Satisficers, in contrast to maximizers, are content with attaining a reasonable target that they set for themselves. We develop a new prescriptive analytics tool called robust satisficing that uses data to help a satisficer achieve her target expected reward or consumption as well as possible under ambiguous risks and prediction uncertainty. It builds upon the robustness optimization framework recently proposed by Long et al. [2021], and we extend it to incorporate aspects of predictive analytics. The extension is non-trivial. We adopt linear regression as the underlying predictive model and propose a new estimator uncertainty and residual ambiguity set to characterize the relations between the underlying regression coefficients, which is uncertain but non-stochastic, and the stochastic random variables representing residuals that have ambiguous distributions. The robust satisficing model is also useful in allocating resources for multiple satisficing agents to meet their expected reward targets. We present some useful robust satisficing models that can be solved efficiently, and provide tractable approximations to tackle adaptive linear optimization problems. The simulation studies for newsvendor problems elucidate the benefits of the robust satisficing framework in helping the firm attain the target expected profits, mitigate shortfalls, and limit target surplus, if desired. The robust satisficing model can also improve solutions over one that is obtained by solving a baseline empirical optimization model using estimated parameters. The improvement is also more pronounced when data availability is limited. Paradoxically, maximizers can also benefit from the analytics of robust satisficing.

Key words: Robust satisficing, predictive analytics, prescriptive analytics, robustness optimization
A prescriptive analytics model uses data and leverages on predictive analytics to help managers make better decisions. Because of future uncertainty, the inputs to such a model usually involve parameters that are uncertain at the time when decisions are made. In such environment, managers are generally concerned with the prospect of not meeting their performance targets. As articulated by Simon (1955), target satisficing, as opposed to utility maximizing, is prevalent in human decision making, especially in complex situations facing risks and uncertainty.

While most prescriptive analytics tools appeal to maximizers, there are limited tools that cater for satisficers. In the spirit of corporate responsibility, such tools can also be used to facilitate firms in their quests for more sustainable growths through satisficing at sustainable expected consumption targets. 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 uncertainty of achieving a satisfactory target. Charnes and Cooper (1963) propose satisficing frameworks via maximizing the probability of target attainment, but such models are computationally intractable and require knowledge of the underlying probability distribution that is often impossible to know. incidentally the family of satisficing decision criteria axiomatize by Brown and Sim (2009) have an embedded preference for diversification, which is sensible and consistent with risk-aversion behavior. Unlike the success probability criterion, such preference also leads to computational tractability for convex problems. Some interesting 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). However, these papers do not leverage on data-driven predictive models.

Data holds information of the historical realizations of the uncertain parameters and other observable factors (side information) that could be used to predict the statistics of the future outcomes. There has been an emergence of prescriptive analytics tools that incorporate observable factors in optimization models (see, e.g., Gao et al. 2017, Bertsimas et al. 2019a, Ho and Hanasusanto 2019). Having side information can improve decision making and is useful in many 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). A common approach in addressing data-driven optimization problems is to incorporate predicted outcomes in the optimization model (see, e.g., Ferreira et al. 2016, Glaeser et al. 2019). There is also an emerging stream of research in joint prediction and optimization or decision-aware learning. This stream of research includes, inter alia, incorporating information of the decision process to the prediction stage (see, e.g., Tulabandhula and Rudin 2013, Elmachtoub and Grigas 2021), using machine learning methods to improve empirical optimization by reweighting data.
points (see, e.g., Bertsimas and Kallus 2020, Kallus and Mao 2020). Some of these ideas are applied in the newsvendor problem (Liyanage and Shanthikumar 2005, Ban and Rudin 2019). As far as we know, Zhu et al. (2021) propose the only satisficing model that incorporates prediction from data using observable factors and uncertainty associated with the prediction. However, the model assume a stochastic-free reward function and does not naturally extend to evaluate expected reward for functions that are not affine in its uncertain parameters.

In this paper, we develop a new prescriptive analytics tool that uses data to help a satisficing decision maker achieve her target expected reward or consumption as well as possible under risk ambiguity and prediction uncertainty. Our robust satisficing framework is based on the satisficing decision criterion of Brown and Sim (2009). It is an extension of the robustness optimization framework of Long et al. (2021), which has demonstrated to have better solution stability and improved efficient frontiers compared to a data-driven robust optimization model. The extension is non-trivial; we incorporate an embedded prediction model in the form of linear regression, and cater for estimation uncertainty associated with the prediction model. Similar to the data-driven robust optimization (see e.g., Mohajerin Esfahani and Kuhn 2018), the robust satisficing framework adopts the Wasserstein metric to characterize the penalty associated with probability distributions deviating away from the referenced empirical distribution. As a prescriptive analytics tool, the robust satisficing framework incorporates observable factors and characterizes uncertain parameters as random variables with ambiguous distributions that depend on the observable factors. Specifically, we propose a new estimator uncertainty and residual ambiguity set to characterize the relations between the underlying regression coefficients, which is uncertain but non-stochastic, and the stochastic random variables representing residuals that have ambiguous distributions. The robust satisficing model is also useful in allocating resources for multiple satisficing agents to meet their expected reward targets.

We present some useful robust satisficing models that can be solved efficiently and provide tractable approximations to tackle adaptive linear optimization problems. The simulation studies on newsvendor problems elucidate the benefits of the robust satisficing framework in helping the firm attain the desired target expected profits, mitigate shortfalls, and control target surplus, if desired. Incidentally, in the presence of risk ambiguity and prediction uncertainty, our robust satisficing model can also improve solutions over one that is obtained by solving a baseline empirical optimization model using estimated parameters. As articulated in Smith and Winkler (2006) Optimizer’s Curse, a naive optimization model whose objective is based on the empirical distribution would likely underestimate the actual objective under the true distribution. In our simulation studies, we observe improvement in the true objective value if we set targets close to the optimal objective of the empirical optimization model, which has an effect of alleviating the optimizer’s
curse. The improvement is also more pronounced when data availability is limited. Paradoxically, maximizers can also benefit from the analytics of robust satisficing.

**Notation.** We denote by \([S] := \{1, 2, \ldots, S\}\) the set of positive indices up to \(S\). We use boldface glyphs, such as \(x \in \mathbb{R}^D\) and \(A \in \mathbb{R}^{M \times N}\) to denote vectors and matrices; we denote by \(x_i\) the \(i\)th element of vector \(x\) and by \(A_{ij}\) the element located in the \(i\)th row and \(j\)th column of matrix \(A\). The vector \(0\) of the appropriate dimension is the the vector of 0s. We use \(\tilde{v} \sim P \in \mathbb{P}_0(\mathbb{R}^I)\) to denote an \(I\)-dimensional random variable \(\tilde{v}\) governed by a probability distribution \(P\), where \(\mathbb{P}_0(\mathbb{R}^I)\) represents the set of all probability distributions in \(\mathbb{R}^I\). For a set \(V \subseteq \mathbb{R}^I\), the term \(\mathbb{P}[\tilde{v} \in V]\) represents the probability of \(\tilde{v}\) lying in the set \(V\) evaluated on the distribution \(P\). For a probability distribution \(P\), we use \(\mathbb{E}_P[\cdot]\) to signify the corresponding expectation.

2. The maximizers’ model

We consider a maximizer who aims to optimize her expected reward or consumption, with the help of data to predict the future outcomes. To set up the model for the maximizer, we consider the optimization problem with decision variable \(x \in X \subseteq \mathbb{R}^D\), and reward function \(f(x, z) : X \times Z \mapsto \mathbb{R}\), where the input to the second argument is subject to uncertainty. We will focus on the case that the function is Lipschitz continuous with respect to the second argument so that there exists a function \(\Lambda(x) : X \mapsto \mathbb{R}\) such that

\[
|f(x, z_1) - f(x, z_2)| \leq \Lambda(x) \|z_1 - z_2\|
\]

for all \(z_1, z_2 \in Z\). The uncertain parameters of the problem are collectively denoted by the random variable \(\tilde{z}\) over the support \(Z \subseteq \mathbb{R}^N\), which, unless otherwise stated, is a closed, convex and bounded set. Data is available in the form of historical realizations of \(\tilde{z}\) alongside observable factors that can be used to predict the statistics of the future outcomes. At the time when the decision has to be made, the decision maker observes and/or decides on a set of \(L\) factors, denoted by \(y \in Y \subseteq \mathbb{R}^L\), which can be used to predict and/or influence the mean of \(\tilde{z}\). Here, \(Y\) is a bounded set. For generality, we let \(y\) be part of the decision variables that could influence the expected reward.

We focus on the ubiquitous linear regression as the underlying predictive model. For convenience of notation, we define \(\bar{y} := \left(\begin{array}{c} 1 \\ y \end{array}\right)\) so that the mean of \(\tilde{z}\) over its true distribution, \(\mathbb{P}^*\) is given by

\[
\mathbb{E}_{\mathbb{P}^*}[\tilde{z}_n] = w_n^\top \bar{y},
\]

for some unobservable and non-stochastic coefficients \(w_n^* \in \mathbb{R} \times \mathcal{W}_n, \mathcal{W}_n \subseteq \mathbb{R}^L, n \in [N]\). We also define the set of feasible coefficient matrices

\[
\mathcal{W} := \left\{ W \in \mathbb{R}^{N \times (L+1)} \mid W = \begin{bmatrix} w_1^\top \\ \vdots \\ w_N^\top \end{bmatrix} \text{ for some } w_n \in \mathbb{R} \times \mathcal{W}_n, n \in [N] \right\},
\]
so that we can express
\[
\tilde{z} = W^* \bar{y} + \tilde{\epsilon}
\]
for some coefficient matrix $W^* \in \mathcal{W}$ and a vector of random residuals, $\tilde{\epsilon}$ that has zero mean under the distribution $\mathbb{P}^*$. To show the dependency explicitly, we define the following function map
\[
\xi : \mathcal{W} \times \mathbb{R}^p \times \mathbb{R}^N \mapsto \mathbb{R}^N,
\]
\[
\xi(W, y, \epsilon) := W \bar{y} + \epsilon.
\]
Given the feasible decision set, $\bar{\mathcal{X}} \subseteq \mathcal{X} \times \mathcal{Y}$, ideally, the maximizer should solve the following risk-neutral optimization problem,
\[
\begin{align*}
Z^* &= \max_{(x, y) \in \bar{\mathcal{X}}} \mathbb{E}_{\mathbb{P}^*}[f(x, \xi(W^*, y, \bar{\epsilon}))] \\
\text{s.t.} & \quad (x, y) \in \bar{\mathcal{X}},
\end{align*}
\]
though in practice, the coefficient matrix $W^*$ is unobservable and the true distribution $\mathbb{P}^*$ is not precisely known. Hence, there is distribution ambiguity in evaluating the expected reward, as well as estimation uncertainty in the coefficient matrix that would prohibit the modeler from obtaining the true optimal solution.

**Stochastic-free optimization model**

To avoid having to evaluate the expected reward function, it is common to replace $\tilde{z}$ by its mean value by solving the following stochastic-free optimization problem
\[
\begin{align*}
Z^*_D &= \max_{(x, y) \in \bar{\mathcal{X}}} f(x, W^* \bar{y}) \\
\text{s.t.} & \quad (x, y) \in \bar{\mathcal{X}},
\end{align*}
\]
which does not take into account of the residuals’ risks. The stochastic-free model coincides with the expected reward function if the function $f(x, z)$ is affine in $z$, but underestimate the expected reward if the function is convex. Likewise, when the function is concave in $z$, the actual expected reward would be overestimated by the stochastic-free model. Nevertheless, because of the simpler analysis, stochastic-free models lead to more tractable optimization models. These include the classical non-stochastic robust optimization models
\[
\begin{align*}
Z_R &= \max_{W \in \mathcal{U}} \min_{x \in \mathcal{X}} f(x, W \bar{y}) \\
\text{s.t.} & \quad (x, y) \in \bar{\mathcal{X}},
\end{align*}
\]
for some uncertainty set $\mathcal{U} \subseteq \mathcal{W}$ that may contain $W^*$. Recent works on joint prediction and optimization have focused on the stochastic-free models, see e.g., the shortest-path problem illustrated in Elmachtoub and Grigas (2021), and the newsvendor with pricing problem of Zhu et al. (2021).
Estimate, predict and optimize

The predictive model via linear regression requires the modeler to determine an estimation of the uncertain coefficient matrix from data. After which, she/he would optimize the problem using the predicted outcomes based on the matrix estimates. To do so, the modeler should have the access to $S$ historical outcomes associated with the uncertain parameters, denoted by $\hat{r}_s \in Z, s \in [S]$ and the respective observable factors, denoted by $\hat{v}_s \in V, s \in [S]$. We assume that the augmented factors $\hat{v}_s := \left( \frac{1}{\hat{v}_s} \right), s \in [S]$ spans dimension $L + 1$ and we define

$$V := \frac{1}{S} \sum_{s \in [S]} \hat{v}_s \hat{v}_s^T, \quad \phi_n := \frac{1}{S} \sum_{s \in [S]} \hat{v}_s[r_s]_n, \quad n \in [N],$$

for which $V$ is an invertible matrix. For a given coefficient matrix $W \in W$, it is also convenient to define the function mapping $\hat{\varepsilon} : W \times [S] \mapsto \mathbb{R}^N$,

$$\hat{\varepsilon}(W, s) := \hat{r}_s - W \hat{v}_s,$$

to show the dependency of the empirical residuals on the coefficient matrix. Likewise, we also define the following function map $\hat{\xi} : W \times \mathbb{R}^L \times [S] \mapsto \mathbb{R}^N$,

$$\hat{\xi}(W, y, s) := \xi(W, y, \hat{\varepsilon}(W, s)) = W(\hat{y} - \hat{v}_s) + \hat{r}_s$$

to show how the empirical samples would change if the historical observable factors, $\hat{v}_s$ were to be replaced by the current factor $y$ under $W$. Because the exact coefficient matrix $W^*$ is unknown, we can only provide an estimate matrix, $\hat{W}$ by evaluating how well we can reduce the variability of the residuals. In the ordinary least squares (OLS) estimation, the optimal estimate is determined by minimizing the sum of the empirical residuals’ squares. Incidentally, the OLS coefficients require empirical residuals to be completely uncorrected with the historical factors, so that

$$\frac{1}{S} \sum_{s \in [S]} \hat{v}_s \hat{\varepsilon}_n(\hat{W}^{\text{OLS}}, s) = 0 \quad \forall n \in [N].$$

This yields the unique OLS estimators as follows,

$$\hat{w}_n^{\text{OLS}} := V^{-1} \phi_n \quad \forall n \in [N].$$

More generally, we need not confine to the OLS estimator of the coefficient matrix.

