappings of the form

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

are optimal in Problems (12), (14), and (17).

Proof. The proof is relegated to Online Appendix A.

Remark 2. 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 $\tau \ge Z_0$ for the decision maker, which is important for adjusting robustness.

Remark 3. 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) and Long et al. (2023) because their corresponding recourse function has the following two-stage representation:

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

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 $\boldsymbol{d} \ne \boldsymbol{0}$ then we can formulate

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

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 $d_r \ne 0$ (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,

$$\mathbb{C}^\epsilon_{\mathbb{P}}[\tilde{\boldsymbol{v}}^\top \boldsymbol{x}(\tilde{\boldsymbol{u}})],$$

where CVaR is defined as

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

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

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

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

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

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:

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

Downloaded from informs.org by [2001:da8:d800:95::25] on 20 November 2025, at 19:27 . For personal use only, all rights reserved.

Observe that, because $\kappa \geq 0$,

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

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

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

Here, 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)). Because the purpose of the recourse adaptation mapping is purely computational, it would not be necessary to consider its interpretability.

4.1. 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 (Ben-Tal et al. 2004, Chen et al. 2008, Goh and Sim 2010, Kuhn et al. 2011, Georghiou et al. 2015, Zhen et al. 2018, Bertsimas et al. 2019b, Chen and Sim 2024), 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 subclass of bilinear mappings:

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

Accordingly, we consider lifted biaffine recourse adaptation as follows:

$$\begin{aligned} \min \; & \kappa \\ \text{s.t.} \; & \frac{1}{S}\sum_{s\in[S]} t_s \leq \tau \\ & \boldsymbol{d}^\top(\boldsymbol{y}_{\ell s}(\boldsymbol{u},\boldsymbol{v},\sigma,\nu) + \bar{\boldsymbol{y}}_{\ell s}(\boldsymbol{u},\boldsymbol{v})) - \kappa(\sigma+\nu) \leq t_s \\ & \qquad \forall (\boldsymbol{u},\boldsymbol{v},\sigma,\nu) \in \bar{\mathcal{Z}}_{\ell s}, \ell \in [L], s \in [S] \\ & \boldsymbol{F}(\boldsymbol{v})\boldsymbol{x}_\ell(\boldsymbol{u}) + \boldsymbol{B}(\boldsymbol{y}_{\ell s}(\boldsymbol{u},\boldsymbol{v},\sigma,\nu) + \bar{\boldsymbol{y}}_{\ell s}(\boldsymbol{u},\boldsymbol{v})) \geq \boldsymbol{f}(\boldsymbol{v}) \\ & \qquad \forall (\boldsymbol{u},\boldsymbol{v},\sigma,\nu) \in \bar{\mathcal{Z}}_{\ell s}, \ell \in [L], s \in [S] \\ & \boldsymbol{x} \in \mathcal{X}, \boldsymbol{t} \in \mathbb{R}^S, \kappa \geq 0 \\ & \boldsymbol{y}_{\ell s} \in \mathcal{L}^{n_u+n_v+2,n_y}, \bar{\boldsymbol{y}}_{\ell s} \in \mathcal{B}^{n_u,n_v,n_y} \qquad \forall \ell \in [L], s \in [S]. \end{aligned} \tag{19}$$

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 $n_y \geq 2$. For instance, suppose the cost function is given by

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

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

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

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

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

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

Downloaded from informs.org by [2001:da8:d800:95::25] on 20 November 2025, at 19:27 . For personal use only, all rights reserved.

$$
\left.
\begin{aligned}
&\boldsymbol{q}_{\ell s}^{k1} = -\sum_{i\in[n_u]} \boldsymbol{e}_i \boldsymbol{e}_k^\top (\boldsymbol{F}^0 \boldsymbol{x}_\ell^i + \boldsymbol{B}\boldsymbol{y}_{\ell s}^i)\\
&\boldsymbol{q}_{\ell s}^{k2} = \sum_{j\in[n_v]} \boldsymbol{e}_j \boldsymbol{e}_k^\top (\boldsymbol{f}^j - \boldsymbol{F}^j \boldsymbol{x}_\ell^0 - \boldsymbol{B}\boldsymbol{y}_{\ell s}^{n_u+j})\\
&\boldsymbol{Q}_{\ell s}^{k3} = -\sum_{i\in[n_u]}\sum_{j\in[n_v]} \boldsymbol{e}_j \boldsymbol{e}_k^\top (\boldsymbol{F}^j \boldsymbol{x}_\ell^i + \boldsymbol{B}\bar{\boldsymbol{y}}_{\ell s}^{ij}) \boldsymbol{e}_i^\top\\
&q_{\ell s}^{k4} = \boldsymbol{e}_k^\top \boldsymbol{B}\boldsymbol{y}_{\ell s}^{n_u+n_v+1}\\
&q_{\ell s}^{k5} = \boldsymbol{e}_k^\top \boldsymbol{B}\boldsymbol{y}_{\ell s}^{n_u+n_v+2}\\
&q_{\ell s}^{k6} = \boldsymbol{e}_k^\top (\boldsymbol{F}^0 \boldsymbol{x}_\ell^0 + \boldsymbol{B}\boldsymbol{y}_{\ell s}^0 - \boldsymbol{f}^0)
\end{aligned}
\right\} \forall \ell \in [L], s \in [S], k \in [n_f]
$$

$$
\begin{aligned}
&\boldsymbol{x} \in \mathcal{X}, \boldsymbol{t} \in \mathbb{R}^S, \kappa \geq 0\\
&\boldsymbol{y}_{\ell s} \in \mathcal{L}^{n_u+n_v+2, n_y}, \bar{\boldsymbol{y}}_{\ell s} \in \mathcal{B}^{n_u, n_v, n_y} \qquad \forall \ell \in [L], s \in [S].
\end{aligned}
$$

5. 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:

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

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):

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

Note that the number of constraints of Problem (22) is reduced to the magnitude of $S = \sum_{\ell\in[L]} |\mathcal{S}_\ell|$. The reformulation procedure follows the approximations developed in Section 4.

