The Dao of Robustness

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

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

We propose a framework for optimization under uncertainty called robustness optimization, which is similar in purpose to, but philosophically different from, robust optimization. Unlike robust optimization approaches, we do not restrict nature to an uncertainty set but allow her to take its cause and even render solutions infeasible. Among these solutions, we favor those with the least adversarial impact on the model under uncertainty. Moreover, the decision maker does not have to size the uncertainty set, but instead specifies an acceptable target, or loss of optimality compared to the baseline model, as a tradeoff for the model’s ability to withstand greater uncertainty. We axiomatize the decision criterion associated with the robustness optimization, termed as the adversarial impact measure, which relates to the maximum level of model infeasibility that may occur relative to the magnitude of deviation from the baseline uncertainty. We also provide a representation theorem of the decision criterion and uncover different types of adversarial impact measures. Similar to robust optimization, we show that robustness optimization via minimizing the adversarial impact can also be done in a tractable way, i.e., it preserves the complexity of the underlying problems including, inter alia, linear, discrete, data-driven and dynamic optimization problems. We also provide computational studies to show that for the same price of robustness, the solutions to our robustness optimization models can withstand greater impact of uncertainty compared to classical robust optimization models, and doing so without incurring additional computational effort.

Key words: Robust optimization, robustness optimization, data-driven optimization, stochastic optimization, dynamic optimization


1. Introduction

“The Dao follows the way of nature (道法自然)” – Lao Tzu

Optimization under uncertainty, despite its importance and ubiquity in real-world problems, has been a perennial difficulty in prescriptive analytics. The overarching challenge in optimization under uncertainty is formulating an efficiently, or at least practically, solvable model that would sufficiently mitigate the adverse impact of uncertainty. A tractable optimization problem under
uncertainty is often associated with a model of uncertainty, which restricts the nature of uncertainty, and if nature were to confine as such, the solution to the model would also be optimal with respect to the model of uncertainty.

In a stochastic linear optimization problem, we often assume a model of uncertainty with discrete distribution to avoid high dimensional integration and would enable us to solve the problem using large-scale linear optimization techniques (see, e.g., Kall and Wallace 1994, Shapiro et al. 2009, Birge and Louveaux 2011). In a dynamic optimization problem, we would often assume an underlying Markov process to reduce the state-space of the problem in an attempt to circumvent the “curse of dimensionality”. In a robust optimization problem, the uncertainty model assumes an adversarial nature that selects the parameters from an uncertainty set that would have the worst impact on the model (see, e.g., Soyster 1973, Ben-Tal and Nemirovski 1998, El Ghaoui et al. 1998). The budgeted uncertainty set proposed by Bertsimas and Sim (2004) retains the linear optimization framework, which can be extended to solving discrete optimization problems. In particular, Bertsimas and Sim (2003) show that the same model of uncertainty also retains the computational complexity of combinatorial optimization problems. In distributionally robust optimization, which generalizes robust optimization, the adversarial nature selects the worst probability distributions over an ambiguity set of probability distributions (see, e.g., Delage and Ye 2010, Wiesemann et al. 2014, Bertsimas et al. 2019, Chen et al. 2019). Similar to robust optimization, the computational tractability of the distributionally robust optimization problem depends on the characterization of the ambiguity set.

The fact remains that we can only be certain about the uncertainty of nature, and this bags the question - what if we were wrong about the model of uncertainty? Indeed, it has well been known that solving an optimization model that assumes a deterministic state of nature could result in unacceptably fragile solutions (see, e.g., Ben-Tal et al. 2009), yet few decision makers would be willing to pay the price of a fully robust solution that could withstand the worst case impact. For better performance in objective, most would be willing to tolerate some levels of model infeasibility when mitigating uncertain outcomes.