**Definition 1 (Feasible baseline estimator).** We define a feasible baseline estimator of the coefficient matrix, as a matrix $\hat{W} \in W$ that satisfies

$$\xi(\hat{W}, y, s) \in Z \quad \forall y \in Y, s \in [S] \quad \text{and} \quad \frac{1}{S} \sum_{s \in [S]} \hat{\varepsilon}(\hat{W}, s) = 0.$$
These conditions ensure that any feasible baseline estimator of the coefficient matrix must be associated with zero-mean empirical residuals, and that the predicted outcomes based on historical realizations would not violate the support set regardless of the decisions $y \in Y$. Under the definition, observe that the OLS estimator of the coefficient matrix, $\hat{W}^{OLS}$, may not be a feasible baseline estimator. In the subsequent section, we show how we can obtain a feasible baseline estimator.

For given a baseline estimator, $\hat{W}$, Problem (1) can be approximated by solving the baseline empirical optimization problem as follows:

$$
\hat{Z} = \max \frac{1}{S} \sum_{s \in [S]} f \left( x, \hat{\xi}(\hat{W}, y, s) \right)
$$

s.t. $(x, y) \in \bar{X}$.

The optimum objective of the baseline empirical optimization problem, $\hat{Z}$ is unlikely to be the same as the expected reward under the actual distribution that is unknown to the decision maker. Disappointment occurs whenever the actual expected reward is less than $\hat{Z}$, which is likely to occur due to the optimizer’s curse. Using $\hat{Z}$ as a reference, the prudent satisficer would choose a reasonable reward target $\tau$ that can be achieved empirically, i.e., $\tau \leq \hat{Z}$. Our robust satisficing model will seek to attain the target as well as possible under risk ambiguity and prediction uncertainty.

3. The satisficer’s model

In contrast to the maximizer, a satisficer, who is concerned with risk ambiguity and prediction uncertainty, would desire a solution that could attain her target expected reward under the actual risk. We note that when the reward function is associated with consumption, target surplus would be associated with over-consumption, which, depending on the context of applications, may not necessarily be desirable. We propose a model that is based on the Long et al. (2021) robustness optimization model, which we adapt in our context as the robust satisficing model as follows:

$$
\min k
$$

s.t. $\tau - \mathbb{E}_P \left[ f(x, \hat{z}) \right] \leq k \Delta_W(P, \hat{P}) \quad \forall P \in \mathcal{P}_0(Z)$

$x \in \mathcal{X}, \ k \geq 0,$

where

$$
\Delta_W(P, \hat{P}) := \inf_{Q \in \mathcal{P}_0(\mathbb{R}^2)} \left\{ \mathbb{E}_Q \left[ \| \hat{z} - \hat{u} \| \right] \mid (\hat{z}, \hat{u}) \sim Q, \hat{z} \sim P, \hat{u} \sim \hat{P} \right\},
$$

is the Wasserstein metric of type-1 and $\hat{P} | \hat{u} = \hat{z} | = 1/S, \ s \in [S]$ with $\hat{z}_s$ representing the corresponding empirical sample. The prudent satisficer specifies $\tau \leq \hat{Z}$, which represents her target
expected reward that has to be achievable by the empirical optimization model. Equivalently, we can express Problem (5) as

$$\min k$$

subject to

$$\tau - \mathbb{E}_p [f(x, \tilde{z})] \leq k \mathbb{E}_p [\|\tilde{z} - \hat{z}\|] \quad \forall P \in \hat{F}$$

$$x \in \mathcal{X},$$

where $$\hat{F}$$ is defined as

$$\hat{F} := \left\{ P \in \mathcal{P}_0 (\mathbb{R}^N \times [S]) \left| P[\tilde{s} = s] = 1/S \right. \right\}.$$  

Note that to describe the discrete empirical distributions, it is convenient to introduce the random scenario $$\tilde{s}$$ that takes values in $$[S]$$ with equal probability.

**Dissatisfaction measure**

[Long et al. (2021)] motivate their robustness optimizing problem by introducing the underlying decision criterion termed as the *fragility measure*, which we adapt to our context as *dissatisfaction measure*. We define $$\mathcal{L}$$ as the set of random variables representing target shortfalls, $$\tau - f(x, \tilde{z})$$, for all $$x \in \mathcal{X}$$ so that $$\tilde{v} \in \mathcal{L}$$ implies there exists $$x \in \mathcal{X}$$ such that $$\tilde{v}(\omega) = \tau - f(x, z(\omega))$$, for all $$\omega \in \Omega$$. The dissatisfaction measure evaluated on the target shortfalls is a functional $$\rho : \mathcal{L} \mapsto [0, +\infty]$$ that has the following representation

$$\rho(\tilde{v}) = \min k$$

subject to

$$\mathbb{E}_p [\tilde{v}] \leq k \mathbb{E}_p [\|\tilde{z} - \hat{z}\|] \quad \forall P \in \hat{F}$$

$$k \geq 0,$$

so that Problem (6) is equivalent to

$$\min \rho(\tau - f(x, \tilde{z}))$$

subject to $$x \in \mathcal{X}.$$  

Identical to [Long et al. (2021)], we show that the dissatisfaction measure has desirable characteristics that aptly describe robust satisficing preferences, which we present as follows.

**Definition 2 (Properties of Dissatisfaction Measure).** The properties of a dissatisfaction measure $$\rho : \mathcal{L} \mapsto [0, +\infty]$$ is consistent with the robust satisficing preferences, that for any $$\tilde{v}, \tilde{v}_1, \tilde{v}_2 \in \mathcal{L}$$:

1. **Monotonicity:** If $$\tilde{v}_1 \geq \tilde{v}_2$$, then $$\rho(\tilde{v}_1) \geq \rho(\tilde{v}_2)$$.
2. **Positive homogeneity:** For any $$\lambda \geq 0$$, we have $$\rho(\lambda \tilde{v}) = \lambda \rho(\tilde{v})$$.
3. **Subadditivity:** $$\rho(\tilde{v}_1 + \tilde{v}_2) \leq \rho(\tilde{v}_1) + \rho(\tilde{v}_2)$$.  

4. **Contentment**: If \( \bar{v} \leq 0 \), then \( \rho(\bar{v}) = 0 \).

5. **Prudence**: If \( \mathbb{E}_\hat{P}[\bar{v}] > 0 \), then \( \rho(\bar{v}) = \infty \).

The first three properties of the dissatisfaction measures are consistent with the coherent risk measure of [Artzner et al. (1999)]. Contentment is associated with the satisficing preference that if a risky position can attain the desired target under any circumstances, then it should be the most preferred. Prudence describes the situation that when the risky position fails to attain the target in the empirical evaluation, then either the risky position has to be avoided, or the satisficer has to lower her target accordingly.

The following result is the consequence of the Lipschitz continuity of the \( f \) function with respect to its second argument.

**Theorem 1 (Attainment feasibility).** For any \( \bar{v} \in \mathcal{L} \) such that \( \mathbb{E}_\hat{P}[\bar{v}] \leq 0 \), \( \rho(\bar{v}) < \infty \).

**Proof.** For any \( \bar{v} \in \mathcal{L} \) such that \( \mathbb{E}_\hat{P}[\bar{v}] \leq 0 \), there exists \( x \in \mathcal{X} \) such that
\[
\mathbb{E}_\hat{P}[f(x, \hat{z})] \geq \tau.
\]
Observe that for all \( P \in \hat{F} \),
\[
\tau - \mathbb{E}_P[f(x, \hat{z})] \leq \mathbb{E}_\hat{P}[f(x, \hat{z})] - \mathbb{E}_P[f(x, \hat{z})] = \mathbb{E}_\hat{P}[f(x, \hat{z}) - f(x, \hat{z})] \leq \Lambda(x) \mathbb{E}_\hat{P}[\|\hat{z} - \hat{z}\|].
\]
Hence, \( \rho(\bar{v}) \leq \Lambda(x) < \infty \). \( \square \)

Theorem 1 implies that if the baseline empirical optimization problem of Problem (4) is feasible and has objective value \( \hat{Z} \), then its optimal solution would also be feasible in the corresponding robust satisficing problem, as long as the \( \tau \leq \hat{Z} \). Such result is important as it demonstrates that if the target is the same as the objective value achievable by the baseline empirical optimization problem, then we would expect the solutions to coincide. Moreover, solving the baseline empirical optimization problem first would provide the satisficer with the range of expected reward targets that would yield feasible solutions to the robust satisficing problem.

**Characterization of estimator uncertainty and residual ambiguity**

We generalize the robust satisficing framework to incorporate aspects of predictive analytics, with the goal of attaining a target risk-based objective under distribution ambiguity and estimation uncertainty. Specifically, we extend the risk-based robust satisficing framework to include parameter uncertainty so that the distribution of \( \hat{z} = \xi(W^*, y, \hat{\epsilon}) \) is directly dependent on the observed factor \( y \in \mathcal{Y} \), uncertain but non-stochastic matrix, \( W \in \mathcal{W} \) as well as the ambiguous distribution associated with the random residual \( \hat{\epsilon} \). We first define the baseline factor-residual covariance as
\[
\hat{\sigma}_n := \frac{1}{S} \sum_{s \in [S]} \hat{v}_n \hat{\epsilon}_n(\hat{W}, s) \quad \forall n \in [N].
\]
Definition 3 (Characterization of residual ambiguity). For a given $W$, we can characterize the distributional ambiguity of the residuals $(\tilde{\epsilon}, \tilde{s}) \sim \mathbb{P}$ to satisfy the following conditions,

1. **Uniformly distributed scenarios**: $\mathbb{P}[\tilde{s} = s] = 1/S$, for all $s \in S$.
2. **Support feasibility**: $\mathbb{P}\left[\xi(W, v, \tilde{\epsilon}) \in \mathcal{Z} \quad \forall v \in \mathcal{Y}\right] = 1$.
3. **Residuals with zero means**: $\mathbb{E}_{P}[\tilde{\epsilon}] = 0$.
4. **Factor-residual covariance specification**: $\mathbb{E}_{P}[\hat{v}_s \tilde{\epsilon}_n] = \hat{\sigma}_n$, for all $n \in [N]$.

The first three properties are straightforward to motivate, while the last property defines the statistical properties of the residuals associated with the choice of coefficient matrix. Since, $\hat{\sigma}_n$ is associated with the estimator matrix, $\hat{W}$, the last property constrains the residuals so that the distribution of residuals under ambiguity would share some common statistical properties with the empirical residuals associated with the baseline estimator, $\hat{W}$. Indirectly, this specification helps to constrain $W$ within the vicinity of $\hat{W}$. If the baseline estimator is OLS, then the last property would be the strict exogeneity condition with $\mathbb{E}_{P}[\hat{v}_s \tilde{\epsilon}_n] = 0$, $\forall n \in [N]$.

The generality of the last property allows us to use other baseline estimators beyond OLS.

