**Manuscript**

# The Analytics of Robust Satisficing

## Predict, Optimize, Satisfice, then Fortify

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

Qingshen Tang
Nanyang Business School, Nanyang Technological University, qinshen.tang@gmail.com

Minglong Zhou
Department of Management Science, School of Management, Fudan University
minglong.zhou@outlook.com

Taozeng Zhu
Institute of Supply Chain Analytics, Dongbei University of Finance and Economics
taozeng.zhu@gmail.com

We introduce a novel approach to prescriptive analytics that leverages robust satisficing techniques to determine optimal decisions in situations of risk ambiguity and prediction uncertainty. Our decision model relies on a reward function that incorporates uncertain parameters, which can be partially predicted using available side information. However, the accuracy of the linear prediction model depends on the quality of regression coefficient estimates derived from the available data. To achieve a desired level of fragility, we begin by establishing a target relative to the predict-then-optimize objective and solve a residual-based robust satisficing model. Next, we solve a new estimation-fortified robust satisficing model that minimizes the influence of estimation uncertainty while ensuring that the estimated fragility of the solution in achieving a less ambitious guarding target falls below the level for the desired target. Our approach is supported by statistical justifications, and we propose tractable models for various scenarios, such as saddle functions, two-stage linear optimization problems, and decision-dependent predictions. We demonstrate the effectiveness of our approach through case studies involving a wine portfolio investment problem and a multi-product pricing problem using real-world data. Our numerical studies show that our approach outperforms the predict-then-optimize approach in achieving higher expected rewards and at lower risks when evaluated on the actual distribution. Notably, we observe significant improvements over the benchmarks, particularly in cases of limited data availability.

*Key words*: Robust optimization, robust satisficing, predictive analytics, prescriptive analytics
*History*: April 19, 2023.

---

## 1. Introduction

A prescriptive analytics model uses data and leverages predictive analytics to help managers make better decisions to improve their objectives, such as increasing profits or reducing costs. The inputs to such a model usually involve some uncertain parameters at the time of decision-making. Historical data contains information about past realizations of the uncertain parameters and other side information that have predictive powers of the parameters' future outcomes. With the ever-growing importance of analytics, there has been an increase in research on incorporating additional information into optimization models (see, *e.g.*, Gao et al. 2017, Ho and Hanasusanto 2019, Bertsimas et al. 2023), which have been effective in enhancing the quality of solutions in diverse applications such as inventory management (Ban and Rudin 2019), vehicle pre-allocation (Hao et al. 2020), product procurement (Ban et al. 2019), and portfolio optimization (Kannan et al. 2020, Nguyen et al. 2021).

A major stream of prescriptive analytics models follows a two-step predict-then-optimize approach (*e.g.,* Craig and Raman 2016, Ferreira et al. 2016, Glaeser et al. 2019, Siegel and Wagner 2021). The basic predict-then-optimize approach first forecasts the expected values of the observable parameters using available side information, and subsequently uses them as inputs to the optimization problem. However, it has been well established that the approach of separating prediction from optimization would often yield solutions that perform poorly (see, for example, Liyanage and Shanthikumar 2005, Smith and Winkler 2006, Ramamurthy et al. 2012, Mundru 2019, Liu et al. 2021). Therefore, researchers have proposed improvements to the predict-then-optimize model by incorporating the decision model's objective into the parameter estimation process for the prediction model (Tulabandhula and Rudin 2013, Loke et al. 2021, Elmachtoub and Grigas 2022). When applied in stochastic optimization, the predict-then-optimize approach estimates the conditional distribution of observable parameters before incorporating the predicted distribution in the optimization phase (Hannah et al. 2010, Bertsimas and Kallus 2020, Grigas et al. 2021, Srivastava et al. 2021, Kannan et al. 2022). For instance, Srivastava et al. (2021) present a generic stochastic optimization model that uses side information to estimate the conditional expectation through kernel regression.

The ultimate goal of a prescriptive analytics model is to improve decision-making; the manager is only concerned with the impact of their objectives resulting from implementing the decision recommended by the prescriptive analytics model. Smith and Winkler (2006), who coin the term, *Optimizer's Curse*, note that the naive predict-then-optimize model is optimistically biased and can perform poorly in the out-of-sample performance. To address overfitting problems in data-driven optimization, Mohajerin Esfahani and Kuhn (2018) propose the seminal data-driven robust optimization model that characterizes its ambiguity set using the Wasserstein distance metric. Gao et al. (2017), Blanchet et al. (2019), and Shafieezadeh-Abadeh et al. (2019) also establish that when applied to regression and classification models, these data-driven robust optimization approaches bear similarities with regularization techniques. Consequently, the idea of incorporating conditional distributions using side information has also been adopted in data-driven robust optimization (see, *e.g.*, Hao et al. 2020, Nguyen et al. 2021, Wang et al. 2021, Bertsimas and Van Parys 2022, Esteban-Pérez and Morales 2022). Kannan et al. (2020) propose a distributionally robust optimization model under ambiguity sets constructed with empirical prediction residuals. Bertsimas et al. (2023) use side information to estimate the conditional probability distribution of the uncertain outcomes before incorporating it into the data-driven robust optimization model. Nguyen et al. (2021) present a data-driven robust optimization model for conditional decision-making using side information, and they illustrate the advantages of their model in portfolio optimization problems.

Most prescriptive analytics tools are focused on utility maximization. Schwartz et al. (2011) argue that utility maximization may be self-deceptive when the decisions are made under probabilistic ambiguity and estimation uncertainty. They propose robust satisficing as an alternative to utility maximization. The goal is to maximize the robustness to the uncertainty of achieving a satisfactory target. As articulated by Simon (1955), target satisficing, as opposed to utility maximization, is prevalent in human decision making, especially in complex situations facing risks and uncertainty. In an empirical study, Mao (1970) concluded that managers would consider risk as "the prospect of not meeting some target rate of return." More recently, a survey of 74 executive MBA students conducted by Chen and Tang (2022) found that over 90% of them reported that their companies set profit targets in every accounting period. Brown and Sim (2009) axiomatize a family of satisficing decision criteria that have an embedded preference for diversification, which is sensible and consistent with risk-aversion behavior. Some applications that are based on the satisficing criteria include routing under travel times uncertainty (Zhang et al. 2021), project selection with uncertain returns (Hall et al. 2015), and project management in meeting cost targets (Goh and Hall 2013).

In the context of satisficing, Long et al. (2023) propose a data-driven robust satisficing model that determines the best possible here-and-now decision to achieve a target expected reward under risk ambiguity. In this paper, we incorporate a prediction model using the residual-based approach of Kannan et al. (2020) and propose new robust satisficing models that mitigate both risk ambiguity and prediction uncertainty. We adopt a linear prediction model due to its widespread use in the literature (see, *e.g.,* Tulabandhula and Rudin 2013, Liu et al. 2021, Elmachtoub and Grigas 2022, Ho-Nguyen and Kılınç-Karzan 2022). When side information is available in the form of predictions made by other learners, the linear model can be interpreted as an ensemble model where the linear coefficients represent the weights of individual predictions. The accuracy of the linear prediction model depends on how well the regression coefficients can be estimated from data. The joint prediction and robust optimization model proposed by Zhu et al. (2022) is relevant to our work.

While they incorporate a prediction model and account for estimation uncertainty, they do not include risk ambiguity and can only focus on deterministic optimization models.

In our approach, we first establish the target relative to the predict-then-optimize objective and solve a residual-based robust satisficing model to obtain an acceptable level of fragility. Subsequently, we solve a new estimation-fortified robust satisficing model that minimizes the influence of estimation uncertainty as much as possible while ensuring that the estimated fragility of the solution in achieving a less ambitious guarding target falls below the level for the desired target. We provide statistical justifications for the robust satisficing models and formulate tractable models when the reward function is a saddle function, and for a two-stage linear optimization problem. We also propose tractable models when the prediction is decision-dependent. We use real data in case studies featuring a wine portfolio investment problem and a multi-product pricing problem. Through these numerical studies, we elucidate the benefits of our approach over the predict-then-optimize approach; when evaluated on the actual distribution, our approach achieves higher expected rewards and at lower risk. We observe consistent and significant improvement over the benchmarks, and the improvements are more pronounced when there is limited data.

**Notation.** We use boldface lowercase letters for vectors (*e.g.*, $\boldsymbol{\theta}$), and calligraphic letters for sets (*e.g.*, $\mathcal{X}$). We use $[N]$ to denote the running index $\{1,2,\ldots,N\}$ for $N$ a known integer. We let $\mathbb{R}^N$ ($\mathbb{R}^N_+$) represent the space of $N$-dimensional real (non-negative) number, and $\mathbb{S}^N$ ($\mathbb{S}^N_{++}$) represent the space of $N$ by $N$ symmetric (positive definite) matrices. We use $\boldsymbol{A}\succeq\boldsymbol{0}$ to denote a positive semidefinite matrix $\boldsymbol{A}$. A random variable $\tilde{v}$ is denoted with a tilde sign such as $\tilde{v}\sim\mathbb{P}$, $\mathbb{P}\in\mathcal{P}_0$, where $\mathcal{P}_0$ represents the set of all possible distributions. For a multivariate random variable, we use $\mathcal{P}_0(\mathcal{Z})$ to represent the set of all distributions for the multivariate random variable with support $\mathcal{Z}\subseteq\mathbb{R}^N$. Specifically, we use $\tilde{\boldsymbol{z}}\sim\mathbb{P}$, $\mathbb{P}\in\mathcal{P}_0(\mathcal{Z})$ to define $\tilde{\boldsymbol{z}}$ as a multivariate random variable with support $\mathcal{Z}$ and distribution $\mathbb{P}$. We use $\mathbb{E}_{\mathbb{P}}[\tilde{v}]$ to denote the expectation of a random variable, $\tilde{v}\sim\mathbb{P}$, over its distribution. Finally, $\boldsymbol{0}$ ($\boldsymbol{1}$) denotes the vector of all zeros (ones) and $\boldsymbol{e}_i$ denotes the $i$th basis vector. The dimensions of these vectors should be clear from the context.

## 2. Data-driven decision model with prediction

We consider a data-driven decision model with a reward function $f(\boldsymbol{x},\boldsymbol{z}):\mathcal{X}\times\mathcal{Z}\to\mathbb{R}$, where the input to the first argument is the here-and-now decision $\boldsymbol{x}\in\mathcal{X}\subseteq\mathbb{R}^D$, and the input to the second argument $\boldsymbol{z}\in\mathcal{Z}\subseteq\mathbb{R}^N$ represents the uncertain input parameters whose risky outcomes can be partially predicted from the current side information. The input parameters of the problem are collectively denoted by the random variable $\tilde{\boldsymbol{z}}$ with an unobservable distribution $\tilde{\boldsymbol{z}}\sim\mathbb{P}^\star$, $\mathbb{P}^\star\in\mathcal{P}_0(\mathcal{Z})$. The current side information is represented by a $J$-dimensional vector $\boldsymbol{u}\in\mathcal{U}\subseteq\mathbb{R}^J$. The linear prediction map $\boldsymbol{h}(\boldsymbol{w},\boldsymbol{u}):\mathbb{R}^M\times\mathcal{U}\to\mathbb{R}^N$, which is linear in $\boldsymbol{w}\in\mathbb{R}^M$, provides the prediction of $\tilde{\boldsymbol{z}}$ such that

$$\tilde{\boldsymbol{z}}:=\boldsymbol{h}(\boldsymbol{w}^\star,\boldsymbol{u})+\boldsymbol{H}(\boldsymbol{u})\tilde{\boldsymbol{\epsilon}},$$

for some unobservable regression coefficients $\boldsymbol{w}^\star\in\mathcal{W}\subseteq\mathbb{R}^M$ and random prediction errors, $\tilde{\boldsymbol{\epsilon}}$. The matrix $\boldsymbol{H}(\boldsymbol{u})\in\mathbb{S}^N_{++}$ is a positive definite matrix for all $\boldsymbol{u}\in\mathcal{U}$, which we use to model some elementary form of heteroscedasticity such as $\boldsymbol{H}(\boldsymbol{u})=\sum_{n\in[N]}u_n\boldsymbol{e}_n\boldsymbol{e}_n^\top$, for $\boldsymbol{u}\in\mathbb{R}^J_{++}$. To obtain a tractable model, the matrix $\boldsymbol{H}(\boldsymbol{u})$ should be deterministic and independent of the unobservable parameters. In the case of homoscedasticity, we have $\boldsymbol{H}(\boldsymbol{u})=\boldsymbol{I}$.

To illustrate the versatility of the linear prediction map, we can also extend it to cluster-wise prediction, such as a linear regression tree. In this case, $\mathcal{U}$ is a partition of $L$ sets of clusters, $\mathcal{U}=\mathcal{U}_1\cup\cdots\cup\mathcal{U}_L$, and

$$\boldsymbol{h}(\boldsymbol{w},\boldsymbol{u})=\sum_{\ell\in[L]:\boldsymbol{u}\in\mathcal{U}_\ell}\boldsymbol{h}_\ell(\boldsymbol{w}_\ell,\boldsymbol{u}),$$

where $\boldsymbol{w}=(\boldsymbol{w}_1,\ldots,\boldsymbol{w}_L)$, and $\boldsymbol{h}_\ell(\boldsymbol{w}_\ell,\boldsymbol{u})$, $\ell\in[L]$, are separate linear maps of $\boldsymbol{w}_\ell$. In a simple prediction tree, we have $\boldsymbol{h}_\ell(\boldsymbol{w}_\ell,\boldsymbol{u})=\boldsymbol{w}_\ell$, $\boldsymbol{w}_\ell\in\mathbb{R}^N$, $\ell\in[L]$. The side information can also be the prediction outputs from $L$ different experts or machine learning models $\boldsymbol{u}_1,\ldots,\boldsymbol{u}_L\in\mathbb{R}^N$, such as a forest of prediction trees, which can then be affinely combined to achieve a superior ensemble prediction as follows:

$$\boldsymbol{h}(\boldsymbol{w},\boldsymbol{u})=\sum_{\ell\in[L]}w_\ell\boldsymbol{u}_\ell+w_{L+1}\boldsymbol{1}_N,$$

where $\boldsymbol{u}=(\boldsymbol{u}_1,\ldots,\boldsymbol{u}_L)$.

We can estimate the regression coefficients and the statistics of the prediction errors from the historical records of the input parameters and the side information. Specifically, we denote $\tilde{\boldsymbol{u}}$ as the $J$-dimensional random variable associated with the side information and $\tilde{\boldsymbol{v}}$ as the random outcome associated with the input parameters without revealing the side information. The unobservable joint distribution is denoted by $(\tilde{\boldsymbol{u}},\tilde{\boldsymbol{v}})\sim\mathbb{Q}^\star$, $\mathbb{Q}^\star\in\mathcal{P}_0(\mathcal{U}\times\mathcal{Z})$. The residual associated with the regression coefficients $\boldsymbol{w}\in\mathcal{W}$ is a composite random variable given by

$$\tilde{\boldsymbol{\zeta}}(\boldsymbol{w}):=\boldsymbol{H}(\tilde{\boldsymbol{u}})^{-1}(\tilde{\boldsymbol{v}}-\boldsymbol{h}(\boldsymbol{w},\tilde{\boldsymbol{u}})).$$

Assumption 1. *We assume that the random residual $\tilde{\boldsymbol{\zeta}}(\boldsymbol{w}^\star)$, evaluated at the actual regression coefficients $\boldsymbol{w}^\star\in\mathcal{W}$, is statistically independent of the side information $\tilde{\boldsymbol{u}}$.*

Under the assumption, the distribution of $\tilde{\boldsymbol{\zeta}}(\boldsymbol{w}^\star)$ does not depend on the realization of $\tilde{\boldsymbol{u}}$. Hence, the prediction errors and residuals have the same distribution.

Given the realization of the side information, $\tilde{\boldsymbol{u}} = \boldsymbol{u}$, the decision maker should ideally solve the following optimization problem,

$$\begin{aligned} Z^\star &= \max_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}_{\mathbb{Q}^\star}\left[f(\boldsymbol{x},\tilde{\boldsymbol{v}})|\tilde{\boldsymbol{u}}=\boldsymbol{u}\right] \\ &= \max_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}_{\mathbb{P}^\star}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right]. \end{aligned} \tag{1}$$

However, the actual regression coefficients $\boldsymbol{w}^\star$ and the actual distributions $\mathbb{Q}^\star$, $\mathbb{P}^\star$ are unobservable, which would prohibit the modeler from obtaining the actual optimal solution.

**Predict-then-optimize model**

To solve Problem (1) approximately, we first need to estimate the regression coefficients for the linear prediction model from the data. The data set $\mathcal{D} = \{(\hat{\boldsymbol{u}}_s, \hat{\boldsymbol{v}}_s)\}_{s\in[S]}$ comprises $S$ realizations of the historical input parameters, $\hat{\boldsymbol{v}}_s \in \mathcal{Z}$ and the respective realizations of the side information $\hat{\boldsymbol{u}}_s \in \mathcal{U}$. With a given prediction loss function, $\rho$, such as the sum of the squared prediction errors with regularization penalty, the estimation problem solves the following optimization problem,

$$\hat{\boldsymbol{w}} \in \arg\min_{\boldsymbol{w}\in\mathcal{W}} \rho(\boldsymbol{w};\mathcal{D}) \tag{2}$$

to determine the estimate, $\hat{\boldsymbol{w}}$. If the estimation process is consistent, then $\hat{\boldsymbol{w}}$ would not be too far away from $\boldsymbol{w}^\star$ when there are sufficient samples. However, when the data is collected over time, we may not have sufficient samples for this purpose. For example, in the case of ordinary least square estimation, to achieve a medium effect of the coefficient of determination, for example, $R^2 = 0.07$, Green (1991) proposes the number of samples $S \geq 50 + 8\times N\times(1+J+N)$. Suppose $N = 10$, $J = 2$; we would require $S \geq 1090$, which translates to over 20 years of weekly collected samples. In many practical circumstances, we may not have the number of samples to achieve the desired prediction accuracy.

For a given vector of regression coefficients, $\boldsymbol{w} \in \mathcal{W}$, it is also convenient to define the map $\hat{\boldsymbol{\zeta}} : \mathcal{W}\times[S] \to \mathbb{R}^N$,

$$\hat{\boldsymbol{\zeta}}(\boldsymbol{w}, s) := \boldsymbol{H}(\hat{\boldsymbol{u}}_s)^{-1}\left(\hat{\boldsymbol{v}}_s - \boldsymbol{h}(\boldsymbol{w},\hat{\boldsymbol{u}}_s)\right),$$

to show the dependency of the empirical residuals on the regression coefficients. Likewise, we also define the following map $\hat{\boldsymbol{z}} : \mathcal{W}\times\mathcal{U}\times[S] \to \mathbb{R}^N$,

$$\hat{\boldsymbol{z}}(\boldsymbol{w},\boldsymbol{u},s) := \boldsymbol{h}(\boldsymbol{w},\boldsymbol{u}) + \boldsymbol{H}(\boldsymbol{u})\hat{\boldsymbol{\zeta}}(\boldsymbol{w},s) = \boldsymbol{h}(\boldsymbol{w},\boldsymbol{u}) - \boldsymbol{g}(\boldsymbol{w},\boldsymbol{u},\boldsymbol{u}_s) + \boldsymbol{H}(\boldsymbol{u})\boldsymbol{H}(\hat{\boldsymbol{u}}_s)^{-1}\hat{\boldsymbol{v}}_s,$$

where the map $\boldsymbol{g} : \mathcal{W}\times\mathcal{U}\times\mathcal{U}$, which is linear in $\boldsymbol{w}$, is defined as

$$\boldsymbol{g}(\boldsymbol{w},\boldsymbol{u},\hat{\boldsymbol{u}}_s) := \boldsymbol{H}(\boldsymbol{u})\boldsymbol{H}(\hat{\boldsymbol{u}}_s)^{-1}\boldsymbol{h}(\boldsymbol{w},\hat{\boldsymbol{u}}_s).$$

Hence, $\hat{\boldsymbol{z}}(\boldsymbol{w},\boldsymbol{u},s)$ describes how the predicted empirical samples would change if the historical side information, $\hat{\boldsymbol{u}}_s$, were to be replaced by the current side information $\boldsymbol{u}$ under $\boldsymbol{w}$. In the case of homoscedasticity, we have $\boldsymbol{g}(\boldsymbol{w},\boldsymbol{u},\hat{\boldsymbol{u}}_s)=\boldsymbol{h}(\boldsymbol{w},\hat{\boldsymbol{u}}_s)$. Accordingly, for a given $\boldsymbol{w}\in\mathcal{W}$, we define the corresponding predicted empirical distribution $\hat{\mathbb{P}}_{\boldsymbol{w}}\in\mathcal{P}_0(\mathbb{R}^N)$ such that $\hat{\mathbb{P}}_{\boldsymbol{w}}\left[\tilde{\boldsymbol{z}}=\hat{\boldsymbol{z}}(\boldsymbol{w},\boldsymbol{u},s)\right]=1/S$ for all $s\in[S]$.

For a given estimate of the regression coefficients $\hat{\boldsymbol{w}}$, Problem (1) can be approximated by solving the predict-then-optimize model as follows:

$$\begin{aligned}\hat{Z}=\max\ & \mathbb{E}_{\hat{\mathbb{P}}_{\hat{\boldsymbol{w}}}}\left[f\left(\boldsymbol{x},\tilde{\boldsymbol{z}}\right)\right]\\ \text{s.t.}\ & \boldsymbol{x}\in\mathcal{X}\\ =\max\ & \frac{1}{S}\sum_{s\in[S]}f\left(\boldsymbol{x},\hat{\boldsymbol{z}}(\hat{\boldsymbol{w}},\boldsymbol{u},s)\right)\\ \text{s.t.}\ & \boldsymbol{x}\in\mathcal{X},\end{aligned}\tag{3}$$

where the objective function is evaluated on the corresponding predicted empirical distribution associated with the estimated regression coefficients, $\hat{\boldsymbol{w}}$. Kannan et al. (2022) provide an in-depth analysis of this approach, including the results on convergence rate and finite sample guarantee.

Indeed, how well Problem (3) would approximate Problem (1) would depend on, among other things, how well we can estimate the regression coefficients of the linear prediction map. We also note that while the predict-then-optimize model appears reasonable, it may result in suboptimal solutions. For instance, Liyanage and Shanthikumar (2005) show that an unbiased estimate for the optimal order quantity obtained from separating estimation and optimization in a newsvendor problem leads to suboptimal solutions. This is because estimates are obtained by minimizing loss function that only focuses on the fitting accuracy to historical data; yet higher fitting accuracy does not translate into better performance in the subsequent optimization (see, *e.g.*, Mundru 2019, Loke et al. 2021, Elmachtoub and Grigas 2022). We also observe this phenomenon in our computational studies. Therefore, it is essential to account for the influence of prediction inaccuracy on the reward function.

## 3. Robust satisficing models

To account for risk ambiguity when evaluating the reward function using the predicted empirical distribution, we first consider the hugely popular data-driven robust optimization model of Mohajerin Esfahani and Kuhn (2018):

$$\begin{aligned}\hat{Z}_\Gamma=\max\ & \inf_{\mathbb{P}\in\mathcal{B}_\Gamma}\mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right]\\ \text{s.t.}\ & \boldsymbol{x}\in\mathcal{X},\end{aligned}\tag{4}$$

where the ambiguity set is defined as follows:

$$\mathcal{B}_\Gamma := \left\{ \mathbb{P} \in \mathcal{P}_0(\mathcal{Z}) \;\middle|\; \Delta(\mathbb{P}, \hat{\mathbb{P}}) \le \Gamma \right\},$$

and

$$\Delta(\mathbb{P}_1, \mathbb{P}_2) := \inf_{\mathbb{Q} \in \mathcal{P}_0(\mathbb{R}^N \times \mathbb{R}^N)} \left\{ \mathbb{E}_{\mathbb{Q}} \left[ \|\tilde{\boldsymbol{z}} - \tilde{\boldsymbol{\xi}}\| \right] \;\middle|\; (\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{\xi}}) \sim \mathbb{Q}, \tilde{\boldsymbol{z}} \sim \mathbb{P}_1, \tilde{\boldsymbol{\xi}} \sim \mathbb{P}_2 \right\},$$

is the type-1 Wasserstein distance metric.