Robust optimization approach addresses this problem by fashioning an uncertainty set that allows the parameters to vary within a prescribed neighbourhood from their deterministic nominal values. It is also well established in robust optimization that under some probabilistic assumptions, modest sized uncertainty sets could provide very high probabilistic protection of a linear constraint with affine perturbation against infeasibility (see, e.g., Ben-Tal and Nemirovski 1998, Bertsimas and Sim 2004). However, if we were wrong about the probabilistic assumptions, a robust optimization model with a modest sized uncertainty set may not sufficiently protect the model against uncertain outcomes.
The same issues arise in stochastic and dynamic optimization where we assume an uncertainty model comprising random variables with known probability distributions. When using sample average approximation (SAA) as a tractable approximation to stochastic optimization problems, the out-of-sample performance can be rather inferior compared to the results reflected in the SAA model (Kleywegt et al. 2002). In the data-driven setting, due to the *optimizer’s curse* (Smith and Winkler 2006), if we solve a stochastic optimization problem using an empirical distribution from a dataset and test the out-of-sample outcome on another, we should also expect inferior results. Distributionally robust optimization models with probability-distance-based ambiguity sets (see, e.g., Ben-Tal et al. 2013, Mohajerin Esfahani and Kuhn 2018) can provide “regularized” solutions, which can effectively mitigate the optimizer’s curse. However, it may be difficult to determine the right probability distance parameter for the ambiguity set. Such parameter is usually chosen via cross-validation techniques (Mohajerin Esfahani and Kuhn 2018), provided there are sufficient data to do so effectively.

In this paper, we propose a framework for optimization under uncertainty known as *robustness optimization*. Unlike robust optimization approaches, we allow nature to take its cause, even renders solutions infeasible, and favor those with the least adversarial impact on the model under uncertainty. In our proposed robustness optimization model, the decision maker specifies a target, or an acceptable loss of optimality compared to the baseline model, as a tradeoff for the model’s ability to withstand greater uncertainty. We believe that articulating preference for robustness from a target-driven perspective is also more interpretable for the decision maker. More interestingly, we also axiomatize the decision criterion associated with the robustness optimization, termed as the *adversarial impact measure*, which relates to the maximum level of model infeasibility that may occur relative to the magnitude of deviation from the baseline uncertainty. We also provide a representation theorem of the decision criterion, and with this representation, we uncover different types of adversarial impact measures. We remark that this measure belongs to the class of *satisficing* (Simon 1959) measure proposed in Brown and Sim (2009) and Brown et al. (2012). Using the adversarial impact measure as a target-based decision criterion, we show how it be applied in choices under uncertainty.

Our work is also related to the stream of research along the lines of *comprehensive robust optimization* of Ben-Tal et al. (2006) or more recently termed as *globalized robust optimization* proposed by Ben-Tal et al. (2017), which permits the modeler to specify how the magnitude of constraint violation should be bounded when the realization of the uncertainty occurs outside the uncertainty set. This has been extended to *probabilistic enveloping constraints* (Xu et al. 2012) and *soft robust model* (Ben-Tal et al. 2010), which has shown to be related to convex measure of risks (Foellmer and Schied 2004). Though our work is related to this stream, we do not require the modeler to
determine how to control the magnitude of constraint violation whenever it arises, but instead, the magnitude of violation itself is subject to optimization. Ben-Tal et al. (2017) introduce a GRC-sum model (Section 5.2) to also minimize the magnitude of violation, which is similar to one of our nonstochastic robustness optimization models, though we do not require the modeler to specify an additional uncertainty set for which feasibility must hold for the model. Instead, we emphasize the role of targets in our robustness optimization problem and extend this model further to include distributional ambiguity. Justified by the properties in our representation theorem, our work serves as a normative framework and uncovers various adversarial impact measures to cater different needs, e.g., the decision maker can construct a robustness optimization model using probability distance measures, utility functions, or convex risk measures, under general distributional ambiguity. Moreover, we show that robustness optimization via minimizing the adversarial impact can also be done in a tractable way, i.e., preserving the complexity of the underlying problems including, among others, linear, discrete, data-driven and dynamic optimization problems. We also provide computational studies to show that for the same objective values attained, the solutions to our robustness optimization models can withstand greater impact of uncertainty compared to classical robust optimization models without incurring additional computational effort.