Similar to the representation of Problem (6), we now characterize the distribution of the random variable $(\tilde{\epsilon}, \tilde{s}) \sim \mathbb{P}$ as a function of $W \in \mathcal{W}$ in the following residual ambiguity set,

$$\mathcal{F}(W) := \left\{ P \in \mathcal{P}_0(\mathbb{R}^N \times [S]) \left| \begin{array}{c}
(\tilde{\epsilon}, \tilde{s}) \sim \mathbb{P} \\
\mathbb{E}_P[\tilde{\epsilon}] = 0 \\
\mathbb{E}_{P}[\hat{v}_s \tilde{\epsilon}_n] = \hat{\sigma}_n \\
\mathbb{P}[\tilde{s} = s] = 1/S \\
\mathbb{P}[(\tilde{\epsilon}, W) \in \mathcal{Z}_s \mid \tilde{s} = s] = 1, \forall s \in [S]
\end{array}\right. \right\}, \quad (10)$$

where

$$\mathcal{Z}_s := \left\{ (\epsilon, W) \left| \begin{array}{c}
\xi(W, v, \epsilon) \in \mathcal{Z} \\
\xi(W, v, s) \in \mathcal{Z} \\
\forall v \in \mathcal{Y}
\end{array}\right. \right\}.$$ 

Note that it has well been known in robust optimization that if the sets $\mathcal{Y}$ and $\mathcal{Z}$ are polyhedron with concise representations, then $\mathcal{Z}_s$ would also be a polyhedron with a concise representation.

To this regard, as an extension of Long et al. (2021) robustness optimization model of Problem (6), we consider the following generalized robust satisficing model with estimation uncertainty,

$$\begin{align*}
\min_{k} & \quad \tau - \mathbb{E}_{P}\left[f(x, \xi(W, y, \tilde{\epsilon}))\right] \\
\text{s.t.} & \quad k \mathbb{E}_{P}\left[\Vert \xi(W, y, \tilde{\epsilon}) - \hat{\xi}(W, \tilde{y}, s)\Vert \right] \quad \forall P \in \mathcal{F}(W), W \in \mathcal{W} \quad (11)
\end{align*}$$

$(x, y) \in \bar{X}$, $k \geq 0$, 

or equivalently
\[
\min k \\
\text{s.t. } \tau - \mathbb{E}_P[f(x, \xi(\hat{W}, y, \tilde{\epsilon}))] \leq k \mathbb{E}_P[\|\hat{\epsilon} - \hat{\epsilon}(\hat{W}, \tilde{s})\|] \quad \forall P \in \mathcal{F}(\hat{W}), \hat{W} \in \mathcal{W} 
\]
\[
(x, y) \in \hat{X}, \quad k \geq 0. 
\]  

**Proposition 1.** In the absence of factors, i.e. \(L = 0\), Problem (11) (or Problem (12)) is equivalent to the robust satisficing model of Problem (6).

Proposition 1 implies that the new robust satisficing framework is a generalization of Long et al. (2021). To motivate the robust satisficing model from the decision criterion perspective, we introduce the following estimator uncertainty and residual ambiguity set as follows:
\[
\hat{\mathcal{G}} := \left\{ P \in \mathcal{P}_0(\mathcal{W} \times \mathbb{R}^N \times [S]) \, \bigg| \, \exists \hat{w} \in \mathcal{W}, Q \in \mathcal{F} (\hat{W}) : \right. \\
\left. (\hat{\epsilon}, \tilde{s}) \sim Q, (\hat{W}, \hat{\epsilon}, \tilde{s}) \sim P \right\}.
\]

As far as we know, such hybrid uncertainty and ambiguity set is new and has not been considered even in recent distributionally robust optimization models. The latest algebraic modeling software RSOME by Chen et al. (2020), which introduces the scenario-wise ambiguity set does not readily support its reformulation. Nevertheless, future version of RSOME will support the new hybrid uncertainty and ambiguity set.

Let \(\hat{L}\) be the set of random variables representing the random target shortfalls \(\tau - f(x, \xi(\hat{W}, y, \tilde{\epsilon}))\), so that \(\hat{v} \in \hat{L}\) implies there exists \((x, y) \in \hat{X}\) such that \(v(\omega) = \tau - f(x, \xi(\hat{W}(\omega), y, \epsilon(\omega)))\) for all \(\omega \in \Omega\). Hence, we define \(\hat{\rho} : \hat{L} \mapsto [0, +\infty]\) on the space of target shortfalls as the following dissatisfaction measure,
\[
\hat{\rho}(\hat{v}) = \min k \\
\text{s.t. } \mathbb{E}_P[\hat{v}] \leq k \mathbb{E}_P[\|\hat{\epsilon} - \hat{\epsilon}(\hat{W}, \tilde{s})\|] \quad \forall P \in \hat{\mathcal{G}} \\
k \geq 0, 
\]
which we can easily verify to satisfy the properties that are consistent with robust satisficing preferences of Definition 2. The property of attainment feasibility is more elusive to obtain as we present in the following result.

**Theorem 2 (Attainment feasibility).** For any
\[
\Phi \geq 1 + \max_{y \in \gamma, s \in [S]} \sum_{j \in [S]} |\hat{v}_s^T V^{-1}(y - \hat{v}_j)|,
\]
we have
\[
\mathbb{E}_P[\|\xi(\hat{W}, y, \tilde{\epsilon}) - \hat{\xi}(\hat{W}, y, \tilde{s})\|] \leq \Phi \mathbb{E}_P[\|\hat{\epsilon} - \hat{\epsilon}(\hat{W}, \tilde{s})\|] \quad \forall P \in \mathcal{F}(\hat{W}), \hat{W} \in \mathcal{W}.
\]
Consequently, for any \((x, y) \in \bar{X}\) such that
\[
\frac{1}{S} \sum_{s \in [S]} f(x, \hat{\xi}(\hat{W}, y, s)) \geq \tau,
\]
we have
\[
\bar{\rho}(\tau - f(x, \xi(\hat{W}, y, \hat{\epsilon}))) \leq \Phi \Lambda(x).
\]

Multiple satisficing agents

We can extend the robust satisficing framework for multiple satisficing agents, each with a different reward function and expected target to attain as follows

\[
\min w^\top k \\
\text{s.t.} \quad \tau_i - \mathbb{E}_p \left[f_i(x, \xi(\hat{W}, y, \hat{\epsilon}))\right] \leq k_i \mathbb{E}_p \left[\|\hat{\epsilon} - \hat{\epsilon}(\hat{W}, \hat{s})\|\right], \quad \forall \mathbb{P} \in \hat{\mathcal{G}}, \ i \in [I] \tag{14}
\]

for some chosen weight parameters \(w_i, i \in [I]\). When considering resource allocation for multiple agents, the goal of the robust satisficing model is to ensure that every satisficing agent could attain her desired expected reward target, without explicitly favouring agents with surpluses. Hence, within the constraints of resources, the robust satisficing model ensures that the allocations of resources are adequate and fair in meeting agents’ expected reward targets as well as possible under risk ambiguity and prediction uncertainty.

Obtaining the baseline estimator

We observe that the baseline estimator is also the unique solution when the penalty function associated with the robust satisficing problem is minimized.

**Theorem 3.** The solution

\[
(\hat{W}, \hat{\mathbb{P}}) = \arg \inf_{W \in W, \mathbb{P} \in \mathcal{F}(W)} \mathbb{E}_p \left[\|\hat{\epsilon} - \hat{\epsilon}(W, \hat{s})\|\right]
\]

for which

\[
\hat{\mathbb{P}} \left[\hat{\epsilon} = \hat{\epsilon}(\hat{W}, \hat{s})\right] = 1
\]

is the unique solution such that

\[
\mathbb{E}_p \left[\|\hat{\epsilon} - \hat{\epsilon}(\hat{W}, \hat{s})\|\right] = 0.
\]

Our robust satisficing model may allow the modeler to use any estimator \(W^\dagger\) for the coefficient matrix obtained from the plethora of approaches such as OLS, LASSO, and so forth, provided it is also a feasible baseline estimator under Definition 1. If this is not the case, we can obtain a feasible baseline estimator whose corresponding empirical residuals would yield the least Wasserstein distance from the empirical residuals under \(W^\dagger\) by solving the following optimization problem:
\[
\inf_{S} \frac{1}{S} \sum_{s \in [S]} \| (W - W^t) \hat{v}_s \|
\]
\[
\text{s.t. } \frac{1}{S} \sum_{s \in [S]} (\hat{r}_s - W \hat{v}_s) = 0
\]
\[
(\hat{r}_s - W \hat{v}_s, W) \in Z_s \quad \forall s \in [S]
\]
\[
W \in \mathcal{W}.
\]

In particular, we can use the optimal solution of \(W\) in Problem (15) as the baseline estimator \(\hat{W}\), for which the corresponding baseline factor-residual covariance, \(\hat{\sigma}_n, n \in [N]\) can be determined.

### 4. Tractable robust satisficing models

In this section, we present tractable robust satisficing models. Using distributionally robust optimization techniques, we can reformulate Problem (11) as a standard non-stochastic robust optimization model. To obtain tractable models, we assume that \(\hat{X}\) is a tractable convex set and the reward function \(f(x, z)\) is a saddle function. Specifically, for a given \(x \in \hat{X}\), the function is lower-semicontinuous and convex in \(z \in Z\), and for a given \(z \in Z\), the function is upper-semicontinuous and concave in \(x \in \hat{X}\). Moreover, either one of the following conditions holds for the reward function.

- The set \(Y\) is a singleton.
- For given \((W, \epsilon)\), the reward function \(f(x, W \bar{y} + \epsilon)\) is jointly concave and upper-semicontinuous in \((x, y) \in \hat{X}, W \bar{y} + \epsilon \in Z\).

Moving forward, we will derive explicit formulation that would enable us to express our robust satisficing model as an explicit deterministic optimization problem that can be solved by commercial solvers. The following assumption will greatly simplify our exposition.

**Assumption 1.** Whenever applicable, we will assume that the technical conditions for strong duality holds such as when addressing the duality of a conic optimization problem, minmax exchange, and so forth. This will enable the distributionally robust counterpart to be expressed as a deterministic optimization problem (see, for instance, \cite{Shapiro2009, Bertsekas2009, MohajerinEsfahani2018, Bertsimas2019}).

**Theorem 4.** Under Assumption 1, we can formulate Problem (11) as the following robust optimization problem

\[
\min k
\]
\[
\text{s.t. } \frac{1}{S} \sum_{s \in [S]} \left( f(x, W \bar{y} + \epsilon_s) + k \| \epsilon_s - \hat{r}_s + W \hat{v}_s \| \right) \geq \tau \quad \forall (\epsilon_1, \ldots, \epsilon_S, W) \in \bar{Z},
\]
\[
(x, y) \in \hat{X}, \quad k \geq 0,
\]
where
\[
\tilde{Z} := \left\{ (\epsilon_1, \ldots, \epsilon_S, W) \begin{array}{l}
\frac{1}{S} \sum_{s \in [S]} \epsilon_s = 0 \\
\frac{1}{S} \sum_{s \in [S]} \hat{v}_s \epsilon_s = \hat{\sigma}_n \forall n \in [N] \\
(\epsilon_s, W) \in Z_s \quad \forall s \in [S] \\
W \in W
\end{array} \right\}.
\]

Consequently, we can formulate the robust satisficing problem as a convex deterministic optimization problem via standard robust optimization reformulation techniques (Ben-Tal et al. 2015).

Stochastics-free reward function

A robust satisficing model with a stochastic-free reward function can be simplified as follows

\[
\min k \\
\text{s.t. } \tau - f(x, W \hat{y}) \leq \frac{k}{S} \sum_{s \in [S]} \left( \hat{v}_s^T D (W - \hat{W}) \hat{v}_s - \gamma_s + \nu_s^T (\hat{r}_s - W \hat{v}_s) \right) \forall W \in W
\]

We present the following useful results to transform a robust satisficing model to a classical robust optimization problem.