Comparing Problem (22) with Problem (10), the key distinction is the way we handle possible deviations of historical samples. In the global model, Problem (10), we allow any sample to deviate over the entire support and fall in other leaves. The model accounts for potential shifts in leaf weights from the empirical one by explicitly considering the worst-case deviations of the samples. In the localized model, Problem (22), we introduce a weight vector $\boldsymbol{w}$ and model the uncertainty via a mixture model where the random side information and outcomes in each leaf node follow one probability distribution. We account for possible shifts in leaf weights by accounting for the uncertainty in weight vector $\boldsymbol{w}$. Under this model of uncertainty, we simplify the robust counterpart by only allowing a sample to deviate over its dedicated leaf. This effectively separates the main problem into smaller pieces that calculate the worst-case (regularized) expected cost for each leaf, denoted as r_ℓ, $\ell \in [L]$. Then we reweight these costs and hedge against potential distribution shifts via $\boldsymbol{w} \in \mathcal{W}$. Problems (10) and (22) are equivalent when $L = 1$. The localized model is a simplification for the purpose of enhancing tractability facing large-scale and computationally challenging problems. In our numerical study, we demonstrate that the localized model does improve computational time at the cost of a certain level of out-of-sample performance.

To establish the feasibility of Problem (22), we focus on the 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 $\tau \geq Z_0$ for some $\kappa_\tau \leq \max\{\bar{L}, \max_{\ell\in[L]}\{|Z_{0,\ell}|\}/\theta\}$.*

Proof. The proof is relegated to Online Appendix A.

Theorem 4 closely follows Theorem 1. The only difference is the objective value κ_τ is also affected by the hyper-parameter θ, which represents our sensitivity to deviations of $\boldsymbol{w}$ from $\hat{\boldsymbol{w}}$. When $\tau = Z_0$, the optimal solution to the localized model is also optimal to the corresponding empirical optimization problem.

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 $\tau \geq Z_0$ and $\Gamma \geq 0$, the optimal solution $\boldsymbol{x}$ of*

Downloaded from informs.org by [2001:da8:d800:95::25] on 20 November 2025, at 19:27 . For personal use only, all rights reserved.

Problem (21)*, which is inherently random as it depends on the independent samples* $\{(\hat{\boldsymbol{u}}_s, \hat{\boldsymbol{v}}_s), s \in [S]\}$*, satisfies*

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

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 Online Appendix A.

The finite sample guarantee and insights in Theorem 5 is consistent with Proposition 1. Specifically, Theorem 5 demonstrates that the theoretical likelihood limit diminishes exponentially toward 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. There are also differences between Theorem 5 and Proposition 1. The right-hand side of the probability bound in Theorem 5 illustrates that target violation is caused by a joint effect of the two uncertainties in the mixture model. The first term accounts for the deviation between the true probability distribution $\mathbb{P}^*_\ell$ and the localized empirical distribution $\hat{\mathbb{P}}_\ell$ for each leaf node ℓ. The second term considers the deviation between mixture weights $\boldsymbol{w}^*$ and $\hat{\boldsymbol{w}}$ and characterizes a probability upper bound. We take an infimum convolution to tighten the bound for any magnitude of target violation parameterized by Γ, leading to a novel bound to the robust satisficing literature.

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 data set.

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,

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

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:

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

where the cost function is $g(\boldsymbol{x}, \boldsymbol{v}) = \sum_{n \in [n_x]} c_n v_n x_n$, 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

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

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:

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

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 nonincreasing 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. On the contrary, the global robust satisficing model (9) cannot be reduced to such a form, and consequently it cannot be solved efficiently even if the nominal combinatorial optimization problem is polynomial time solvable.

Downloaded from informs.org by [2001:da8:d800:95::25] on 20 November 2025, at 19:27 . For personal use only, all rights reserved.

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. The main model in our numerical study is the robust satisficing model with side information under a tree-based affine response policy. We adopt the tractable approximation derived from Theorem 2, that is, Problem (18). In the main text, we adopt ℓ_1-norm in the robust satisficing model for simplicity. We relegate additional comparisons between ℓ_1-norm and ℓ_2-norm to Online Appendix B.3. We denote the solution to this model for a tree with L leaf nodes as "RTGlobal" policy. We also refer to the solution obtained from the corresponding localized tree-based model, Problem (22), as the "RTLocal" policy. We set the target τ in Model (18) and corresponding localized tree-based model as $Z_0^{\text{Affine}}+\alpha$, where $\alpha>0$ denotes a relative target margin, and Z_0^{Affine} is the optimal value of the following empirical optimization problem:

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

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:

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

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

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

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

6.1. Simulated Data

In this section, we conduct numerical experiments based on the simulated data. The results demonstrate the benefits of adopting our proposed model. We also draw additional numerical experiments to illustrate the efficiency of our proposed localized model.