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

### 2. Framework for Robustness Optimization

To motivate our framework for what we call *robustness optimization*, we first consider the following deterministic *baseline optimization problem*,

$$
Z_0 = \max \mathbf{c}^T \mathbf{x} \\
\text{s.t. } f(\mathbf{x}, \tilde{\mathbf{z}}) \leq 0 \\
\mathbf{x} \in \mathcal{X},
$$

for a given function $f : \mathcal{X} \times \mathcal{Z} \mapsto \mathbb{R}$, where the input to the second argument is subject to uncertainty. The uncertain parameters of the problem are collectively denoted by the random variable $\tilde{\mathbf{z}}$ over the support $\mathcal{Z} \subseteq \mathbb{R}^N$, which for all practical purposes, is assumed bounded. For the deterministic problem, only a chosen baseline value $\tilde{\mathbf{z}} \in \mathcal{Z}$ is considered in the constraint.
To better protect the constraint against infeasibility, robust optimization solves the following problem,

$$Z_r = \max c^\top x$$

$$\text{s.t. } f(x, z) \leq 0, \forall z \in \mathcal{U}(r)$$

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

where $\mathcal{U}(r)$ is typically a norm-based adjustable uncertainty set,

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

The absolutely robust model is given by

$$Z_\infty = \max c^\top x$$

$$\text{s.t. } f(x, z) \leq 0 \forall z \in \mathcal{Z}$$

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

though this approach may be perceived as overly conservative, especially whenever $Z_\infty$ is significantly lower than $Z_0$.

The most commonly adopted budgeted uncertainty set of Bertsimas and Sim (2004) is a norm-based adjustable uncertainty set with $\mathcal{Z} = [-1, 1]^N$ and $\ell_1$-norm for which the baseline value is $\hat{z} = 0$. The parameter $r$ is intuitively interpreted as the maximum number of uncertain parameters that nature chooses to deviate from their baseline values.

The robust optimization problem necessitate the solution $x$ to remain feasible for any realization of $z$ in the uncertainty set. It is important to note that robust optimization model does not specify how the constraint should behave whenever the uncertain parameters deviate from the uncertainty set. To understand the potential issue, consider the following robust continuous knapsack problem under the budgeted uncertainty set,

$$\max \sum_{j \in [r]} x_j$$

$$\text{s.t. } z^\top x \leq r, \forall z \in \mathcal{U}(r)$$

$$x \in [0, 1]^N.$$

Observe that for any parameter $r \in [N]$, there would be multiple optimal solutions to Problem (3), with $x_j = 1$ for all $j \in [r]$, though the most preferred optimal solution should also require $x_j = 0$ for all $j \in [N]\{r\}$. The issue of having multiple optimal solutions in such robust optimization problems has also been observed in Iancu and Trichakis (2014). We could resolve this issue if the uncertain constraint is evaluated for all possible realizations of $z \in \mathcal{Z}$, taking into account of how different realizations would directly impact on the feasibility of the constraint.
We now introduce our robustness optimization model:

\[
\begin{align*}
\min & \quad k \\
\text{s.t.} & \quad f(\mathbf{x}, \mathbf{z}) \leq k \| \mathbf{z} - \hat{\mathbf{z}} \|, \forall \mathbf{z} \in \mathcal{Z} \\
& \quad \mathbf{c}^\top \mathbf{x} \geq \tau \\
& \quad \mathbf{x} \in \mathcal{X}, k \geq 0.
\end{align*}
\]

(4)

There are philosophical differences between robust optimization and robustness optimization. Unlike robust optimization, the robustness optimization model does not only decouple the model from a restricted nature, but allows nature to vary over the entire support, while controlling the level of infeasibly as much as possible whenever nature deviates from the baseline value. Observe that in robust optimization, the norm relating to the deviation of uncertain parameters from their baseline values is articulated at the uncertainty set and not within the model. In contrast, the same norm in robustness optimization is directly associated with the feasibility of the model and relates to how the magnitude of deviation may affect the degree of infeasibility of the problem. However, the maximum allowable constraint violation is proportional to how much \( \mathbf{z} \) would deviate from the baseline value \( \hat{\mathbf{z}} \). The proportionality factor \( k \) effectively describes the adversarial impact on the model. As it decreases in value, the less magnitude of infeasibility the model could occur, relative to how far the uncertain outcome deviates from the baseline value. Absolute robustness is ensured when the adversarial impact is zero. Robustness of a model is measured by how much the adversarial impact can be reduced for the model.