**Theorem 5.** Under Assumption [1], Problem (17) is equivalent to the following robust optimization problem,

\[
\min k \\
\text{s.t. } \tau - f(x, W \hat{y}) \leq \sum_{s \in [S]} \left( \hat{v}_s^T D (W - \hat{W}) \hat{v}_s - \gamma_s + \nu_s^T (\hat{r}_s - W \hat{v}_s) \right) \forall W \in W
\]

In particular, if \( Z = \mathbb{R}^N \), then Problem (17) is equivalent to the following robust optimization problem,

\[
\min k \\
\text{s.t. } \tau - f(x, W \hat{y}) \leq \sum_{s \in [S]} \left( \hat{v}_s^T D (W - \hat{W}) \hat{v}_s \right) \forall W \in W
\]

\[
\| D^T \hat{v}_s \|_* \leq 1 \quad \forall s \in [S] \\
D \in \mathbb{R}^{(L+1) \times N}
\]

\[
(x, y) \in \bar{X}, k \geq 0.
\]
5. Linear adaptive robust satisficing problem

We can extend the framework to address adaptive linear optimization under risk ambiguity and estimation uncertainty, which opens up to a plethora of interesting applications. Here $x \in \mathcal{X} \subseteq \mathbb{R}^{D_1}$ represents the here-and-now variables before the realization of the random variable $\tilde{z}$ and $f(x, z)$ represents the total two-stage reward function by solving a linear optimization problem after $z$ has been realized as follows:

$$f(x, z) = \max \ c^T(z)x + d^T y$$

subject to $A(z)x + By \leq b(z)$

$y \in \mathbb{R}^{D_2}$,

where

$$A(z) := A_0 + \sum_{i \in [N]} A_i z_i, \quad b(z) := b_0 + \sum_{i \in [N]} b_i z_i, \quad c(z) := c_0 + \sum_{i \in [N]} c_i z_i$$

are affine mappings of $z$. Long et al. (2021) discuss the following robustness adaptive linear optimization model without considering estimation uncertainty:

$$\min_k$$

subject to $\tau - \mathbb{E}_P [c^T(\tilde{z})x + d^T y(\tilde{z})] \leq k \mathbb{E}_P [\|\tilde{z} - \hat{z}_i\|] \quad \forall P \in \hat{F}$

$A(z)x + By(z) \leq b(z)$

$y \in \mathcal{R}^{N, D_2}$

$x \in \mathcal{X}$, $k \geq 0$,

(18)

where the family of recourse functions is defined as

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

However, adaptive optimization model are generally difficult to solve because the recourse decision $y$ can be any arbitrary function of the uncertain parameter $z$. A popular method to tractably solve such an adaptive linear optimization problem is to use affine recourse adaptation (see, e.g., Ben-Tal et al. 2004, Kuhn et al. 2011, Bertsimas et al. 2019b, Chen et al. 2020). In particular, Chen et al. (2020) propose a tractable scenario-wise lifted affine recourse adaptation to solve the problem approximately that performs almost as well as the exact model for a multi-item newsvendor problem. Consider the following class of affine recourse functions

$$\mathcal{L}^{N, P} := \{y \in \mathcal{R}^{N, P} \mid y(z) = y_0 + \sum_{i \in [N]} y_i z_i \text{ for some } y_i \in \mathbb{R}^P, i \in [N] \cup \{0\}\}.$$
Sim et al. (2021) apply a scenario-wise lifted affine recourse adaptation to approximate Problem (18) with the following model:

\[
\begin{align*}
\min & \quad k \\
\text{s.t.} & \quad \tau - \mathbb{E}_P [c^\top (\tilde{z}) x + d^\top y_s (\tilde{z}, \tilde{u})] \leq k \mathbb{E}_P [\tilde{u}] \quad \forall P \in \tilde{F}_{\text{lifted}} \\
& \quad A(z) x + B y_s (z, u) \leq b(z) \quad \forall (z, u) \in \tilde{Z}_s, s \in [S] \\
& \quad y_s \in \mathcal{L}^{N+1, D_2} \quad \forall s \in [S] \\
& \quad x \in \mathcal{X}, \ k \geq 0,
\end{align*}
\]

(19)

where \( \tilde{Z}_s := \{ (z, u) \in \mathcal{Z} \times \mathbb{R} \mid u \geq \|z - \hat{z}_s\| \} \), and

\[
\tilde{F}_{\text{lifted}} := \left\{ P \in \mathcal{P}_0 (\mathcal{Z} \times \mathbb{R} \times [S]) \mid \begin{aligned}
& (\tilde{z}, \tilde{u}, \tilde{s}) \sim P \\
& P[\tilde{s} = s] = 1/S \\
& P[(\tilde{z}, \tilde{u}) \in \tilde{Z}_s \mid \tilde{s} = s] = 1
\end{aligned} \right\}.
\]

When the adaptive optimization problem has complete recourse, \( i.e., \) for any right hand side vector, \( t \), there exists a feasible recourse \( w \) such that \( B w \leq t \). Long et al. (2021) show that Problem (19) would be feasible as long as \( \tau \leq \hat{\tau} \).

Under our framework, we can extend their result to the case with estimation uncertainty. For a tractable approximation, \( \tilde{z} \) cannot be decision dependent, \( i.e., \mathcal{Y} = \{ v \} \) and we only optimize over \( x \in \mathcal{X} \) as in following robustness adaptive linear optimization model:

\[
\begin{align*}
\min & \quad k \\
\text{s.t.} & \quad \tau - \mathbb{E}_P \left[ c^\top (\xi (\hat{W}, v, \tilde{\epsilon})) x + d^\top y_s (\xi (\hat{W}, v, \tilde{\epsilon})) \right] \leq k \mathbb{E}_P [\|\tilde{\epsilon} - \hat{\epsilon} (\hat{W}, s)\|] \quad \forall P \in \hat{G} \\
& \quad A(z) x + B y(z) \leq b(z) \quad \forall z \in \mathcal{Z} \\
& \quad y \in \mathcal{R}^{N, D_2} \\
& \quad x \in \mathcal{X}, \ k \geq 0.
\end{align*}
\]

(20)

Consider the following scenario-wise lifted affine recourse adaptation:

\[
\begin{align*}
\min & \quad k \\
\text{s.t.} & \quad \tau - \mathbb{E}_P \left[ c^\top (\xi (\hat{W}, v, \tilde{\epsilon})) x + d^\top y_s (\xi (\hat{W}, v, \tilde{\epsilon}), \tilde{u}_1) \right] \leq k \mathbb{E}_P [\tilde{u}_2] \quad \forall P \in \hat{G}_{\text{lifted}} \\
& \quad A(z) x + B y_s (z, u_1) \leq b(z) \quad \forall (z, u_1) \in \tilde{Z}_s, s \in [S] \\
& \quad y_s \in \mathcal{L}^{N+1, D_2} \quad \forall s \in [S] \\
& \quad x \in \mathcal{X}, \ k \geq 0,
\end{align*}
\]

(21)
where the lifted ambiguity set $\hat{\mathcal{G}}_{\text{lifted}}$ is defined as:

$$
\hat{\mathcal{G}}_{\text{lifted}} := \left\{ \mathbb{P} \in \mathcal{P}_0(\mathcal{W} \times \mathbb{R}^N \times [S] \times \mathbb{R}) \mid \begin{array}{l}
\exists \mathcal{Q} \in \hat{\mathcal{G}}:
\mathcal{Q} = \mathcal{G},
\mathbb{W}, \mathbb{e}, \tilde{s}, \tilde{u}_1, \tilde{u}_2 \sim \mathbb{P}\\
\mathbb{E}_\mathbb{P}[\tilde{u}_1] \leq \Phi \mathbb{E}_\mathbb{P}[\tilde{u}_2]\\
\mathbb{P}\left[\tilde{u}_1 \geq \|\hat{\xi}(-\mathbb{W}, \mathbb{v}, \mathbb{e}) - \hat{\xi}(\mathbb{W}, \mathbb{v}, \tilde{s})\| \mid \tilde{s} = s\right] = 1 \forall s \in [S]\\
\mathbb{P}\left[\tilde{u}_2 \geq \|\hat{\epsilon} - \hat{\epsilon}(\mathbb{W}, \tilde{s})\| \mid \tilde{s} = s\right] = 1 \forall s \in [S]\end{array}\right\},
$$

Note that we have incorporated the lifted random variables $\tilde{u}_1$ and $\tilde{u}_2$ that are associated with the epigraphs of $\|\hat{\xi}(\mathbb{W}, \mathbb{v}, \mathbb{e}) - \hat{\xi}(\mathbb{W}, \mathbb{v}, \tilde{s})\|$ and $\|\hat{\epsilon} - \hat{\epsilon}(\mathbb{W}, \tilde{s})\|$ respectively. By Theorem 2, we require $\mathbb{E}_\mathbb{P}[\tilde{u}_1] \leq \Phi \mathbb{E}_\mathbb{P}[\tilde{u}_2]$, encoding the underlying relations between $\mathbb{E}_\mathbb{P}\left[\|\hat{\xi}(\mathbb{W}, \mathbb{v}, \mathbb{e}) - \hat{\xi}(\mathbb{W}, \mathbb{v}, \tilde{s})\|\right]$ and $\mathbb{E}_\mathbb{P}\left[\|\hat{\epsilon} - \hat{\epsilon}(\mathbb{W}, \tilde{s})\|\right]$ for any $\mathbb{P} \in \hat{\mathcal{G}}_{\text{lifted}}$. The following proposition states that our affine recourse adaptation is built on an equivalent model to Model (20), indicating the validity of the scenario-wise lifted affine recourse adaptation.

**Proposition 2.** The set of marginal distributions in $\hat{\mathcal{G}}_{\text{lifted}}$ onto $(\mathbb{W}, \mathbb{e}, \tilde{s})$ is equivalent to $\hat{\mathcal{G}}$.

The scenario-wise lifted affine recourse adaptation in Model (19) is still a feasible one for Model (20) as long as $\tau \leq \hat{Z}$, when the adaptive linear optimization problem has complete recourse.

**Theorem 6 (Attainment feasibility under complete recourse).** Suppose the adaptive linear optimization problem has complete recourse, then the scenario-wise lifted affine recourse adaptation model (21) is feasible whenever $\tau \leq \hat{Z}$.

Theorem 6 extends the corresponding result of Long et al. (2021) to the case with estimation uncertainty. Model (21) can achieve the same in-sample expected reward as the empirical optimization with baseline estimator $\hat{\mathbb{W}}$ when $\tau = \hat{Z}$. As we will show in our simulation study, greater confidence of target attainment compared to the baseline solution can be achieved if $\tau < \hat{Z}$.

Following Theorem 4, Problem (21) can be equivalently written as the following robust linear optimization model:

$$
\begin{align*}
\min k \\
\text{s.t.} \quad & \frac{1}{S} \sum_{s \in [S]} \left( c^\top (\xi(W, v, \epsilon_s))x + d^\top y_s(\xi(W, v, \epsilon_s), u_{1s}) + ku_{2s} \right) \geq \tau \quad \forall(\epsilon_1, \ldots, \epsilon_S, u_1, u_2, W) \in \hat{Z}, \\
A(z)x + By_s(z, u_1) & \leq b(z) \quad \forall(\epsilon, u_1) \in \hat{Z}_s, s \in [S] \\
y_s & \in \mathcal{L}^{N+1,D_2} \quad \forall s \in [S] \\
x & \in \bar{X}, \quad k \geq 0,
\end{align*}
$$

Sim et al.: The Analytics of Robust Satisficing 17
where the joint uncertainty set is given by

$$\tilde{Z} := \left\{ (\epsilon_1, \ldots, \epsilon_S, u_1, u_2, W) \mid \begin{align*}
&\frac{1}{S} \sum_{s \in [S]} \epsilon_s = 0 \\
&\frac{1}{S} \sum_{s \in [S]} \hat{v}_s[\epsilon_s]_n = \hat{\sigma}_n \quad \forall n \in [N] \\
&(\xi(W, v, \epsilon_s), u_{1s}) \in \tilde{Z}_s \quad \forall s \in [S] \\
&u_{2s} \geq \|\epsilon_s - \hat{\epsilon}(W, s)\| \quad \forall s \in [S] \\
&\sum_{s \in [S]} (u_{1s} - \Phi u_{2s}) \leq 0 \\
&W \in \mathcal{W}
\right\}.$$  

In Appendix A, we present an application of a newsvendor problem and show that our model is practically implementable and can help the firm achieve the desired profit targets.

References


A. An application in a newsvendor problem

Consider a retailer selling one product with a retail price $p$ and an ordering cost $c$. The retailer’s profit is $f(x, z) = (p - c)x - p(x - z)^+ = p\min\{x, z\} - cx$, where $x$ is the order quantity and $\tilde{z}$ is the stochastic demand. The retailer has $S$ historical observations pairs, $(\hat{r}_s, \hat{v}_s)_{s \in [S]}$, where $\hat{v}_s$ may contains information such as weather, seasonality, and promotions. We do not consider decisions in factors, i.e., $Y = \{\bar{v}\}$. Hence, $\bar{v}$ corresponds to the observed exogenous factors, which already contains the intercept. Under scenario-wise lifted affine recourse adaptation, the retailer can solve the following robust satisficing (RS) model to achieve a target profit $\tau$ as much as possible under ambiguity and estimation uncertainty:

$$
\begin{align*}
\min_k & \\
\text{s.t.} & E_p [p y_s (\xi (W, \bar{v}, \hat{\epsilon}, \hat{u}_1))] + k E_p [\hat{u}_2] - cx \geq \tau & \forall P \in \bar{F}(W), W \in \mathcal{W} \\
y_s(z, u_1) & \leq x & \forall (z, u_1) \in \bar{Z}_s, s \in [S] \\
y_s(z, u_1) & \leq z & \forall (z, u_1) \in \bar{Z}_s, s \in [S] \\
y_s & \in \mathcal{L}^{2,1} & \forall s \in [S] \\
x & \geq 0.
\end{align*}
$$

(22)

Following our results on adaptive linear optimization, Problem (22) can be reformulated into a robust linear optimization model:

$$
\begin{align*}
\min_k & \\
\text{s.t.} & \frac{1}{S} \sum_{s \in [S]} (ry_s (Wz + \epsilon_s, u_{1s}) + ku_{2s}) \geq \tau + cx & \forall (\epsilon_1, \ldots, \epsilon_S, u_1, u_2, W) \in \bar{Z} \\
y_s(z, u_1) & \leq x & \forall (z, u_1) \in \bar{Z}_s, s \in [S] \\
y_s(z, u_1) & \leq z & \forall (z, u_1) \in \bar{Z}_s, s \in [S] \\
y_s & \in \mathcal{L}^{2,1} & \forall s \in [S] \\
x & \geq 0,
\end{align*}
$$

where the joint uncertainty set is given by

$$
\bar{Z} := \left\{ (\epsilon_1, \ldots, \epsilon_S, u_1, u_2, W) \left| \begin{array}{l}
\frac{1}{S} \sum_{s \in [S]} \epsilon_s = 0 \\
\frac{1}{S} \sum_{s \in [S]} \hat{v}_s [\epsilon_s]_n = \hat{\sigma}_n & \forall n \in [N] \\
(Wz + \epsilon_s, u_{1s}) \in \bar{Z}_s & \forall s \in [S] \\
u_{2s} \geq \|\epsilon_s - \hat{\epsilon} (W, s)\| & \forall s \in [S] \\
\sum_{s \in [S]} (u_{1s} - \Phi u_{2s}) \leq 0 \\
W \in \mathcal{W}
\end{array} \right. \right\}.
$$
The above robust linear optimization model can be reformulated to a linear optimization model using standard robust optimization techniques. For brevity, we omit the explicit formulation here. The key message is that our robust satisficing model can incorporate estimation uncertainty without increasing the computational complexity, compared to the robustness optimization (RnO) approach in [Long et al. (2021)], which is a special case of our robust satisficing model without the consideration of estimation uncertainty.