6.1.1. Baseline Setting. 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:

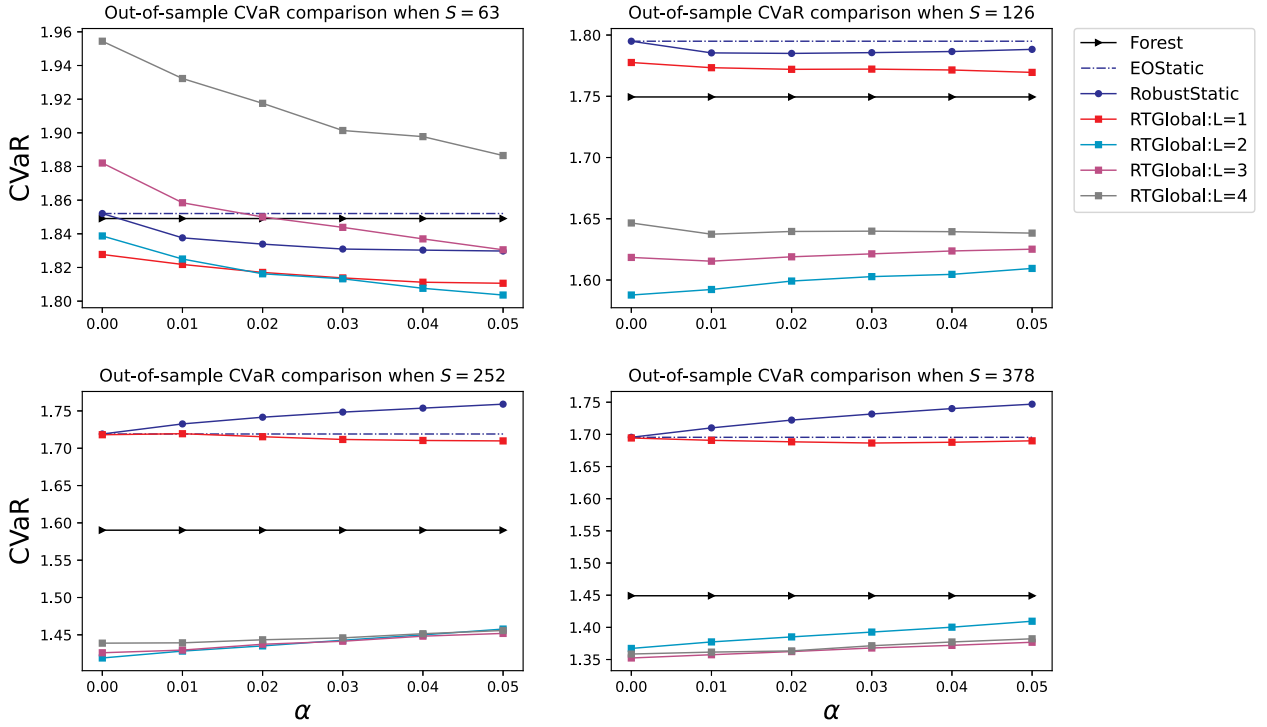
$$\begin{aligned} \tilde{v}_1 &= 1+0.2\exp(\tilde{u}_1) - LN(0,1-0.5\mathbb{1}_{[-3,-1]}(\tilde{u}_2)) \\ \tilde{v}_2 &= 1-0.2\tilde{u}_1 - LN(0,1-0.5\mathbb{1}_{[-1,1]}(\tilde{u}_2)) \\ \tilde{v}_3 &= 1+0.2|\tilde{u}_1| - LN(0,1-0.5\mathbb{1}_{[1,3]}(\tilde{u}_2)), \end{aligned}$$

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

To mimic daily returns over the past 3 months, 6 months, 1 year, or 1.5 years, 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 four. We choose a small number to make the tree structure interpretable. We train the tree in our model based on the algorithm described in Section 2. We compute the out-of-sample CVaR for each policy using the test set $\{\hat{\tilde{\boldsymbol{u}}}_s,\hat{\tilde{\boldsymbol{v}}}_s\}_{s\in[10000]}$. We repeat this procedure 100 times and report the average results in Figure 1.

We observe that the RTGlobal policy with only one leaf node performs the best when the training size is small ($S=63$) because it avoids overfitting the limited data. However, as the training size increases, there is little improvement in the performance of the RTGlobal policy with only one leaf node, suggesting underfitting when the training size is relatively large ($S\ge 126$). In

Figure 1. (Color online) CVaR Comparison Under the Setting of Kallus and Mao (2023)



contrast, the performance of the RTGlobal policy with more than one leaf node improves significantly as the number of training samples increases. When the training size is relatively large ($S \geq 126$), the RTGlobal policy with more than one leaf node outperforms the RobustStatic policy, demonstrating the benefits of using side information. Perhaps surprisingly, the the RTGlobal policy with more than one leaf node 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, even when $\alpha = 0$. 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.

6.1.2. 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 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

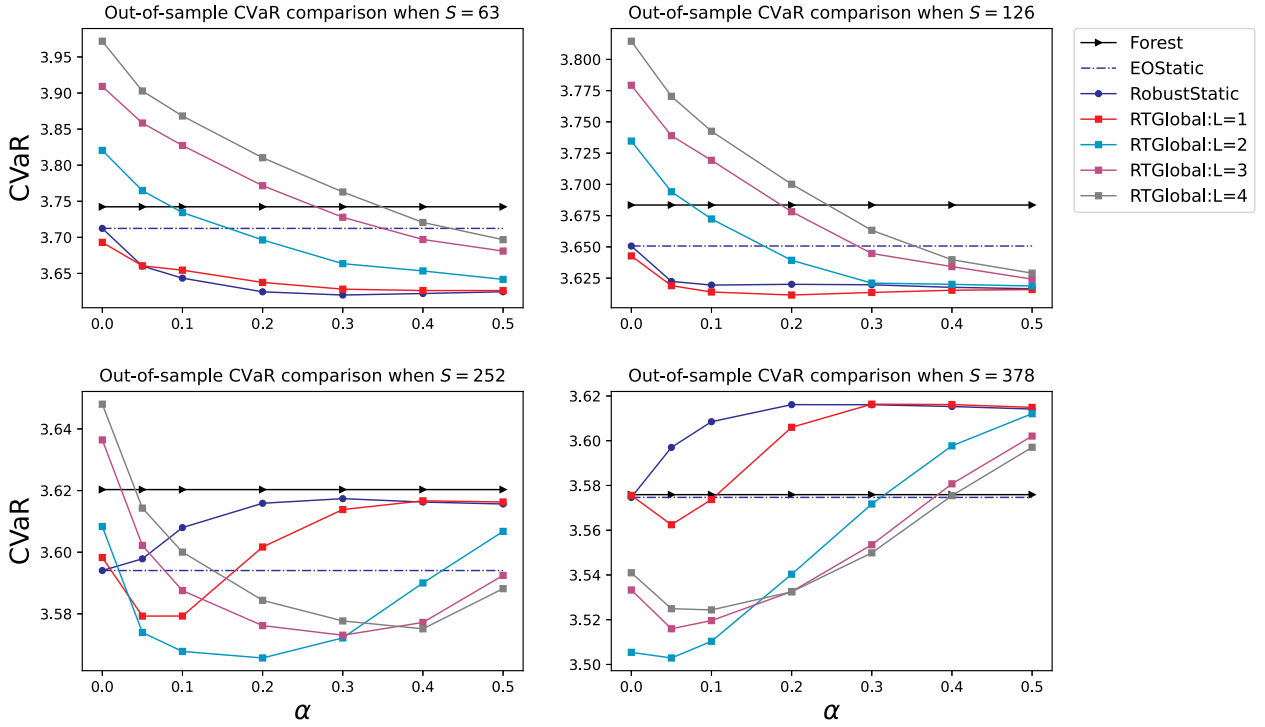
$$\tilde{v}_1 = 1 + 0.2\exp(\tilde{u}_1) - LN(0, 1 - 0.5\mathbb{1}_{[-3,-1]}(\tilde{u}_2)) - LN(0, 0.8)$$
$$\tilde{v}_2 = 1 - 0.2\tilde{u}_1 - LN(0, 1 - 0.5\mathbb{1}_{[-1,1]}(\tilde{u}_2)) - LN(0, 0.8)$$
$$\tilde{v}_3 = 1 + 0.2|\tilde{u}_1| - LN(0, 1 - 0.5\mathbb{1}_{[1,3]}(\tilde{u}_2)) - LN(0, 0.8).$$