Suppose the actual data-generating distribution $\mathbb{P}^\star$, $\tilde{\boldsymbol{z}} \sim \mathbb{P}^\star$ is a light-tailed distribution and $\mathbb{P}^S$ is the distribution that governs the distribution of independent samples $\hat{\boldsymbol{z}}_1, \ldots, \hat{\boldsymbol{z}}_S$ drawn from $\mathbb{P}^\star$, which constitutes the empirical distribution $\hat{\mathbb{P}}$. Fournier and Guillin (2015) show that

$$\mathbb{P}^S \left[ \Delta(\mathbb{P}^\star, \hat{\mathbb{P}}) > \Gamma \right] \le \begin{cases} c_1 \exp(-c_2 S \Gamma^{\max\{N,2\}}) & \text{if } 0 < \Gamma \le 1, \\ c_1 \exp(-c_2 S \Gamma^\alpha) & \text{if } \Gamma > 1, \end{cases} \tag{5}$$

for some $\alpha > 1$ that characterizes the light tail distribution, and positive constants, $c_1$ and $c_2$. Consequently, Mohajerin Esfahani and Kuhn (2018) observe that if $\Gamma$ is chosen in the order of $S^{-1/\max\{N,2\}}$, then the underlying data-generating distribution $\mathbb{P}^\star$ would be contained in the Wasserstein ball with high probability. Tighter results with milder dependence on $N$ have been established in Blanchet et al. (2019), Shafieezadeh-Abadeh et al. (2019), and Si et al. (2020). Laudably, the consistency analysis of Gao (2022) resolves the "curse of dimensionality" by reducing the dependence on $N$ and showing that a non-asymptotic rate of $S^{-1/2}$ would suffice to achieve consistency for the data-driven robust optimization problem.

Using the same metric, Long et al. (2023) propose the robust satisficing model that is specified by a target expected reward, $\tau$ as follows:

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

The robust satisficing model constrains the target shortfall by employing the fragility parameter, $\kappa$, which is multiplied by the Wasserstein distance that measures the discrepancy between an ambiguous distribution and the empirical distribution.

**Consistency analysis of robust satisficing**

Consistency analysis in the robust satisficing framework requires optimizing the expected reward asymptotically by fine-tuning the target as the dataset expands. Though Long et al. (2023) have yet to analyze consistency, they show that the optimal solution to the robust satisficing problem with target $\tau$ has the following profile of confidence bounds:

$$\mathbb{P}^S[\tau - \mathbb{E}_{\mathbb{P}^\star}[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})] > K_\tau r] \le \mathbb{P}^S \left[ \Delta(\mathbb{P}^\star, \hat{\mathbb{P}}) > r \right] \quad \forall r \ge 0. \tag{7}$$

As a target reference for analyzing consistency, we focus on the empirical optimization problem,

$$\hat{Z} = \max_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}_{\hat{\mathbb{P}}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right], \tag{8}$$

for which its analysis of convergence to optimality has well been studied (see, *e.g.*, Kleywegt et al. 2002). Suppose $\hat{\boldsymbol{x}}$ is the optimal solution to the robust satisficing problem for $\tau = \hat{Z}$, and $\Gamma$ is the size of the Wasserstein ball such that

$$\mathbb{P}^S\left[\Delta(\mathbb{P}^\star,\hat{\mathbb{P}}) > \Gamma\right] \leq \epsilon,$$

for some $\epsilon > 0$, then we have the following confidence bound:

$$\mathbb{P}^S\left[\hat{Z} - \mathbb{E}_{\mathbb{P}^\star}\left[f(\hat{\boldsymbol{x}},\tilde{\boldsymbol{z}})\right] > K_{\hat{Z}}\Gamma\right] \leq \mathbb{P}^S\left[\Delta(\mathbb{P}^\star,\hat{\mathbb{P}}) > \Gamma\right] \leq \epsilon.$$

Hence, by choosing $\tau = \hat{Z}$, consistency can be achieved if, as the sample size expands, $K_{\hat{Z}}\Gamma$ diminishes to zero at a rate that preserves the confidence bound of the Wasserstein ball. We also require that $K_{\hat{Z}} \leq \bar{L}$ for some positive constant $\bar{L}$, which is the case if the reward function is Lipschitz continuous (see Theorem 3). We can also optimize for a target $\tau \leq \hat{Z}$ that strengthens this confidence bound.

THEOREM 1. *Suppose we choose $\tau = \hat{Z} - \hat{\eta}\Gamma$, where $\hat{\eta}$ is the optimal solution to the following minimization problem,*

$$\hat{K} = \min_{\eta\geq 0}\{\eta + K_{\hat{Z}-\eta\Gamma}\},$$

*so that $\hat{K} \leq K_{\hat{Z}}$. Then the optimal solution of the robust satisficing problem with target $\tau$ satisfies the confidence bound:*

$$\mathbb{P}^S\left[\hat{Z} - \mathbb{E}_{\mathbb{P}^\star}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right] > \hat{K}\Gamma\right] \leq \epsilon.$$

*Moreover, the optimal solution coincides with the corresponding robust optimization solution with size parameter $\Gamma$ such that $\hat{Z}_\Gamma = \hat{Z} - \hat{K}\Gamma \leq \tau$.*

Theorem 1 introduces an adaptive adjustment of the target in consistency analysis as the dataset grows, ultimately resulting in the asymptotic optimization of the expected reward through robust satisficing. More importantly, this finding suggests that the consistency analysis of robust satisficing can be adapted to align with any of the well-established consistency analysis of data-driven robust optimization problems documented in the current literature. Notably, the results from Gao (2022) indicate that choosing the target $\tau = \hat{Z} - \hat{\eta}\Gamma$, with $\Gamma$ changing at a rate of $S^{-1/2}$, is adequate for achieving consistency.

Due to the challenge in determining $\Gamma$ from theoretical bounds, the consistency analysis suggests an alternative method for setting $\tau$. We propose choosing $\tau = \hat{Z} - \chi\hat{\sigma}/\sqrt{S}$, where $\hat{\sigma}$ represents the sample standard deviation of the empirical reward associated with the optimal solution to the empirical optimization problem (8), and $\chi \geq 0$ serves as a confidence parameter. Compared to the size of the Wasserstein ball, $\Gamma$, the confidence parameter, $\chi$, is also more interpretable and intuitive to specify, and can also be established using cross-validation techniques, whenever it is feasible to so do.

**Fragility and target trade-offs**

Beyond consistency, it is crucial to develop optimization models accounting for changing conditions and imperfect data, with the aim of enhancing robustness and reducing solution fragility. This includes handling limited data samples or non-independent samples due to imperfect data collection. For instance, DeMiguel et al. (2009) found that the mean-variance strategy required an estimation window of over 3,000 months to outperform the equal-weighted portfolio, indicating a scarcity of samples that are statistically independent, especially when they are collected over time.

Long et al. (2023) interpret the objective of the robust satisficing model as the fragility level of the optimal solution in its ability to achieve the specified target. Specifically, the fragility of a solution $\boldsymbol{x} \in \mathcal{X}$ in achieving the target $\tau$, evaluated on the referenced distribution $\mathbb{Q} \in \mathcal{P}_0(\mathbb{R}^N)$, is defined as follows:

$$\begin{aligned} \varkappa(\boldsymbol{x}, \tau, \mathbb{Q}) := \min \ & \kappa \\ \text{s.t. } & \tau - \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right] \leq \kappa \Delta(\mathbb{P}, \mathbb{Q}) \quad \forall \mathbb{P} \in \mathcal{P}_0(\mathcal{Z}), \\ & \kappa \geq 0. \end{aligned} \tag{9}$$

Hence, $K_\tau = \min_{\boldsymbol{x} \in \mathcal{X}} \{\varkappa(\boldsymbol{x}, \tau, \hat{\mathbb{P}})\}$. Observe that the lower the fragility of a solution, the less the expected target shortfalls would be impacted by the disparity of the referenced distribution from a different distribution. In particular, when $\varkappa(\boldsymbol{x}, \tau, \hat{\mathbb{P}}) = 0$, then the target $\tau$ would always be attainable by the solution, *i.e.*,

$$\tau \leq f(\boldsymbol{x}, \boldsymbol{z}) \qquad \forall \boldsymbol{z} \in \mathcal{Z},$$

which also implies that

$$\varkappa(\boldsymbol{x}, \tau, \mathbb{Q}) = 0 \qquad \forall \mathbb{Q} \in \mathcal{P}_0(\mathbb{R}^N).$$

We remark that increasing the number of samples may not necessarily reduce the fragility of the optimal solution since it depends on the extent of the expected reward falling short of the target due to distributional shift. A carefully chosen target value can enhance the solution's resilience to distributional changes, and cross-validation may be used to determine this value. We also explore the benefits of using a finite sample confidence bound for less ideal sampling environment to justify the robust satisficing model and understand its performance under realistic conditions.

THEOREM 2. *Consider the random variable $\tilde{\boldsymbol{z}} \sim \mathbb{P}^\star$ and the joint random variable $(\tilde{\boldsymbol{z}}, \hat{\boldsymbol{z}}_1, \ldots, \hat{\boldsymbol{z}}_S) \sim \mathbb{Q}$, $\mathbb{Q} \in \mathcal{P}_0(\mathcal{Z}^{S+1})$. Suppose for all $s \in [S]$, $\mathbb{E}_{\mathbb{Q}}[\|\tilde{\boldsymbol{z}} - \hat{\boldsymbol{z}}_s\|] \le \mu$ and $\boldsymbol{\Sigma}$ is the covariance matrix of the random variable, $(\mathbb{E}_{\tilde{\boldsymbol{z}}\sim\mathbb{P}^\star}[\|\tilde{\boldsymbol{z}} - \hat{\boldsymbol{z}}_1\|], \ldots, \mathbb{E}_{\tilde{\boldsymbol{z}}\sim\mathbb{P}^\star}[\|\tilde{\boldsymbol{z}} - \hat{\boldsymbol{z}}_S\|])$. Then, the optimal solution to the robust satisficing problem (6) in which $\tau \le \hat{Z}$ has the following profile of confidence bounds:*

$$\mathbb{Q}\left[\tau - \mathbb{E}_{\mathbb{P}^\star}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right] > \mu K_\tau r\right] \le \frac{\mathbf{1}^\top \boldsymbol{\Sigma} \mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Sigma} \mathbf{1} + \mu^2 S^2 (r-1)^2} \qquad \forall r \ge 1.$$

EXAMPLE 1. To provide a numerical illustration of Theorem 2, we consider a 50-dimensional multivariate standard normal distributed random variable $\tilde{\boldsymbol{z}} \sim \mathbb{P}^\star$ and use the $\ell_2$-norm to specify the Wasserstein metric. By simulation, we obtain the mean and standard deviation of $\mathbb{E}_{\tilde{\boldsymbol{z}}\sim\mathbb{P}^\star}\left[\|\tilde{\boldsymbol{z}} - \tilde{\boldsymbol{\xi}}\|_2\right]$, where $(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{\xi}}) \sim \mathbb{P}^\star \times \mathbb{P}^\star$, which are approximately $\mu = 9.95$ and $\sigma = 1.0$, respectively. Suppose $\boldsymbol{\Sigma} = \sigma^2(\rho \mathbf{1}\mathbf{1}^\top + (1-\rho)\boldsymbol{I})$ and $\rho = 0.1$, the optimal solutions to the robust satisficing problem (6) with $S = 100$ would have the following profile of confidence bounds:

$$\mathbb{Q}\left[\tau - \mathbb{E}_{\mathbb{P}^\star}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right] > \mu K_\tau r\right] \le \begin{cases} 1.0 & \text{if } r = 1.0 \\ 1.2 \times 10^{-3} & \text{if } r = 1.5 \\ 3.0 \times 10^{-4} & \text{if } r = 2.0. \end{cases}$$

In the aforementioned example, a proportional reduction in the fragility of the solution can have a substantial impact on its robustness against significant deviations from the target expected reward. While it may be practically prohibitive or costly to improve the confidence of target attainment by collecting more samples or reducing the dependency of samples, we can still reduce the fragility by setting a less ambitious target, $\tau$ and letting the robust satisficing model determine the least fragile solution for that target accordingly.

PROPOSITION 1. *Suppose $\mathcal{X}$ is a convex set and the reward function $f(\boldsymbol{x}, \boldsymbol{z})$ is concave in $\boldsymbol{x} \in \mathcal{X}$ for all $\boldsymbol{z} \in \mathcal{Z}$. Let $\tau_\lambda = (1-\lambda)\underline{Z} + \lambda \hat{Z}$, $\lambda \in [0,1]$, where $\underline{Z}$ is the worst-case reward given by*

$$\underline{Z} = \max_{\boldsymbol{x}\in\mathcal{X}} \min_{\boldsymbol{z}\in\mathcal{Z}} f(\boldsymbol{x}, \boldsymbol{z}).$$

*Then,*

$$K_{\tau_\lambda} \le \lambda K_{\hat{Z}}.$$

Proposition 1 highlights the trade-offs between setting ambitious targets and the fragility of solutions in achieving those targets. Setting a target that is always achievable, such as $\tau = \underline{Z}$, may be overly conservative, while choosing an overly ambitious target like $\tau = \hat{Z}$ would likely lead to disappointment in the outcome. For instance, under the condition of Theorem 2, the optimal solution to the robust satisficing problem (6) in which $\tau = \tau_\lambda$ has the following profile of confidence bounds:

$$\begin{aligned} \mathbb{Q}\left[\tau_\lambda - \mathbb{E}_{\mathbb{P}^\star}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right] > \lambda \mu K_{\hat{Z}} r\right] &\le \mathbb{Q}\left[\tau_\lambda - \mathbb{E}_{\mathbb{P}^\star}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right] > \mu K_{\tau_\lambda} r\right] \\ &\le \frac{\mathbf{1}^\top \boldsymbol{\Sigma} \mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Sigma} \mathbf{1} + \mu^2 S^2 (r-1)^2} \qquad \forall r \ge 1. \end{aligned}$$

Hence, selecting a target $\tau$ somewhere between these extremes can proportionally reduce the fragility of the solution, rendering the trade-offs manageable and explainable even under imperfect sampling conditions. The computational studies of Long et al. (2023), Sim et al. (2021), as well as in the present paper, show that by setting the target close to the objective of the empirical optimization model, $\hat{Z}$, we can often improve the out-of-sample performance over the empirical optimization model.

**Residual-based robust satisficing model**

Following from Kannan et al. (2020), we can also extend the robust satisficing model of Long et al. (2023) to incorporate the prediction model by replacing the empirical distribution $\hat{\mathbb{P}}$, with the predicted empirical distribution $\hat{\mathbb{P}}_{\hat{\boldsymbol{w}}}$ as follows,

$$\begin{aligned} K_\tau = \min\ & \kappa \\ \text{s.t.}\ & \tau - \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right] \le \kappa \Delta(\mathbb{P},\hat{\mathbb{P}}_{\hat{\boldsymbol{w}}}) \quad \forall \mathbb{P}\in\mathcal{P}_0(\mathcal{Z}), \\ & \boldsymbol{x}\in\mathcal{X}, \kappa\ge 0, \end{aligned} \tag{10}$$

which can be equivalently reformulated as the following robust optimization problem,

$$\begin{aligned} K_\tau = \min\ & \kappa \\ \text{s.t.}\ & \tau \le \frac{1}{S}\sum_{s\in[S]} \left(f(\boldsymbol{x},\boldsymbol{z}_s) + \kappa\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\hat{\boldsymbol{w}},\boldsymbol{u},s)\|\right) \quad \forall \boldsymbol{z}_s\in\mathcal{Z}, s\in[S], \\ & \boldsymbol{x}\in\mathcal{X}, \kappa\ge 0. \end{aligned} \tag{11}$$

Similar to Kannan et al. (2020), we call Problem (10) the residual-based robust satisficing model. We assume that Problem (11) is solvable and focus on the reward function $f(\boldsymbol{x},\boldsymbol{z})$ that is Lipschitz continuous with respect to $\boldsymbol{z}$, with maximum Lipschitz constant of $\bar{L}$ for all $\boldsymbol{x}\in\mathcal{X}$.

THEOREM 3. *The residual-based robust satisficing problem* (11) *is feasible for all* $\tau\le\hat{Z}$ *and* $K_\tau\le\bar{L}$. *Moreover, suppose* $\hat{\mathbb{P}}_{\hat{\boldsymbol{w}}}\in\mathcal{P}_0(\mathcal{Z})$ *and* $\tau=\hat{Z}$, *then its optimal solution is also optimal in the predict-then-optimization problem* (3).

We next show how the Wasserstein distance metric might deviate when the regression coefficients differ from $\hat{\boldsymbol{w}}$.

PROPOSITION 2. *For all* $\boldsymbol{w}\in\mathcal{W}$, $\mathbb{P}\in\mathcal{P}_0(\mathcal{Z})$,

$$|\Delta(\mathbb{P},\hat{\mathbb{P}}_{\hat{\boldsymbol{w}}}) - \Delta(\mathbb{P},\hat{\mathbb{P}}_{\boldsymbol{w}})| \le \frac{1}{S}\sum_{s\in[S]} \|\boldsymbol{h}(\boldsymbol{w}-\hat{\boldsymbol{w}},\boldsymbol{u}) - \boldsymbol{g}(\boldsymbol{w}-\hat{\boldsymbol{w}},\boldsymbol{u},\hat{\boldsymbol{u}}_s)\|.$$

Consequently, we introduce the estimation distance metric for evaluating how much regression coefficients $\hat{\boldsymbol{w}}$ might deviate from their true values $\boldsymbol{w}^*$.

Definition 1 (Estimation distance metric). An estimation distance metric $\Upsilon : \mathbb{R}^M \to \mathbb{R}_+$ is a norm or a semi-norm that satisfies

$$\bar{\theta} = \max_{\Upsilon(\boldsymbol{r})=1} \left\{ \frac{1}{S} \sum_{s \in [S]} \| \boldsymbol{h}(\boldsymbol{r}, \boldsymbol{u}) - \boldsymbol{g}(\boldsymbol{r}, \boldsymbol{u}, \hat{\boldsymbol{u}}_s) \| \right\} < \infty, \tag{12}$$

so that for all $\boldsymbol{w} \in \mathcal{W}$, $\mathbb{P} \in \mathcal{P}_0(\mathcal{Z})$,

$$|\Delta(\mathbb{P}, \hat{\mathbb{P}}_{\hat{\boldsymbol{w}}}) - \Delta(\mathbb{P}, \hat{\mathbb{P}}_{\boldsymbol{w}})| \leq \bar{\theta} \Upsilon(\boldsymbol{w} - \hat{\boldsymbol{w}}).$$

Note that unlike a norm, a semi-norm does not have the positive-definite property and $\Upsilon(\boldsymbol{r}) = 0$ does not necessarily imply $\boldsymbol{r} = \boldsymbol{0}$. As a special case, we define the nominal estimation distance metric as

$$\bar{\Upsilon}(\boldsymbol{r}) := \frac{1}{S} \sum_{s \in [S]} \| \boldsymbol{h}(\boldsymbol{r}, \boldsymbol{u}) - \boldsymbol{g}(\boldsymbol{r}, \boldsymbol{u}, \hat{\boldsymbol{u}}_s) \|, \tag{13}$$

which yields $\bar{\theta} = 1$. If the estimation problem is identifiable, then the estimation distance metric

$$\Upsilon(\boldsymbol{r}) = \frac{1}{S} \sum_{s \in [S]} \| \boldsymbol{h}(\boldsymbol{r}, \hat{\boldsymbol{u}}_s) \|, \tag{14}$$

would be a norm. We call this the prediction-based estimation distance metric because it is derived from evaluating the empirical prediction inaccuracy associated with the estimation uncertainty. Unlike the nominal estimation distance metric, the prediction-based estimation distance metric does not depend on the current side information, $\boldsymbol{u}$.

Proposition 3. *Under Assumption 1, let $\mathbb{Q}^\star \in \mathcal{P}_0(\mathcal{U} \times \mathcal{Z})$ be the joint distribution of the random variable $(\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}})$, $(\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}}) \sim \mathbb{Q}^\star$ and after observing $\boldsymbol{u}$, we have $\tilde{\boldsymbol{z}} \sim \mathbb{P}^\star$. Accordingly, we let $\mathbb{Q}^S$ be the distribution that governs the distribution of $S$ samples of $(\hat{\boldsymbol{u}}_s, \hat{\boldsymbol{v}}_s)$, for $s \in [S]$ independently drawn from $\mathbb{Q}^\star$. The vector of regression coefficients $\hat{\boldsymbol{w}}$ is obtained from solving Problem (2) using $(\hat{\boldsymbol{u}}_s, \hat{\boldsymbol{v}}_s)$, $s \in [S]$. We also let $\mathbb{P}^S$ denote the distribution that governs the distribution $\hat{\boldsymbol{z}}(\boldsymbol{w}^\star, \boldsymbol{u}, s)$ for $s \in [S]$. The optimal solution of the residual-based robust satisficing problem (10) in which $\tau \leq \hat{Z}$, has the following profile of confidence bounds:*

$$\mathbb{Q}^S \left[ \tau - \mathbb{E}_{\mathbb{P}^\star} \left[ f(\boldsymbol{x}, \tilde{\boldsymbol{z}}) \right] > K_\tau (r + \bar{\theta} \Upsilon(\boldsymbol{w}^\star - \hat{\boldsymbol{w}})) \right] \leq \mathbb{P}^S \left[ \Delta(\mathbb{P}^\star, \hat{\mathbb{P}}_{\boldsymbol{w}^\star}) > r \right] \qquad \forall r \geq 0.$$

Proposition 3 provides insight into how the probability of target attainment in the residual-based robust satisficing model is connected to the finite sample confidence guarantee of the Wasserstein ball, as analyzed in Fournier and Guillin (2015), as well as the accuracy of the estimated regression coefficients evaluated using the estimation distance metric. By minimizing $\kappa$ to the lowest value of $K_\tau$, the residual-based robust satisficing model is doing its best to reduce the impact of risk ambiguity and estimation uncertainty on the target shortfall.

We note that the accuracy of the prediction model is a crucial factor that affects the target shortfall, and significant estimation uncertainty can profoundly impact the target shortfall. This contrasts the robust satisficing model proposed by Long et al. (2023), which does not consider a prediction model. Ignoring estimation uncertainty can be problematic when the accuracy of predictions depends on how well the underlying regression coefficients can be estimated from available data. When the estimation uncertainty is high, it would increase the risks of target shortfalls at larger magnitudes.

We note that, following the same reasoning as in Theorem 1, the consistency analysis of the residual-based robust satisficing problem can be related to the analysis of its corresponding residual-based robust optimization problem, as examined in Kannan et al. (2020). Additionally, Theorem 2 can be extended to investigate the optimal solution of the residual-based robust satisficing model under suboptimal sampling conditions. However, addressing estimation uncertainty would yield even greater benefits.