Observe that the constraint of Problem (1) is implied at \( \mathbf{z} = \hat{\mathbf{z}} \), hence, any solution that is infeasible in the deterministic baseline problem would also be so in the robustness optimization problem. Moreover, any solution that is robustly feasible within the neighborhood of the baseline value would be also feasible in the robustness optimization problem.

**Proposition 1.** For a given solution \( \mathbf{x} \in \mathcal{X} \), satisfying, \( \mathbf{c}^\top \mathbf{x} \geq \tau, \sup_{\mathbf{z} \in \mathcal{Z}} f(\mathbf{x}, \mathbf{z}) < \infty \) and

\[ f(\mathbf{x}, \mathbf{z}) \leq 0, \quad \forall \mathbf{z} \in \mathcal{U}, \]

where \( \mathcal{U} \) contains \( \hat{\mathbf{z}} \) in its interior, then \( \mathbf{x} \) would be also feasible in the robustness optimization problem (4).

Because the robustness optimization model does not restrict \( \mathbf{z} \) to an uncertainty set, it can also eliminate the degeneracy issue we face in the robust knapsack problem (3). In particular, with \( \tau = r \), the absolutely robust solution, for which \( k = 0 \) is achieved uniquely when \( x_i = 1 \) for \( i \in [r] \) and zero otherwise.
Another important difference from the robust optimization model is that the decision maker specifies the parameter \( \tau \), as opposed to the size of the uncertainty set \( r \) for the robust optimization model. Without other assumptions on the underlying uncertainty, it would not be intuitive for the decision maker to pin down the value of \( r \) for the robust optimization problem. A common approach is to relate \( r \) to some probabilistic guarantees of feasibility, based on additional distributional assumptions of the underlying uncertainty. For instance, Bertsimas and Sim (2004) provides a probabilistic guarantee for choosing \( r \) under the assumption that the components of \( \tilde{z} \) are independently distributed with mean exactly at \( 0 \), and this applies only to a linear constraint under affine perturbation. Even so, the bound can be rather weak. In contrast, for the robustness optimization problem, the input parameter \( \tau \), \( \tau \in [Z_\infty, Z_0] \) is directly related to the objective value of the model the decision maker is addressing, and can be interpreted as the target objective she is willing to accept, relative to a reference – the baseline optimization objective, \( Z_0 \). The model will then determine the most robust solution within the achievable target. Therefore, the robustness optimization benchmarks the baseline problem as a reference. While one may argue that target, \( \tau \) is hard to specify, it should at least be more tangible and intuitive to specify compared to the parameter \( r \) used in the robust optimization problem (2). Moreover, any approach such as cross validation techniques to determine the parameter \( r \) can also be applied to determine the parameter \( \tau \).

Despite the differences between the robust and robustness optimization problems, there is one important similarity - both approaches have the same computational complexity. In particular, the robust counterpart in the robust optimization problem is

\[
f(x, z) \leq 0, \quad \forall z \in Z : \|z - \tilde{z}\| \leq r,
\]

while the robust counterpart for the robustness optimization problem can be expressed as

\[
f(x, z) - k\xi \leq 0, \quad \forall z \in Z, \xi \in \mathbb{R} : \|z - \tilde{z}\| \leq \xi,
\]

which can be optimized efficiently depending on the function \( f(x, z) \) and the support set \( Z \) (see, e.g., Ben-Tal and Nemirovski 1998, El Ghaoui et al. 1998, Ben-Tal et al. 2015).