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 RobustStatic and RTGlobal 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 RTGlobal with only one leaf node, outperform complex policies when the training size is relatively small ($S \leq 126$). 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

Downloaded from informs.org by [2001:da8:d800:95::25] on 20 November 2025, at 19:27 . For personal use only, all rights reserved.

Figure 2. (Color online) CVaR Comparison in a Low Signal-Noise Ratio Setting



observation indicates that RTGlobal with more than one leaf node can better utilize side information as the training size increases.

We also provide a summary of the computation time for the RTGlobal model to investigate its scalability in Table 2. As anticipated, the computation time increases with the sample size (S), the number of stocks (n_x), the number of covariates (n_u), and the number of leaves (L). Although the time required to solve the original robust satisficing model with tree-based affine policies for the case ($S = 378, n_x = 12, n_u = 10, L = 4$) is somewhat long, it remains under six minutes.

6.1.3. Comparison Between RTGlobal and RTLocal Policies. In Section 5, we propose a localized tree-based model for the purpose of enhancing computational efficiency. This model treats each leaf separately and evaluates localized robust counterparts rather than a global one, potentially compromising out-of-sample performance. Here, we present comparisons between RTGlobal and RTLocal policies to elucidate the tradeoff between computation time and model performance. RTLocal becomes equivalent to RTGlobal when $L = 1$; thus, we only consider cases where $L \geq 2$. In the RTLocal model, we set parameter $\theta \in \{1, 10\}$. We present results with $\theta = 1$ in the main text and relegate the results with $\theta = 10$ to Online Appendix B.1.

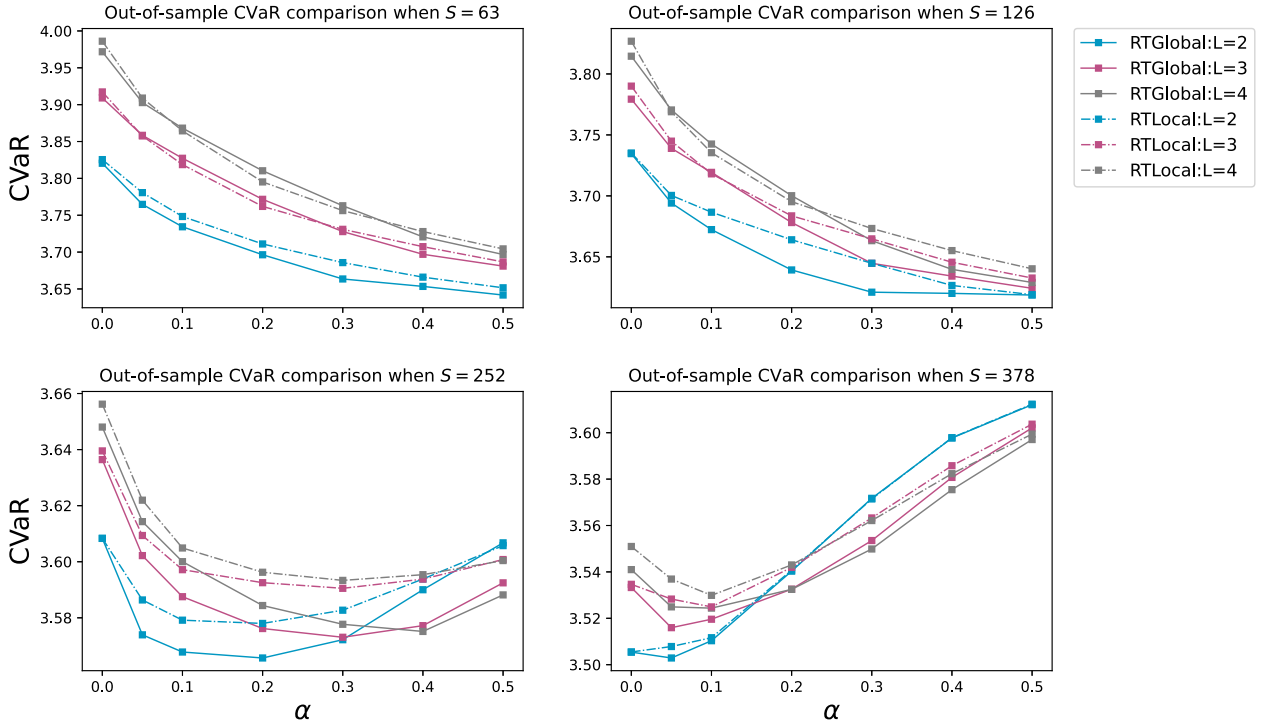
Figure 3 presents the out-of-sample CVaR of RTGlobal and RTLocal policies. For a fixed number of leaves, L, RTGlobal policies generally outperform RTLocal policies in the sense that the performance curves of the RTLocal model mostly lie above those of the RTGlobal model. This indicates that RTGlobal model achieves better robustness for the same parameter α, that is, RTGlobal obtains a lower out-of-sample CVaR. From another angle, RTGlobal sets a lower parameter α than RTLocal when both models achieve the same