**Estimation-fortified robust satisficing model**

For a given solution $\boldsymbol{x} \in \mathcal{X}$, and target $\tau$, the actual fragility $\varkappa(\boldsymbol{x}, \tau, \hat{\mathbb{P}}_{\boldsymbol{w}^*})$ with respect to $\hat{\mathbb{P}}_{\boldsymbol{w}^\star}$ is unobservable. Nevertheless, we can still evaluate the estimated fragility, $\varkappa(\boldsymbol{x}, \tau, \hat{\mathbb{P}}_{\hat{\boldsymbol{w}}})$ with respect to the predicted empirical distribution $\hat{\mathbb{P}}_{\hat{\boldsymbol{w}}}$. However, although the optimal solutions generated by the residual-based robust satisficing model may result in the lowest estimated fragility, the actual fragility may still be higher. To provide extended flexibility for reducing the influence of estimation uncertainty, we introduce a less ambitious guarding target denoted by $\underline{\tau} \leq \tau$ and propose a new model. We explore solutions in a neighborhood where the estimated fragility with respect to the guarding target is bounded by $K_\tau$. Our aim is to find the most robust solutions within this neighborhood that are least affected by estimation uncertainty. The model, referred to as the estimation-fortified robust satisficing model, is formulated as follows:

$$
\begin{aligned}
\Theta_{\tau,\underline{\tau}} = \min\ & \theta \\
\text{s.t. }\ & \underline{\tau} - \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right] \leq K_\tau \left(\Delta(\mathbb{P}, \hat{\mathbb{P}}_{\boldsymbol{w}}) + \theta \Upsilon(\hat{\boldsymbol{w}} - \boldsymbol{w})\right) \quad \forall \mathbb{P} \in \mathcal{P}_0(\mathcal{Z}), \boldsymbol{w} \in \mathcal{W}, \\
& \boldsymbol{x} \in \mathcal{X}, \theta \geq 0,
\end{aligned}
\tag{15}
$$

which can be reformulated equivalently as the following robust optimization problem,

$$
\begin{aligned}
\Theta_{\tau,\underline{\tau}} = \min\ & \theta \\
\text{s.t. }\ & \underline{\tau} \leq \frac{1}{S} \sum_{s \in [S]} \left(f(\boldsymbol{x}, \boldsymbol{z}_s) + K_\tau \left(\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\boldsymbol{w}, \boldsymbol{u}, s)\| + \theta \Upsilon(\hat{\boldsymbol{w}} - \boldsymbol{w})\right)\right) \\
& \qquad \forall \boldsymbol{w} \in \mathcal{W}, \boldsymbol{z}_s \in \mathcal{Z}, s \in [S], \\
& \boldsymbol{x} \in \mathcal{X}, \theta \geq 0.
\end{aligned}
\tag{16}
$$

THEOREM 4. *The estimation-fortified robust satisficing problem (15) is feasible for $\underline{\tau} \leq \tau \leq \hat{Z}$ and $\Theta_{\tau,\underline{\tau}} \leq \bar{\theta}$. Suppose Problem (16) is solvable, then the following properties hold:*

- *For any feasible solution $\boldsymbol{x}$,*

$$\varkappa(\boldsymbol{x}, \underline{\tau}, \mathbb{P}_{\hat{\boldsymbol{w}}}) \leq K_\tau.$$

- *If $\Theta_{\tau,\underline{\tau}} = 0$, then the optimal solution $\hat{\boldsymbol{x}}$ satisfies*

$$\varkappa(\hat{\boldsymbol{x}}, \underline{\tau}, \mathbb{P}_{\boldsymbol{w}}) \leq K_\tau \qquad \forall \boldsymbol{w} \in \mathcal{W}.$$

- *Under the same conditions of Proposition 3, the optimal solution $\hat{\boldsymbol{x}}$ has the following profile of confidence bounds:*

$$\mathbb{Q}^S\left[\underline{\tau} - \mathbb{E}_{\mathbb{P}^\star}\left[f(\hat{\boldsymbol{x}}, \tilde{\boldsymbol{z}})\right] > K_\tau(r + \Theta_{\tau,\underline{\tau}}\Upsilon(\boldsymbol{w}^\star - \hat{\boldsymbol{w}}))\right] \leq \mathbb{P}^S\left[\Delta(\mathbb{P}^\star, \hat{\mathbb{P}}_{\boldsymbol{w}^\star}) > r\right] \qquad \forall r \geq 0.$$

Theorem 4 shows that the feasible solution of the estimation-fortified robust satisficing model guarantees an estimated fragility for the guarding target $\underline{\tau}$ is at most $K_\tau$. This means that when we set $\underline{\tau} = \tau$, the solution of the estimation-fortified robust satisficing model also solves the residual-based robust satisficing problem. However, by adjusting the distance between the guarding target and the desired target, the estimation-fortified robust satisficing model has the potential of improving the robustness of the solution in achieving the desired target by accounting for various degrees of estimation uncertainty. Therefore, the estimation-fortified robust satisficing model complements the robust satisficing model by offering a more flexible approach to managing estimation uncertainty. Due to Theorem 3, we note that setting $\tau = \hat{Z}$ enables the estimation-fortified robust satisficing model to improve the predict-then-optimize solution against estimation uncertainty.

It is important to note that the estimation-fortified robust satisficing model does not guarantee a solution that always achieves the guarding target. When the objective is set to zero, the estimated fragility of the optimal solutions with respect to the guarding target is bounded by a constant value, represented by $K_\tau$, regardless of the regression coefficients. However, even with this bound, it is not certain that the guarding target will almost surely be achieved by these optimal solutions. It is worth noting that if $K_\tau = 0$, then $\Theta_{\tau,\underline{\tau}} = 0$, and it is unnecessary to solve the estimation-fortified robust satisficing model if we have already found a robust solution that can almost surely achieve the target, $\tau$. Nonetheless, whenever $K_\tau > 0$, the estimation-fortified robust satisficing model complements the residual-based robust satisficing model, making its solution less sensitive to estimation uncertainty in achieving the target $\tau$.

In addition, Theorem 4 offers statistical support for the estimation-fortified robust satisficing model by relating to the convergence of the Wasserstein metric in an ideal sampling environment. This model minimizes the influence of estimation uncertainty by reducing $\theta$ to the lowest value of $\Theta_{\tau,\underline{\tau}}$, which helps limit the risk of falling short of the guarding target $\underline{\tau}$.

**Predict, optimize, satisfice, then fortify**

In formulating the estimation-fortified robust satisficing problem, one should prescribe model parameters based on outputs from three baseline models. We summarize the four-step process as follows.

1. **Predict.** Solve the estimation model, Problem (2), to obtain the baseline estimates $\hat{\boldsymbol{w}}$ of the regression coefficients for the linear prediction model.
2. **Optimize.** Solve the predict-then-optimize model, Problem (3), to determine the baseline return, $\hat{Z}$.
3. **Satisfice.** Set a target $\tau \leq \hat{Z}$. Solve the residual-based robust satisficing problem (10) and obtain the minimum estimated fragility level $K_\tau$. If $K_\tau = 0$, then return the corresponding optimal solution and the process stops here.
4. **Fortify.** Set a guarding target $\underline{\tau} \leq \tau$. Solve the estimation-fortified robust satisficing problem (15) for the guarding target $\underline{\tau}$ and estimated fragility level $K_\tau$. Return the corresponding optimal solution.

The final solution of the four-step process minimizes the influence of estimation uncertainty as much as possible while ensuring that the estimated fragility of the solution in achieving a less ambitious guarding target, $\underline{\tau}$ falls below the level for the desired target, $\tau$. To enhance the model's out-of-sample performance in practical scenarios, it is advisable to determine the desired values of the target parameters $\underline{\tau} \leq \tau \leq \hat{Z}$ through cross-validation, whenever feasible. As we will see in our computational studies, the four-step process significantly improves the out-of-sample expected reward over the predict-then-optimization solution.

## 4. Tractable optimization models

To obtain a tractable exact formulation for Problem (16), we focus on the case where the reward function $f(\boldsymbol{x}, \boldsymbol{z})$ is a saddle function.

Definition 2 (Saddle reward function). We say that the reward function $f(\boldsymbol{x}, \boldsymbol{z})$ is a saddle function if and only if for a given $\boldsymbol{x} \in \mathcal{X}$, the function is lower-semicontinuous and convex in $\boldsymbol{z} \in \mathcal{Z}$, and for a given $\boldsymbol{z} \in \mathcal{Z}$, the function is concave in $\boldsymbol{x} \in \mathcal{X}$.

We can reformulate the robust optimization problem (16) as a deterministic modest-sized convex optimization problem (see, *e.g.*, Ben-Tal et al. 2015).

Unfortunately, solving the estimation-fortified robust satisficing model would lead to a more complex model than the residual-based robust satisficing model. Even with $f(\boldsymbol{x}, \boldsymbol{z}) = \min\{\boldsymbol{a}^\top \boldsymbol{z}, 0\}$ and $\mathcal{Z} = \mathbb{R}^N$, the exact representation of the robust constraints is

$$\underline{\tau} \leq \frac{1}{S} \sum_{s \in [S]} \left( \min\{\boldsymbol{a}^\top \boldsymbol{z}_s, 0\} + K_\tau \left( \|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\boldsymbol{w}, \boldsymbol{u}, s)\| + \theta \Upsilon(\hat{\boldsymbol{w}} - \boldsymbol{w}) \right) \right)$$
$$\forall \boldsymbol{w} \in \mathcal{W}, \boldsymbol{z}_s \in \mathbb{R}^N, s \in [S],$$

and it is equivalent to

$$\tau \leq \min_{\substack{\boldsymbol{w}\in\mathcal{W}\\ \boldsymbol{z}_s\in\mathbb{R}^N, s\in[S]}} \left\{ \frac{1}{S}\sum_{s\in\mathcal{S}} \boldsymbol{a}^\top \boldsymbol{z}_s + \frac{K_\tau}{S}\sum_{s\in[S]} \left(\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\boldsymbol{w},\boldsymbol{u},s)\| + \theta\Upsilon(\hat{\boldsymbol{w}}-\boldsymbol{w})\right) \right\} \quad \forall \mathcal{S}\subseteq[S],$$

which has an exponential number of robust linear constraints. In contrast, the constraint of the residual-based robust satisficing problem has the following exact representation

$$\tau \leq \frac{1}{S}\sum_{s\in[S]} \left(\min\{\boldsymbol{a}^\top \boldsymbol{z}_s, 0\} + \kappa\left(\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\hat{\boldsymbol{w}},\boldsymbol{u},s)\|\right)\right) \qquad \forall \boldsymbol{z}_s\in\mathbb{R}^N, s\in[S],$$

which does not require to evaluate over $\boldsymbol{w}\in\mathcal{W}$, and could be tractably computed as follows,

$$\tau \leq \frac{1}{S}\sum_{s\in[S]} \left(\min\left\{ \min_{\boldsymbol{z}_s\in\mathbb{R}^N} \left\{\boldsymbol{a}^\top \boldsymbol{z}_s + \kappa\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\hat{\boldsymbol{w}},\boldsymbol{u},s)\|\right\}, 0\right\}\right).$$

We will next present how we can provide a tractable safe approximation to handle reward functions that are concave piecewise affine in the uncertain parameters.

**Linear optimization with recourse**

We can extend the framework to address adaptive linear optimization under risk ambiguity and estimation uncertainty, which opens up a plethora of useful applications. Here, $\boldsymbol{x}\in\mathcal{X}\subseteq\mathbb{R}^D$ represents the here-and-now variables, and $f(\boldsymbol{x},\boldsymbol{z})$ represents the total two-stage reward function by solving a linear optimization problem after $\boldsymbol{z}\in\mathcal{Z}$ has been realized as follows:

$$\begin{aligned} f(\boldsymbol{x},\boldsymbol{z}) = \max\ & \boldsymbol{d}^\top \boldsymbol{y} \\ \text{s.t.}\ & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \leq \boldsymbol{b}(\boldsymbol{z}), \\ & \boldsymbol{y}\in\mathbb{R}^{D_2}, \end{aligned} \tag{17}$$

where $\boldsymbol{A}:\mathbb{R}^N\to\mathbb{R}^{M\times D}$ and $\boldsymbol{b}:\mathbb{R}^N\to\mathbb{R}^M$ are affine maps of $\boldsymbol{z}$ as defined below:

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

We assume that Problem (17) has complete recourse, *i.e.*, for every $\boldsymbol{t}\in\mathbb{R}^M$, there exists $\boldsymbol{y}\in\mathbb{R}^{D_2}$ such that $\boldsymbol{B}\boldsymbol{y}\leq\boldsymbol{t}$. The corresponding predict-then-optimize model is given by:

$$\begin{aligned} \hat{Z} = \max\ & \frac{1}{S}\sum_{s\in[S]} \boldsymbol{d}^\top \boldsymbol{y}_s \\ \text{s.t.}\ & \boldsymbol{A}(\hat{\boldsymbol{z}}(\hat{\boldsymbol{w}},\boldsymbol{u},s))\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}_s \leq \boldsymbol{b}(\hat{\boldsymbol{z}}(\hat{\boldsymbol{w}},\boldsymbol{u},s)) && \forall s\in[S], \\ & \boldsymbol{y}_s\in\mathbb{R}^{D_2} && \forall s\in[S], \\ & \boldsymbol{x}\in\mathcal{X}. \end{aligned}$$

Although the residual-based robust satisficing problem is a harder problem, we can adopt the tractable safe approximation of Long et al. (2023) as follows,

$$\begin{aligned} \bar{K}_\tau = \min\ & \kappa \\ \text{s.t. } & \tau \leq \frac{1}{S}\sum_{s\in[S]} \left(\boldsymbol{d}^\top \boldsymbol{y}_s(\boldsymbol{z}_s,\phi_s) + \kappa\phi_s\right) && \forall (\boldsymbol{z}_s,\phi_s) \in \bar{\mathcal{Z}}_s, s\in[S], \\ & \boldsymbol{A}(\boldsymbol{z}_s)\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}_s(\boldsymbol{z}_s,\phi) \leq \boldsymbol{b}(\boldsymbol{z}_s) && \forall (\boldsymbol{z}_s,\phi_s) \in \bar{\mathcal{Z}}_s, s\in[S], \\ & \boldsymbol{y}_s \in \mathcal{L}^{N+1,D_2} && \forall s\in[S], \\ & \boldsymbol{x}\in\mathcal{X},\ k\geq 0, \end{aligned} \tag{18}$$

where

$$\bar{\mathcal{Z}}_s := \{(\boldsymbol{z}_s,\phi_s)\in\mathcal{Z}\times\mathbb{R} \mid \phi_s \geq \|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\hat{\boldsymbol{w}},\boldsymbol{u},s)\|\},$$

and

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

Note that since this is a safe approximation, we have $\bar{K}_\tau \geq K_\tau$. More importantly, if Problem (17) has complete recourse, then the approximation is also feasible for any target $\tau \leq \hat{Z}$. Moreover, in the same vein of Theorem 3, we can also establish that if $\hat{\mathbb{P}}_{\hat{\boldsymbol{w}}} \in \mathcal{P}_0(\mathcal{Z})$ and $\tau = \hat{Z}$, then its optimal solution is also optimal in the predict-then-optimization problem.

The estimation-fortified robust satisficing problem (16) has an equivalent adaptive linear optimization problem

$$\begin{aligned} \Theta_{\tau,\underline{\tau}} = \min\ & \theta \\ \text{s.t. } & \underline{\tau} \leq \frac{1}{S}\sum_{s\in[S]} \left(\boldsymbol{d}^\top \boldsymbol{y}(\boldsymbol{z}_s) + \bar{K}_\tau\left(\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\boldsymbol{w},\boldsymbol{u},s)\| + \theta\Upsilon(\hat{\boldsymbol{w}}-\boldsymbol{w})\right)\right) \\ & \qquad\qquad\qquad\qquad\qquad\qquad \forall \boldsymbol{w}\in\mathcal{W}, \boldsymbol{z}_s\in\mathcal{Z}, s\in[S], \\ & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) \leq \boldsymbol{b}(\boldsymbol{z}) \qquad \forall \boldsymbol{z}\in\mathcal{Z}, \\ & \boldsymbol{y}:\mathbb{R}^N\to\mathbb{R}^{D_2}, \\ & \boldsymbol{x}\in\mathcal{X}, \theta\geq 0. \end{aligned} \tag{19}$$

Accordingly, we propose a compatible safe tractable approximation of the estimation-fortified robust satisficing problem.

$$\begin{aligned} \bar{\Theta}_{\tau,\underline{\tau}} = \min\ & \theta \\ \text{s.t. } & \underline{\tau} \leq \frac{1}{S}\sum_{s\in[S]} \left(\boldsymbol{d}^\top \boldsymbol{y}_s(\boldsymbol{z}_s,\phi_s) + \bar{K}_\tau\eta\right) && \forall \left(\eta,(\boldsymbol{z}_s,\phi_s)_{s\in[S]}\right)\in\bar{\mathcal{Z}}(\theta), \\ & \boldsymbol{A}(\boldsymbol{z}_s)\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}_s(\boldsymbol{z}_s,\phi_s) \leq \boldsymbol{b}(\boldsymbol{z}_s) && \forall (\boldsymbol{z}_s,\phi_s)\in\bar{\mathcal{Z}}_s, s\in[S], \\ & \boldsymbol{y}_s\in\mathcal{L}^{N+1,D_2} && \forall s\in[S], \\ & \boldsymbol{x}\in\mathcal{X},\ \theta\in[0,\bar{\vartheta}], \end{aligned} \tag{20}$$

where

$$\bar{\mathcal{Z}}(\theta) := \left\{ \left(\eta, (\boldsymbol{z}_s, \phi_s)_{s\in[S]}\right) \;\middle|\; \begin{array}{l} \eta \in \mathbb{R}, (\boldsymbol{z}_s, \phi_s) \in \bar{\mathcal{Z}}_s \;\forall s \in [S] \\ \eta \geq \dfrac{1}{S} \displaystyle\sum_{s\in[S]} \left(\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\boldsymbol{w}, \boldsymbol{u}, s)\| + \theta \Upsilon(\hat{\boldsymbol{w}} - \boldsymbol{w})\right) \\ \eta \geq \dfrac{\theta}{\bar{\vartheta} S} \displaystyle\sum_{s\in[S]} \phi_s \\ \text{for some } \boldsymbol{w} \in \mathcal{W} \end{array} \right\},$$

for any fixed $\bar{\vartheta}$, $\bar{\vartheta} \geq \bar{\theta}$. Note that we can choose $\bar{\vartheta} = 1$ for any estimation distance metric that dominates the nominal metric, *i.e.*, for all $\boldsymbol{r} \in \mathbb{R}^M$, $\Upsilon(\boldsymbol{r}) \geq \bar{\Upsilon}(\boldsymbol{r})$.

THEOREM 5. *Problem* (20) *is a conservative approximation of the estimation-fortified robust satisficing problem* (19). *Moreover, any feasible solution of Problem* (18) *would also be feasible in Problem* (20). *Furthermore, if* $\underline{\tau} = \tau$, *then any feasible solution of Problem* (20) *would also be optimal in solution in Problem* (18).

Theorem 5 demonstrates the compatibility of the safe approximations for both robust satisficing models, which is essential for the four-step process. Although not an exact approach, if $\hat{\mathbb{P}}_{\hat{\boldsymbol{w}}} \in \mathcal{P}_0(\mathcal{Z})$ and $\underline{\tau} = \tau = \hat{Z}$, the safe approximation of the estimation-fortified robust satisficing problem could still provide optimal solutions that are also optimal in the predict-then-optimization problem.

Observe that unlike the safe approximation of the residual-based robust satisficing problem, Problem (20) has an uncertainty set $\bar{\mathcal{Z}}(\theta)$ that is dependent on the decision variable $\theta$. Since the set $\bar{\mathcal{Z}}(\theta)$ is non-increasing in $\theta$, we can solve Problem (20) by binary search on $\theta$. In each sub-problem, we solve the following maximization problem

$$\begin{aligned} T_\tau(\theta) = \max\ & t \\ \text{s.t. } & t \leq \frac{1}{S} \sum_{s\in[S]} \left(\boldsymbol{d}^\top \boldsymbol{y}_s(\boldsymbol{z}_s, \phi_s) + \bar{K}_\tau \eta\right) && \forall \left(\eta, (\boldsymbol{z}_s, \phi_s)_{s\in[S]}\right) \in \bar{\mathcal{Z}}(\theta), \\ & \boldsymbol{A}(\boldsymbol{z}_s)\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}_s(\boldsymbol{z}_s, \phi_s) \leq \boldsymbol{b}(\boldsymbol{z}_s) && \forall (\boldsymbol{z}_s, \phi_s) \in \bar{\mathcal{Z}}_s, s \in [S], \\ & \boldsymbol{y}_s \in \mathcal{L}^{N+1, D_2} && \forall s \in [S], \\ & \boldsymbol{x} \in \mathcal{X}, \end{aligned} \tag{21}$$

so that $\bar{\Theta}_{\tau,\underline{\tau}} = \min\{\theta | \underline{\tau} \leq T_\tau(\theta), \theta \in [0, \bar{\vartheta}]\}$.

**Decision-dependent prediction**

There are many well-motivated examples that incorporate decisions as inputs to the prediction, such as multi-product pricing problems. We can extend the model when $\boldsymbol{u}$ is part of the here-and-now decisions from the joint feasibility set $(\boldsymbol{x}, \boldsymbol{u}) \in \mathcal{Y}$, $\mathcal{Y} \subseteq \mathcal{X} \times \bar{\mathcal{U}}$, $\bar{\mathcal{U}} \subseteq \mathcal{U}$. Suppose $\bar{\mathcal{U}}$ is a discrete set; we can exhaustively enumerate $\boldsymbol{u} \in \bar{\mathcal{U}}$ and solve each robust satisficing problem using the prediction from $\boldsymbol{u}$. We can also explore mixed integer optimization methods to formulate the problem. Suppose

$\bar{\mathcal{U}}$ is a convex set, to obtain a tractable convex optimization problem, we have to ensure that the uncertain parameters associated with the robust satisficing model are not dependent on $\boldsymbol{u}$. We require $\mathcal{Z} = \mathbb{R}^N$ and homoscedasticity, so that $\boldsymbol{g}(\boldsymbol{w}, \boldsymbol{u}, \hat{\boldsymbol{u}}_s) = \boldsymbol{h}(\boldsymbol{w}, \hat{\boldsymbol{u}}_s)$.

The residual-based robust satisficing has the following equivalent representation,

$$\begin{aligned} K_\tau = \min\ & \kappa \\ \text{s.t. } & \tau \leq \frac{1}{S}\sum_{s\in[S]} \left(f(\boldsymbol{x}, \boldsymbol{z}_s) + \kappa \|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\hat{\boldsymbol{w}}, \boldsymbol{u}, s)\|\right) \ \forall \boldsymbol{z}_s \in \mathbb{R}^N, s \in [S], \\ & (\boldsymbol{x}, \boldsymbol{u}) \in \mathcal{Y}, \kappa \geq 0. \end{aligned}$$

With the change of variables, $\boldsymbol{z}_s = \boldsymbol{h}(\hat{\boldsymbol{w}}, \boldsymbol{u}) + \boldsymbol{\epsilon}_s$, we obtain

$$\begin{aligned} K_\tau = \min\ & \kappa \\ \text{s.t. } & \tau \leq \frac{1}{S}\sum_{s\in[S]} \Big(f(\boldsymbol{x}, \boldsymbol{h}(\hat{\boldsymbol{w}}, \boldsymbol{u}) + \boldsymbol{\epsilon}_s) + \kappa \Big(\|\boldsymbol{\epsilon}_s - \hat{\boldsymbol{\zeta}}(\hat{\boldsymbol{w}}, s)\|\Big)\Big) \ \forall \boldsymbol{\epsilon}_s \in \mathbb{R}^N, s \in [S], \\ & (\boldsymbol{x}, \boldsymbol{u}) \in \mathcal{Y}, \kappa \geq 0. \end{aligned} \tag{22}$$

Accordingly, the estimation-fortified robust satisficing problem is,

$$\begin{aligned} \Theta_{\tau,\underline{\tau}} = \min\ & \theta \\ \text{s.t. } & \underline{\tau} \leq \frac{1}{S}\sum_{s\in[S]} \left(f(\boldsymbol{x}, \boldsymbol{z}_s) + K_\tau \left(\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\boldsymbol{w}, \boldsymbol{u}, s)\| + \theta \Upsilon(\hat{\boldsymbol{w}} - \boldsymbol{w})\right)\right) \ \forall \boldsymbol{z}_s \in \mathbb{R}^N, \boldsymbol{w} \in \mathcal{W}, s \in [S], \\ & (\boldsymbol{x}, \boldsymbol{u}) \in \mathcal{Y}, \theta \geq 0. \end{aligned}$$

Ostensibly, we cannot consider the nominal estimation distance metric for the estimation-fortified robust satisficing model because of its dependency on $\boldsymbol{u}$. If the estimation problem is identifiable, the prediction-based estimation distance metric is also a viable option.