Note also that while we focus on uncertainty at the constraint, we can also consider cases where the uncertain parameters appear at the objective function as follows:

<table>
<thead>
<tr>
<th>Baseline</th>
<th>Robust</th>
<th>Robustness (( \tau \in [Z_0, Z_\infty] ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Z_0 = \min f(x, \tilde{z}) )</td>
<td>( Z_r = \min \max_{z \in U(t)} f(x, z) )</td>
<td>( \min k )</td>
</tr>
<tr>
<td>s.t. ( x \in \mathcal{X} )</td>
<td>s.t. ( x \in \mathcal{X} )</td>
<td>s.t. ( f(x, z) - \tau \leq k|z - \tilde{z}|, \forall z \in Z )</td>
</tr>
<tr>
<td>( x \in \mathcal{X}, k \geq 0 )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
We emphasize that robustness optimization should not be misconstrued as a relaxation of robust optimization. For instance, the robust constraint of Problem (2) indicates

\[
\begin{align*}
  f(x, z) &\leq 0, \quad \forall z \in \mathcal{U}(r) \\
  f(x, z) &\leq +\infty, \quad \forall z \in \mathcal{Z}\backslash\mathcal{U}(r),
\end{align*}
\]

while the robustness constraint in Problem (4) implies that

\[
\begin{align*}
  f(x, z) &\leq k\|z - \hat{z}\|, \quad \forall z \in \mathcal{U}(r) \\
  f(x, \hat{z}) &\leq k\|z - \hat{z}\|, \quad \forall z \in \mathcal{Z}\backslash\mathcal{U}(r).
\end{align*}
\]

Hence, while the robustness optimization model appears to be a relaxation of robust optimization for \(z \in \mathcal{U}(r)\), it is more constrained than the robust optimization model when the realization is outside the uncertainty set. Note that \(\mathcal{U}(r)\) is typically a small set compared to the support \(\mathcal{Z}\), and hence we are actually tightening the robustness constraint for a much larger uncertainty set \(\mathcal{Z}\backslash\mathcal{U}(r)\), which the robust optimization model simply ignores.

We remark that Model (4) is a special case of the GRC-sum model proposed in Section 5.2 of Ben-Tal et al. (2017), which for a single constraint has the following representation,

\[
\begin{align*}
  \min k \\
  \text{s.t. } f(x, z) &\leq k\|z - u\|, \quad \forall z \in \mathcal{Z}, u \in \mathcal{U} \\
  x &\in \mathcal{X}, k \geq 0,
\end{align*}
\]

for a given uncertainty set \(\mathcal{U} \subseteq \mathcal{Z}\). Observe that in contrast to Model (4), the solutions that are feasible in the GRC-sum model would also be robustly feasible as follows

\[
\begin{align*}
  f(x, z) &\leq 0, \quad \forall z \in \mathcal{U}.
\end{align*}
\]

While the GRC-sum model is more general in representation, it requires the modeler to specify an uncertainty set that would necessarily lead to a larger optimization problem compared to Problem (4). Our robustness optimization model has the same computational complexity as the robust optimization and we emphasize the role of targets, instead of sizing uncertainty sets, in the decision making procedure. Moreover, our numerical studies suggest that even without enforcing an additional uncertainty set, for the same computational effort, the solutions to our robustness optimization models already dominate those obtained by classical robust optimization models in their ability to withstand greater impact of uncertainty. In subsequent sections, we also generalize this model further to include distributional ambiguity.
Stochastic baseline optimization problem
We can extend the framework to consider stochastic uncertainty in which the baseline problem is associated with a random variable \( \tilde{z} \) with a distribution \( \hat{P} \in \mathcal{P}_0(\mathcal{Z}) \):

\[
Z_0 = \max c^\top x \\
\text{s.t. } \mathbb{E}_{\hat{P}}[f(x, \tilde{z})] \leq 0 \\
x \in \mathcal{X}.
\]

The distributionally robust formulation to the above baseline problem is

\[
Z_r = \max c^\top x \\
\text{s.t. } \mathbb{E}_{\hat{P}}[f(x, \tilde{z})] \leq 0, \forall \hat{P} \in \mathcal{B}(r) \\
x \in \mathcal{X},
\]

where the probability-distance-based ambiguity set is given by

\[
\mathcal{B}(r) = \left\{ \hat{P} \in \mathcal{P}_0(\mathcal{Z}) \mid \tilde{z} \sim \hat{P}, \Delta(\hat{P}, \hat{P}) \leq r \right\},
\]

where \( \Delta(\hat{P}, \hat{P}) \) represents the probability distance of distributions \( P \) from \( \hat{P} \) defined as follows:

**Definition 1.** A probability distance function, \( \Delta(\hat{P}, \hat{P}) \) is a nonnegative function on the domain of probability distributions such that \( \Delta(\hat{P}, \hat{P}) = 0 \) if \( P = \hat{P} \).