Table 2. Computational Time (s) of the RTGlobal Model on Large Instances When $\alpha = 0.2$

		$L = 2$			$L = 3$			$L = 4$		
Sample size	No. stocks	$n_u = 2$	$n_u = 6$	$n_u = 10$	$n_u = 2$	$n_u = 6$	$n_u = 10$	$n_u = 2$	$n_u = 6$	$n_u = 10$
$S = 252$	$n_x = 6$	1.23	4.60	14.26	2.08	8.13	24.88	4.21	18.44	58.20
	$n_x = 9$	1.93	7.62	32.79	3.60	13.50	59.70	8.31	26.29	121.42
	$n_x = 12$	2.88	12.05	50.00	5.13	20.09	90.56	10.21	33.36	150.11
$S = 378$	$n_x = 6$	1.94	8.01	25.43	3.39	15.81	50.83	6.53	25.28	154.83
	$n_x = 9$	3.20	15.22	61.06	5.49	26.32	107.06	11.96	51.79	209.94
	$n_x = 12$	4.88	24.43	101.15	8.34	37.87	179.80	21.85	58.18	320.84

Downloaded from informs.org by [2001:da8:d800:95::25] on 20 November 2025, at 19:27 . For personal use only, all rights reserved.

Figure 3. (Color online) CVaR Comparison Between RTGlobal and RTLocal Policies When $\theta = 1$



out-of-sample CVaR, leading to a smaller optimality gap in empirical CVaR. These observations align with the intuition behind the two models.

The performance difference between RTGlobal and RTLocal policies depends on the samples. When there are few samples, the true distribution is likely to deviate much from the empirical distribution. Consequently, policies obtained from both models would perform poorly, and the difference is marginal. When there are sufficient samples, the empirical optimization model tends to be optimal, and robustness hardly leads to improvements. Then, the difference between RTGlobal and RTLocal models will be insignificant. In this numerical experiment, the difference is noticeable when we have a medium sample size, $S = 252$.

RTLocal provides a reasonable approximation of RTGlobal. The immediate question to investigate is the improvement in computation times. We record the computation times of the experiments presented in Figure 3, which are conducted on a Macbook Pro with 16 GB memory. All RTGlobal and RTLocal policies are solved using the Gurobi solver 11.0.2 in Python. The computation times are summarized in Table 3. The hyper-parameter α has no significant effect on the

Table 3. Comparison of Computational Time (s)

		$S = 63$			$S = 126$			$S = 252$			$S = 378$		
Parameter value	Model	$L = 2$	$L = 3$	$L = 4$	$L = 2$	$L = 3$	$L = 4$	$L = 2$	$L = 3$	$L = 4$	$L = 2$	$L = 3$	$L = 4$
$\alpha = 0.1$	Global	0.88	1.48	2.69	4.22	5.48	6.62	8.29	19.62	45.17	16.82	52.63	56.11
	Local: $\theta = 1.0$	0.35	0.43	0.42	0.74	0.85	0.89	1.73	1.85	1.91	3.12	3.04	3.22
	Local: $\theta = 10.0$	0.35	0.43	0.42	0.74	0.85	0.89	1.73	1.85	1.91	3.12	3.05	3.09
	Static	0.01			0.02			0.06			0.10		
$\alpha = 0.3$	Global	0.85	1.47	2.25	2.20	4.70	6.51	5.33	11.05	37.08	10.03	33.53	38.86
	Local: $\theta = 1.0$	0.35	0.41	0.41	0.76	0.87	0.92	1.82	1.91	1.98	3.32	3.21	3.36
	Local: $\theta = 10.0$	0.35	0.41	0.40	0.76	0.86	0.90	1.79	1.93	1.92	3.31	3.17	3.18
	Static	0.01			0.02			0.05			0.10		
$\alpha = 0.5$	Global	0.87	1.57	2.29	1.74	4.01	7.68	5.44	14.59	30.58	10.75	20.67	41.98
	Local: $\theta = 1.0$	0.35	0.42	0.41	0.77	0.89	0.92	1.84	1.95	2.00	3.42	3.33	3.58
	Local: $\theta = 10.0$	0.35	0.42	0.40	0.76	0.87	0.91	1.83	1.95	1.95	3.42	3.30	3.32
	Static	0.01			0.02			0.05			0.10		
	Time to train a tree	0.40	0.75	1.07	0.95	1.79	2.59	2.05	3.98	5.86	3.35	6.42	9.44
	Forest	9.14			21.18			51.20			96.54		

Downloaded from informs.org by [2001:da8:d800:95::25] on 20 November 2025, at 19:27 . For personal use only, all rights reserved.