With the change of variables, $\boldsymbol{z}_s = \boldsymbol{h}(\boldsymbol{w}, \boldsymbol{u}) + \boldsymbol{\epsilon}_s$, we obtain

$$\begin{aligned} \Theta_{\tau,\underline{\tau}} = \min\ & \theta \\ \text{s.t. } & \underline{\tau} \leq \frac{1}{S}\sum_{s\in[S]} \Big(f(\boldsymbol{x}, \boldsymbol{h}(\boldsymbol{w}, \boldsymbol{u}) + \boldsymbol{\epsilon}_s) + K_\tau \Big(\|\boldsymbol{\epsilon}_s - \hat{\boldsymbol{\zeta}}(\boldsymbol{w}, s)\| + \theta \Upsilon(\hat{\boldsymbol{w}} - \boldsymbol{w})\Big)\Big) \\ & \qquad\qquad \forall \boldsymbol{\epsilon}_s \in \mathbb{R}^N, \boldsymbol{w} \in \mathcal{W}, s \in [S], \\ & (\boldsymbol{x}, \boldsymbol{u}) \in \mathcal{Y}, \theta \geq 0. \end{aligned} \tag{23}$$

For Problem (23) to be tractable convex optimization problems, the reward function has to be a saddle function, and $f(\boldsymbol{x}, \boldsymbol{h}(\boldsymbol{w}, \boldsymbol{u}) + \boldsymbol{\epsilon}_s)$ has to be jointly concave in $(\boldsymbol{x}, \boldsymbol{u})$ for all $\boldsymbol{w} \in \mathcal{W}$. In our second application, we provide an example of multi-product pricing to illustrate this.

Given that predictions are decision-dependent, we cannot assume that empirical samples are drawn independently. Nevertheless, Theorem 2 offer valuable insights into the robustness of the robust satisficing solutions in the less ideal sampling environment. The robust satisficing models strive to minimize the negative impact of risk ambiguity and estimation uncertainty, achieving an empirically attainable target, $\tau \leq \hat{Z}$. Although cross-validation is inapplicable due to decision-dependent predictions, we can set $\tau = \hat{Z}$ to compare with the predict-then-optimize solution. To establish a guarding target, we can define $\underline{\tau} = \hat{Z} - \chi\hat{\sigma}/\sqrt{S}$, where $\hat{\sigma}$ represents the sample standard deviation of the predict-then-optimize solution's estimated reward, and $\chi \geq 0$ corresponds to the specified confidence level, such as $\chi = 2$. This method proves effective for the multi-product pricing problem explored in Section 6.

## 5. Application I: Wine portfolio investment problem

We consider a stylized wine portfolio investment problem, which determines the portfolio of wines to procure from a collection of $N$ different types of wines to achieve an expected revenue of $\tau$ when the portfolio of wines is sold at the end of the investment period. It has well been known that the future price of wine highly depends on some observable factors such as regions and weather parameters (see, *e.g.*, Oczkowski 2016, Hekimoğlu et al. 2017). Hence, each wine type $n \in [N]$ can be associated with a vector of side information, $\boldsymbol{u}_n$, that could be used to predict its future price. We can employ the standard log-linear specification for modeling the price of wine $n$, denoted as $p_n$ (see, *e.g.*, Oczkowski 2016, Hekimoğlu and Kazaz 2020):

$$\tilde{z}_n = \log p_n = \boldsymbol{w}^\top \boldsymbol{u}_n + \tilde{\epsilon}_n,$$

which is also known as the "Bordeaux Equation" developed by Ashenfelter et al. (1995) and Ashenfelter (2008). A collection of $S$ observation pairs $(\hat{v}_s, \hat{\boldsymbol{u}}_s)$, $s \in [S]$, where $\hat{v}_s = \log(\hat{p}_s)$ and $\hat{p}_s$ is the price of wine that is associated with side information $\hat{\boldsymbol{u}}_s$.

**Predict-then-optimize model.** Given a normalized budget of one, the predict-then-optimize (PO) model would solve the following optimization problem,

$$\begin{aligned} \hat{Z} = \max \ & \sum_{n \in [N]} x_n \exp\left(\hat{\boldsymbol{w}}^\top \boldsymbol{u}_n\right) \hat{r} \\ \text{s.t. } & \boldsymbol{c}^\top \boldsymbol{x} \leq 1, \\ & \boldsymbol{x} \geq \boldsymbol{0}, \end{aligned} \tag{24}$$

where $c_n$, $n \in [N]$ is the unit cost of type $n$ wine and

$$\hat{r} = \frac{1}{S} \sum_{s \in [S]} \hat{p}_s \exp\left(-\hat{\boldsymbol{w}}^\top \hat{\boldsymbol{u}}_s\right)$$

is the average price variation that is attributed to the empirical residuals. We note that this simplification of Problem (3) applies only when the uncertain parameters appear linearly in the objective and the empirical residuals are perfectly correlated among different wine types.

**Residual-based robust models.** Kannan et al. (2020) use Wasserstein-based ambiguity sets constructed from sample residuals and propose the residual-based robust optimization (RO) model as follows:

$$\begin{aligned}&\max\ \tau\\ &\text{s.t.}\ \ \mathbb{E}_{\mathbb{P}}\left[\sum_{n\in[N]} x_n\exp(\tilde{z}_n)\right]\geq\tau\quad\forall\mathbb{P}\in\mathcal{F}_\Gamma,\\ &\qquad \boldsymbol{x}\in\mathcal{X},\end{aligned}$$

where

$$\mathcal{F}_\Gamma:=\left\{\mathbb{P}\in\mathcal{P}_0(\mathbb{R}^N)\ \middle|\ \Delta(\mathbb{P},\hat{\mathbb{P}}_{\hat{\boldsymbol{w}}})\leq\Gamma\right\},$$

and $\hat{\mathbb{P}}_{\hat{\boldsymbol{w}}}$ is the predicted empirical distribution where $\tilde{\boldsymbol{\xi}}\sim\hat{\mathbb{P}}$ indicates

$$\hat{\mathbb{P}}\left[\tilde{\xi}_n=\hat{\boldsymbol{w}}^\top\boldsymbol{u}_n+\hat{v}_s-\hat{\boldsymbol{w}}^\top\hat{\boldsymbol{u}}_s\right]=\frac{1}{S}\qquad\forall n\in[N],s\in[S].$$

Under $\ell_1$-norm, the residual-based RO is equivalent to the following problem,

$$\begin{aligned}&\max\ \tau-\kappa\Gamma\\ &\text{s.t.}\ \ \tau\leq\frac{1}{S}\sum_{s\in[S]}\left(\sum_{n\in[N]}\min_{z\in\mathbb{R}}\left\{x_n\exp(z)+\kappa|z-(\hat{\boldsymbol{w}}^\top\boldsymbol{u}_n+\hat{v}_s-\hat{\boldsymbol{w}}^\top\hat{\boldsymbol{u}}_s)|\right\}\right),\\ &\qquad \boldsymbol{x}\in\mathcal{X},\kappa\geq 0.\end{aligned}$$

For a given revenue target and under $\ell_1$-norm, $\tau<\hat{Z}$, the residual-based robust satisficing (RS) model is given as follows:

$$\begin{aligned}K_\tau=&\min\ \kappa\\ &\text{s.t.}\ \tau\leq\frac{1}{S}\sum_{s\in[S]}\sum_{n\in[N]}\left(x_n\exp\left(z_{s,n}\right)+\kappa\left(|z_{s,n}-(\hat{\boldsymbol{w}}^\top\boldsymbol{u}_n+\hat{v}_s-\hat{\boldsymbol{w}}^\top\hat{\boldsymbol{u}}_s)|\right)\right)\ \forall\boldsymbol{z}_s\in\mathbb{R}^N,s\in[S],\\ &\qquad \boldsymbol{c}^\top\boldsymbol{x}\leq 1,\\ &\qquad \boldsymbol{x}\geq\boldsymbol{0},\kappa\geq 0.\end{aligned}\tag{25}$$

**Estimation-fortified robust satisficing model.** We use an $\ell_2$-norm estimation distance metric, so the corresponding estimation-fortified (EF) robust satisficing model is

$$\begin{aligned}\Theta_{\tau,\underline{\tau}}=&\min\ \theta\\ &\text{s.t.}\ \underline{\tau}\leq\frac{1}{S}\sum_{s\in[S]}\sum_{n\in[N]}\left(x_n\exp\left(z_{s,n}\right)+K_\tau|z_{s,n}-(\boldsymbol{w}^\top\boldsymbol{u}_n+\hat{v}_s-\boldsymbol{w}^\top\hat{\boldsymbol{u}}_s)|\right)+\theta K_\tau\|\hat{\boldsymbol{w}}-\boldsymbol{w}\|_2\\ &\qquad\qquad \forall\boldsymbol{w}\in\mathbb{R}^J,\boldsymbol{z}_s\in\mathbb{R}^N,s\in[S],\\ &\qquad \boldsymbol{c}^\top\boldsymbol{x}\leq 1,\\ &\qquad \boldsymbol{x}\geq\boldsymbol{0},\theta\geq 0.\end{aligned}\tag{26}$$

THEOREM 6. *The RS model (25) has the following explicit formulation,*

$$\begin{aligned}
K_\tau = \min\ & \kappa \\
\text{s.t. } & \frac{1}{S}\sum_{s\in[S]}\sum_{n\in[N]}\Big(\phi_{s,n}(\hat{\boldsymbol{w}}^\top \boldsymbol{u}_n + \hat{v}_s - \hat{\boldsymbol{w}}^\top \hat{\boldsymbol{u}}_s) - \beta_{s,n}\Big) \geq \tau, \\
& \phi_{s,n}\log(\phi_{s,n}/x_n) \leq \phi_{s,n} + \beta_{s,n} && \forall s\in[S], n\in[N], \\
& \|\boldsymbol{\phi}_s\|_\infty \leq \kappa && \forall s\in[S], \\
& \boldsymbol{c}^\top \boldsymbol{x} \leq 1, \\
& \boldsymbol{x} \geq \boldsymbol{0}, \kappa \geq 0, \\
& \boldsymbol{\phi}_s \in \mathbb{R}^N_+, \boldsymbol{\beta}_s \in \mathbb{R}^N && \forall s\in[S].
\end{aligned} \tag{27}$$

*The EF model (26) has the following explicit formulation,*

$$\begin{aligned}
\Theta_{\tau,\underline{\tau}} = \min\ & \theta \\
\text{s.t. } & \frac{1}{S}\sum_{s\in[S]}\sum_{n\in[N]}(\phi_{s,n}\hat{v}_s - \beta_{s,n}) - \boldsymbol{\eta}^\top \hat{\boldsymbol{w}} \geq \underline{\tau}, \\
& \frac{1}{S}\sum_{s\in[S]}\sum_{n\in[N]}\phi_{s,n}(\boldsymbol{u}_n - \hat{\boldsymbol{u}}_s) + \boldsymbol{\eta} = \boldsymbol{0}, \\
& \phi_{s,n}\log(\phi_{s,n}/x_n) \leq \phi_{s,n} + \beta_{s,n} && \forall s\in[S], n\in[N], \\
& \|\boldsymbol{\phi}_s\|_\infty \leq K_\tau && \forall s\in[S], \\
& \|\boldsymbol{\eta}\|_2 \leq \theta K_\tau, \\
& \boldsymbol{c}^\top \boldsymbol{x} \leq 1, \\
& \boldsymbol{x} \geq \boldsymbol{0}, \theta \geq 0, \boldsymbol{\eta} \in \mathbb{R}^J, \\
& \boldsymbol{\phi}_s \in \mathbb{R}^N_+, \boldsymbol{\beta}_s \in \mathbb{R}^N && \forall s\in[S].
\end{aligned} \tag{28}$$

We follow the convention that $0\log(0/t) = 0$ if $t \geq 0$. Model (28) is a conic optimization problem involving exponential cones, which is currently supported in the commercial solver MOSEK (ApS 2021).

**Case study**

In this case study, we use the data on wine prices from Chapter 1 of The Analytics Edge (Bertsimas et al. 2016, p7), as shown in Table 1. This data set contains the historical log of prices of vintage and weather data from 1952 to 1980 (the prices of wine in 1954 and 1956 are missing and do not appear in the data set). Due to the limited data availability, we randomly select the data of $N = 5$ products as testing data and the rest as training data so that $S = 22$. Unfortunately, we do not have the price of each wine at different vintage years, which could be used to infer the cost of wine at the point of investment. Hence, in our numerical study, we generate the cost $c_n = \vartheta_n p_n$ for $n \in [N]$, where $\vartheta_n$ is randomly generated from a uniform distribution in $[0.3, 0.6]$. We then use the training sample to obtain estimates $\hat{\boldsymbol{w}}$. Subsequently, we solve the PO model (24), and

**Table 1 Vintage and weather data from The Analytics Edge (Bertsimas et al. 2016, p7).**

| Vintage | Price | Log of Price | Winter Rain (ml) | AGST (°C ) | Harvest Rain (ml) | Age of Vintage (yrs) |
|---|---|---|---|---|---|---|
| 1952 | 0.3684 | -0.99868 | 600 | 17.1167 | 160 | 31 |
| 1953 | 0.6348 | -0.45440 | 690 | 16.7333 | 80 | 30 |
| 1955 | 0.4458 | -0.80796 | 502 | 17.1500 | 130 | 28 |
| 1957 | 0.2211 | -1.50926 | 420 | 16.1333 | 110 | 26 |
| 1958 | 0.1797 | -1.71655 | 582 | 16.4167 | 187 | 25 |
| 1959 | 0.6584 | -0.41800 | 485 | 17.4833 | 187 | 24 |
| 1960 | 0.1388 | -1.97491 | 763 | 16.4167 | 290 | 23 |
| 1961 | 1.0000 | 0.00000 | 830 | 17.3333 | 38 | 22 |
| 1962 | 0.3310 | -1.10572 | 697 | 16.3000 | 52 | 21 |
| 1963 | 0.1685 | -1.78098 | 608 | 15.7167 | 155 | 20 |
| 1964 | 0.3059 | -1.18435 | 402 | 17.2667 | 96 | 19 |
| 1965 | 0.1063 | -2.24194 | 602 | 15.3667 | 267 | 18 |
| 1966 | 0.4726 | -0.74943 | 819 | 16.5333 | 86 | 17 |
| 1967 | 0.1913 | -1.65388 | 714 | 16.2333 | 118 | 16 |
| 1968 | 0.1054 | -2.25018 | 610 | 16.2000 | 292 | 15 |
| 1969 | 0.1167 | -2.14784 | 575 | 16.5500 | 244 | 14 |
| 1970 | 0.4044 | -0.90544 | 622 | 16.6667 | 89 | 13 |
| 1971 | 0.2724 | -1.30031 | 551 | 16.7667 | 112 | 12 |
| 1972 | 0.1014 | -2.28879 | 536 | 14.9833 | 158 | 11 |
| 1973 | 0.1561 | -1.85700 | 376 | 17.0667 | 123 | 10 |
| 1974 | 0.1108 | -2.19958 | 574 | 16.3000 | 184 | 9 |
| 1975 | 0.3007 | -1.20168 | 572 | 16.9500 | 171 | 8 |
| 1976 | 0.2534 | -1.37264 | 418 | 17.6500 | 247 | 7 |
| 1977 | 0.1070 | -2.23503 | 821 | 15.5833 | 87 | 6 |
| 1978 | 0.2704 | -1.30769 | 763 | 15.8167 | 51 | 5 |
| 1979 | 0.2145 | -1.53960 | 717 | 16.1667 | 122 | 4 |
| 1980 | 0.1359 | -1.99582 | 578 | 16.0000 | 74 | 3 |

obtain $\boldsymbol{x}^{\mathrm{PO}}$ and $\hat{Z}$. The revenues generated by investment decisions are then evaluated on the same test data set. We repeat this process for 100 iterations and derive the simulated expected return, defined as the expected revenue subtracted by the budget of one. We also evaluate the out-of-sample performance improvement on expected return between the two models. For example, the percentage improvement of the EF model against the PO model is denoted as $\frac{\mathrm{EF}-\mathrm{PO}}{\mathrm{PO}}\%$.

We solve all the models using RSOME (a python-based algebraic modeling package, see Chen and Xiong (2023) and https://xiongpengnus.github.io/rsome/), which can handle exponential cones with the MOSEK solver.

**Diversification of wine portfolio.** The PO model would recommend investing in a single wine type that exhibits the highest predicted expected return. This approach is both naive and risky.

**Table 2** **An example of wine investment portfolios obtained by different models.**

| Vintage | Cost | Predicted Return | Actual Return | Cost $\cdot\, x^\star_{\text{PO}}$ | Cost $\cdot\, x^\star_{\text{RS}}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | $\varphi=0.60$ | $\varphi=0.70$ | $\varphi=0.80$ | $\varphi=0.90$ | $\varphi=1.0$ |
| 1959 | 0.243 | 1.489 | 2.711 | 0.000 | 0.200 | 0.182 | 0.000 | 0.000 | 0.000 |
| 1962 | 0.152 | 2.491 | 2.172 | 1.000 | 0.200 | 0.231 | 0.440 | 0.614 | 1.000 |
| 1963 | 0.097 | 1.517 | 1.742 | 0.000 | 0.200 | 0.190 | 0.128 | 0.000 | 0.000 |
| 1965 | 0.046 | 1.471 | 2.287 | 0.000 | 0.200 | 0.170 | 0.000 | 0.000 | 0.000 |
| 1966 | 0.203 | 1.990 | 2.330 | 0.000 | 0.200 | 0.227 | 0.432 | 0.386 | 0.000 |
| Realized return | | | | 2.172 | 2.248 | 2.244 | 2.185 | 2.233 | 2.172 |

While it is possible to incorporate risk aversion into the PO model, diversification may not occur because the predicted empirical returns among different types of wines are perfectly correlated. In contrast, the robust optimization models will yield diversified wine investment portfolios. For instance, in Table 2, we show that reducing $\varphi$ can lead to a more diversified wine investment portfolio in the RS model. Interestingly, when the target is sufficiently low, the optimally robust portfolio becomes an equal-weighted portfolio, which despite its simplicity, is known to perform robustly well in actual portfolio optimization (DeMiguel et al. 2009).

**PO versus RO and RS.** We vary $\Gamma \in \{0, 0.1, \ldots, 1.9, 2.0\}$ and solve the RO model (27) to obtain $\boldsymbol{x}^{\text{RO}}$ for each $\Gamma$. For the RS model, we let $\tau = \varphi \hat{Z}$ and vary the target multiplier $\varphi \in \{0.70, 0.71, \ldots, 0.99, 0.999\}$ to obtain the corresponding $\boldsymbol{x}^{\text{RS}}$ for each $\varphi$. Note that since the RS model would be infeasible when $\tau > \hat{Z}$, to avoid numerical issues in the optimization problem, we set the largest value of $\varphi$ to 0.999 instead of 1.0. Note that if $\varphi \to 1.0$ ($\Gamma \to 0$), then the solution $\boldsymbol{x}^{\text{RS}}$ ($\boldsymbol{x}^{\text{RO}}$) converges to $\boldsymbol{x}^{\text{PO}}$. In Figure 1, we plot the out-of-sample performance improvement against PO on the expected return achieved by RO (left) and RS (right). We note that RO's performance improvement curve exhibits a less smooth trajectory compared to RS. Additionally, we can find a reasonable range of $\varphi \in [0.8, 1]$ where RS achieves a notable improvement in out-of-sample performance over the PO model. This range of parameters leading to improvements aligns with the ranges reported in recent computation studies of robust satisficing models (Long et al. 2023). Improvement by RO occurs for a range of radius parameters $\Gamma \in [0, 0.5]$. However, we note that the range of radius parameters leading to improvement is specific to the problem at hand. Without cross-validation, it would be challenging to identify suitable values of $\Gamma$ that would significantly improve over the predict-and-optimization solution.

**PO versus RS and EF.** To demonstrate the benefits of addressing estimation uncertainty, we compare the performance of the solutions obtained by PO, RS, and EF models. For the RS model, we record its maximum percentage improvement on the expected return against PO, as

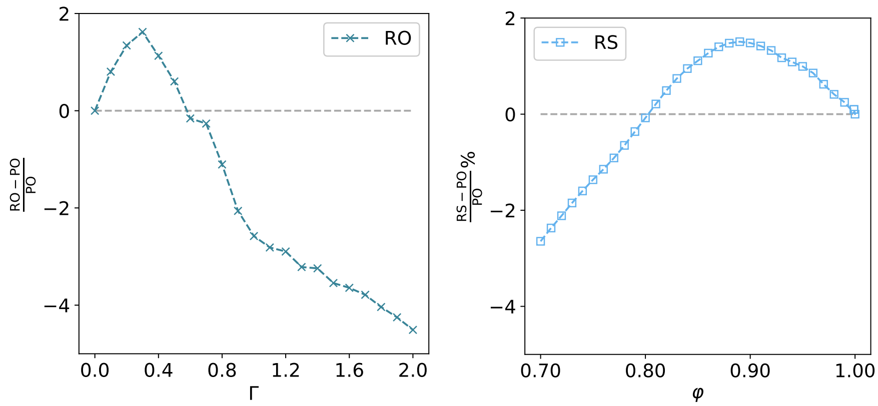


**Figure 1** **Out-of-sample performance improvement on the expected return of RO (left) and RS (right) against PO.**

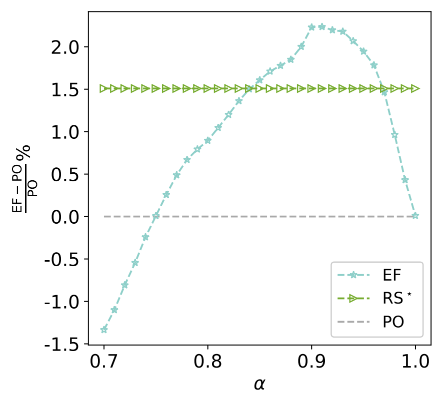


**Figure 2** **Out-of-sample percentage of improvement on the expected return of EF against PO (dashed line with $\star$) and the maximum percentage of improvement on the expected return of RS against PO (dashed line with $\triangleright$).**

shown in the right panel of Figure 1. For the EF model, we let $\tau = \hat{Z}$ and solve Problem (27) to obtain $K_{\hat{Z}}$. Subsequently, given $K_\tau = K_{\hat{Z}}$, we vary the guarding target $\underline{\tau} = \alpha\hat{Z}$ for some tuning parameter $\alpha \in \{0.70, 0.71, \ldots 0.99, 0.999\}$ and solve Problem (28) to obtain the $\boldsymbol{x}^{\text{EF}}$ for each $\alpha$. Note that $\alpha \to 1.0$ leads to the optimal PO wine portfolio investment decisions. The percentage improvements on the expected return of EF with respect to PO are presented in Figure 2, which also includes the maximum percentage improvement achieved by RS with respect to PO.

We observe that EF outperforms PO in terms of the out-of-sample expected returns for most instances of $\alpha \in [0.75, 1]$. Besides, there exists a range of $\alpha$ values such that EF has a higher expected return than RS. These findings highlight the performance benefits obtained by fortifying the solution against estimation uncertainty.

**Return-risk profiles of EF and RS.** We extend more numerical experiments that benchmark RS under various target parameters $\varphi \in \{0.80, 0.81, \ldots, 0.99, 0.999\}$. For each RS with target $\tau =$

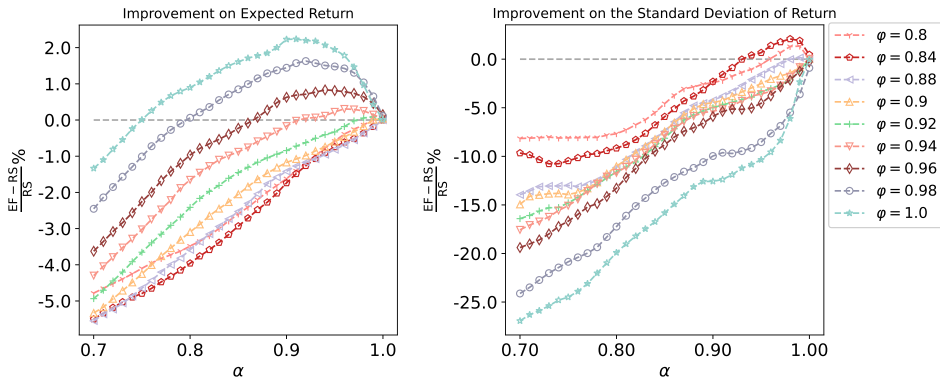


**Figure 3 Out-of-sample percentage of improvement on the expected return (left) and the percentage of improvement on the standard deviation of expected return (right) of EF against RS for different $\varphi$.**

$\phi\hat{Z}$, we obtain a series of solutions from the corresponding EF model under a set of parameter values $\alpha \in \{0.70, 0.71, \ldots, 0.999\}$. We report the out-of-sample improvement achieved by EF with respect to RS in terms of the expected return and its standard deviation in Figure 3. For clarity of the plot, we only include results for $\varphi \in \{0.80, 0.84, 0.88, 0.90, 0.92, \ldots, 0.98, 0.999\}$. For targets $\tau \geq 0.9\hat{Z}$, we observe that decreasing the guarding target for a fixed target always leads to a reduction in the standard deviation. Furthermore, we have identified certain regions where setting a marginally lower guarding target in EF can lead to an improvement in the expected reward over RS. In these regions, the EF model dominates the RS model with the same target in terms of risk profile, achieving higher expected returns and lower standard deviation. In particular, when compared to the predict-and-optimize solution, which corresponds to both the RS and EF models at $\varphi = \alpha = 1$, the EF model with $\varphi = 1$ and $\alpha = 0.9$ can generate portfolios that exhibit more than a 2% increase in expected returns, while simultaneously decreasing the standard deviations by 13%. However, we do not observe the dominance of EF over RS when $\varphi$ falls below 0.90. An intuitive explanation is that when the target falls below this level, the impact of risk ambiguity addressed in the RS model would have compensated for the influence of estimation uncertainty that the EF model tries to address.