We conduct a simulation study to illustrate that a satisficer can always better achieve the target profit using our approach. Specifically, robust satisficing model consistently leads to a lower target shortfall whenever target is not achieved, and it leads to a lower target surplus whenever target is attained out-of-sample. In addition, we illustrate that a maximizer can also benefit from our robust satisficing approach. Specifically, by accepting a small loss of in-sample profit, the robust satisficing model can improve the out-of-sample profit compared to the empirical optimization (EO) model. [Long et al. (2021)] show that a one-product newsvendor problem can be solved to optimality under the robustness optimization approach and their scenario-wise lifted affine recourse adaptation. This result is not necessarily true when we incorporate estimation uncertainty in the robust satisficing model. Nevertheless, in terms of target satisficing, we show in our simulation study that the solution of our robust satisficing model outperforms that of the robustness optimization model.

Simulation

The retail price $p$ is set as 10, and the unit cost $c$ is 7. We let the demand be non-negative, i.e., $Z = \mathbb{R}_+$. We do not impose further restrictions on $W$ and let $W = \mathbb{R}^{L+1}$, where the number of factors is set as $L+1 = 30$. The baseline estimator is chosen to be the OLS estimator.

In each experiment, the true coefficient vector $w^* \in \mathbb{R}^{L+1}$ is randomly generated. Among the 30 elements in the true coefficient vector, we let 20 of them be non-zero, and the rest of the factors are not relevant. Then, among the twenty relevant coefficients, five are generated uniformly from $[-10, -1]$, and fifteen are generated uniformly from $[1, 10]$. The factors of a data record, $v$, is generated as follows. First, the first element is set to $v_0 = 1$, representing the intercept term. Then, $v_\ell$, $\ell \in [L]$, are drawn from independent and identically distributed uniform distributions on $[10, 20]$. The demand associated with an observation of factors, $v$, is given as

$$\tilde{z} = w^* \mathbf{v} + \tilde{\epsilon},$$

where $\tilde{\epsilon}$ is a zero-mean random noise, and we let $\tilde{\epsilon} \sim N(0, \sigma^2)$. In this simulation, we set $\sigma = 20$. Note that the demand realizations are on the scale of a few hundreds; hence, $\sigma = 20$ is still a realistic setting. We generate $S$ samples, $\{(\tilde{v}_1, \tilde{r}_1), \ldots, (\tilde{v}_S, \tilde{r}_S)\}$, as the training data. The number of samples $S$ is chosen to be a multiple of the number of factors, i.e., $S = \beta(L+1)$ for some $\beta > 0$. 

In our simulation, we vary $\beta \in \{2, 3, 5\}$ to reflect a small data to medium data setting. At the point of decision-making, we generate the observed factors $\bar{v}$, but the true demand is unknown. To test the out-of-sample performance of the solutions given factors $\bar{v}$, we generate 1,000 demand realizations based on the true data generating model as the test data.

We consider two benchmarks—empirical optimization model and Long et al. (2021) robustness optimization model. Both benchmarks use the OLS estimates, $\hat{w}$. In any instance, we first solve the empirical optimization model and get the baseline objective $\hat{Z}$. By Theorem 6, both robust satisficing model and robustness optimization model are feasible whenever $\tau \leq \hat{Z}$. We solve the two models over a range of targets $\tau = \alpha \hat{Z}$ for $\alpha \in [\alpha, 1]$. We refer to $\alpha$ as the normalized target. For each model $m \in \{EO, RnO, RS\}$ and normalized target $\alpha$, we attain the profit of model $m$ under realized demand with index $t \in [1000]$, $\pi^m_t(\alpha)$. Then, the out-of-sample average profit of model $m$ is $\sum_{t \in [1000]} \pi^m_t(\alpha)/1000$, and the out-of-sample target deviation is $\sum_{t \in [1000]} (\alpha \hat{Z} - \pi^m_t(\alpha))/1000$. Target deviation represents target shortfall whenever it is positive, and it represents target surplus whenever it is negative. Satisfiers aim to maximize the robustness to uncertainty of achieving the target. Specifically, they want the absolute value of the target deviation to be small. Maximizers aim to achieve higher expected profits. In this simulation, we illustrate that our proposed RS model would always better satisfy the satisfiers’ objective, while also being able to help maximizers achieve a higher profit, compared to benchmark models. For better illustration, for each instance, we normalize the out-of-sample profits of the three models by that of the EO model; we also normalize the out-of-sample target deviation for each normalized target by that corresponding target.

**Summary of results** We present the results using instances with $\beta = 2$. We generate 100 random instances, each with observed factors $\bar{v}_i$, for $i \in [100]$. Each element of the observed factors, $\bar{v}_{i\ell}$, for $i \in [100]$ and $\ell \in [L]$, is randomly drawn from a uniform distribution on $[10, 20]$, which has the same support set as the factors in the historical sample. The average normalized out-of-sample metrics are summarized in Figure 1. On the left, we plot the average profit with respect to the normalized target $\alpha$. As we can see, by accepting a small loss of in-sample profit, we can achieve a higher out-of-sample average profit than the EO model. More importantly, we focus on the plot of average target deviation on the right. Our model has the best frontier in terms of target deviation. With large targets, we have a low level of target shortfall. With small targets, we have low target surplus. Note that in above experiment, $\bar{v}$ is generated from the same support set as the historical samples. In this case, we observe that the OLS would often overestimate and underestimate the mean demand with a close probability for different random instances. However, if we generate $\bar{v}$ from a different support as the historical samples, then OLS may underestimate or overestimate
more consistently, which would enable our model to provide more significant improvement as we illustrate in the next experiment.

Now, we randomly generate $\bar{v}_1, \ldots, \bar{v}_{100}$ from a different support as the historical sample. Specifically, we generate $\bar{v}_{i\ell}$ from a uniform distribution on $[0, 10]$, for $i \in [100]$, $\ell \in [L]$. By design, the 100 observed factors consistently lie on one side of the historical samples, and OLS would overestimate the true demand more frequently in this particular experiment. The average out-of-sample normalized metrics are summarized in Figure 2. Again, our model consistently achieves the best frontier in terms of target deviation, indicating the robustness of our solutions to attaining the prescribed target while controlling target surplus. In addition, because the OLS would overestimate the true demand more frequently in this particular experiment, our model can potentially achieve a better expected out-of-sample average profit.

Besides leading to a superior frontier on target deviation, our model also leads to less risky ordering decisions. In Figure 3, we plot the standard deviations of out-of-sample profits of the three models, normalized by that of the EO model. Our model has the best frontier on the percentage reduction of standard deviation of out-of-sample profits with respect to the EO model. This indicates the robust satisficing model not only aims to achieve the target profit in expectation, but also attain it much more safely than benchmark models.

We have tested more instances, and the above observations are consistent for $\beta \in \{2, 3, 5\}$ and other cost parameter setting. We always achieve a superior frontier on target deviation and reduction of stand deviation compared to the two benchmarks in this problem setting. The main message we want to convey is that a satisficer would be able to best satisfy his prescribed target using our proposed model. We can also improve out-of-sample average profit whenever the OLS estimates overestimate the true demand. For brevity and to avoid repetition, we do not present the figures for other experiments.
Figure 1  Newsvendor problem: Experiment 1 (averaging over 100 random instances).

Figure 2  Newsvendor problem: Experiment 2 (averaging over 100 random instances).
Figure 3  Normalized standard deviation of out-of-sample profits for experiment 1 (left) and experiment 2 (right).
B. Proof of results

Proof of Proposition 1
When \( L = 0 \), Problems (9) and (11) can be, respectively, rewritten as

\[
\min_{x \in X, k \geq 0} \frac{1}{k}
\begin{align*}
\text{s.t.} \quad & \quad \tau - \mathbb{E}_p [f(x, \hat{z})] \leq k \mathbb{E}_p [\|\hat{z} - \hat{r}_s\|] \quad \forall p \in \tilde{F},
\end{align*}
\]

and

\[
\min_{x \in X, k \geq 0} \frac{1}{k}
\begin{align*}
\text{s.t.} \quad & \quad \tau - \mathbb{E}_p [f(x, W + \hat{e})] \leq k \mathbb{E}_p [\|\hat{e} - W - \hat{r}_s\|] \quad \forall p \in \mathcal{F}(W), \; W \in \mathbb{R}^N.
\end{align*}
\]

Note that for any \( W \in \mathcal{W} \) and \( \mathbb{P}^\dagger \in \mathcal{F}(W) \), there exists \( p \in \tilde{F} \) such that \( \mathbb{P}[\hat{z} \in \mathcal{E}] = \mathbb{P}^\dagger[W + \hat{e} \in \mathcal{E}] \), for all \( \mathcal{E} \subseteq \mathcal{Z} \). Similarly, for any \( p \in \tilde{F} \), let \( W = \mathbb{E}_p[\hat{z}] \) and there also exists \( \mathbb{P}^\dagger \in \mathcal{F}(W) \) such that \( \mathbb{P}^\dagger[W + \hat{e} \in \mathcal{E}] = \mathbb{P}[\hat{z} \in \mathcal{E}] \), for all \( \mathcal{E} \subseteq \mathcal{Z} \). This completes the proof.

Proof of Theorem 2

Note that for any \( W \in \mathcal{W} \), \( p \in \mathcal{F}(W) \),

\[
\mathbb{E}_p \left[ \|\hat{\xi}(W, y, \hat{e}) - \hat{\xi}(W, y, \hat{s})\| \right] - \Phi \mathbb{E}_p [\|\hat{e} - \hat{\xi}(W, \hat{s})\|]
= \mathbb{E}_p \left[ \|W\hat{y} + \hat{e} - W(\hat{y} - \hat{v}_s) - \hat{r}_s\| - \Phi \|\hat{e} - \hat{r}_s + W\hat{v}_s\| \right].
\]

It suffices to show that

\[
\sup_{W \in \mathcal{W}, p \in \mathcal{F}(W)} \mathbb{E}_p \left[ \|W\hat{y} + \hat{e} - W(\hat{y} - \hat{v}_s) - \hat{r}_s\| - \Phi \|\hat{e} - \hat{r}_s + W\hat{v}_s\| \right] \leq 0.
\]

Let \( d_n := \left( \begin{array}{c} 0 \\ \alpha_n \end{array} \right) = \frac{1}{S} \sum_{s \in [S]} \hat{v}_s \hat{e}_n(W, s), \; n \in [N], \; \mathcal{Z}_s(W) := \{\epsilon, (\epsilon, W) \in \mathcal{Z}_s\}, \; s \in [S], \) and

\[
\mathcal{F}_s(W) := \left\{ p \in \mathcal{P}_0(\mathbb{R}^N) \left| p \mid \hat{e} = \bar{P}, \; \mathbb{P}[(\hat{e}, W) \in \mathcal{Z}_s] = 1 \right. \right\} \quad \forall s \in [S].
\]