Under this definition, the probability-distance-based ambiguity sets would include \( \phi \)-divergence (Pardo 2006, Ben-Tal et al. 2013) and Wasserstein metric, also known as Kantorovich–Rubinstein metric (Mohajerin Esfahani and Kuhn 2018), which we will discuss in more details in the next section.

Similar to Problem (4), the distributional robustness counterpart of the baseline problem is given by,

\[
\begin{align*}
\min k \\
\text{s.t. } \mathbb{E}_{\hat{P}}[f(x, \tilde{z})] \leq k\Delta(\hat{P}, \hat{P}), \forall \hat{P} \in \mathcal{P}_0(\mathcal{Z}) \\
c^\top x \geq \tau \\
x \in \mathcal{X},
\end{align*}
\]

for some chosen target \( \tau \in [Z_{\infty}, Z_0] \). Note that a deterministic constraint can be expressed as a stochastic expectation constraint if the distribution satisfies \( \hat{P}[\tilde{z} = \hat{z}] = 1 \). Likewise, the distributional robustness optimization model (7) also generalizes the robustness optimization problem (4) for which the probability distance metric is given by

\[
\Delta(\hat{P}, \hat{P}) = \mathbb{E}_{\hat{P}}[\|\tilde{z} - \hat{z}\|].
\]

This is because for the function \( g(x, k, z) := f(x, z) - k\|z - \hat{z}\| \), the distributionally robust counterpart

\[
\mathbb{E}_{\hat{P}}[g(x, k, \tilde{z})] \leq 0, \forall \hat{P} \in \mathcal{P}_0(\mathcal{Z})
\]
is the same as the robust counterpart,
\[ g(x, k, z) \leq 0, \quad \forall z \in Z, \]
since
\[ \sup_{z \in Z} g(x, k, z) \leq \sup_{p \in \mathcal{P}_0(Z)} \mathbb{E}_p [g(x, k, \hat{z})] \leq \sup_{p \in \mathcal{P}_0(Z)} \text{ess sup}_p [g(x, k, \hat{z})] \leq \sup_{z \in Z} g(x, k, z). \]
Interestingly, with the same probability distance metric, the distributionally robust optimization (6) is not necessarily equivalent to the corresponding robust optimization problem (2), unless the function \( f(x, z) \) is concave in \( z \), for which the result can easily be shown using Jensen’s inequality.

The robustness optimization model can also be extended to address distributionally robust baseline optimization problem. However, due to the additional complexity of the model, we relegate the discussion to Appendix B.

**Multiple constraints**
We can extend to multiple constraints, where we first consider the following baseline problem with \( I \) constraints subject to uncertainty.
\[
Z_0 = \max c^\top x \\
\text{s.t. } \mathbb{E}_p [f_i(x, \hat{z})] \leq 0, \; i \in [I] \\
x \in \mathcal{X}.
\]
The corresponding robustness optimization problem minimizes a weighted sum of individual adversarial impacts as follows
\[
\min w^\top \kappa \\
\text{s.t. } \mathbb{E}_p [f_i(x, \hat{z})] \leq \kappa_i \Delta(p, \hat{P}), \; \forall p \in \mathcal{P}_0(Z), \; i \in [I] \\
c^\top x \geq \tau \\
x \in \mathcal{X}, \kappa \geq 0,
\]
where \( w \in \mathbb{R}_+^I \) is a vector of weights specified by the modeler. Similar to goal programming (see Charnes et al. 1955), the weights can be associated to the unit violation costs of the corresponding constraints. The GRC-sum model proposed in Section 5.2 of Ben-Tal et al. (2017) is based on equal weights, which may pose an issue if the constraints are associated with different units of accounting.
In the more general case, while maintaining tractability, the collective adversarial impact can also be evaluated based on the maximum impact over all weights from a given an uncertainty set \( \mathcal{W} \subseteq \mathbb{R}_+^I \times \mathbb{R} \) as follows,
\[
\min \max_{(w, w_0) \in \mathcal{W}} \left\{ w^\top \kappa + w_0 \right\} \\
\text{s.t. } \mathbb{E}_p [f_i(x, \hat{z})] \leq \kappa_i \Delta(p, \hat{P}), \; \forall p \in \mathcal{P}_0(Z), \; i \in [I] \\
c^\top x \geq \tau \\
x \in \mathcal{X}, \kappa \geq 0.
\]
3. A Measure of Adversarial Impact