Downloaded from informs.org by [2001:da8:d800:95::25] on 20 November 2025, at 19:27 . For personal use only, all rights reserved.

computational time. The computation time of the RTGlobal model increases with both the sample size and the number of leaf nodes, whereas, in contrast, the computation time of the RTLocal model is minimally affected by the number of leaf nodes. This is expected and aligns with the analysis in Section 5. For a sample size of $S = 378$ and $L = 4$ leaf nodes, the computational time is less than 60 seconds for the RTGlobal model and 4 seconds for the RTLocal model. Notably, the computation time for the RTLocal model remains efficient in all our test cases. Despite being less efficient than the RTLocal model, the RTGlobal models still enjoy shorter computation times than the stochastic optimization forest method. The computational time of the robust static model is consistently short and minimally affected by the sample size. In the last row of Table 3, we present the computation times to obtain a tree according to the heuristic approach described in Section 2.

The computation times are generally short even for large S and L, rendering this heuristic practically relevant. When the problem scale becomes even larger (e.g., sample size S increases), enumerating all possible splits can be more time-consuming. Nevertheless, we can leverage the technique adopted in the popular catBoost package (Dorogush et al. 2018) to discretize each feature into a fixed amount of bins and only consider the corresponding splits.

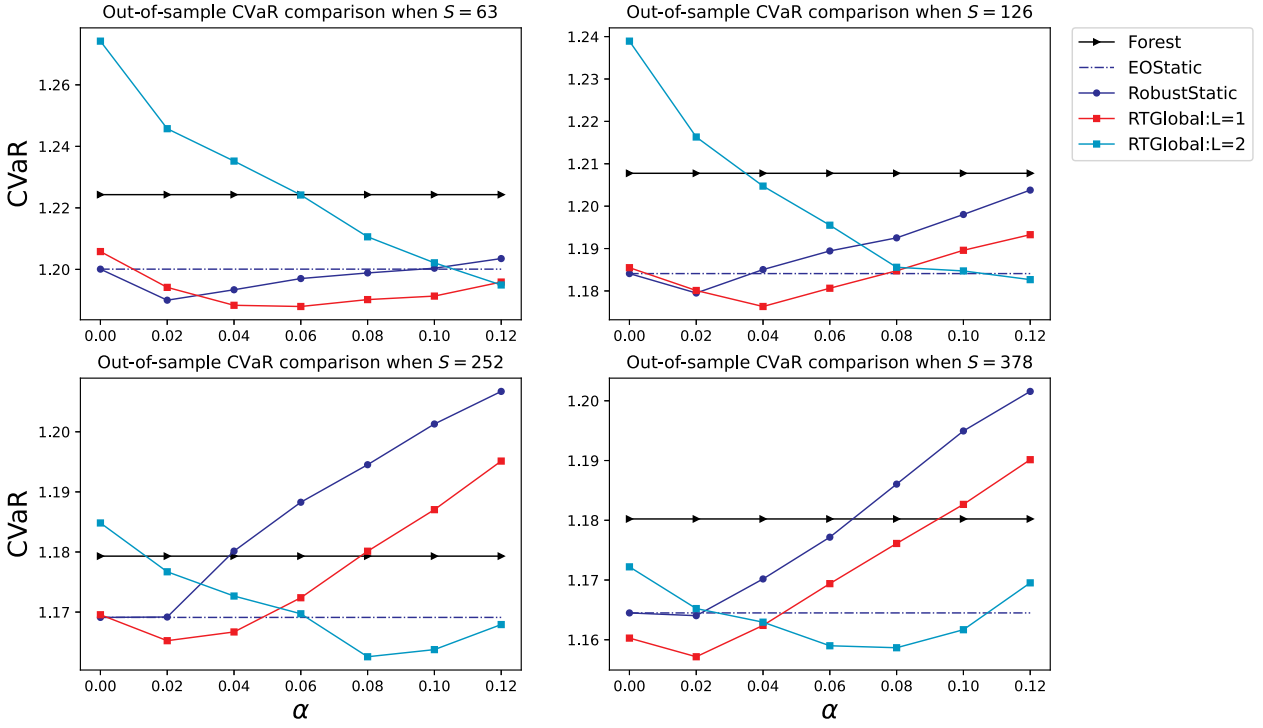
6.2. Real-World Data Set

In the next set of numerical experiments, we consider using real data based on five stocks in the S&P500 from January 1, 2010, to January 1, 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 data set $\mathcal{H}_t \triangleq \{\hat{\boldsymbol{u}}_s, \hat{v}_s\}_{s=t-S}^{t-1}$ 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 4. In this simulation study, we assume $L = 1, 2$.

Similar to the findings in Figure 2, we observe that, when the sample size is relatively small ($S \le 126$), the RTGlobal policy with only one leaf node has the best

Figure 4. (Color online) CVaR Comparison Based on the S&P Data



Downloaded from informs.org by [2001:da8:d800:95::25] on 20 November 2025, at 19:27 . For personal use only, all rights reserved.

performance, indicating the difficulty in leveraging side information by using the complex policies in a low signal-noise ratio setting with limited data. However, as the sample size increases, the performance of the RTGlobal policy with two leaf nodes improves. For the largest training size ($S = 378$), the RTGlobal policy with two leaf nodes outperforms the RobustStatic and RTGlobal policy with one leaf node, highlighting the benefits of incorporating side information even in a low signal-noise ratio setting. Additionally, we note that the RTGlobal policy with two leaf nodes can outperform the Forest policy across all experiments, which is consistent with previous simulation results.

7. 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, whereas 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:

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

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

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

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:

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

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; that is, the response policy is a tree-based static policy with $L = 1$. For both robust satisficing models, we adopt ℓ_1-norm.

Downloaded from informs.org by [2001:da8:d800:95::25] on 20 November 2025, at 19:27 . For personal use only, all rights reserved.