We define

\[
\mathcal{A} := \left\{ (a_1, \ldots, a_S) \in \mathcal{B} \left| \forall s \in [S]: ||a_s(\epsilon)||_\star \leq 1, \forall \epsilon \in \mathbb{R}^N \right. \right\},
\]

where \( \mathcal{B} \) represents some linear topological space for which the measurable function maps \( a_s : \mathbb{R}^N \to \mathbb{R}^N, \; s \in [S] \) are defined. In the spirit of Assumption 1, \( \mathcal{A} \) is a compact convex subset of an appropriate linear topological space, \( \mathcal{B} \). Then, we have

\[
\sup_{p \in \mathcal{F}(W)} \mathbb{E}_p \left[ \|W\hat{y} + \hat{e} - W(\hat{y} - \hat{v}_s) - \hat{r}_s\| - \Phi \|\hat{e} - \hat{r}_s + W\hat{v}_s\| \right]
= \sup_{p \in \mathcal{F}_s(W), a \in [S]} \left\{ \inf_{\delta_n \in \mathbb{R}^{L + 1}, n \in [N]} \frac{1}{S} \sum_{s \in [S]} \mathbb{E}_{\hat{p}_s} \left[ \|W\hat{y} + \hat{e}_s - W(\hat{y} - \hat{v}_s) - \hat{r}_s\| - \Phi \|\hat{e}_s - \hat{r}_s + W\hat{v}_s\| \right] \right\}
\]
\[
\begin{align*}
&\quad + \sum_{n \in [N]} \delta_n^T (\hat{v}_s [\hat{e}_s]_n - d_n) \bigg) \\
= \sup_{P_s \in \mathcal{F}(W), s \in [S]} \left\{ \inf_{\delta_n \in \mathbb{R}^{L+1}, n \in [N]} \frac{1}{S} \sum_{s \in [S]} \mathbb{E}_{\hat{y} \sim \mathcal{A}_s} \left[ \sup_{\|a_s\|_s \leq 1} a_s^T (W \hat{y} + \hat{e}_s - \hat{W} (\hat{y} - \hat{v}_s) - \hat{r}_s) \right. \\
&\quad - \Phi \|\hat{e}_s - \hat{r}_s + W \hat{v}_s\| + \sum_{n \in [N]} \delta_n^T (\hat{v}_s [\hat{e}_s]_n - d_n) \bigg) \right\} \\
= \sup_{P_s \in \mathcal{F}(W), s \in [S]} \left\{ \inf_{\delta_n \in \mathbb{R}^{L+1}, n \in [N]} \left\{ \sup_{(a_1, \ldots, a_s) \in \mathcal{A}} \frac{1}{S} \sum_{s \in [S]} \int_{\epsilon_s \in \mathcal{Z}_s(W)} a_s^T (\epsilon_s) (W \hat{y} + \epsilon_s - \hat{W} (\hat{y} - \hat{v}_s) - \hat{r}_s) \\
&\quad - \Phi \|\epsilon_s - \hat{r}_s + W \hat{v}_s\| + \sum_{n \in [N]} \delta_n^T (\hat{v}_s [\epsilon_s]_n - d_n) \right\} \right\} \\
= \sup_{P_s \in \mathcal{F}(W), s \in [S]} \left\{ \sup_{(a_1, \ldots, a_s) \in \mathcal{A}} \left\{ \inf_{\delta_n \in \mathbb{R}^{L+1}, n \in [N]} \frac{1}{S} \sum_{s \in [S]} \int_{\epsilon_s \in \mathcal{Z}_s(W), s \in [S]} a_s^T (\epsilon_s) (W \hat{y} + \epsilon_s - \hat{W} (\hat{y} - \hat{v}_s) - \hat{r}_s) \\
&\quad - \Phi \|\epsilon_s - \hat{r}_s + W \hat{v}_s\| + \sum_{n \in [N]} \delta_n^T (\hat{v}_s [\epsilon_s]_n - d_n) \right\} \right\} \\
\leq \sup_{P_s \in \mathcal{F}(W), s \in [S]} \left\{ \sup_{(a_1, \ldots, a_s) \in \mathcal{A}} \int_{\epsilon_s \in \mathcal{Z}_s(W), s \in [S]} \frac{1}{S} \sum_{s \in [S]} a_s^T (\epsilon_s) (W \hat{y} + \epsilon_s - \hat{W} (\hat{y} - \hat{v}_s) - \hat{r}_s) \\
&\quad - \Phi \|\epsilon_s - \hat{r}_s + W \hat{v}_s\| + \sum_{n \in [N]} \delta_n^T (\epsilon_1, \ldots, \epsilon_s) (\hat{v}_s [\epsilon_s]_n - d_n) \right\} \right\}
\end{align*}
\]

where the first equality is due to the Lagrangian relaxation and using the law of total probability, the second equality is due to the definition of dual norm, the fourth equality is due to Sion's minimax theorem [Sion 1958], the last inequality is due to the relaxation of the inner optimization, and choosing the following function maps,

\[
\delta_0 (\epsilon_1, \ldots, \epsilon_s) := V^{-1} \left[ \sum_{j \in [S]} [a_j (\epsilon_j)]_n (\hat{y} - \hat{v}_j) \right] \quad \forall \epsilon_s \in \mathcal{Z}_s(W), s \in [S], n \in [N].
\]

We will show that

\[
\sup_{W \in \mathcal{W}} \left\{ \sup_{P_s \in \mathcal{F}(W)} \int_{\epsilon_s \in \mathcal{Z}_s(W), s \in [S]} \frac{1}{S} \sum_{s \in [S]} a_s^T (\epsilon_s) (W \hat{y} + \epsilon_s - \hat{W} (\hat{y} - \hat{v}_s) - \hat{r}_s) \\
&\quad - \Phi \|\epsilon_s - \hat{r}_s + W \hat{v}_s\| + \sum_{n \in [N]} \delta_n^T (\epsilon_1, \ldots, \epsilon_s) (\hat{v}_s [\epsilon_s]_n - d_n) \right\} \leq 0.
\]

It suffices to show that, for all \( \epsilon_s \in \mathcal{Z}_s(W), W \in \mathcal{W}, (a_1, \ldots, a_s) \in \mathcal{A}, s \in [S], \)

\[
\frac{1}{S} \sum_{s \in [S]} a_s^T (\epsilon_s) (W \hat{y} + \epsilon_s - \hat{W} (\hat{y} - \hat{v}_s) - \hat{r}_s) - \Phi \|\epsilon_s - \hat{r}_s + W \hat{v}_s\|
\]
\[ + \sum_{n \in [N]} \bar{\delta}_n^T (\epsilon_1, \cdots, \epsilon_S) (\hat{v}_s[\epsilon_s]_n - d_n) \]

\[ = \frac{1}{S} \sum_{s \in [S]} \left( a_s^T (\epsilon_s) \hat{W} (\hat{v}_s - \hat{y}) - a_s^T (\epsilon_s) \hat{r}_s \right) + \frac{1}{S} \sum_{s \in [S]} \left( a_s^T (\epsilon_s) \epsilon_s + \sum_{n \in [N]} \delta_n^T (\epsilon_1, \cdots, \epsilon_S) \hat{v}_s[\epsilon_s]_n \right) \]

\[ + \frac{1}{S} \sum_{s \in [S]} a_s^T (\epsilon_s) W \hat{y} - \frac{1}{S} \sum_{s \in [S]} \Phi \| \epsilon_s - \hat{r}_s + W \hat{v}_s \| - \sum_{n \in [N]} \delta_n^T (\epsilon_1, \cdots, \epsilon_S) d_n \]

\[ \leq 0. \]

For the first term, we have

\[ = \frac{1}{S} \sum_{s \in [S]} \left( \sum_{n \in [N]} [a_s(\epsilon_s)]_n (\hat{v}_s - \hat{y})^T \hat{w}_n \right) - a_s^T (\epsilon_s) \hat{r}_s \]

\[ = \frac{1}{S} \sum_{s \in [S]} \left( \sum_{n \in [N]} [a_s(\epsilon_s)]_n (\hat{v}_s - \hat{y})^T \hat{w}_n \right) - a_s^T (\epsilon_s) \hat{r}_s \]

\[ = \frac{1}{S} \sum_{s \in [S]} \left( \sum_{n \in [N]} [a_s(\epsilon_s)]_n (\hat{v}_s - \hat{y})^T \hat{w}_n \right) - a_s^T (\epsilon_s) \hat{r}_s \]

\[ = \frac{1}{S} \sum_{s \in [S]} \left( \sum_{n \in [N]} [a_s(\epsilon_s)]_n (\hat{v}_s - \hat{y})^T \hat{w}_n \right) - a_s^T (\epsilon_s) \hat{r}_s \]

\[ = \frac{1}{S} \sum_{s \in [S]} \left( \sum_{n \in [N]} [a_s(\epsilon_s)]_n (\hat{v}_s - \hat{y})^T \hat{w}_n \right) - a_s^T (\epsilon_s) \hat{r}_s \]

\[ = \frac{1}{S} \sum_{s \in [S]} \left( \sum_{n \in [N]} [a_s(\epsilon_s)]_n (\hat{v}_s - \hat{y})^T \hat{w}_n \right) - a_s^T (\epsilon_s) \hat{r}_s \]

\[ = \sum_{n \in [N]} \delta_n^T (\epsilon_1, \cdots, \epsilon_S) d_n - \frac{1}{S} \sum_{j \in [S]} \left( \hat{v}_j^T \left[ \delta_1^T (\epsilon_1, \cdots, \epsilon_S), \cdots, \delta_N^T (\epsilon_1, \cdots, \epsilon_S) \right] + a_j^T (\epsilon_j) \right) \hat{r}_j, \]

where the second equality is due to \( \frac{1}{S} \sum_{s \in [S]} \hat{v}_s (\hat{r}_s - \hat{w}_n^T \hat{v}_s) = d_n, \forall n \in [N] \), which implies

\[ \hat{w}_n = V^{-1} \left( \sum_{s \in [S]} (\hat{v}_s[\hat{r}_s]_n - d_n) \right) \forall n \in [N]. \]

For the second term, we have

\[ \frac{1}{S} \sum_{s \in [S]} \left( a_s^T (\epsilon_s) \epsilon_s + \sum_{n \in [N]} \delta_n^T (\epsilon_1, \cdots, \epsilon_S) \hat{v}_s[\epsilon_s]_n \right) \]

\[ = \frac{1}{S} \sum_{s \in [S]} \left( \frac{\hat{v}_s^T}{\sqrt{S}} \left[ \delta_1^T (\epsilon_1, \cdots, \epsilon_S), \cdots, \delta_N^T (\epsilon_1, \cdots, \epsilon_S) \right] + a_s^T (\epsilon_s) \right) \epsilon_s. \]
In the following, we consider the third term,
\[
\frac{1}{S} \sum_{s \in [S]} a_s^\top (\epsilon_s) W \hat{y} = \frac{1}{S} \sum_{s \in [S]} (a_s^\top (\epsilon_s) W (\hat{y} - \hat{v}_s) + a_s^\top (\epsilon_s) W \hat{v}_s)
\]
\[
= \frac{1}{S} \sum_{s \in [S]} \left\{ \left( \sum_{n \in [N]} [a_s(\epsilon_s)]_n (\hat{y} - \hat{v}_s)^\top w_n \right) + a_s^\top (\epsilon_s) W \hat{v}_s \right\}
\]
\[
= \frac{1}{S} \sum_{s \in [S]} \left\{ \left( \sum_{n \in [N]} [a_s(\epsilon_s)]_n (\hat{y} - \hat{v}_s)^\top V^{-1} \left[ \sum_{j \in [S]} \hat{w}_j \hat{v}_j^\top \right] w_n \right) + a_s^\top (\epsilon_s) W \hat{v}_s \right\}
\]
\[
= \frac{1}{S} \sum_{j \in [S]} \left\{ \sum_{n \in [N]} \left[ \sum_{s \in [S]} (a_s(\epsilon_s))_n (\hat{y} - \hat{v}_s)^\top V^{-1} (\hat{w}_j \hat{v}_j^\top) w_n \right] + a_j^\top (\epsilon_j) W \hat{v}_j \right\}
\]
\[
= \frac{1}{S} \sum_{j \in [S]} \left\{ \sum_{n \in [N]} \left[ \sum_{s \in [S]} \left( \hat{v}_j^\top (\delta_n(\epsilon_1, \ldots, \epsilon_S) w_n) \hat{v}_j \right) + a_j^\top (\epsilon_j) W \hat{v}_j \right] \right\}
\]
\[
= \frac{1}{S} \sum_{j \in [S]} \left\{ \left[ \hat{v}_j^\top (\delta_1(\epsilon_1, \ldots, \epsilon_S), \ldots, \delta_n(\epsilon_1, \ldots, \epsilon_S)) \right] \hat{v}_j + a_j^\top (\epsilon_j) W \hat{v}_j \right\}
\]
\[
\leq \frac{1}{S} \sum_{j \in [S]} \left\{ \left[ \delta_1(\epsilon_1, \ldots, \epsilon_S), \ldots, \delta_n(\epsilon_1, \ldots, \epsilon_S) \right] ^\top \hat{v}_s + a_s(\epsilon_s) \right\} \leq 0,
\]
where the first inequality is due to the definition of dual norm, and the last inequality is due to
\[
\left\| \left[ \delta_1(\epsilon_1, \ldots, \epsilon_S), \ldots, \delta_n(\epsilon_1, \ldots, \epsilon_S) \right] ^\top \hat{v}_s + a_s(\epsilon_s) \right\|.
\]
\[
\leq 1 + \max_{\|a_s\| \leq 1, s \in [S]} \left\{ \sum_{j \in [S]} a_j \hat{v}_j^\top V^{-1} (\hat{v} - \hat{v}_j) \right\}.
\]
\[
\leq 1 + \max_{\|a_s\| \leq 1, s \in [S]} \left\{ \sum_{j \in [S]} |a_j| \hat{v}_j^\top V^{-1} (\hat{v} - \hat{v}_j) \right\},
\]
\[
= 1 + \max_{\|a_s\| \leq 1, s \in [S]} \left\{ \sum_{j \in [S]} |a_j| \hat{v}_j^\top V^{-1} (\hat{v} - \hat{v}_j) \right\} \leq \Phi.
\]