We now elicit the decision criterion associated with the robustness optimization problem, which we call the adversarial impact measure. Let \((\Omega, \mathcal{F})\) be a measurable space with \(\sigma\)-algebra \(\mathcal{F}\). The baseline probability distribution is given by \(\hat{P}, \hat{P} \in \mathcal{P}_0\). Denote \(\mathcal{L} = L^\infty(\Omega, \mathcal{F}, \hat{P})\) as the space of bounded measurable real valued functions. Let \(\hat{v} \in \mathcal{L}\) denote the random variable representing uncertain outcomes of a model’s constraint, with positive values being infeasible.

**Definition 2 (Adversarial Impact Measure).** The functional \(\rho : \mathcal{L} \mapsto [0, +\infty]\) is an adversarial impact measure if and only if it has the following representation

\[
\rho(\hat{v}) = \min k \\
\text{s.t. } E_{\hat{P}}[\hat{v}] \leq k\Delta(\hat{P}, \hat{P}), \ \forall \hat{P} \in \mathcal{P}_0 \\
k \geq 0,
\]

for some probability distance function \(\Delta\).

With this definition, Problem (7) is equivalent to

\[
\min \rho(f(x, \hat{z})) \\
\text{s.t. } c^T x \geq \tau \\
x \in \mathcal{X}.
\]

Minimizing the adversarial impact measure over all solutions is effectively choosing the most robust solution for which uncertainty has the least negative impact on the model. We next show that the adversarial impact measure is associated with salient properties consistent with coherent decision making in achieving robustness of the solution, while ensuring tractability of the problem when the criterion is to be minimized.

**Theorem 1 (Properties of Adversarial Impact Measure).** The functional \(\rho : \mathcal{L} \mapsto [0, +\infty]\) is an adversarial impact measure if and only if it is lower semi-continuous and satisfies the following properties:

1. **Monotonicity:** If \(\hat{v}_1 \geq \hat{v}_2\), then \(\rho(\hat{v}_1) \geq \rho(\hat{v}_2)\).
2. **Positive Homogeneity:** For any \(\lambda \geq 0\), we have \(\rho(\lambda \hat{v}) = \lambda \rho(\hat{v})\).
3. **Subadditivity:** \(\rho(\hat{v}_1 + \hat{v}_2) \leq \rho(\hat{v}_1) + \rho(\hat{v}_2)\).
4. **Absolute Robustness:** If \(\hat{v} \leq 0\), then \(\rho(\hat{v}) = 0\).
5. **Baseline Infeasibility:** If \(E_{\hat{P}}[\hat{v}] > 0\), then \(\rho(\hat{v}) = \infty\).

Moreover, the probability distance \(\Delta\) associated with \(\rho\) is given by

\[
\Delta(\hat{P}, \hat{P}) = \sup_{\hat{v} \in \mathcal{L}} \{E_{\hat{P}}[\hat{v}] \mid \rho(\hat{v}) \leq 1\}.
\]
The first three properties of the adversarial impact measures coincide with three out of the four axioms of coherent monetary risk measures proposed by Artzner et al. (1999). The monotonicity property requires that the greater the severity of infeasibility, the larger the value of adversarial impact. Positive homogeneity property dictates that the adversarial impact scales accordingly with the underlying uncertainty. Consequently, the adversarial impact would have the same “units” as the constraint. Likewise, the property of subadditivity is synonymous with the preference for risk pooling, which is associated with uncertainty aversion. It implies that the collective adversarial impact of the combined uncertainty in meeting one aggregate constraint should be smaller than the sum of the adversarial impacts if the feasibility of the uncertain constraints are considered separately. Incidentally, positive homogeneity and subadditivity imply convexity, which is also an important precursor to obtaining a tractable optimization model when the adversarial impact measure is to be minimized.