These findings highlight the importance of carefully selecting the target and guarding target to achieve the desired balance between risk and reward. In the EF-Opt model, we consider only the $\alpha$ parameter that yields the best out-of-sample expected reward for each target $\tau = \varphi\hat{Z}$. The left panel of Figure 4 plots the return-risk frontiers and compares the EF-Opt and RS models for various $\varphi$. By minimizing the impact of estimation uncertainty, the EF models can further extend the reach of the efficient frontier and achieve higher expected returns. In the right panel of Figure 4, we plot the corresponding optimal $\alpha^*$ under different values of $\varphi$ used in plotting the

EF-Opt frontier. From the right panel of Figure 4, we propose the following heuristic of setting the parameter value $\alpha$ for each $\varphi$:

$$\alpha(\varphi) = \min\left\{1, \alpha + \frac{1-\varphi}{1-\underline{\varphi}}(1-\underline{\alpha})\right\}.$$

Here, $\alpha(1) = \underline{\alpha}$ and $\alpha(\underline{\varphi}) = 1$. For this example, we can use $\underline{\alpha} = 0.9$ and $\underline{\varphi} = 0.9$ to set $\alpha$ heuristically, potentially leading to higher expected revenues at similar risks as observed from the return-risk frontiers.

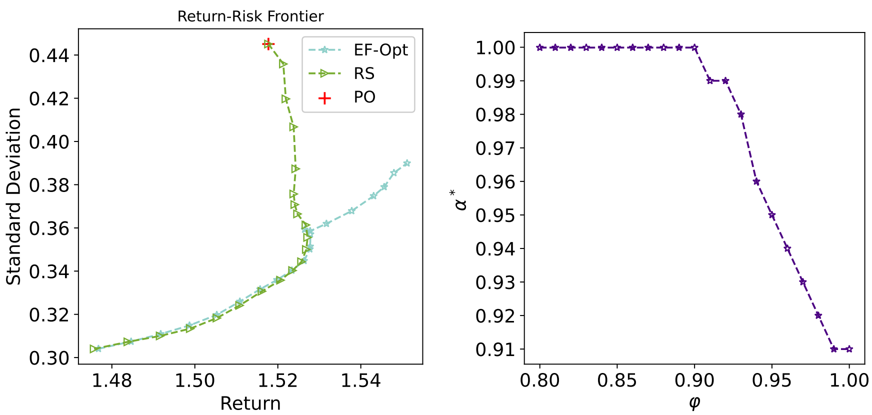


**Figure 4** **Return-risk frontiers of EF-Opt and RS for $\varphi \in [0.80, 1]$ and the best $\alpha^*$ with respect to $\varphi$.**

## 6. Application II: Multi-product pricing problem

In the second application, we investigate a classical multi-product pricing problem (Lim et al. 2008, Kluberg and Perakis 2012, Federgruen and Hu 2015) using historical data that we can use to forecast the demands of products from their prices. A retailer sells $N$ products in the market with demand $\tilde{z}_n$ for product $n \in [N]$, which is a function of price $\boldsymbol{p} \in [\underline{\boldsymbol{p}}, \bar{\boldsymbol{p}}] \subseteq \mathbb{R}^N$ as follows,

$$\tilde{z}_n = w_n + \boldsymbol{q}_n^\top \boldsymbol{p} + \tilde{\epsilon}_n,$$

where $\boldsymbol{q}_n, w_n$, $n \in [N]$, are some unobservable but deterministic parameters. This linear demand model is widely used in pricing problems (see, *e.g.,* Thiele 2009, Lu et al. 2014) and is shown to be "somewhat surprisingly, quite useful" (Besbes and Zeevi 2015). The total expected revenue $\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{p}, \boldsymbol{w} + \boldsymbol{Q}\boldsymbol{p} + \tilde{\boldsymbol{\epsilon}})]$ can be expressed as

$$\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{p}, \boldsymbol{w} + \boldsymbol{Q}\boldsymbol{p} + \tilde{\boldsymbol{\epsilon}})] = \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{p}^\top \boldsymbol{w} + \boldsymbol{p}^\top \boldsymbol{Q}\boldsymbol{p} + \boldsymbol{p}^\top \tilde{\boldsymbol{\epsilon}}\right].$$

The retailer has $S$ historical observation pairs, $\{(\hat{\boldsymbol{p}}_s, \hat{\boldsymbol{v}}_s), s \in [S]\}$, which enable her to solve the robust satisficing model that incorporates prediction under endogenous pricing.

We consider the multiproduct residual-based robust satisficing (RS) problem as follows:

$$\begin{aligned} K_\tau = \min\ & \kappa \\ \text{s.t. } & \tau \leq \frac{1}{S}\sum_{s\in[S]}\left(f(\boldsymbol{p}, \hat{\boldsymbol{w}} + \hat{\boldsymbol{Q}}\boldsymbol{p} + \boldsymbol{\epsilon}_s) + \kappa\left\|\boldsymbol{\epsilon}_s - \hat{\boldsymbol{v}}_s + \hat{\boldsymbol{w}} + \hat{\boldsymbol{Q}}\hat{\boldsymbol{p}}_s\right\|\right) \ \forall \boldsymbol{\epsilon}_s \in \mathbb{R}^N, s\in[S], \\ & \boldsymbol{p} \in [\underline{\boldsymbol{p}}, \bar{\boldsymbol{p}}], \kappa \geq 0. \end{aligned} \tag{29}$$

We use the prediction-based estimation distance metric,

$$\Upsilon(\boldsymbol{w}, \boldsymbol{Q}) = \frac{1}{S}\sum_{s\in[S]}\|\boldsymbol{w} + \boldsymbol{Q}\hat{\boldsymbol{p}}_s\|,$$

which yields reasonably good performance in our numerical study. We solve the following multi-product estimation-fortified (EF) robust satisficing model,

$$\begin{aligned} \Theta_{\tau,\underline{\tau}} = \min\ & \theta \\ \text{s.t. } & \underline{\tau} \leq \frac{1}{S}\sum_{s\in[S]}\left(f(\boldsymbol{p}, \boldsymbol{w} + \boldsymbol{Q}\boldsymbol{p} + \boldsymbol{\epsilon}_s) + K_\tau\left(\|\boldsymbol{\epsilon}_s - \hat{\boldsymbol{v}}_s + \boldsymbol{w} + \boldsymbol{Q}\hat{\boldsymbol{p}}_s\| + \theta\|\boldsymbol{w} - \hat{\boldsymbol{w}} + (\boldsymbol{Q} - \hat{\boldsymbol{Q}})\hat{\boldsymbol{p}}_s\|\right)\right) \\ & \qquad \forall \boldsymbol{\epsilon}_s \in \mathbb{R}^N, (\boldsymbol{w}, \boldsymbol{Q}) \in \mathcal{W}, s \in [S], \\ & \boldsymbol{p} \in [\underline{\boldsymbol{p}}, \bar{\boldsymbol{p}}], \theta \geq 0, \end{aligned} \tag{30}$$

where $\mathcal{W}$ is specified as

$$\mathcal{W} = \left\{(\boldsymbol{w}, \boldsymbol{Q}) \in \mathbb{R}^{N\times(1+N)} \;\middle|\; \boldsymbol{Q} + \boldsymbol{Q}^\top \preceq \boldsymbol{0}\right\}.$$

Here, we mandate $\boldsymbol{Q} + \boldsymbol{Q}^\top \preceq \boldsymbol{0}$ (*i.e.,* negative semidefinite), which is a common assumption in multi-product pricing literature (see, for example, Kluberg and Perakis 2012, Federgruen and Hu 2015, Armstrong and Vickers 2018).

THEOREM 7. *When $\hat{\boldsymbol{Q}} + \hat{\boldsymbol{Q}}^\top \preceq \boldsymbol{0}$, the multiproduct RS problem (29) can be reformulated as the following explicit semidefinite optimization problem:*

$$\begin{aligned} K_\tau = \min\ & \kappa \\ s.t.\ & \tau \leq \frac{1}{2}(\hat{\boldsymbol{Q}} + \hat{\boldsymbol{Q}}^\top) \bullet \boldsymbol{Y} + \frac{1}{S}\sum_{s\in[S]}\boldsymbol{p}^\top\left(\hat{\boldsymbol{v}}_s - \hat{\boldsymbol{Q}}\hat{\boldsymbol{p}}_s\right), \\ & \begin{pmatrix} \boldsymbol{Y} & \boldsymbol{p} \\ \boldsymbol{p}^\top & 1 \end{pmatrix} \succeq \boldsymbol{0}, \\ & \|\boldsymbol{p}\|_* \leq \kappa, \\ & \boldsymbol{Y} \in \mathbb{S}^N, \boldsymbol{p} \in [\underline{\boldsymbol{p}}, \bar{\boldsymbol{p}}], \kappa \geq 0. \end{aligned} \tag{31}$$

*The multiproduct EF problem (30) can be reformulated as the following explicit semidefinite optimization problem:*

$$
\begin{aligned}
\Theta_{\tau,\underline{\tau}} = \min\ & \theta \\
\text{s.t.}\ & \frac{1}{S}\sum_{s\in[S]} \left(\boldsymbol{p}^\top \hat{\boldsymbol{v}}_s + \boldsymbol{t}_s^\top \hat{\boldsymbol{w}} + \boldsymbol{t}_s^\top \hat{\boldsymbol{Q}}\hat{\boldsymbol{p}}_s\right) \geq \underline{\tau}, \\
& \begin{pmatrix} \boldsymbol{Y} & \boldsymbol{p} \\ \boldsymbol{p}^\top & 1 \end{pmatrix} \succeq \boldsymbol{0}, \\
& \boldsymbol{Y} = \frac{1}{S}\sum_{s\in[S]} (\boldsymbol{p}+\boldsymbol{t}_s)\hat{\boldsymbol{p}}_s^\top, \\
& \frac{1}{S}\sum_{s\in[S]} \boldsymbol{t}_s = \boldsymbol{0}, \\
& \|\boldsymbol{p}\|_* \leq K_\tau, \\
& \|\boldsymbol{t}_s\|_* \leq K_\tau \theta & \forall s\in[S], \\
& \boldsymbol{t}_s \in \mathbb{R}^N & \forall s\in[S], \\
& \boldsymbol{Y}\in\mathbb{S}^N, \boldsymbol{p}\in[\underline{\boldsymbol{p}},\bar{\boldsymbol{p}}], \\
& \theta \geq 0.
\end{aligned}
\tag{32}
$$

Theorem 7 states that the residual-based multiproduct pricing problem can be solved as a semidefinite programming problem. The estimation-fortified multiproduct pricing problem can be cast as a convex conic optimization problem, despite the presence of endogenous pricing decisions in demand estimation. Specifically, if we choose the $\ell_1$-norm for the Wasserstein and the prediction discrepancy metrics, then Problem (32) would have only one semidefinite constraint and the rest are linear constraints. In contrast, if we use the $\ell_2$-norm, there would be one semidefinite constraint and $S+1$ second-order-cone constraints. As such, the resulting Problem (32) can be practically solved by available solvers such as MOSEK and SDPT3 (Toh et al. 1999).

**Benchmark models.** We benchmark the performance of our proposed model against the predict-then-optimize (PO) model:

$$
\hat{Z} = \max_{\boldsymbol{p}\in[\underline{\boldsymbol{p}},\bar{\boldsymbol{p}}]} \left\{ \boldsymbol{p}^\top \hat{\boldsymbol{Q}}\boldsymbol{p} + \boldsymbol{p}^\top \hat{\boldsymbol{w}} + \frac{1}{S}\sum_{s\in[S]} \boldsymbol{p}^\top \hat{\boldsymbol{\epsilon}}_s \right\}, \tag{33}
$$

where $\hat{\boldsymbol{\epsilon}}_s = \hat{\boldsymbol{v}}_s - \hat{\boldsymbol{Q}}\hat{\boldsymbol{p}}_s - \hat{\boldsymbol{w}}$ is the empirical residual and $\hat{\boldsymbol{Q}}$ and $\hat{\boldsymbol{w}}$ are estimated parameters using regression approaches such as least square (LS), LASSO, or Ridge regression. We can formulate Problem (33) as a convex quadratic optimization problem if $\hat{\boldsymbol{Q}} + \hat{\boldsymbol{Q}}^\top$ is a negative semidefinite matrix. Otherwise, Problem (33) would not be a convex optimization problem, and we do not know how to solve such problems optimally with a reasonable computational effort. Note that even though the actual $\boldsymbol{Q}^\star + \boldsymbol{Q}^{\star\top}$ is negative semidefinite, the estimated $\hat{\boldsymbol{Q}} + \hat{\boldsymbol{Q}}^\top$ may not necessarily be a negative semidefinite matrix. As such, in our case study, we impose a negative semidefinite constraint on $\hat{\boldsymbol{Q}} + \hat{\boldsymbol{Q}}^\top$ when we obtain their estimates from data. For instance, in the case of LS, we solve the following constrained least square optimization problem:

$$(\hat{\boldsymbol{w}}, \hat{\boldsymbol{Q}}) = \underset{\boldsymbol{w};\, \boldsymbol{Q}+\boldsymbol{Q}^\top \preceq \boldsymbol{0}}{\arg\min} \left\{ \frac{1}{S} \sum_{s \in [S]} \| \boldsymbol{w} + \boldsymbol{Q}\hat{\boldsymbol{p}}_s - \hat{\boldsymbol{v}}_s \|_2^2 \right\}. \tag{34}$$

**Data-driven and confidence-adjusted guarding target.** Because of the endogenous outcomes, it would not be possible to perform cross-validation on the target parameter in the robust satisficing problem. To benchmark against the predict-then-optimize solution, we set $\tau = \hat{Z}$. The retailer may set the guarding target relative to the expected revenue generated by PO via a confidence level that the actual expected revenue might exceed this target. Specifically, for a given confidence level, $\chi$, we propose the setting guarding target $\underline{\tau}$, $\underline{\tau} = \hat{Z} - \chi\hat{\sigma}/\sqrt{S}$, where $\boldsymbol{p}^{\mathrm{PO}}$ is the optimal price of the PO problem (33), $\hat{\sigma} = \sqrt{\sum_{s\in[S]} \left( (\hat{\boldsymbol{\epsilon}}_s - \sum_{s\in S} \hat{\boldsymbol{\epsilon}}_s / S)^\top \boldsymbol{p}^{\mathrm{PO}} \right)^2 / (S-1)}$ is the sample standard deviation, and $\chi$ is a parameter directly associated with the confidence level, *e.g.,* $\chi \in \{1, 2, 3, 4\}$. In this way, we can obtain a guarding target revenue that is closely related to the historical data and the confidence level specified by the decision maker, regardless of the size of the data. As we will demonstrate in our case study comparison, for an appropriate range of confidence intervals, the EF model with a confidence-adjusted guarding target significantly improves the average revenue against the PO model in out-of-sample tests.

In what follows, we conduct a case study to demonstrate the effectiveness of our proposed EF model. We first demonstrate that higher prediction accuracy in the prediction model may not lead to better performance in the subsequent optimization model. Specifically, we compare PO models with prediction estimates obtained from LS, Ridge, and LASSO regression. By choosing PO with LS as a benchmark, we explore the performance of the EF model according to expected revenue. The case study was conducted in CVX (a MATLAB-based algebraic modeling package, see in http://cvxr.com/cvx/), using the MOSEK solver.

### Case study

We investigate the value of our proposed model by applying it to a real-world data set of an international cosmetics company. This data set was acquired from Perakis et al. (2023), where they selected twelve skin care products based on the sales volume. For more details of the selection of the products, please refer to Perakis et al. (2023). Similar to Perakis et al. (2023), we sequentially partition the twelve selected products into five groups, each containing four products. For example, group 1 contains products 1 - 4, group 2 includes products 3 - 6, and so forth. Hence, we emulate the random selection of four products in each group. This data set spans nearly a decade (from August 2007 to September 2016), and we have 3055 samples to estimate the parameters $\boldsymbol{w}$ and $\boldsymbol{Q}$.

Given the span of the data, we assume that the estimates ($\boldsymbol{w}$ and $\boldsymbol{Q}$) are the actual parameters (i.e., $\boldsymbol{w}^\star$ and $\boldsymbol{Q}^\star$). Using data sets, when the parameters are estimated via ordinary least squares, we verify that $\boldsymbol{Q}^\star + \boldsymbol{Q}^{\star\top}$ are negative semidefinite matrices.

Subsequently, we randomly select $S \in \{50, 100, 150, 200, 250\}$ price samples from historical data, denoted as $(\hat{\boldsymbol{p}}_s)_{s\in[S]}$, and its associated demands $(\hat{\boldsymbol{v}}_s)_{s\in[S]}$. As such, we have $5 \times 5 = 25$ experiments with varying product groups and sample sizes. For each experiment, we conduct 100 iterations and evaluate the average results.

Within each iteration, we first select $S$ training samples from the data and estimate $\hat{\boldsymbol{Q}}$ and $\hat{\boldsymbol{w}}$ by solving Problem (34). The nominal estimates $\hat{\boldsymbol{Q}}$ and $\hat{\boldsymbol{w}}$ then serve as the input parameters to the RS model. Then, we solve the PO model to obtain $\boldsymbol{p}^{\rm PO}$, $\hat{Z}$, and $\sigma_S$. We also solve the PO model with actual parameters $\boldsymbol{w}^\star$ and $\boldsymbol{Q}^\star$ to obtain the optimal price $\boldsymbol{p}^\star$. We denote this model as the true model and will use it to normalize the outputs, including the mean revenue and standard deviation of other models. We then solve the RS model on the training data with $\tau = \hat{Z}$ and solve the EF model with $\underline{\tau} = \hat{Z} - \chi\sigma_S$ and obtain the optimal prices.

For a given price, $\boldsymbol{p}$, we can obtain the expected revenue and the standard deviation of the revenue as $\boldsymbol{p}^\top\boldsymbol{\mu}(\boldsymbol{p})$ and $\sqrt{\boldsymbol{p}^\top\boldsymbol{\Sigma}\boldsymbol{p}}$, respectively. Here, $\boldsymbol{\mu}(\boldsymbol{p}) = \boldsymbol{w}^\star + \boldsymbol{Q}^\star\boldsymbol{p}$ is the mean demand that associates with price $\boldsymbol{p}$, and $\boldsymbol{\Sigma}$ is the covariance matrix of $\tilde{\boldsymbol{\epsilon}}$, the sample of which can be derived from $\hat{\boldsymbol{\epsilon}}_s = \hat{\boldsymbol{v}}_s - \boldsymbol{Q}^\star\hat{\boldsymbol{p}}_s - \boldsymbol{w}^\star$ for $s \in [S]$ over the entire data set. We then normalize the expected revenue and standard deviation of the EF model with corresponding counterparts of the true model. Finally, we average over all the iterations to obtain the average normalized revenue and standard deviation.

**The perils of the PO model.** We first investigate the performance of the PO model with estimates obtained from LS, Ridge, and LASSO regression. The performance comparison is summarized in Figure 5, where we also include the prediction accuracy of the three estimates. The metrics of LASSO and Ridge regressions are normalized with respect to the corresponding metrics of the PO model.

Figure 5 demonstrates that the PO model with estimates from Ridge regression leads to lower expected revenue on average compared to that with estimates from LS regression—as shown in the upper left panel of the figure—though the former improves the prediction accuracy across all the groups and samples, as elucidated by the upper right panel. Because we do not observe sparsity in $\boldsymbol{w}^\star$ and $\boldsymbol{Q}^\star$ from our data set, it is reasonable to believe that LASSO regression is not a good candidate as the underlying prediction model. As we illustrate in the lower left panel, the PO model with estimates from LASSO regression has poorer performance compared to the PO model with estimates obtained from LS regression. Incidentally, the difference is the most significant when

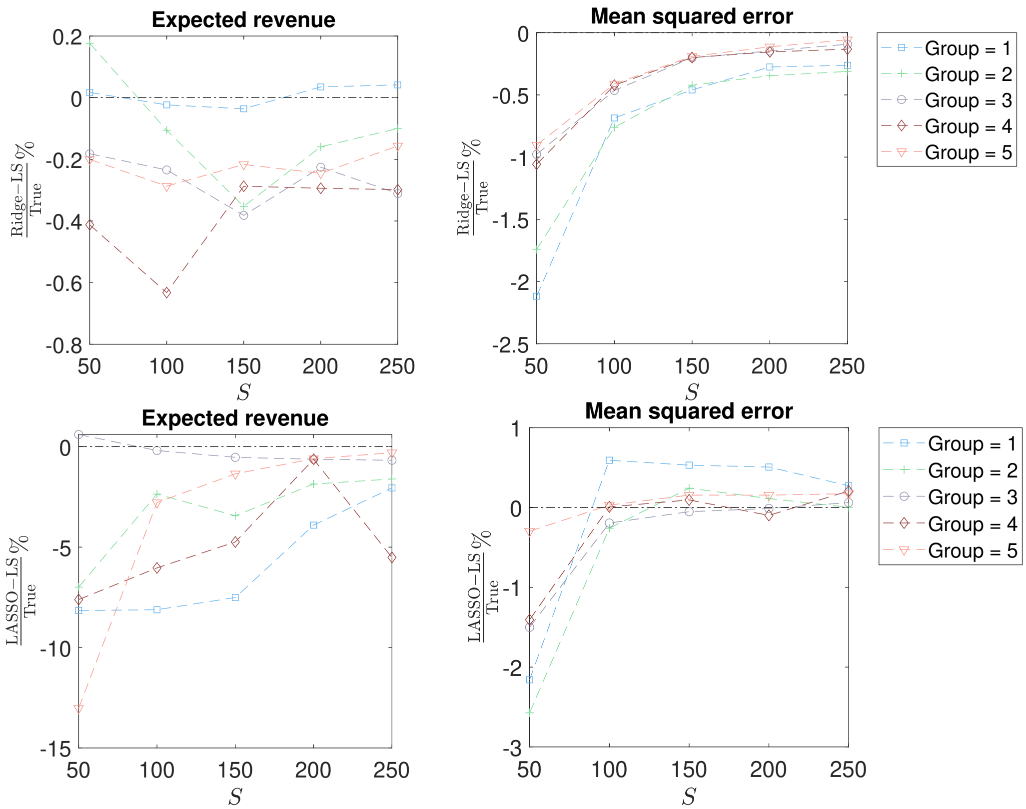


**Figure 5** **Out-of-sample percentage of improvement of PO with Ridge estimates against PO with LS estimates on expected revenue (upper left) and mean squared error (upper right), and that of PO with LASSO estimates against PO with LS estimates on expected revenue (lower left) and mean squared error (lower right).**

the sample size is small. In this scenario, LASSO regression has the best improvement in terms of prediction mean squared error compared to LS regression. This comparison exemplifies that higher prediction accuracy may not necessarily contribute to better performance in the subsequent optimization model. As such, in what follows, we only benchmark on the PO model with estimates from LS regression.

**Computation time of RS and EF.** We first demonstrate the computational tractability of both the RS and EF model. Table 3 summarizes the average computation times of the RS and EF models under different numbers of products and sample sizes when $\chi = 1.0$. We can see that the computation time for EF mildly increases with the number of products $N$ and modestly increases with the sample size $S$, and the computation time for RS is around 0.35 seconds across all the settings, indicating that both the RS and EF models are computationally scalable.