Therefore, we have
\[
\text{first term} + \text{second term} + \text{third term} - \text{fourth term} - \text{fifth term}
\]
\[
\leq \frac{1}{S} \sum_{s \in [S]} \left\{ \left[ \delta_1(\epsilon_1, \ldots, \epsilon_S), \ldots, \delta_n(\epsilon_1, \ldots, \epsilon_S) \right] ^\top \hat{v}_s + a_s(\epsilon_s) \right\} \leq 0,
\]
where the first inequality is due to the definition of dual norm, and the last inequality is due to
\[
\left\| \left[ \delta_1(\epsilon_1, \ldots, \epsilon_S), \ldots, \delta_n(\epsilon_1, \ldots, \epsilon_S) \right] ^\top \hat{v}_s + a_s(\epsilon_s) \right\|.
\]

The second statement follows from the proof of Theorem 1, which completes the proof.
Proof of Theorem 3

Observe that \( \mathbb{E}_P[\|\xi(W, y, \hat{e}) - \xi(W, y, \hat{e}(W, \hat{s}))\|] \geq 0 \), for all \( P \in \mathcal{F}(W) \), \( W \in \mathcal{W} \). For \( \hat{P} \in \mathcal{P}_0(\mathbb{R}^N \times \{S\}) \) such that \( \hat{P}[\hat{e} = \hat{e}(W, \hat{s})] = 1 \), we should have

\[
\mathbb{E}_{\hat{P}} \left[ \|\xi(W, \hat{e}) - \xi(W, \hat{e}(W, \hat{s}))\| \right] = 0,
\]

which implies \( (W, \hat{P}) \) is an optimal solution to the following optimization problem,

\[
\inf_{W \in \mathcal{W}, P \in \mathcal{F}(W)} \mathbb{E}_P[\|\xi(W, y, \hat{e}) - \xi(W, y, \hat{e}(W, \hat{s}))\|].
\]

Next we show the uniqueness. Denote \( \mathcal{F}_{\text{opt}} \) as the set of all optimal solutions to the above problem. Then we have

\[
\mathcal{F}_{\text{opt}} := \left\{ (W, P) \left| \mathbb{E}_P[\|\xi(W, y, \hat{e}) - \xi(W, y, \hat{e}(W, \hat{s}))\|] = 0 \right. \quad \mathbb{P} \in \mathcal{F}(W), \quad W \in \mathcal{W} \right\}.
\]

We first show that \( W = \hat{W} \) holds if \( (W, P) \in \mathcal{F}_{\text{opt}} \). For any \( (W, P) \in \mathcal{F}_{\text{opt}} \), we have \( \mathbb{P}[\xi(W, y, \hat{e}) = \xi(W, y, \hat{e}(W, \hat{s}))] = 1 \), \( \mathbb{E}_P[\hat{e}] = 0 \), and \( \mathbb{E}_P[\hat{e}_n] = \hat{\sigma}_n \) for all \( n \in [N] \). Suppose there exists \( W \neq \hat{W} \) and \( P \in \mathcal{F}(W) \) such that \( (W, P) \in \mathcal{F}_{\text{opt}} \).

It follows that

\[
\frac{1}{S} \sum_{s \in [S]} \left[ \hat{v}_s \left( [\hat{r}_s]_n - w_n^T \hat{w}_s \right) \right] = \left( \begin{array}{c} 0 \\ \hat{\sigma}_n \end{array} \right), \quad \forall n \in [N].
\]

From these equations, we obtain

\[
w_n = \left( \sum_{s \in [S]} \hat{v}_s \hat{v}_s^T \right)^{-1} \left[ \left( \sum_{s \in [S]} \hat{v}_s [\hat{r}_s]_n \right) - S \left( \begin{array}{c} 0 \\ \hat{\sigma}_n \end{array} \right) \right]
\]

\[
= \left( \sum_{s \in [S]} \hat{v}_s \hat{v}_s^T \right)^{-1} \left[ \left( \sum_{s \in [S]} \hat{v}_s [\hat{r}_s]_n \right) - \sum_{s \in [S]} \hat{v}_s \left( [\hat{r}_s]_n - w_n^T \hat{w}_s \right) \right]
\]

\[
= \left( \sum_{s \in [S]} \hat{v}_s \hat{v}_s^T \right)^{-1} \left[ \left( \sum_{s \in [S]} \hat{v}_s [\hat{r}_s]_n \right) - \sum_{s \in [S]} \hat{v}_s \left( [\hat{r}_s]_n - w_n^T \hat{w}_s \right) \right]
\]

\[
= \left( \sum_{s \in [S]} \hat{v}_s \hat{v}_s^T \right)^{-1} \left( \sum_{s \in [S]} \hat{v}_s \hat{v}_s^T \right) \hat{w}_n
\]

\[
= \hat{w}_n,
\]

for all \( n \in [N] \). Here the existence of inverse matrix stems from the fact that \( \hat{v}_s, s \in [S] \) span dimension of \( L + 1 \), and the third equality is because of \( \sum_{s \in [S]} \hat{e}(\hat{W}, s) = 0 \). Yet \( w_n = \hat{w}_n \) for all \( n \in [N] \) contradicts with \( W \neq \hat{W} \). Therefore, for any \( (W, P) \in \mathcal{F}_{\text{opt}} \), we have \( W = \hat{W} \), which further implies that \( \mathcal{F}_{\text{opt}} = \{(\hat{W}, \hat{P})\} \).
Proof of Theorem 4

Under Assumption 1 we have

\[
\inf_{\delta \in \mathcal{F}(W)} \mathbb{E}_\pi \left[ f(\delta, W \bar{y} + \hat{e}) + k \| \hat{e} - \hat{r} + W \hat{v}_s \| \right] = \sup_{D \in \mathbb{R}^{(L+1) \times N}} \left\{ \inf_{(\epsilon_s, W) \in \mathbb{Z}_s, s \in [S]} \frac{1}{S} \sum_{s \in [S]} \left( f(\epsilon_s, W \bar{y} + \epsilon_s) - \hat{v}_s^T D(\epsilon_s - \hat{r}_s + \hat{W} \hat{v}_s) \right) + k \| \epsilon_s - \hat{r}_s + W \hat{v}_s \| \right\},
\]

where the first equality is due to the strong duality guaranteed by Assumption 1 and the second equality stems from Sion’s minimax theorem [Sion 1958]. The assertion now follows if we apply the above result to Problem (11).

\[ \square \]

Proof of Theorem 5

Let \( d_n := (0, \sigma_n) = \frac{1}{S} \sum_{s \in [S]} \hat{v}_s e_n(W, s), n \in [N] \). For the first statement, we have

\[
\inf_{\delta \in \mathcal{F}(W)} \mathbb{E}_\pi \left[ \| \xi(W, y, \hat{e}) - \xi(W, y, \hat{e}(W, \hat{s})) \| \right] = \sup_{\delta_n \in \mathbb{R}_{L+1}, n \in [N]} \left\{ \inf_{(\epsilon_s, W) \in \mathbb{Z}_s, s \in [S]} \frac{1}{S} \sum_{s \in [S]} \left( \| \epsilon_s - \hat{r}_s + W \hat{v}_s \| + \sum_{n \in [N]} \delta_n \| \hat{v}_s(\epsilon_s)_n - d_n \| \right) \right\},
\]

where the equality is due to the strong duality guaranteed by Assumption 1. Let \( \mathbb{I}(\epsilon_s|\mathbb{Z}_s(W)) \) denote indicator functions on \( \mathbb{Z}_s(W), s \in [S], \) defined by

\[
\mathbb{I}(\epsilon_s|\mathbb{Z}_s(W)) := \begin{cases} 0, & \text{if } \epsilon_s \in \mathbb{Z}_s(W) \\ \infty, & \text{otherwise} \end{cases},
\]

for all \( s \in [S], \) where \( \mathbb{Z}_s(W) := \{ \epsilon | (\epsilon, W) \in \mathbb{Z}_s \}, s \in [S]. \) Then, we have

\[
\sup_{\delta_n \in \mathbb{R}_{L+1}, n \in [N]} \left\{ \inf_{(\epsilon_s, W) \in \mathbb{Z}_s, s \in [S]} \frac{1}{S} \sum_{s \in [S]} \left( \| \epsilon_s - \hat{r}_s + W \hat{v}_s \| + \sum_{n \in [N]} \delta_n \| \hat{v}_s(\epsilon_s)_n - d_n \| \right) \right\}
\]
\[
\begin{align*}
&= \sup_{\delta_n \in \mathbb{R}^{L+1}, n \in [N]} \left\{ \inf_{(\epsilon_s, W) \in Z_s, \epsilon_s \in [S]} \frac{1}{S} \sum_{s \in [S]} \left( \| \epsilon_s - \hat{r}_s + W \hat{v}_s \| + \sum_{n \in [N]} \delta_n^T \hat{v}_s \epsilon_n \right) \\
&\quad - \sum_{n \in [N]} \delta_n^T \left( \frac{1}{S} \sum_{s \in [S]} \hat{v}_s \epsilon_n \right) \right\} \\
&= \sup_{\delta_n \in \mathbb{R}^{L+1}, n \in [N]} \left\{ \inf_{(\epsilon_s, W) \in Z_s, \epsilon_s \in [S]} \frac{1}{S} \sum_{s \in [S]} \left( \| \epsilon_s - \hat{r}_s + W \hat{v}_s \| + \sum_{n \in [N]} \delta_n^T \hat{v}_s \epsilon_n - \sum_{n \in [N]} \delta_n^T \hat{v}_s \epsilon_n \right) \right\} \\
&= \sup_{D \in \mathbb{R}^{(L+1) \times N}} \left\{ \inf_{\epsilon_s \in \mathbb{R}^N, \epsilon_s \in [S]} \frac{1}{S} \sum_{s \in [S]} \left( \| \epsilon_s - \hat{r}_s + W \hat{v}_s \| - \hat{v}_s^T D \epsilon_s - \hat{r}_s + W \hat{v}_s \right) \right\} \\
&= \sup_{D \in \mathbb{R}^{(L+1) \times N}} \left\{ -\sup_{\epsilon_s \in \mathbb{R}^N, \epsilon_s \in [S]} \frac{1}{S} \sum_{s \in [S]} \left( \hat{v}_s^T D \epsilon_s - \hat{r}_s + W \hat{v}_s \right) \right\} \\
&= \sup_{D \in \mathbb{R}^{(L+1) \times N}} \left\{ \frac{1}{S} \sum_{s \in [S]} \left( \hat{v}_s^T D \epsilon_s - \hat{r}_s + W \hat{v}_s \right) - \inf_{\| \mu_s \|_s \leq 1, \mu_s, \nu_s \in \mathbb{R}^N} \left\{ \mathbb{I}^* \left( \nu_s | Z_s(W) \right) \right\} \right\} \\
&\quad - \nu_s^T \left( \hat{r}_s - W \hat{v}_s \right) \bigg| \mu_s + \nu_s = D^T \hat{v}_s \bigg) \\
&= \sup \frac{1}{S} \sum_{s \in [S]} \left( \hat{v}_s^T D \epsilon_s - \hat{r}_s + W \hat{v}_s \right) - \gamma_s + \nu_s^T \left( \hat{r}_s - W \hat{v}_s \right) \\
\text{s.t.} & \mu_s + \nu_s = D^T \hat{v}_s \quad \forall s \in [S] \\
&\| \mu_s \|_s \leq 1, \mu_s, \nu_s \in \mathbb{R}^N \quad \forall s \in [S] \\
&\gamma_s \geq \nu_s^T \epsilon \quad \forall \epsilon \in Z_s(W), \ s \in [S] \\
&\| \mu_s \|_s \leq 1, \mu_s, \nu_s \in \mathbb{R}^N \quad \forall s \in [S] \\
&\gamma \in \mathbb{R}^S, D \in \mathbb{R}^{(L+1) \times N},
\end{align*}
\]

where the first equality is due to the definition of \( d_n, n \in [N] \), the third equality is using the change-of-variable trick \( D := [-\delta_1, \cdots, -\delta_N] \), the fourth equality is due to the definition of indicator function, the sixth equation is using the \( \inf \) to \( \sup \) conversion, the seventh equality is using the sum-rules for conjugate functions (see, e.g., Rockafellar [1970], Ben-Tal et al. [2015]), and the eighth...
equality follows by the inf to sup conversion, and the last equality is using the epigraph to represent
the support function \( \Gamma(\nu_s | Z_s(W)) := \sup_{e \in Z_s(W)} \{ \nu_s^e \} \), \( s \in [S] \). The assertion now follows if we apply the above result to Problem (17).