Table 4. Revenue Improvement in 10 Random Sample Instances

Performance	1	2	3	4	5	6	7	8	9	10	Average
R^S	33.76	33.66	33.85	33.48	33.82	33.81	33.50	33.85	33.78	33.63	33.72
R^N	30.38	32.30	30.38	31.71	30.38	32.22	32.31	32.07	32.28	32.20	31.62
$\frac{R^S-R^N}{R^N}$	11.13%	4.20%	11.42%	5.59%	11.34%	4.96%	3.68%	5.57%	4.65%	4.44%	6.70%

7.1. 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 $n_d = 4$ nonoverlapping subsets of the same length and denote the ith 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, $v^i(u, \tilde{\epsilon}^i) = w_0^i + w_1^i u + \tilde{\epsilon}^i$, where $i \in [n_d]$. We randomly generate $\bar{w}$ from the uniform distribution on $[0,1]^J$ and set $w_1^i = (1 + 0.2i)\bar{w}$ for each $i \in [n_d]$. For the intercept w_0^i, we set $w_0^1 = 10 \times \mathbf{1}$ and $w_0^i = w_0^{i-1} + \underline{u}_i w_1^{i-1} - \underline{u}_i w_1^i$ for $i \in \{2, \ldots, n_d\}$. The error term $\tilde{\epsilon}^i$ has a zero mean, and its variance is $\boldsymbol{\delta}_e^i = 0.1\boldsymbol{\mu}_v^i$, where $\boldsymbol{\mu}_v^i = w_0^i + w_1^i \mu_u$. The support set of the demand is a box, $\mathcal{I}_v \triangleq [w^1 + w_1^1 \underline{u}_1, w^{n_d} + w_1^{n_d} \overline{u}_{n_d}]$. Next, we set the supply capacity as $q = 0.5 \times \mathbf{1}^\top (w_0^{n_d} + w_1^{n_d} \overline{u}_{n_d})$. 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 distribution $\mathcal{N}(\mathbf{0}, \mathrm{diag}(\boldsymbol{\delta}_e^{j^*}))$ 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 $\hat{v}_s = v^{j^*}(\hat{u}_s, \hat{\boldsymbol{\epsilon}}_s) = w_0^{j^*} + w_1^{j^*} \hat{u}_s + \hat{\boldsymbol{\epsilon}}_s$. We truncate demand samples so that they fall into the support set $\mathcal{I}_v$. Then, $\{(\hat{u}_s, \hat{v}_s), s \in [S]\}$ constitutes our training samples. We generate 10,000 test samples with the same procedure to evaluate the out-of-sample performance.

7.2. 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). We set the target parameter as $\tau = Z_0 + \alpha\delta_0$ for a target margin parameter $\alpha > 0$, where δ_0 is the standard deviation of the sample revenues. Calibrating the target parameter τ then reduces to tuning the target margin α. We pin down an appropriate target margin, $\alpha \in [0, 4]$, via a fivefold cross-validation combined with a golden search procedure. We relegate the details of this algorithm to Online Appendix C. 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 10 different training data sets and test them on the same set of test data of size 10,000.

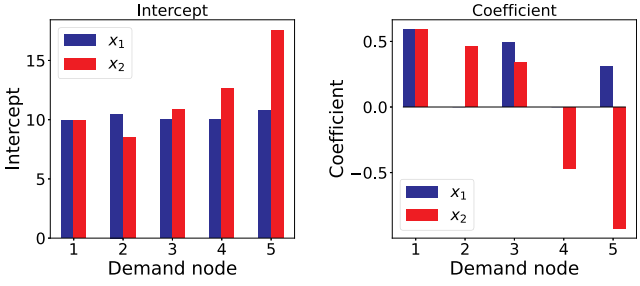
Let R^S and R^N denote the out-of-sample revenue of the robust satisficing models with tree-based affine policy and the benchmark robust satisficing model (with static policy ignoring side information), respectively. We compute the revenue ratio between the two models, $\frac{R^S-R^N}{R^N}$, and summarize the results in Table 4. We note that the average revenue ratio over the 10 random instances is 106.7%, indicating an improvement in revenue of 6.7%. 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.

7.3. 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 5. 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 nonnegligible magnitudes of the coefficients indicate the strong influence of precipitation information on the allocation quantities.

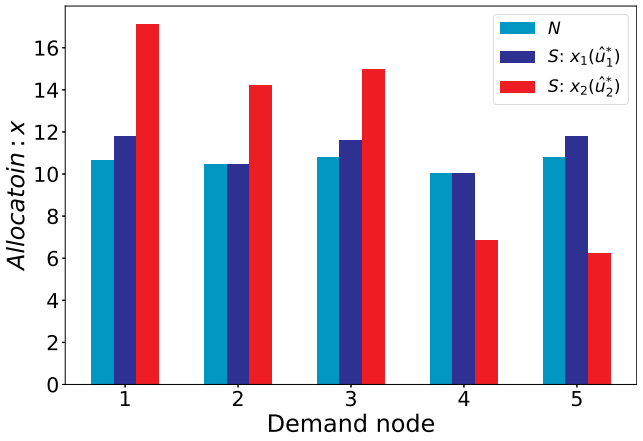
Note that the mean demands at all demand regions increase with precipitation u. When u falls within the first leaf node, the overall demands do not overwhelm the supply capacity q. Hence, it is reasonable to see positive coefficients; that is, we should deploy more

Figure 5. (Color online) Tree-Based Affine Allocation Policy



Downloaded from informs.org by [2001:da8:d800:95::25] on 20 November 2025, at 19:27 . For personal use only, all rights reserved.

Figure 6. (Color online) Allocation Quantities of Robust Satisficing with (S) and Without (N) Side Information



taxis to all demand regions 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 6. Specifically, we plot the allocation quantities for two instances: $\hat{u}_1^*$ and $\hat{u}_2^*$, which lie in leaf nodes 1 and 2, 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 6. The most notable difference is that our policy prioritizes demands at the first three demand regions when u lies in the second leaf node because of 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.

Acknowledgments

The authors thank the editors and reviewers for constructive and insightful comments that significantly enhanced this paper. The authors are listed in alphabetical order with equal contribution.