**Performance comparison between EF and PO.** We now compare the performance of the PO model with the EF model for $\chi = 1$. The performance comparison of the out-of-sample average revenue is summarized in Figure 6. We observe that EF outperforms PO in most of the experiments

**Table 3 Computation time (s) of the RB and EF models under different groups and sample sizes.**

| Group size (N) | 50 | | 100 | | 150 | | 200 | | 250 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | RS | EF | RS | EF | RS | EF | RS | EF | RS | EF |
| 4 | 0.39 | 1.45 | 0.37 | 2.16 | 0.35 | 2.93 | 0.35 | 3.66 | 0.36 | 4.72 |
| 6 | 0.37 | 1.42 | 0.36 | 2.32 | 0.35 | 3.04 | 0.36 | 3.93 | 0.34 | 4.75 |
| 8 | 0.35 | 1.38 | 0.39 | 2.43 | 0.38 | 3.42 | 0.37 | 4.22 | 0.36 | 5.18 |
| 10 | 0.36 | 1.68 | 0.35 | 2.68 | 0.33 | 3.75 | 0.34 | 4.93 | 0.34 | 6.12 |
| 12 | 0.35 | 1.76 | 0.38 | 2.84 | 0.36 | 4.00 | 0.35 | 5.21 | 0.34 | 6.60 |

(20/25). The average improvement is 4.30% over all the experiments, and the improvement can be as much as 13.90% for Group 2 when $S=50$. This demonstrates the benefits of accounting for the estimation uncertainty. Figure 6 also illustrates how the expected revenue changes with the size of the training sample. In general, the revenue improvement of EF over PO increases as the sample size decreases. This is partly due to the substantially low prediction accuracy when the sample size is small, while EF is robust to adversarial scenarios. Thus, the improvement is more pronounced when limited data is available for prediction.

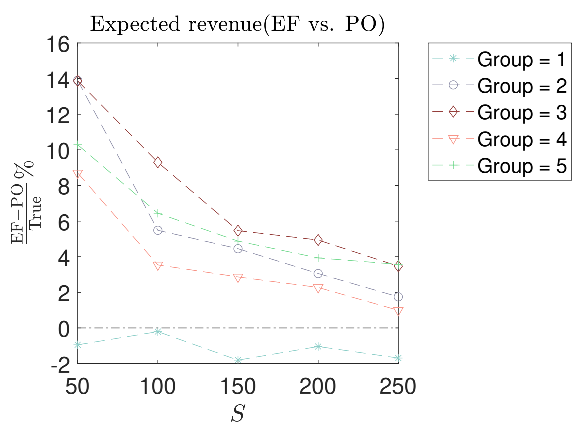


**Figure 6 Percentage of improvement on expected revenue of EF against PO in the out-of-sample test for $\chi = 1$.**

**The impact of confidence level $\chi$.** We conduct the experiments under other confidence levels $\chi \in \{1,2,3,4\}$. The performance comparisons are summarized in the left panel of Figure 7, where each box describes the out-of-sample distribution of the percentage improvement in expected revenue of EF over PO. For each $\chi \in \{1,2,3,4\}$, the test instances include experiments on all five product groups and sample sizes $S \in \{50,100,150,200,250\}$, *i.e.*, there are 25 data points for each box. From the box plot, we can see that EF consistently outperforms both PO in terms of the median of the expected revenues. This further confirms the robustness of our EF model

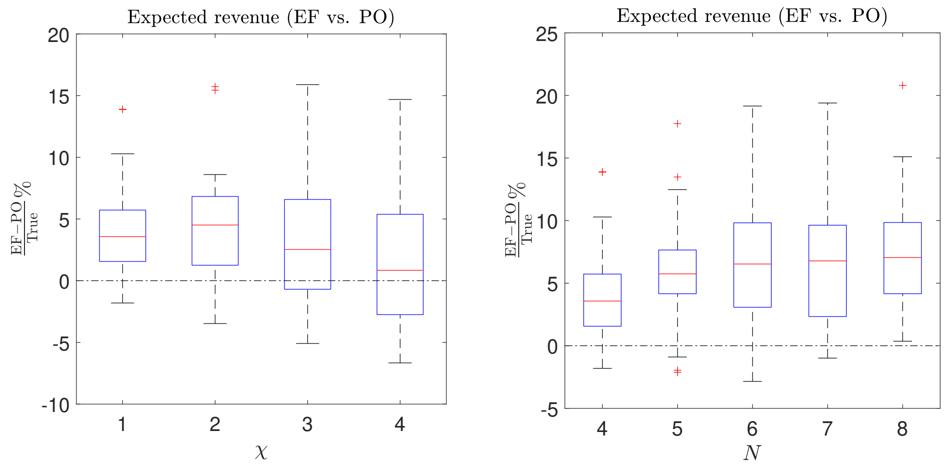


**Figure 7** **Percentage of improvement on expected revenue of EF against PO in the out-of-sample test varying $\chi \in \{1,2,3,4\}$ (left) and $N \in \{4,5,6,7,8\}$ (right).**

after accounting for the estimation uncertainty of the regression coefficients. In addition, performance improvement over the predict-then-optimize approach is not very sensitive to the choice of confidence level, $\chi$.

**The impact of the number of products $N$.** All previous experiments are conducted with each group containing four products. We now explore how the number of products in each group, $N \in \{4,5,6,7,8\}$, affects the performance of the EF and PO models. For example, if $N = 5$, group 1 contains products 1-5, group 2 includes products 3-7, and so forth. We fix $\chi = 1$ and conduct experiments for all five product groups and sample sizes $S \in \{50,100,150,200,250\}$ as before; hence, there are in total 25 experiments for each $N \in \{4,5,6,7,8\}$. We summarize the percentage improvement on the expected revenue of EF against PO in the right panel of Figure 7. We can see that EF consistently outperforms PO in terms of the median of the expected value, the improvement is more salient when there are more products. This further substantiates the benefits of incorporating estimation uncertainty in the robust satisficing model.

In summary, our estimation-fortified robust satisficing model could offer solutions to the multi-product pricing problem that result in higher expected revenues compared to the solutions obtained by the predict-then-optimize approach.

## 7. Conclusion

In conclusion, we have proposed a novel prescriptive analytics approach that leverages an estimation-fortified robust satisficing model to manage risk ambiguity and prediction uncertainty in decision-making. Our approach incorporates uncertain parameters in the form of linear predictions using side information. The accuracy of the prediction depends on the estimation of regression coefficients, which is almost never correct under limited data. Our proposed approach aims to address such estimation uncertainty. Through case studies on wine portfolio investment and multi-product pricing problems, we have demonstrated the effectiveness of our approach in achieving higher expected rewards compared to the predict-then-optimize approach when evaluated on the actual distribution. Our approach consistently outperforms benchmarks, particularly when data is limited. Overall, the proposed estimation-fortified robust satisficing model offers a tractable and statistically justified approach for various management decision-making scenarios when facing saddle reward functions, two-stage linear optimization problems, and decision-dependent predictions.

## References

ApS, Mosek. 2021. Mosek modeling cookbook.

Armstrong, Mark, John Vickers. 2018. Multiproduct pricing made simple. *Journal of Political Economy* **126**(4) 1444–1471.

Ashenfelter, Orley. 2008. Predicting the quality and prices of bordeaux wine. *The Economic Journal* **118**(529) F174–F184.

Ashenfelter, Orley, David Ashmore, Robert Lalonde. 1995. Bordeaux wine vintage quality and the weather. *Chance* **8**(4) 7–14.

Ban, Gah-Yi, Jérémie Gallien, Adam J Mersereau. 2019. Dynamic procurement of new products with covariate information: The residual tree method. *Manufacturing & Service Operations Management* **21**(4) 798–815.

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

Ben-Tal, Aharon, Dick Den Hertog, Jean-Philippe Vial. 2015. Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical Programming* **149**(1) 265–299.

Bertsimas, D., A.K. O'Hair, W.R. Pulleyblank. 2016. *The Analytics Edge*. Dynamic Ideas LLC.

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

Bertsimas, Dimitris, Christopher McCord, Bradley Sturt. 2023. Dynamic optimization with side information. *European Journal of Operational Research* **304**(2) 634–651.

Bertsimas, Dimitris, Bart Van Parys. 2022. Bootstrap robust prescriptive analytics. *Mathematical Programming* **195** 39–78.

Besbes, Omar, Assaf Zeevi. 2015. On the (surprising) sufficiency of linear models for dynamic pricing with demand learning. *Management Science* **61**(4) 723–739.

Blanchet, Jose, Yang Kang, Karthyek Murthy. 2019. Robust wasserstein profile inference and applications to machine learning. *Journal of Applied Probability* **56**(3) 830–857.

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

Chen, Lucy Gongtao, Qinshen Tang. 2022. Supply chain performance with target-oriented firms. *Manufacturing & Service Operations Management* **24**(3) 1714–1732.

Chen, Zhi, Peng Xiong. 2023. RSOME in Python: An open-source package for robust stochastic optimization made easy. *INFORMS Journal of Computing* .

Craig, Nathan C, Ananth Raman. 2016. Improving store liquidation. *Manufacturing & Service Operations Management* **18**(1) 89–103.

DeMiguel, Victor, Lorenzo Garlappi, Raman Uppal. 2009. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The Review of Financial Studies* **22**(5) 1915–1953.

Elmachtoub, Adam N, Paul Grigas. 2022. Smart "predict, then optimize". *Management Science* **68**(1) 9–26.

Esteban-Pérez, Adrián, Juan M Morales. 2022. Distributionally robust stochastic programs with side information based on trimmings. *Mathematical Programming* **195** 1069–1105.

Federgruen, Awi, Ming Hu. 2015. Multi-product price and assortment competition. *Operations Research* **63**(3) 572–584.

Ferreira, Kris Johnson, Bin Hong Alex Lee, David Simchi-Levi. 2016. Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & Service Operations Management* **18**(1) 69–88.

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

Gao, Rui. 2022. Finite-sample guarantees for wasserstein distributionally robust optimization: Breaking the curse of dimensionality. *Operations Research* .

Gao, Rui, Xi Chen, Anton J Kleywegt. 2017. Wasserstein distributional robustness and regularization in statistical learning. Available at arXiv: 1712.06050.

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

Goh, Joel, Nicholas G Hall. 2013. Total cost control in project management via satisficing. *Management Science* **59**(6) 1354–1372.

Green, Samuel B. 1991. How many subjects does it take to do a regression analysis. *Multivariate Behavioral Research* **26**(3) 499–510.

Grigas, Paul, Meng Qi, Max Shen. 2021. Integrated conditional estimation-optimization. *Available at arXiv:2110.12351* .

Hall, Nicholas G, Daniel Zhuoyu Long, Jin Qi, Melvyn Sim. 2015. Managing underperformance risk in project portfolio selection. *Operations Research* **63**(3) 660–675.

Hannah, Lauren, Warren Powell, David M Blei. 2010. Nonparametric density estimation for stochastic optimization with an observable state variable. *Advances in Neural Information Processing Systems*. 820–828.

Hao, Zhaowei, Long He, Zhenyu Hu, Jun Jiang. 2020. Robust vehicle pre-allocation with uncertain covariates. *Production and Operations Management* **29**(4) 955–972.

Hekimoğlu, Mert Hakan, Burak Kazaz. 2020. Analytics for wine futures: Realistic prices. *Production and Operations Management* **29**(9) 2096–2120.

Hekimoğlu, Mert Hakan, Burak Kazaz, Scott Webster. 2017. Wine analytics: Fine wine pricing and selection under weather and market uncertainty. *Manufacturing & Service Operations Management* **19**(2) 202–215.

Ho, Chin Pang, Grani A Hanasusanto. 2019. On data-driven prescriptive analytics with side information: A regularized Nadaraya-Watson approach. Tech. rep., Technical report, March.

Ho-Nguyen, Nam, Fatma Kılınç-Karzan. 2022. Risk guarantees for end-to-end prediction and optimization processes. *Management Science* **68**(12) 8680–8698.

Kannan, Rohit, Güzin Bayraksan, James R Luedtke. 2020. Residuals-based distributionally robust optimization with covariate information. Available at arXiv: 2012.01088.

Kannan, Rohit, Güzin Bayraksan, James R Luedtke. 2022. Data-driven sample average approximation with covariate information. *Available at arXiv:2207.13554* .

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

Kluberg, Jonathan, Georgia Perakis. 2012. Generalized quantity competition for multiple products and loss of efficiency. *Operations Research* **60**(2) 335–350.

Lim, Andrew, J George Shanthikumar, Thaisiri Watewai. 2008. Robust multi-product pricing. *Available at SSRN: 1078012* .

Liu, Sheng, Long He, Zuo-Jun Max Shen. 2021. On-time last-mile delivery: Order assignment with travel-time predictors. *Management Science* **67**(7) 4095–4119.

Liyanage, Liwan H, J George Shanthikumar. 2005. A practical inventory control policy using operational statistics. *Operations Research Letters* **33**(4) 341–348.

Loke, Gar Goei, Qinshen Tang, Yangge Xiao. 2021. Decision-driven regularization: A blended model for predict-then-optimize. *Available at SSRN: 3623006* .

Long, Daniel Zhuoyu, Melvyn Sim, Minglong Zhou. 2023. Robust satisficing. *Operations Research* **71**(1) 61–82.

Lu, Ye, Youhua Chen, Miao Song, Xiaoming Yan. 2014. Optimal pricing and inventory control policy with quantity-based price differentiation. *Operations Research* **62**(3) 512–523.

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

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

Mundru, Nishanth. 2019. Predictive and prescriptive methods in operations research and machine learning: an optimization approach. Ph.D. thesis, Massachusetts Institute of Technology.

Nguyen, Viet Anh, Fan Zhang, Jose Blanchet, Erick Delage, Yinyu Ye. 2021. Robustifying conditional portfolio decisions via optimal transport. *Available at arXiv: 2103.16451* .

Oczkowski, Edward. 2016. The effect of weather on wine quality and prices: An australian spatial analysis. *Journal of Wine Economics* **11**(1) 48–65.

Perakis, Georgia, Melvyn Sim, Qinshen Tang, Peng Xiong. 2023. Robust pricing and production with information partitioning and adaptation. *Management Science* **69**(3) 1398–1419.

Ramamurthy, Vivek, J George Shanthikumar, Zuo-Jun Max Shen. 2012. Inventory policy with parametric demand: Operational statistics, linear correction, and regression. *Production and Operations Management* **21**(2) 291–308.

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

Shafieezadeh-Abadeh, Soroosh, Daniel Kuhn, Peyman Esfahani Mohajerin. 2019. Regularization via mass transportation. *Journal of Machine Learning Research* **20**(103) 1–68.

Si, Nian, Jose Blanchet, Soumyadip Ghosh, Mark Squillante. 2020. Quantifying the empirical wasserstein distance to a set of measures: Beating the curse of dimensionality. *Advances in Neural Information Processing Systems* **33** 21260–21270.

Siegel, Andrew F, Michael R Wagner. 2021. Profit estimation error in the newsvendor model under a parametric demand distribution. *Management Science* **67**(8) 4863–4879.

Sim, Melvyn., Long. Zhao, Minglong. Zhou. 2021. Tractable robust supervised learning models. *Available at SSRN: 3981205* .

Simon, Herbert A. 1955. A behavioral model of rational choice. *The Quarterly Journal of Economics* 99–118.

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

Srivastava, Prateek R, Yijie Wang, Grani A Hanasusanto, Chin Pang Ho. 2021. On data-driven prescriptive analytics with side information: A regularized nadaraya-watson approach. *Available at arXiv: 2110.04855* .

Thiele, Aurélie. 2009. Multi-product pricing via robust optimisation. *Journal of Revenue and Pricing Management* **8**(1) 67–80.

Toh, Kim-Chuan, Michael J Todd, Reha H Tütüncü. 1999. SDPT3—a MATLAB software package for semidefinite programming, version 1.3. *Optimization Methods and Software* **11**(1-4) 545–581.

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

Wang, Tianyu, Ningyuan Chen, Chun Wang. 2021. Distributionally robust prescriptive analytics with wasserstein distance. *Available at arXiv:2106.05724* .

Zhang, Yu, Zhenzhen Zhang, Andrew Lim, Melvyn Sim. 2021. Robust data-driven vehicle routing with time windows. *Operations Research* **69**(2) 469–485.

Zhu, Taozeng, Jingui Xie, Melvyn Sim. 2022. Joint estimation and robustness optimization. *Management Science* **68**(3) 1659–1677.

# A. Proof of Results

*Proof of Theorem 1.* Observe that the optimal solution of the robust satisficing problem with $\tau = \hat{Z} - \hat{\eta}\Gamma$ also satisfies the following bound:

$$\begin{aligned}\mathbb{P}^S\left[\hat{Z} - \mathbb{E}_{\mathbb{P}^\star}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right] > \hat{K}\Gamma\right] &= \mathbb{P}^S\left[\hat{Z} - \mathbb{E}_{\mathbb{P}^\star}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right] > (\hat{\eta} + K_\tau)\Gamma\right] \\ &= \mathbb{P}^S\left[\tau - \mathbb{E}_{\mathbb{P}^\star}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right] > K_\tau\Gamma\right] \\ &\le \mathbb{P}^S\left[\Delta(\mathbb{P}^\star, \hat{\mathbb{P}}) > \Gamma\right] \le \epsilon.\end{aligned}$$

We note that

$$\hat{K} = \min_{\eta \ge 0}\{\eta + K_{\hat{Z} - \eta\Gamma}\} \le K_{\hat{Z}}.$$

For the second part of the theorem, Mohajerin Esfahani and Kuhn (2018) show that the data-driven robust optimization problem (4) has an equivalent formulation,

$$\begin{aligned}\hat{Z}_\Gamma = \max\ & \tau - \Gamma\kappa \\ \text{s.t.}\ & \tau \le \frac{1}{S}\sum_{s\in[S]}\left(f(\boldsymbol{x}, \boldsymbol{z}_s) + \kappa\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}_s\|\right)\ \forall \boldsymbol{z}_s \in \mathcal{Z}, s \in [S], \\ & \boldsymbol{x} \in \mathcal{X}, \kappa \ge 0, \tau \in \mathbb{R}.\end{aligned}$$

Observe that the solution $\tau$ must satisfy

$$\tau \le \frac{1}{S}\sum_{s\in[S]} f(\boldsymbol{x}, \hat{\boldsymbol{z}}_s),$$

which implies $\tau \le \hat{Z}$. Hence, with a change of variable, $\tau = \hat{Z} - \eta\Gamma$, $\eta \ge 0$, we obtain the following equivalent formulation,

$$\begin{aligned}\hat{Z}_\Gamma = \max\ & \hat{Z} - \Gamma(\eta + \kappa) \\ \text{s.t.}\ & \hat{Z} - \eta\Gamma \le \frac{1}{S}\sum_{s\in[S]}\left(f(\boldsymbol{x}, \boldsymbol{z}_s) + \kappa\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}_s\|\right)\ \forall \boldsymbol{z}_s \in \mathcal{Z}, s \in [S], \\ & \boldsymbol{x} \in \mathcal{X}, \kappa \ge 0, \eta \ge 0.\end{aligned}$$

We also note that

$$\begin{aligned}\hat{K} = \min\ & \eta + \kappa \\ \text{s.t.}\ & \hat{Z} - \eta\Gamma \le \frac{1}{S}\sum_{s\in[S]}\left(f(\boldsymbol{x}, \boldsymbol{z}_s) + \kappa\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}_s\|\right)\ \forall \boldsymbol{z}_s \in \mathcal{Z}, s \in [S], \\ & \boldsymbol{x} \in \mathcal{X}, \kappa \ge 0, \eta \ge 0.\end{aligned}$$

Therefore $\hat{Z}_\Gamma = \hat{Z} - \Gamma\hat{K}$. Now observe that, since $\hat{Z}_\Gamma = \hat{Z} - \Gamma\hat{\eta} - \Gamma K_{\hat{Z} - \hat{\eta}\Gamma}$, we have

$$\begin{aligned}\hat{Z}_\Gamma = \max\ & \hat{Z} - \hat{\eta}\Gamma - \Gamma\kappa \\ \text{s.t.}\ & \hat{Z} - \hat{\eta}\Gamma \le \frac{1}{S}\sum_{s\in[S]}\left(f(\boldsymbol{x}, \boldsymbol{z}_s) + \kappa\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}_s\|\right)\ \forall \boldsymbol{z}_s \in \mathcal{Z}, s \in [S], \\ & \boldsymbol{x} \in \mathcal{X}, \kappa \ge 0,\end{aligned}$$

which has the same set of optimal solutions as the robust satisficing problem,

$$\begin{aligned} K_\tau = \min\ & \kappa \\ \text{s.t. } & \tau \leq \frac{1}{S}\sum_{s\in[S]}\left(f(\boldsymbol{x},\boldsymbol{z}_s) + \kappa\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}_s\|\right)\ \forall \boldsymbol{z}_s \in \mathcal{Z}, s\in[S], \\ & \boldsymbol{x}\in\mathcal{X}, \kappa \geq 0, \end{aligned} \tag{35}$$

in which $\tau = \hat{Z} - \hat{\eta}\Gamma$. □

*Proof of Theorem 2.* Note that the optimal solutions to the robust satisfiicing problem (6) satisfy,

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

or equivalently,

$$\tau \leq \frac{1}{S}\sum_{s\in[S]}\left(f(\boldsymbol{x},\boldsymbol{z}_s) + K_\tau\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}_s\|\right) \qquad \forall \boldsymbol{z}_s\in\mathcal{Z}, s\in[S].$$

Therefore, for any $\boldsymbol{z}\in\mathcal{Z}$,

$$\tau - f(\boldsymbol{x},\boldsymbol{z}) \leq \frac{1}{S}\sum_{s\in[S]} K_\tau\|\boldsymbol{z} - \hat{\boldsymbol{z}}_s\|,$$

which implies,

$$\tau - \mathbb{E}_{\mathbb{P}^\star}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right] \leq \frac{1}{S}\sum_{s\in[S]} K_\tau \mathbb{E}_{\mathbb{P}^\star}\left[\|\tilde{\boldsymbol{z}} - \hat{\boldsymbol{z}}_s\|\right].$$

Hence, for all $r\geq 1$,

$$\begin{aligned} \mathbb{Q}\left[\tau - \mathbb{E}_{\mathbb{P}^\star}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right] > \mu K_\tau r\right] &\leq \mathbb{Q}\left[\frac{1}{S}\sum_{s\in[S]}\mathbb{E}_{\tilde{\boldsymbol{z}}\sim\mathbb{P}^\star}\left[\|\tilde{\boldsymbol{z}} - \hat{\boldsymbol{z}}_s\|\right] \geq \mu r\right] \\ &\leq \mathbb{Q}\left[\sum_{s\in[S]}\left(\mathbb{E}_{\tilde{\boldsymbol{z}}\sim\mathbb{P}^\star}\left[\|\tilde{\boldsymbol{z}} - \hat{\boldsymbol{z}}_s\|\right] - \mathbb{E}_{\mathbb{Q}}\left[\|\tilde{\boldsymbol{z}} - \hat{\boldsymbol{z}}_s\|\right]\right) \geq S\mu(r-1)\right] \\ &\leq \frac{\mathbf{1}^\top\boldsymbol{\Sigma}\mathbf{1}}{\mathbf{1}^\top\boldsymbol{\Sigma}\mathbf{1} + \mu^2 S^2 (r-1)^2}, \end{aligned}$$

where the last inequality follows from one-sided Chebyshev's inequality.

□

*Proof of Proposition 1.* Observe that $K_{\underline{Z}} = 0$. Moreover, since $\mathcal{X}$ is convex and $f(\boldsymbol{x},\boldsymbol{z})$ is concave in $\boldsymbol{x}\in\mathcal{X}$ for all $\boldsymbol{z}\in\mathcal{Z}$, this implies that $K_\tau$ is convex and non-decreasing in $\tau$. Therefore, for any $\tau_\lambda = (1-\lambda)\underline{Z} + \lambda\hat{Z}$, $\lambda\in[0,1]$ we have

$$\begin{aligned} K_{\tau_\lambda} &\leq (1-\lambda)K_{\underline{Z}} + \lambda K_{\hat{Z}} \\ &= \lambda K_{\hat{Z}}. \end{aligned}$$

□

*Proof of Theorem 3.* For any $\tau \le \hat{Z}$, there exists some $\bar{\boldsymbol{x}}$ such that

$$\tau \le \frac{1}{S}\sum_{s\in[S]} f(\bar{\boldsymbol{x}}, \hat{\boldsymbol{z}}_s),$$

where we define $\hat{\boldsymbol{z}}_s = \hat{\boldsymbol{z}}(\hat{\boldsymbol{w}}, \boldsymbol{u}, s)$ for simplicity. Then, for any $\kappa \ge \bar{L}$, we have

$$\frac{1}{S}\sum_{s\in[S]} (f(\bar{\boldsymbol{x}}, \boldsymbol{z}_s) + \kappa\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}_s\|) \ge \frac{1}{S}\sum_{s\in[S]} \left(f(\bar{\boldsymbol{x}}, \hat{\boldsymbol{z}}_s) - \bar{L}\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}_s\| + \kappa\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}_s\|\right) \ge \tau.$$

Hence, $\bar{\boldsymbol{x}}$ is feasible with some $K_\tau \le \bar{L}$.