For the second statement, we have

\[
\inf_{P \in \mathcal{F}(W)} \mathbb{E}_P[\|\xi(W, y, \hat{e}) - \xi(W, y, \hat{e}(W, s))\|]
\]

\[
= \sup_{\delta \in \mathbb{R}^{L+1}, n \in [N]} \left\{ \inf_{\epsilon_s \in \mathbb{R}^N, s \in [S]} \frac{1}{S} \sum_{s \in [S]} \left( \|\epsilon_s - \hat{r}_s + W\hat{\delta}_s\| + \sum_{n \in [N]} \delta_n^T(\hat{\delta}_s[n] - d_n) \right) \right\}
\]

\[
= \sup_{\delta \in \mathbb{R}^{L+1}, n \in [N]} \left\{ \inf_{\epsilon_s \in \mathbb{R}^N, s \in [S]} \frac{1}{S} \sum_{s \in [S]} \left( \|\epsilon_s - \hat{r}_s + W\hat{\delta}_s\| + \sum_{n \in [N]} \delta_n^T(\hat{\delta}_s[n] - \sum_{n \in [N]} \delta_n^T\hat{\delta}_n(W, s)) \right) \right\}
\]

\[
= \sup_{\delta \in \mathbb{R}^{L+1}, n \in [N]} \left\{ \inf_{\epsilon_s \in \mathbb{R}^N, s \in [S]} \frac{1}{S} \sum_{s \in [S]} \left( \|\epsilon_s - \hat{r}_s + W\hat{\delta}_s\| + \sum_{n \in [N]} \delta_n^T(\hat{\delta}_s[n] - \hat{\delta}_s + \hat{W}\hat{\delta}_s) \right) \right\}
\]

\[
= \sup_{\delta \in \mathbb{R}^{L+1}, n \in [N]} \left\{ \inf_{\epsilon_s \in \mathbb{R}^N, s \in [S]} \frac{1}{S} \sum_{s \in [S]} \left( -\hat{\delta}_s^T[\delta_1, \ldots, \delta_N] (W - \hat{W})\hat{\delta}_s \right) \right\}
\]

\[
= \sup_{\delta \in \mathbb{R}^{L+1}, n \in [N]} \left\{ \frac{1}{S} \sum_{s \in [S]} \left( \hat{\delta}_s^T[\delta_1, \ldots, \delta_N] (W - \hat{W})\hat{\delta}_s \right) \right\}
\]

\[
\text{s.t. } \left\| -[\delta_1, \ldots, \delta_N]^T \hat{\delta}_s \right\|_s \leq 1 \quad \forall s \in [S]
\]

\[
= \sup_{D \in \mathbb{R}^{(L+1)xN}} \frac{1}{S} \sum_{s \in [S]} \left( \hat{\delta}_s^T D(W - \hat{W})\hat{\delta}_s \right)
\]

\[
\text{s.t. } \left\| D^T\hat{\delta}_s \right\|_s \leq 1 \quad \forall s \in [S],
\]

where the first equality is due to the strong duality guaranteed by Assumption\( \square \), the sixth equality stems from the inf to sup conversion, the seventh equality comes from the conjugate of the norm, and the last equality is due to the change-of-variable trick \( D := [-\delta_1, \ldots, -\delta_N] \). This statement now follows if we apply the above result to Problem (17). This completes the proof. \( \square \)

**Proof of Proposition 2**

By definition, for any \( P \in \tilde{\mathcal{G}}_{\text{lifted}} \), there exists some \( Q \in \tilde{\mathcal{G}} \) such that \( (\tilde{W}, \tilde{\epsilon}, \tilde{s}) \sim Q \) and \( (\tilde{W}, \tilde{\epsilon}, \tilde{s}, \tilde{u}_1, \tilde{u}_2) \sim P \). Hence, we have \( \Pi_{(\tilde{W}, \tilde{\epsilon}, \tilde{s})} P = Q \in \tilde{\mathcal{G}} \), which indicates that \( \Pi_{(\tilde{W}, \tilde{\epsilon}, \tilde{s})} \tilde{\mathcal{G}}_{\text{lifted}} \subseteq \tilde{\mathcal{G}} \).
Now we prove the reverse. For any \( P \in \tilde{G} \), we can construct a \( Q \in \mathcal{P}_0(\mathcal{W} \times \mathbb{R}^N \times [S] \times \mathbb{R} \times \mathbb{R}) \) such that

\[
\Pi_{(W,\tilde{\tau},\tilde{\varepsilon})} Q = P
\]

\[
Q \left[ \tilde{u}_1 = \| \xi(W, v, \tilde{\varepsilon}) - \hat{\xi}(W, v, \tilde{s}) \| \mid \tilde{s} = s \right] = 1 \quad \forall s \in [S]
\]

\[
Q \left[ \tilde{u}_2 = \| \tilde{\varepsilon} - \hat{\varepsilon}(W, \tilde{s}) \| \mid \tilde{s} = s \right] = 1 \quad \forall s \in [S].
\]

It follows that \( Q \in \tilde{G}_{\text{lifted}} \). Because \( \Pi_{(W,\tilde{\tau},\tilde{\varepsilon})} Q = P \), we have \( P \in \Pi_{(W,\tilde{\tau},\tilde{\varepsilon})} \tilde{G}_{\text{lifted}} \). Hence, we also have \( \tilde{G} \subseteq \Pi_{(W,\tilde{\tau},\tilde{\varepsilon})} \tilde{G}_{\text{lifted}} \). \( \square \)

**Proof of Theorem 6**

Following the definition of the hybrid uncertainty and ambiguity set \( \tilde{G} \), the lifted hybrid uncertainty and ambiguity set, \( \tilde{G}_{\text{lifted}} \), can also be defined as

\[
\tilde{G}_{\text{lifted}} = \left\{ P \in \mathcal{P}_0(\mathcal{W} \times \mathbb{R}^N \times [S] \times \mathbb{R} \times \mathbb{R}) \mid \exists W \in \mathcal{W}, Q \in \tilde{F}(W) : \begin{align*}
(\tilde{\varepsilon}, \tilde{u}_1, \tilde{u}_2, \tilde{s}) &\sim P \\
E_p [\tilde{u}_1] &\leq \Phi E_p [\tilde{u}_2] \\
E_p [\tilde{\varepsilon}] &\equiv 0 \\
P [\tilde{s} = s] &\equiv 1/S \\
P [(\tilde{\varepsilon}, W) \in Z_s \mid \tilde{s} = s] &\equiv 1 \quad \forall s \in [S] \\
P [\tilde{u}_1 \geq \| \xi(W, v, \tilde{\varepsilon}) - \hat{\xi}(W, v, \tilde{s}) \| \mid \tilde{s} = s] &\equiv 1 \quad \forall s \in [S] \\
P [\tilde{u}_2 \geq \| \tilde{\varepsilon} - \hat{\varepsilon}(W, \tilde{s}) \| \mid \tilde{s} = s] &\equiv 1 \quad \forall s \in [S] 
\end{align*} \right\},
\]

where the inner ambiguity set \( \tilde{F}(W) \) is defined as:

\[
\tilde{F}(W) := \left\{ P \in \mathcal{P}_0(\mathbb{R}^N \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times [S]) \mid \begin{align*}
(\tilde{\varepsilon}, \tilde{u}_1, \tilde{u}_2, \tilde{s}) &\sim P \\
E_p [\tilde{u}_1] &\leq \Phi E_p [\tilde{u}_2] \\
E_p [\tilde{\varepsilon}] &\equiv 0 \\
P [\tilde{s} = s] &\equiv 1/S \\
P [(\tilde{\varepsilon}, W) \in Z_s \mid \tilde{s} = s] &\equiv 1 \quad \forall s \in [S] \\
P [\tilde{u}_1 \geq \| \xi(W, v, \tilde{\varepsilon}) - \hat{\xi}(W, v, \tilde{s}) \| \mid \tilde{s} = s] &\equiv 1 \quad \forall s \in [S] \\
P [\tilde{u}_2 \geq \| \tilde{\varepsilon} - \hat{\varepsilon}(W, \tilde{s}) \| \mid \tilde{s} = s] &\equiv 1 \quad \forall s \in [S] 
\end{align*} \right\}.
\]

Suppose the recourse problem has complete recourse. For any \( \tau \leq \hat{Z} \), by Long et al. [2021], Problem (19) is feasible. Specifically, there exists some \( \bar{x} \in \mathcal{X}, \tilde{k} \geq 0 \), and a scenario-wise lifted affine recourse adaptation \( \bar{y}_s \in \mathcal{L}^{N+1,D_2} \), for \( s \in [S] \), such that

\[
E_p [c^T (\bar{z}) \bar{x} + d^T \bar{y}_s (\bar{z}, \tilde{u})] + \tilde{k} E_p [\tilde{u}] \geq \tau \quad \forall P \in \tilde{F}_{\text{lifted}}
\]

\[
A(z)\bar{x} + B\bar{y}_s (z, u) \leq b(z) \quad \forall (z, u) \in \hat{Z}_s, \forall s \in [S].
\]
For any $W \in \mathcal{W}$ and $P \in \mathcal{F}(W)$, we have that the projection of $P$ onto $(\xi(W, \bar{e}), \bar{u}_1, \bar{s})$ belongs to $\mathcal{F}_{\text{lifted}}$, i.e., $\Pi(\xi(W, \bar{e}), \bar{u}_1, \bar{s})P \in \mathcal{F}_{\text{lifted}}$. To see this, consider the following ambiguity set

$$
\mathcal{F}(W) = \left\{ P \in \mathcal{P}_0(\mathbb{R}^N \times \mathbb{R} \times \mathbb{R} \times [S]) \right\}
\begin{align*}
&\{ (\bar{e}, \bar{u}_1, \bar{u}_2, \bar{s}) \sim P \\ &\mathbb{E}_P[\bar{u}_1] \leq \Phi \mathbb{E}_P[\bar{u}_2] \\ &P[\bar{s} = s] = 1/\bar{s} \\ &P[(\bar{e}, W) \in \mathcal{Z}_s | \bar{s} = s] = 1 \\ &P[\bar{u}_1 \geq \|\xi(W, v, \bar{e}) - \xi(\bar{W}, v, \bar{s})\| | \bar{s} = s] = 1 \\ &P[\bar{u}_2 \geq \|\xi(W, v, \bar{e}) - \xi(W, v, \bar{e}(W, \bar{s}))\| | \bar{s} = s] = 1 \quad \forall s \in [S]
\end{align*}
$$

Clearly, $\mathcal{F}(W) \subseteq \tilde{\mathcal{F}}(W)$. Now, the marginal distribution of $\tilde{\mathcal{F}}(W)$ onto $(\xi(W, \bar{e}), \bar{u}_1, \bar{s})$ is exactly $\mathcal{F}_{\text{lifted}}$. Hence, we have $\Pi(\xi(W, \bar{e}), \bar{u}_1, \bar{s})P \in \mathcal{F}_{\text{lifted}}$ for any $P \in \mathcal{G}_{\text{lifted}}$. Therefore, we have

$$
\begin{align*}
\inf_{P \in \mathcal{G}_{\text{lifted}}} \mathbb{E}_P \left[ c^T(\xi(W, v, \bar{e}))x + d^T \bar{y}_s(\xi(W, v, \bar{e}), \bar{u}_1) + \Phi \tilde{k} \bar{u}_2 \right] \\
\geq \inf_{P \in \mathcal{G}_{\text{lifted}}} \mathbb{E}_P \left[ c^T(\xi(W, v, \bar{e}))x + d^T \bar{y}_s(\xi(W, v, \bar{e}), \bar{u}_1) + \tilde{k} \bar{u}_1 \right] \\
\geq \inf_{P \in \mathcal{F}_{\text{lifted}}} \mathbb{E}_P \left[ c^T(\tilde{z})x + d^T \bar{y}_s(\tilde{z}, \bar{u}_1) + \tilde{k} \bar{u}_1 \right] \\
\geq \tau,
\end{align*}
$$

which gives us

$$
\mathbb{E}_P \left[ c^T(\xi(W, v, \bar{e}))x + d^T \bar{y}_s(\xi(W, v, \bar{e}), \bar{u}_1) + \tilde{k} \mathbb{E}_P[\bar{u}_2] \right] \geq \tau \quad \forall P \in \mathcal{G}_{\text{lifted}}.
$$

Hence, $x$, $\Phi \tilde{k}$, and the scenario-wise lifted affine recourse adaptation $\bar{y}_s$ for $s \in [S]$, constitute a set of feasible solutions to Problem (21).