References

Aghaei S, Gómez A, Vayanos P (2024) Strong optimal classification trees. *Oper. Res.*, ePub ahead of print July 31, https://doi.org/10.1287/opre.2021.0034.

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

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

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

Bertsimas D, de Ruiter FJ (2016) Duality in two-stage adaptive linear optimization: Faster computation and stronger bounds. *INFORMS J. Comput.* 28(3):500–511.

Bertsimas D, Dunn J (2017) Optimal classification trees. *Machine Learn.* 106(7):1039–1082.

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

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

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

Bertsimas D, Van Parys B (2021) Bootstrap robust prescriptive analytics. *Math. Programming* 195(1):39–78.

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

Bertsimas D, McCord C, Sturt B (2023a) Dynamic optimization with side information. *Eur. J. Oper. Res.* 304(2):634–651.

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

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

Blanchet J, Kang Y, Murthy K (2019) Robust Wasserstein profile inference and applications to machine learning. *J. Appl. Probab.* 56(3):830–857.

Chen L, Sim M (2024) Robust CARA optimization. *Oper. Res.*, ePub ahead of print February 26, https://doi.org/10.1287/opre.2021.0654.

Chen L, Ramachandra A, Rujeerapaiboon N (2025) Robust data-driven CARA optimization. Preprint, submitted February 24, https://doi.org/10.2139/ssrn.5130565.

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

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

de Ruiter FJ, Zhen J, Den Hertog D (2023) Dual approach for two-stage robust nonlinear optimization. *Oper. Res.* 71(5):1794–1799.

Dorogush AV, Ershov V, Gulin A (2018) Catboost: Gradient boosting with categorical features support. Preprint, submitted August 24, https://arxiv.org/abs/1810.11363.

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

Esteban-Pérez A, Morales JM (2021) Distributionally robust stochastic programs with side information based on trimmings. *Math. Programming* 195(1):1069–1105.

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

Downloaded from informs.org by [2001:da8:d800:95::25] on 20 November 2025, at 19:27 . For personal use only, all rights reserved.

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

Gao R (2023) Finite-sample guarantees for Wasserstein distributionally robust optimization: Breaking the curse of dimensionality. *Oper. Res.* 71(6):2291–2306.

Gao R, Chen X, Kleywegt AJ (2024) Wasserstein distributionally robust optimization and variation regularization. *Oper. Res.* 72(3): 1177–1191.

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

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

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

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

Hannah L, Powell W, Blei D (2010) Nonparametric density estimation for stochastic optimization with an observable state variable. Lafferty J, Williams C, Shawe-Taylor J, Zemel R, Culotta A, eds. *Advances in Neural Information Processing Systems*, vol. 23 (Curran Associates Inc., Red hook, NY), 820–828.

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

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

Kannan R, Bayraksan G, Luedtke JR (2024) Residuals-based distributionally robust optimization with covariate information. *Math. Programming* 207(1):369–425.

Kuhn D, Wiesemann W, Georghiou A (2011) Primal and dual linear decision rules in stochastic and robust optimization. *Math. 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. *Oper. Res. Lett.* 33(4):341–348.

Loke GG, Tang Q, Xiao Y (2021) Decision-driven regularization: A blended model for predict-then-optimize. Preprint, submitted June 17, https://doi.org/10.2139/ssrn.3623006.

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

Medhat M, Schmeling M (2022) Short-term momentum. *Rev. Financial Stud.* 35(3):1480–1526.

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

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

Nguyen VA, Zhang F, Wang S, Blanchet J, Delage E, Ye Y (2024) Robustifying conditional portfolio decisions via optimal transport. *Oper. Res.*, ePub ahead of print November 4, https://doi.org/10.1287/opre.2021.0243.

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

Shafieezadeh-Abadeh S, Kuhn D, Esfahani PM (2019) Regularization via mass transportation. *J. Machine Learn. Res.* 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. *Adv. Neural Inform. Processing Systems* 33:21260–21270.

Sim M, Zhao L, Zhou M (2021) A new perspective on supervised learning via robust satisficing. Preprint, submitted December 14, https://doi.org/10.2139/ssrn.3981205.

Sim M, Tang Q, Zhou M, Zhu T (2024) The analytics of robust satisficing: Predict, optimize, satisfice, then fortify. *Oper. Res.*, ePub ahead of print October 7, https://doi.org/10.1287/opre.2023.0199.

Smith JE, Winkler RL (2006) The optimizer's curse: Skepticism and postdecision surprise in decision analysis. *Management Sci.* 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. Preprint, submitted October 10, https://arxiv.org/abs/2110.04855.

Tulabandhula T, Rudin C (2013) Machine learning with operational costs. *J. Machine Learn. Res.* 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. Preprint, submitted October 16, https://optimization-online.org/?p=20639.

Zhang L, Yang J, Gao R (2024) Optimal robust policy for feature-based newsvendor. *Management Sci.* 70(4):2315–2329.

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

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 J. Comput.* 34(1):196–210.

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

Li Chen is a lecturer (assistant professor) in the Discipline of Business Analytics at the University of Sydney Business School. He has broad interests in optimization under uncertainty and data-driven decision making.

Melvyn Sim is professor and provost's chair at the Department of Analytics and Operations, NUS Business School. His research interests fall broadly under the categories of decision making and optimization under uncertainty with applications ranging from finance, supply chain management, and healthcare to engineered systems.

Xun Zhang is an associate researcher at the School of Management, University of Science and Technology of China. His research focuses on developing data-driven prescriptive analytics models under uncertainty, with applications in areas such as inventory management, pricing, and resource allocation.

Long Zhao is an assistant professor in the Department of Analytics and Operations at NUS Business School. His research focuses on data-driven decision making, with applications in portfolio optimization and forecast aggregation.

Minglong Zhou is an assistant professor at the Department of Management Science at the School of Management at Fudan University. His research is broadly in the area of decision making under uncertainty with applications in the domains of healthcare, transportation, and supply chain management.