Next, we will verify the second part of this theorem. Let $(\boldsymbol{x}^\dagger, \kappa^\dagger)$ be the optimal solution to Problem (11) for target $\hat{Z}$. Then, from the constraint of Problem (11), we have

$$\begin{aligned}
&\hat{Z} \le \frac{1}{S}\sum_{s\in[S]} \left(f(\boldsymbol{x}^\dagger, \boldsymbol{z}_s) + \kappa^\dagger\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\hat{\boldsymbol{w}}, \boldsymbol{u}, s)\|\right) \ \forall \boldsymbol{z}_s \in \mathcal{Z}, s \in [S],\\
\Longrightarrow\ &\hat{Z} \le \frac{1}{S}\sum_{s\in[S]} \left(f(\boldsymbol{x}^\dagger, \boldsymbol{z}_s) + \kappa^\dagger\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\hat{\boldsymbol{w}}, \boldsymbol{u}, s)\|\right) \ \forall \boldsymbol{z}_s = \hat{\boldsymbol{z}}(\hat{\boldsymbol{w}}, \boldsymbol{u}, s), s \in [S],\\
\Longrightarrow\ &\hat{Z} \le \frac{1}{S}\sum_{s\in[S]} f(\boldsymbol{x}^\dagger, \hat{\boldsymbol{z}}(\hat{\boldsymbol{w}}, \boldsymbol{u}, s)) = \hat{Z},
\end{aligned}$$

which completes the proof. □

*Proof of Proposition 2.* It suffices to show that for all $\boldsymbol{w} \in \mathcal{W}$ and $\boldsymbol{z}_s \in \mathcal{Z}$,

$$\begin{aligned}
&\frac{1}{S}\sum_{s\in[S]} \|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\hat{\boldsymbol{w}}, \boldsymbol{u}, s)\|\\
\le\ &\frac{1}{S}\sum_{s\in[S]} \|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\boldsymbol{w}, \boldsymbol{u}, s)\| + \frac{1}{S}\sum_{s\in[S]} \|\hat{\boldsymbol{z}}(\boldsymbol{w}, \boldsymbol{u}, s) - \hat{\boldsymbol{z}}(\hat{\boldsymbol{w}}, \boldsymbol{u}, s)\|\\
=\ &\frac{1}{S}\sum_{s\in[S]} \|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\boldsymbol{w}, \boldsymbol{u}, s)\| + \frac{1}{S}\sum_{s\in[S]} \|\boldsymbol{h}(\boldsymbol{w} - \hat{\boldsymbol{w}}, \boldsymbol{u}) - \boldsymbol{g}(\boldsymbol{w} - \hat{\boldsymbol{w}}, \boldsymbol{u}, \hat{\boldsymbol{u}}_s)\|,
\end{aligned}$$

and that

$$\begin{aligned}
&\frac{1}{S}\sum_{s\in[S]} \|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\hat{\boldsymbol{w}}, \boldsymbol{u}, s)\|\\
\ge\ &\frac{1}{S}\sum_{s\in[S]} \|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\boldsymbol{w}, \boldsymbol{u}, s)\| - \frac{1}{S}\sum_{s\in[S]} \|\hat{\boldsymbol{z}}(\boldsymbol{w}, \boldsymbol{u}, s) - \hat{\boldsymbol{z}}(\hat{\boldsymbol{w}}, \boldsymbol{u}, s)\|\\
=\ &\frac{1}{S}\sum_{s\in[S]} \|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\boldsymbol{w}, \boldsymbol{u}, s)\| - \frac{1}{S}\sum_{s\in[S]} \|\boldsymbol{h}(\boldsymbol{w} - \hat{\boldsymbol{w}}, \boldsymbol{u}) - \boldsymbol{g}(\boldsymbol{w} - \hat{\boldsymbol{w}}, \boldsymbol{u}, \hat{\boldsymbol{u}}_s)\|.
\end{aligned}$$

Now, note that

$$
\begin{aligned}
&\Delta(\mathbb{P},\hat{\mathbb{P}}_{\boldsymbol{w}}) - \Delta(\mathbb{P},\hat{\mathbb{P}}_{\hat{\boldsymbol{w}}}) \\
= &\inf_{\mathbb{Q}_{1s}, s\in[S]} \left\{ \frac{1}{S}\sum_{s\in[S]} \mathbb{E}_{\mathbb{Q}_{1s}}[\|\tilde{\boldsymbol{z}}_1 - \hat{\boldsymbol{z}}(\boldsymbol{w},\boldsymbol{u},s)\|] \right\} - \inf_{\mathbb{Q}_{2s}, s\in[S]} \left\{ \frac{1}{S}\sum_{s\in[S]} \mathbb{E}_{\mathbb{Q}_{2s}}[\|\tilde{\boldsymbol{z}}_2 - \hat{\boldsymbol{z}}(\hat{\boldsymbol{w}},\boldsymbol{u},s)\|] \right\} \\
\geq &\inf_{\mathbb{Q}_{1s}, s\in[S]} \left\{ \frac{1}{S}\sum_{s\in[S]} \mathbb{E}_{\mathbb{Q}_{1s}}[\|\tilde{\boldsymbol{z}}_1 - \hat{\boldsymbol{z}}(\boldsymbol{w},\boldsymbol{u},s)\| - \|\tilde{\boldsymbol{z}}_1 - \hat{\boldsymbol{z}}(\hat{\boldsymbol{w}},\boldsymbol{u},s)\|] \right\} \\
= &\inf_{\boldsymbol{z}_s, s\in[S]} \left\{ \frac{1}{S}\sum_{s\in[S]} (\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\boldsymbol{w},\boldsymbol{u},s)\| - \|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\hat{\boldsymbol{w}},\boldsymbol{u},s)\|) \right\} \\
\geq &-\frac{1}{S}\sum_{s\in[S]} \|\boldsymbol{h}(\boldsymbol{w}-\hat{\boldsymbol{w}},\boldsymbol{u}) - \boldsymbol{g}(\boldsymbol{w}-\hat{\boldsymbol{w}},\boldsymbol{u},\hat{\boldsymbol{u}}_s)\|.
\end{aligned}
$$

The last equality follows by noting the worst-case distribution is a point-mass. Similarly, we have

$$
\Delta(\mathbb{P},\hat{\mathbb{P}}_{\hat{\boldsymbol{w}}}) - \Delta(\mathbb{P},\hat{\mathbb{P}}_{\boldsymbol{w}}) \geq -\frac{1}{S}\sum_{s\in[S]} \|\boldsymbol{h}(\boldsymbol{w}-\hat{\boldsymbol{w}},\boldsymbol{u}) - \boldsymbol{g}(\boldsymbol{w}-\hat{\boldsymbol{w}},\boldsymbol{u},\hat{\boldsymbol{u}}_s)\|.
$$

This completes the proof. □

*Proof of Proposition 3.* From the definition of the estimation distance metric, we have

$$
\Delta(\mathbb{P}^\star,\hat{\mathbb{P}}_{\hat{\boldsymbol{w}}}) \leq \Delta(\mathbb{P}^\star,\hat{\mathbb{P}}_{\boldsymbol{w}^\star}) + \bar{\theta}\Upsilon(\boldsymbol{w}^\star - \hat{\boldsymbol{w}}).
$$

From the constraint of the residual-based robust satisficing problem (10), we have

$$
\tau - \mathbb{E}_{\mathbb{P}^\star}[f(\boldsymbol{x},\tilde{\boldsymbol{z}})] \leq K_\tau \Delta(\mathbb{P}^\star,\hat{\mathbb{P}}_{\hat{\boldsymbol{w}}}) \leq K_\tau\left(r + \bar{\theta}\Upsilon(\boldsymbol{w}^\star - \hat{\boldsymbol{w}})\right),
$$

for all $r \geq \Delta(\mathbb{P}^\star,\hat{\mathbb{P}}_{\boldsymbol{w}^\star})$. Therefore, for all $r \geq 0$,

$$
\begin{aligned}
&\mathbb{Q}^S\left[\mathbb{E}_{\mathbb{P}^\star}[f(\boldsymbol{x},\tilde{\boldsymbol{z}})] < \tau - K_\tau\left(r + \bar{\theta}\Upsilon(\boldsymbol{w}^* - \hat{\boldsymbol{w}})\right)\right] \\
\leq &\mathbb{P}^S\left[\Delta(\mathbb{P}^\star,\hat{\mathbb{P}}_{\boldsymbol{w}^\star}) > r\right].
\end{aligned}
$$

□

*Proof of Theorem 4.* It suffices to show that the optimal solution of the residual-based robust satisficing problem (11), $\bar{\boldsymbol{x}}$ is also feasible in Problem (16) for $\theta = \bar{\theta}$. From the proof of Proposition 2, we can establish that

$$
\begin{aligned}
\frac{1}{S}\sum_{s\in[S]} \|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\hat{\boldsymbol{w}},\boldsymbol{u},s)\| &\leq \frac{1}{S}\sum_{s\in[S]} \|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\boldsymbol{w},\boldsymbol{u},s)\| + \frac{1}{S}\sum_{s\in[S]} \|\boldsymbol{h}(\boldsymbol{w}-\hat{\boldsymbol{w}},\boldsymbol{u}) - \boldsymbol{g}(\boldsymbol{w}-\hat{\boldsymbol{w}},\boldsymbol{u},\hat{\boldsymbol{u}}_s)\| \\
&\leq \frac{1}{S}\sum_{s\in[S]} \|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\boldsymbol{w},\boldsymbol{u},s)\| + \bar{\theta}\Upsilon(\boldsymbol{w} - \hat{\boldsymbol{w}}) \quad \forall \boldsymbol{w} \in \mathcal{W}.
\end{aligned}
$$

Hence,

$$\begin{aligned}
&\tau \leq \frac{1}{S}\sum_{s\in[S]}\left(f(\bar{\boldsymbol{x}},\boldsymbol{z}_s)+K_\tau\|\boldsymbol{z}_s-\hat{\boldsymbol{z}}(\hat{\boldsymbol{w}},\boldsymbol{u},s)\|\right) && \forall \boldsymbol{z}_s\in\mathcal{Z}, s\in[S]\\
\Longrightarrow\ &\underline{\tau} \leq \frac{1}{S}\sum_{s\in[S]}\left(f(\bar{\boldsymbol{x}},\boldsymbol{z}_s)+K_\tau\|\boldsymbol{z}_s-\hat{\boldsymbol{z}}(\hat{\boldsymbol{w}},\boldsymbol{u},s)\|\right) && \forall \boldsymbol{z}_s\in\mathcal{Z}, s\in[S]\\
\Longrightarrow\ &\underline{\tau} \leq \frac{1}{S}\sum_{s\in[S]}\left(f(\bar{\boldsymbol{x}},\boldsymbol{z}_s)+K_\tau\left(\|\boldsymbol{z}_s-\hat{\boldsymbol{z}}(\boldsymbol{w},\boldsymbol{u},s)\|+\bar{\theta}\Upsilon(\boldsymbol{w}-\hat{\boldsymbol{w}})\right)\right) && \forall \boldsymbol{w}\in\mathcal{W}, \boldsymbol{z}_s\in\mathcal{Z}, s\in[S]\\
\Longrightarrow\ &\underline{\tau} \leq \frac{1}{S}\sum_{s\in[S]}\left(f(\bar{\boldsymbol{x}},\boldsymbol{z}_s)+K_\tau\left(\|\boldsymbol{z}_s-\hat{\boldsymbol{z}}(\boldsymbol{w},\boldsymbol{u},s)\|+\bar{\theta}\Upsilon(\boldsymbol{w}-\hat{\boldsymbol{w}})\right)\right) && \forall \boldsymbol{w}\in\mathcal{W}, \boldsymbol{z}_s\in\mathcal{Z}, s\in[S].
\end{aligned}$$

Hence, $\bar{\boldsymbol{x}}$ is feasible in Problem (16), and $\Theta_{\tau,\underline{\tau}} \leq \bar{\theta}$.

Let $\boldsymbol{x}$ be a feasible solution of Problem (16). Observe that $\Upsilon(\hat{\boldsymbol{w}}-\hat{\boldsymbol{w}})=0$; hence,

$$\begin{aligned}
&\underline{\tau} \leq \frac{1}{S}\sum_{s\in[S]}\left(f(\boldsymbol{x},\boldsymbol{z}_s)+K_\tau\left(\|\boldsymbol{z}_s-\hat{\boldsymbol{z}}(\boldsymbol{w},\boldsymbol{u},s)\|+\theta\Upsilon(\boldsymbol{w}-\hat{\boldsymbol{w}})\right)\right) && \forall \boldsymbol{w}\in\mathcal{W}, \boldsymbol{z}_s\in\mathcal{Z}, s\in[S]\\
\Longrightarrow\ &\underline{\tau} \leq \frac{1}{S}\sum_{s\in[S]}\left(f(\boldsymbol{x},\boldsymbol{z}_s)+K_\tau\left(\|\boldsymbol{z}_s-\hat{\boldsymbol{z}}(\hat{\boldsymbol{w}},\boldsymbol{u},s)\|\right)\right) && \forall \boldsymbol{z}_s\in\mathcal{Z}, s\in[S].
\end{aligned}$$

Therefore, $\varkappa(\boldsymbol{x},\underline{\tau},\hat{\mathbb{P}}_{\hat{\boldsymbol{w}}}) \leq K_\tau$.

Now, let $\hat{\boldsymbol{x}}$ be the optimal solution of Problem (16) in which $\Theta_{\tau,\underline{\tau}}=0$. From the constraint of Problem (16), we have for all $\boldsymbol{w}\in\mathcal{W}$,

$$\underline{\tau} \leq \frac{1}{S}\sum_{s\in[S]}\left(f(\hat{\boldsymbol{x}},\boldsymbol{z}_s)+K_\tau\left(\|\boldsymbol{z}_s-\hat{\boldsymbol{z}}(\boldsymbol{w},\boldsymbol{u},s)\|\right)\right) \quad \forall \boldsymbol{z}_s\in\mathcal{Z}, s\in[S],$$

indicating that $\varkappa(\hat{\boldsymbol{x}},\underline{\tau},\hat{\mathbb{P}}_{\boldsymbol{w}}) \leq K_\tau$.

Now we focus on the last part of the theorem where $\hat{\boldsymbol{x}}$ is the optimal solution of Problem (16). From the constraint of Problem (15), we have

$$\underline{\tau}-\mathbb{E}_{\mathbb{P}^\star}\left[f(\hat{\boldsymbol{x}},\tilde{\boldsymbol{z}})\right] \leq K_\tau\left(\Delta(\mathbb{P}^\star,\hat{\mathbb{P}}_{\boldsymbol{w}^\star})+\Theta_{\tau,\underline{\tau}}\Upsilon(\boldsymbol{w}^\star-\hat{\boldsymbol{w}})\right) \leq K_\tau\left(r+\Theta_{\tau,\underline{\tau}}\Upsilon(\boldsymbol{w}^\star-\hat{\boldsymbol{w}})\right),$$

for all $r \geq \Delta(\mathbb{P}^\star,\hat{\mathbb{P}}_{\boldsymbol{w}^\star})$. Therefore, for all $r \geq 0$,

$$\mathbb{Q}^S\left[\mathbb{E}_{\mathbb{P}^\star}\left[f(\hat{\boldsymbol{x}},\tilde{\boldsymbol{z}})\right] < \underline{\tau}-K_\tau\left(r+\Theta_{\tau,\underline{\tau}}\Upsilon(\boldsymbol{w}^*-\hat{\boldsymbol{w}})\right)\right] \leq \mathbb{P}^S\left[\Delta(\mathbb{P}^\star,\hat{\mathbb{P}}_{\boldsymbol{w}^\star}) > r\right].$$

□

*Proof of Theorem 5.* From Theorem 4, we note that $\Theta_{\tau,\underline{\tau}} \le \bar{\theta} \le \bar{\vartheta}$. Next, we observe that for any $\theta \in [0, \bar{\vartheta}]$,

$$\begin{aligned}
&\frac{1}{S}\sum_{s\in[S]} \|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\boldsymbol{w},\boldsymbol{u},s)\| + \theta\Upsilon(\hat{\boldsymbol{w}} - \boldsymbol{w}) \\
&= \frac{1}{S}\sum_{s\in[S]} \|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\boldsymbol{w},\boldsymbol{u},s)\| + \bar{\theta}\frac{\theta}{\bar{\theta}}\Upsilon(\hat{\boldsymbol{w}} - \boldsymbol{w}) \\
&\ge \frac{1}{S}\sum_{s\in[S]} \|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\boldsymbol{w},\boldsymbol{u},s)\| + \frac{\theta}{\bar{\theta}}\frac{1}{S}\sum_{s\in[S]} \|\boldsymbol{h}(\hat{\boldsymbol{w}} - \boldsymbol{w},\boldsymbol{u}) - \boldsymbol{g}(\hat{\boldsymbol{w}} - \boldsymbol{w},\boldsymbol{u},\hat{\boldsymbol{u}}_s)\| \\
&\ge \frac{\theta}{\bar{\vartheta}}\frac{1}{S}\sum_{s\in[S]} \|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\boldsymbol{w},\boldsymbol{u},s)\| + \frac{\theta}{\bar{\vartheta}}\frac{1}{S}\sum_{s\in[S]} \|\boldsymbol{h}(\hat{\boldsymbol{w}} - \boldsymbol{w},\boldsymbol{u}) - \boldsymbol{g}(\hat{\boldsymbol{w}} - \boldsymbol{w},\boldsymbol{u},\hat{\boldsymbol{u}}_s)\| \\
&\ge \frac{\theta}{\bar{\vartheta}}\frac{1}{S}\sum_{s\in[S]} \|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\boldsymbol{w},\boldsymbol{u},s) + \boldsymbol{h}(\hat{\boldsymbol{w}} - \boldsymbol{w},\boldsymbol{u}) - \boldsymbol{g}(\hat{\boldsymbol{w}} - \boldsymbol{w},\boldsymbol{u},\hat{\boldsymbol{u}}_s)\| \\
&= \frac{\theta}{\bar{\vartheta}}\frac{1}{S}\sum_{s\in[S]} \|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\hat{\boldsymbol{w}},\boldsymbol{u},s)\|.
\end{aligned}$$

The conservative approximation follows from observing that for any feasible solution in Problem (20), $f(\boldsymbol{x},\boldsymbol{z}_s) \ge \boldsymbol{d}^\top \boldsymbol{y}_s(\boldsymbol{z}_s,\phi_s)$ for all $\boldsymbol{z}_s \in \mathcal{Z}, \phi_s \ge \|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\hat{\boldsymbol{w}},\boldsymbol{u},s)\|$, $s \in [S]$. Hence,

$$f(\boldsymbol{x},\boldsymbol{z}_s) \ge \boldsymbol{d}^\top \boldsymbol{y}_s(\boldsymbol{z}_s,\phi_s) \qquad \forall(\eta,(\boldsymbol{z}_s,\phi_s)_{s\in[S]}) \in \bar{\mathcal{Z}}(\theta),$$

and for any feasible $\theta \in [0, \bar{\vartheta}]$,

$$\begin{aligned}
&\underline{\tau} \le \frac{1}{S}\sum_{s\in[S]} \left(\boldsymbol{d}^\top \boldsymbol{y}_s(\boldsymbol{z}_s,\phi_s) + \bar{K}_\tau \eta\right) \quad \forall(\eta,(\boldsymbol{z}_s,\phi_s)_{s\in[S]}) \in \bar{\mathcal{Z}}(\theta) \\
\Longrightarrow\ &\underline{\tau} \le \frac{1}{S}\sum_{s\in[S]} \left(f(\boldsymbol{x},\boldsymbol{z}_s) + \bar{K}_\tau \eta\right) \quad \forall(\eta,(\boldsymbol{z}_s,\phi_s)_{s\in[S]}) \in \bar{\mathcal{Z}}(\theta) \\
\Longrightarrow\ &\underline{\tau} \le \frac{1}{S}\sum_{s\in[S]} f(\boldsymbol{x},\boldsymbol{z}_s) + \bar{K}_\tau\left(\max\left\{\frac{1}{S}\sum_{s\in[S]}(\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\boldsymbol{w},\boldsymbol{u},s)\| + \theta\Upsilon(\hat{\boldsymbol{w}} - \boldsymbol{w})),\right.\right. \\
&\qquad\qquad \left.\left.\frac{1}{S}\sum_{s\in[S]}\frac{\theta}{\bar{\vartheta}}\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\hat{\boldsymbol{w}},\boldsymbol{u},s)\|\right\}\right) \quad \forall \boldsymbol{w} \in \mathcal{W}, \boldsymbol{z}_s \in \mathcal{Z}, s \in [S] \\
\Longrightarrow\ &\underline{\tau} \le \frac{1}{S}\sum_{s\in[S]} \left(f(\boldsymbol{x},\boldsymbol{z}_s) + \bar{K}_\tau\left(\|\boldsymbol{z}_s - \hat{\boldsymbol{z}}(\boldsymbol{w},\boldsymbol{u},s)\| + \theta\Upsilon(\hat{\boldsymbol{w}} - \boldsymbol{w})\right)\right) \\
&\qquad\qquad \forall \boldsymbol{w} \in \mathcal{W}, \boldsymbol{z}_s \in \mathcal{Z}, s \in [S].
\end{aligned}$$

The last inequality follows from the inequality established at the beginning of this proof.

Now, we prove the second part of the theorem. For any feasible solution of Problem (18), $\bar{\boldsymbol{x}} \in \mathcal{X}$ and $\bar{\boldsymbol{y}}_s \in \mathcal{L}^{N+1,D_2}$, for all $s \in [S]$, we have

$$\begin{aligned}
&\tau \le \frac{1}{S}\sum_{s\in[S]} \left(\boldsymbol{d}^\top \bar{\boldsymbol{y}}_s(\boldsymbol{z}_s,\phi_s) + \bar{K}_\tau \phi_s\right) && \forall(\boldsymbol{z}_s,\phi_s) \in \bar{\mathcal{Z}}_s, s \in [S], \\
&\boldsymbol{A}(\boldsymbol{z})\bar{\boldsymbol{x}} + \boldsymbol{B}\bar{\boldsymbol{y}}_s(\boldsymbol{z},\phi) \le \boldsymbol{b}(\boldsymbol{z}) && \forall(\boldsymbol{z},\phi) \in \bar{\mathcal{Z}}_s, s \in [S].
\end{aligned}$$

Next, observe that

$$\eta \geq \frac{1}{S}\sum_{s\in[S]}\frac{\theta}{\bar{\vartheta}}\phi_s \qquad \forall(\eta,(\boldsymbol{z}_s,\phi_s)_{s\in[S]})\in\bar{\mathcal{Z}}(\theta),$$

indicating

$$\eta \geq \frac{1}{S}\sum_{s\in[S]}\phi_s \qquad \forall(\eta,(\boldsymbol{z}_s,\phi_s)_{s\in[S]})\in\bar{Z}_{\bar{\vartheta}}.$$

Therefore, for any $(\eta,(\boldsymbol{z}_s,\phi_s)_{s\in[S]})\in\bar{Z}_{\bar{\vartheta}}$, we have

$$\underline{\tau}\leq\tau\leq\frac{1}{S}\sum_{s\in[S]}\left(\boldsymbol{d}^\top\bar{\boldsymbol{y}}_s(\boldsymbol{z}_s,\phi_s)+\bar{K}_\tau\phi_s\right)\leq\frac{1}{S}\sum_{s\in[S]}\left(\boldsymbol{d}^\top\bar{\boldsymbol{y}}_s(\boldsymbol{z}_s,\phi_s)+\bar{K}_\tau\eta\right).$$

Hence, solutions $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{y}}_s$, $s\in[S]$ are also feasible to Problem (20) for some $\theta\leq\bar{\vartheta}$.

For the third part of the theorem, we note that when $\tau=\underline{\tau}$, then for any feasible solution of Problem (20), $\hat{\boldsymbol{x}}\in\mathcal{X}$, $\theta\in[0,\bar{\vartheta}]$, and $\hat{\boldsymbol{y}}_s\in\mathcal{L}^{N+1,D_2}$, for all $s\in[S]$, we have

$$\begin{aligned}
&\tau\leq\frac{1}{S}\sum_{s\in[S]}\left(\boldsymbol{d}^\top\hat{\boldsymbol{y}}_s(\boldsymbol{z}_s,\phi_s)+\bar{K}_\tau\eta\right) \quad \forall\left(\eta,(\boldsymbol{z}_s,\phi_s)_{s\in[S]}\right)\in\bar{\mathcal{Z}}(\theta),\\
&\boldsymbol{A}(\boldsymbol{z}_s)\hat{\boldsymbol{x}}+\boldsymbol{B}\hat{\boldsymbol{y}}_s(\boldsymbol{z}_s,\phi_s)\leq\boldsymbol{b}(\boldsymbol{z}_s) \qquad \forall(\boldsymbol{z}_s,\phi_s)\in\bar{\mathcal{Z}}_s, s\in[S].
\end{aligned}$$

Moreover,

$$\begin{aligned}
\bar{\mathcal{Z}}(\theta) &= \left\{\left(\eta,(\boldsymbol{z}_s,\phi_s)_{s\in[S]}\right)\ \middle|\ \begin{array}{l}\eta\in\mathbb{R},(\boldsymbol{z}_s,\phi_s)\in\bar{\mathcal{Z}}_s\ \forall s\in[S]\\ \eta\geq\frac{1}{S}\sum_{s\in[S]}\left(\|\boldsymbol{z}_s-\hat{\boldsymbol{z}}(\boldsymbol{w},\boldsymbol{u},s)\|+\theta\Upsilon(\hat{\boldsymbol{w}}-\boldsymbol{w})\right)\\ \eta\geq\frac{\theta}{\bar{\vartheta}S}\sum_{s\in[S]}\phi_s\\ \text{for some } \boldsymbol{w}\in\mathcal{W}\end{array}\right\}\\
&\supseteq \left\{\left(\eta,(\boldsymbol{z}_s,\phi_s)_{s\in[S]}\right)\ \middle|\ \begin{array}{l}\eta\in\mathbb{R},(\boldsymbol{z}_s,\phi_s)\in\bar{\mathcal{Z}}_s\ \forall s\in[S]\\ \eta\geq\frac{1}{S}\sum_{s\in[S]}\left(\|\boldsymbol{z}_s-\hat{\boldsymbol{z}}(\boldsymbol{w},\boldsymbol{u},s)\|+\theta\Upsilon(\hat{\boldsymbol{w}}-\boldsymbol{w})\right)\\ \eta\geq\frac{\theta}{\bar{\vartheta}S}\sum_{s\in[S]}\phi_s\\ \boldsymbol{w}=\hat{\boldsymbol{w}},\eta=\frac{1}{S}\sum_{s\in[S]}\phi_s\end{array}\right\}\\
&= \left\{\left(\eta,(\boldsymbol{z}_s,\phi_s)_{s\in[S]}\right)\ \middle|\ \begin{array}{l}\eta\in\mathbb{R},(\boldsymbol{z}_s,\phi_s)\in\bar{\mathcal{Z}}_s\ \forall s\in[S]\\ \eta=\frac{1}{S}\sum_{s\in[S]}\phi_s\end{array}\right\}.
\end{aligned}$$

Hence, we have

$$\begin{aligned}
&\tau\leq\frac{1}{S}\sum_{s\in[S]}\left(\boldsymbol{d}^\top\hat{\boldsymbol{y}}_s(\boldsymbol{z}_s,\phi_s)+\bar{K}_\tau\eta\right) \quad \forall\left(\eta,(\boldsymbol{z}_s,\phi_s)_{s\in[S]}\right)\in\bar{\mathcal{Z}}(\theta)\\
\Longrightarrow\ &\tau\leq\frac{1}{S}\sum_{s\in[S]}\left(\boldsymbol{d}^\top\hat{\boldsymbol{y}}_s(\boldsymbol{z}_s,\phi_s)+\bar{K}_\tau\phi_s\right)\ \forall(\boldsymbol{z}_s,\phi_s)\in\bar{\mathcal{Z}}_s, s\in[S].
\end{aligned}$$

Therefore, solutions $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}_s$, $s \in [S]$ are also optimal to Problem (18) in which $\kappa = \bar{K}_\tau$.

□

*Proof of Theorem 6.* Consider the following optimization problem,

$$\begin{aligned} \inf\ & \frac{1}{S} \sum_{s\in[S]} \sum_{n\in[N]} \left( x_n \exp\left(z_{s,n}\right) + \kappa \left( |z_{s,n} - (\hat{\boldsymbol{w}}^\top \boldsymbol{u}_n + \hat{v}_s - \hat{\boldsymbol{w}}^\top \hat{\boldsymbol{u}}_s)| \right) \right) \\ \text{s.t. } & \boldsymbol{z}_s \in \mathbb{R}^N \qquad \forall s \in [S], \end{aligned}$$

which is equivalent to

$$\begin{aligned} \inf\ & \frac{1}{S} \sum_{s\in[S]} \left( \boldsymbol{x}^\top \boldsymbol{a}_s + \kappa \|\boldsymbol{b}_s\|_1 \right) \\ \text{s.t. } & \exp\left(z_{s,n}\right) \le a_{s,n} && \forall n \in [N], s \in [S], \\ & b_{s,n} = z_{s,n} - (\hat{\boldsymbol{w}}^\top \boldsymbol{u}_n + \hat{v}_s - \hat{\boldsymbol{w}}^\top \hat{\boldsymbol{u}}_s) && \forall n \in [N], s \in [S], \\ & \boldsymbol{z}_s, \boldsymbol{b}_s \in \mathbb{R}^N, \boldsymbol{a}_s \ge \boldsymbol{0} && \forall s \in [S]. \end{aligned}$$

We can obtain the dual problem of the above problem as follows,

$$\begin{aligned} \sup\ & \sum_{s\in[S]} \sum_{n\in[N]} \left( \phi_{s,n} (\hat{\boldsymbol{w}}^\top \boldsymbol{u}_n + \hat{v}_s - \hat{\boldsymbol{w}}^\top \hat{\boldsymbol{u}}_s) - \beta_{s,n} \right) \\ \text{s.t. } & \|\boldsymbol{\phi}_s\|_\infty \le \kappa / S && \forall s \in [S], \\ & -d_{s,n} \log\left(-S d_{s,n} / x_n\right) + d_{s,n} - \beta_{s,n} \le 0 && \forall s \in [S], n \in [N], \\ & \boldsymbol{d}_s + \boldsymbol{\phi}_s = \boldsymbol{0} && \forall s \in [S], \\ & \boldsymbol{d}_s \in \mathbb{R}^N_-, \boldsymbol{\beta}_s, \boldsymbol{\phi}_s \in \mathbb{R}^N && \forall s \in [S]. \end{aligned}$$

Therefore, for the residual-based robust satisficing model,

$$\begin{aligned} K_\tau = \min\ & \kappa \\ \text{s.t. } & \tau \le \frac{1}{S} \sum_{s\in[S]} \sum_{n\in[N]} \left( x_n \exp\left(z_{s,n}\right) + \kappa \left( |z_{s,n} - (\hat{\boldsymbol{w}}^\top \boldsymbol{u}_n + \hat{v}_s - \hat{\boldsymbol{w}}^\top \hat{\boldsymbol{u}}_s)| \right) \right) \quad \forall \boldsymbol{z}_s \in \mathbb{R}^N, s \in [S], \\ & \boldsymbol{c}^\top \boldsymbol{x} \le 1, \\ & \boldsymbol{x} \ge \boldsymbol{0}, \kappa \ge 0, \end{aligned} \tag{36}$$

we can equivalently and explicitly reformulate it as

$$\begin{aligned} \min\ & \kappa \\ \text{s.t. } & \sum_{s\in[S]} \sum_{n\in[N]} \left( \phi_{s,n} (\hat{\boldsymbol{w}}^\top \boldsymbol{u}_n + \hat{v}_s - \hat{\boldsymbol{w}}^\top \hat{\boldsymbol{u}}_s) - \beta_{s,n} \right) \ge \tau, \\ & \|\boldsymbol{\phi}_s\|_\infty \le \kappa / S && \forall s \in [S], \\ & -d_{s,n} \log\left(-S d_{s,n} / x_n\right) + d_{s,n} - \beta_{s,n} \le 0 && \forall s \in [S], n \in [N], \\ & \boldsymbol{d}_s + \boldsymbol{\phi}_s = \boldsymbol{0} && \forall s \in [S], \\ & \boldsymbol{d}_s \in \mathbb{R}^N_-, \boldsymbol{\beta}_s, \boldsymbol{\phi}_s \in \mathbb{R}^N && \forall s \in [S], \\ & \boldsymbol{c}^\top \boldsymbol{x} \le 1, \\ & \boldsymbol{x} \ge \boldsymbol{0}, \kappa \ge 0. \end{aligned}$$

This model can be further simplified as

$$
\begin{aligned}
&\min\ \kappa\\
&\text{s.t. } \frac{1}{S}\sum_{s\in[S]}\sum_{n\in[N]}\Big(\phi_{s,n}(\hat{\boldsymbol{w}}^\top\boldsymbol{u}_n+\hat{v}_s-\hat{\boldsymbol{w}}^\top\hat{\boldsymbol{u}}_s)-\beta_{s,n}\Big)\ge\tau,\\
&\quad \|\boldsymbol{\phi}_s\|_\infty\le\kappa && \forall s\in[S],\\
&\quad \phi_{s,n}\log(\phi_{s,n}/x_n)\le\phi_{s,n}+\beta_{s,n} && \forall s\in[S],n\in[N],\\
&\quad \boldsymbol{\phi}_s\in\mathbb{R}^N_+,\boldsymbol{\beta}_s\in\mathbb{R}^N && \forall s\in[S],\\
&\quad \boldsymbol{c}^\top\boldsymbol{x}\le 1,\\
&\quad \boldsymbol{x}\ge\boldsymbol{0},\kappa\ge 0.
\end{aligned}
$$

Similarly, for the estimation-fortified robust satisficing model, we first consider the following optimization problem,

$$
\begin{aligned}
&\inf\ \frac{1}{S}\sum_{s\in[S]}\sum_{n\in[N]}\big(x_n\exp(z_{s,n})+K_\tau|z_{s,n}-(\boldsymbol{w}^\top\boldsymbol{u}_n+\hat{v}_s-\boldsymbol{w}^\top\hat{\boldsymbol{u}}_s)|\big)+K_\tau\theta\|\hat{\boldsymbol{w}}-\boldsymbol{w}\|_2\\
&\text{s.t. } \boldsymbol{w}\in\mathbb{R}^J,\\
&\quad \boldsymbol{z}_s\in\mathbb{R}^N \qquad \forall s\in[S].
\end{aligned}
$$

The right-hand side of the first constraint is equivalent to

$$
\begin{aligned}
&\inf\ \frac{1}{S}\sum_{s\in[S]}\Big(\boldsymbol{a}_s^\top\boldsymbol{x}+K_\tau\|\boldsymbol{b}_s\|_1\Big)+K_\tau\theta\|\boldsymbol{\zeta}\|_2\\
&\text{s.t. } \exp(z_{s,n})\le a_{s,n} && \forall n\in[N],s\in[S],\\
&\quad b_{s,n}=z_{s,n}-(\boldsymbol{w}^\top\boldsymbol{u}_n+\hat{v}_s-\boldsymbol{w}^\top\hat{\boldsymbol{u}}_s) && \forall n\in[N],s\in[S],\\
&\quad \boldsymbol{z}_s,\boldsymbol{b}_s\in\mathbb{R}^N,\boldsymbol{a}_s\ge\boldsymbol{0} && \forall s\in[S],\\
&\quad \boldsymbol{\zeta}=\hat{\boldsymbol{w}}-\boldsymbol{w},\\
&\quad \boldsymbol{w},\boldsymbol{\zeta}\in\mathbb{R}^J.
\end{aligned}
$$

We can obtain its dual problem as follows,

$$
\begin{aligned}
&\sup\ \sum_{s\in[S]}\sum_{n\in[N]}(\phi_{s,n}\hat{v}_s-\beta_{s,n})-\boldsymbol{\eta}^\top\hat{\boldsymbol{w}}\\
&\text{s.t. } \sum_{s\in[S]}\sum_{n\in[N]}\phi_{s,n}(\boldsymbol{u}_n-\hat{\boldsymbol{u}}_s)+\boldsymbol{\eta}=\boldsymbol{0},\\
&\quad \|\boldsymbol{\phi}_s\|_\infty\le K_\tau/S && \forall s\in[S],\\
&\quad \|\boldsymbol{\eta}\|_2\le\theta K_\tau,\\
&\quad -d_{s,n}\log(-Sd_{s,n}/x_n)+d_{s,n}-\beta_{s,n}\le 0 && \forall s\in[S],n\in[N],\\
&\quad \boldsymbol{d}_s+\boldsymbol{\phi}_s=\boldsymbol{0} && \forall s\in[S],\\
&\quad \boldsymbol{d}_s\in\mathbb{R}^N_-,\boldsymbol{\beta}_s,\boldsymbol{\phi}_s\in\mathbb{R}^N && \forall s\in[S],\\
&\quad \boldsymbol{\eta}\in\mathbb{R}^J.
\end{aligned}
$$

Hence, we can equivalently reformulate Problem (26) as follows,

$$
\begin{aligned}
\Theta_{\tau,\underline{\tau}} = \min\ & \theta \\
\text{s.t.}\ & \frac{1}{S}\sum_{s\in[S]}\sum_{n\in[N]}\left(\phi_{s,n}\hat{v}_s - \beta_{s,n}\right) - \boldsymbol{\eta}^\top \hat{\boldsymbol{w}} \geq \underline{\tau}, \\
& \frac{1}{S}\sum_{s\in[S]}\sum_{n\in[N]}\phi_{s,n}(\boldsymbol{u}_n - \hat{\boldsymbol{u}}_s) + \boldsymbol{\eta} = \boldsymbol{0}, \\
& \phi_{s,n}\log\left(\phi_{s,n}/x_n\right) \leq \phi_{s,n} + \beta_{s,n} && \forall s\in[S], n\in[N], \\
& \|\boldsymbol{\phi}_s\|_\infty \leq K_\tau && \forall s\in[S], \\
& \|\boldsymbol{\eta}\|_2 \leq \theta K_\tau, \\
& \boldsymbol{c}^\top \boldsymbol{x} \leq 1, \\
& \boldsymbol{x} \geq \boldsymbol{0}, \theta \geq 0, \boldsymbol{\eta} \in \mathbb{R}^J, \\
& \boldsymbol{\phi}_s \in \mathbb{R}^N_+, \boldsymbol{\beta}_s \in \mathbb{R}^N && \forall s\in[S].
\end{aligned}
\tag{37}
$$

This completes the proof. □

*Proof of Theorem 7.* For the residual-based multiproduct pricing problem, given $\kappa$, the robust counterpart of the robust satisficing constraint of Problem (29) is the following convex (quadratic) optimization problem with only equality constraints:

$$
\begin{aligned}
\inf\ & \boldsymbol{p}^\top \hat{\boldsymbol{w}} + \boldsymbol{p}^\top \hat{\boldsymbol{Q}}\boldsymbol{p} + \frac{1}{S}\sum_{s\in[S]}\left(\boldsymbol{p}^\top \boldsymbol{\epsilon}_s + \kappa\|\boldsymbol{b}_s\|\right) \\
\text{s.t.}\ & \boldsymbol{b}_s = \boldsymbol{\epsilon}_s - \hat{\boldsymbol{v}}_s + \hat{\boldsymbol{w}} + \hat{\boldsymbol{Q}}\hat{\boldsymbol{p}}_s && \forall s\in[S], \\
& \boldsymbol{\epsilon}_s \in \mathbb{R}^N && \forall s\in[S].
\end{aligned}
$$

The equivalent dual problem is given by

$$
\begin{aligned}
\sup\ & \boldsymbol{p}^\top \hat{\boldsymbol{Q}}\boldsymbol{p} + \frac{1}{S}\sum_{s\in[S]}\boldsymbol{p}^\top\left(\hat{\boldsymbol{v}}_s - \hat{\boldsymbol{Q}}\hat{\boldsymbol{p}}_s\right) \\
\text{s.t.}\ & \|\boldsymbol{p}\|_* \leq \kappa.
\end{aligned}
$$

This problem can be equivalently written as

$$
\begin{aligned}
\sup\ & \frac{1}{2}(\hat{\boldsymbol{Q}} + \hat{\boldsymbol{Q}}^\top)\bullet \boldsymbol{Y} + \frac{1}{S}\sum_{s\in[S]}\boldsymbol{p}^\top\left(\hat{\boldsymbol{v}}_s - \hat{\boldsymbol{Q}}\hat{\boldsymbol{p}}_s\right) \\
\text{s.t.}\ & \begin{pmatrix} \boldsymbol{Y} & \boldsymbol{p} \\ \boldsymbol{p}^\top & 1 \end{pmatrix} \succeq \boldsymbol{0}, \\
& \boldsymbol{Y} \in \mathbb{S}^N, \|\boldsymbol{p}\|_* \leq \kappa.
\end{aligned}
$$

Substituting this dual problem into Problem (29) yields the final optimization problem.

For the estimation-fortified multiproduct pricing problem, note that Problem (30) has the following equivalent form:

$$\begin{aligned}
&\min\ \theta\\
&\text{s.t.}\ \inf_{(\boldsymbol{\epsilon}_1,\ldots,\boldsymbol{\epsilon}_S,\boldsymbol{w},\boldsymbol{Q})\in\bar{\mathcal{Z}}}\left\{\frac{1}{S}\sum_{s\in[S]}\left(\boldsymbol{p}^\top(\boldsymbol{w}+\boldsymbol{Q}\boldsymbol{p}+\boldsymbol{\epsilon}_s)+\kappa\,\|\boldsymbol{\epsilon}_s-\hat{\boldsymbol{v}}_s+\boldsymbol{w}+\boldsymbol{Q}\hat{\boldsymbol{p}}_s\|+K_\tau\theta\|\hat{\boldsymbol{w}}-\boldsymbol{w}+(\hat{\boldsymbol{Q}}-\boldsymbol{Q})\hat{\boldsymbol{p}}_s\|\right)\right\}\geq\underline{\tau},\\
&\quad\ \boldsymbol{p}\in[\underline{\boldsymbol{p}},\bar{\boldsymbol{p}}],\ \theta\geq 0,\lambda\geq 0,
\end{aligned}$$

where

$$\bar{\mathcal{Z}}=\{(\boldsymbol{\epsilon}_1,\ldots,\boldsymbol{\epsilon}_S,\boldsymbol{w},\boldsymbol{Q})\mid\boldsymbol{\epsilon}_1,\ldots,\boldsymbol{\epsilon}_S\in\mathbb{R}^N,\boldsymbol{w}\in\mathbb{R}^N,\boldsymbol{Q}+\boldsymbol{Q}^\top\preceq\boldsymbol{0}\}.$$

Given $\boldsymbol{p}\in[\underline{\boldsymbol{p}},\bar{\boldsymbol{p}}]$ and $\kappa\geq 0$, the LHS of the robust constraint can be written as

$$\begin{aligned}
&\inf\ \frac{1}{2}\boldsymbol{P}\bullet\boldsymbol{p}\boldsymbol{p}^\top+\boldsymbol{w}^\top\boldsymbol{p}+\frac{1}{S}\sum_{s\in[S]}\left(\boldsymbol{p}^\top\boldsymbol{\epsilon}_s+\ K_\tau(\|\boldsymbol{\eta}_s\|+\theta\|\boldsymbol{\phi}_s\|)\right)\\
&\text{s.t.}\ \boldsymbol{P}=\boldsymbol{Q}+\boldsymbol{Q}^\top,\\
&\quad\ \boldsymbol{\eta}_s=\boldsymbol{\epsilon}_s-\hat{\boldsymbol{v}}_s+\boldsymbol{w}+\boldsymbol{Q}\hat{\boldsymbol{p}}_s && \forall s\in[S],\\
&\quad\ \boldsymbol{\phi}_s=\hat{\boldsymbol{w}}-\boldsymbol{w}+(\hat{\boldsymbol{Q}}-\boldsymbol{Q})\hat{\boldsymbol{p}}_s && \forall s\in[S],\\
&\quad\ \boldsymbol{\epsilon}_s,\boldsymbol{\eta}_s,\boldsymbol{\phi}_s\in\mathbb{R}^N && \forall s\in[S],\\
&\quad\ \boldsymbol{P}\preceq\boldsymbol{0},\boldsymbol{w}\in\mathbb{R}^N,\boldsymbol{Q}\in\mathbb{R}^{N\times N}.
\end{aligned}\tag{38}$$

We write the dual of Problem (38) as the following:

$$\begin{aligned}
&\sup\ \sum_{s\in[S]}\left(-\boldsymbol{r}_s^\top\hat{\boldsymbol{v}}_s+\boldsymbol{t}_s^\top(\hat{\boldsymbol{w}}+\hat{\boldsymbol{Q}}\hat{\boldsymbol{p}}_s)\right)\\
&\text{s.t.}\ \begin{pmatrix}\boldsymbol{Y} & \boldsymbol{p}\\ \boldsymbol{p}^\top & 1\end{pmatrix}\succeq\boldsymbol{0},\\
&\quad\ \boldsymbol{Y}+\sum_{s\in[S]}(\boldsymbol{r}_s-\boldsymbol{t}_s)\hat{\boldsymbol{p}}_s^\top=\boldsymbol{0},\\
&\quad\ \boldsymbol{p}+\sum_{s\in[S]}(\boldsymbol{r}_s-\boldsymbol{t}_s)=\boldsymbol{0},\\
&\quad\ \boldsymbol{p}+S\boldsymbol{r}_s=\boldsymbol{0} && \forall s\in[S],\\
&\quad\ \|\boldsymbol{r}_s\|_*\leq K_\tau/S && \forall s\in[S],\\
&\quad\ \|\boldsymbol{t}_s\|_*\leq K_\tau\theta/S && \forall s\in[S],\\
&\quad\ \boldsymbol{r}_s,\boldsymbol{t}_s\in\mathbb{R}^N && \forall s\in[S],\\
&\quad\ \boldsymbol{Y}\in\mathbb{S}^N.
\end{aligned}$$

Hence, Problem (30) can be reformulated as follows,

$$
\begin{aligned}
\Theta_{\tau,\underline{\tau}} = \min\ & \theta \\
\text{s.t.}\ & \frac{1}{S}\sum_{s\in[S]}\left(\boldsymbol{p}^\top\hat{\boldsymbol{v}}_s + \boldsymbol{t}_s^\top\hat{\boldsymbol{w}} + \boldsymbol{t}_s^\top\hat{\boldsymbol{Q}}\hat{\boldsymbol{p}}_s\right) \geq \underline{\tau}, \\
& \begin{pmatrix} \boldsymbol{Y} & \boldsymbol{p} \\ \boldsymbol{p}^\top & 1 \end{pmatrix} \succeq \boldsymbol{0}, \\
& \boldsymbol{Y} = \frac{1}{S}\sum_{s\in[S]}(\boldsymbol{p}+\boldsymbol{t}_s)\hat{\boldsymbol{p}}_s^\top, \\
& \frac{1}{S}\sum_{s\in[S]}\boldsymbol{t}_s = \boldsymbol{0}, \\
& \|\boldsymbol{p}\|_* \leq K_\tau, \\
& \|\boldsymbol{t}_s\|_* \leq K_\tau\theta & \forall s\in[S], \\
& \boldsymbol{t}_s \in \mathbb{R}^N & \forall s\in[S], \\
& \boldsymbol{Y}\in\mathbb{S}^N, \boldsymbol{p}\in[\underline{\boldsymbol{p}},\bar{\boldsymbol{p}}], \\
& \theta \geq 0.
\end{aligned}
$$

This completes the proof. □