The last two properties elucidate the extreme values associated with the adversarial impact measure. The absolute robustness property asserts that if the model’s constraint is absolutely feasible, then the corresponding adversarial impact measure should be the lowest value at zero. The baseline infeasibility property ensures that any solution with finite adversarial impact would also be feasible in the baseline constraint, i.e., \( E[v] \leq 0 \). These two properties would rule out any monetary risk measure as a candidate for adversarial impact measure.

Theorem 1 shows that any metric that satisfies the salient properties is also an adversarial impact measure, even though it may not have the same explicit probability-distance-based representation. Incidentally, the adversarial impact measure belongs to a class of satisficing measures proposed by Brown and Sim (2009). Similar to the satisficing measures, the adversarial impact measure has also a risk-based representation via normalized convex risk measure.

**Definition 3.** A normalized convex risk measure is a lower semi-continuous functional \( \mu : \mathcal{L} \to \mathbb{R} \), that satisfies the following properties:

1. **Monotonicity**: If \( \tilde{v}_1 \geq \tilde{v}_2 \), then \( \mu(\tilde{v}_1) \geq \mu(\tilde{v}_2) \).
2. **Translation invariance**: For any \( a \in \mathbb{R} \), \( \mu(\tilde{v} + a) = \mu(\tilde{v}) + a \).
3. **Convexity**: For any \( \lambda \in [0, 1] \), \( \mu(\lambda \tilde{v}_1 + (1 - \lambda)\tilde{v}_2) \leq \lambda \mu(\tilde{v}_1) + (1 - \lambda)\mu(\tilde{v}_2) \).
4. **Normalization**: \( \mu(0) = 0 \).

**Proposition 2 (Risk-based Representation).** The functional \( \rho : \mathcal{L} \to [0, +\infty) \) is an adversarial impact measure if and only if there exists some normalized convex risk measure \( \mu : \mathcal{L} \to \mathbb{R} \), satisfying

\[
\mu(\tilde{v}) \geq E_{\tilde{v}}[\tilde{v}] \quad \forall \tilde{v} \in \mathcal{L},
\]

such that

\[
\rho(\tilde{v}) = \inf \{ k > 0 \mid k \mu(\tilde{v}/k) \leq 0 \}.
\]
Note that the property associated with the inequality (10) is implied for convex risk measures that are *law-invariant under* \( \hat{P} \) (see Foellmer and Schied 2004). Proposition 2 implies that apart from its probability-distance-based representation, we can also construct an adversarial impact measure using convex risk measure, which is well studied in the literature (see, e.g., Foellmer and Schied 2002, 2004). Based on these representations, we next explore different types of adversarial impact measures that lead to tractable optimization models.

### Phi-divergence and optimized certainty equivalent

A popular class of probability distance metrics is called \( \phi \)-divergence (see, e.g., Liese and Vajda 2006, Reid and Williamson 2011). Let \( P, \hat{P} \in \mathcal{P}_0 \) be two probability measures. We use \( P \ll \hat{P} \) to denote that \( P \) is absolutely continuous with respect to \( \hat{P} \). The \( \phi \)-divergence of \( P \) from \( \hat{P} \) is defined as

\[
D_{\phi}(P||\hat{P}) := \begin{cases} 
\mathbb{E}_\hat{P} \left[ \phi \left( \frac{dP}{d\hat{P}} \right) \right] & \text{if } P \ll \hat{P} \\
\infty & \text{otherwise,}
\end{cases}
\]

for some convex function \( \phi \) with \( \phi(1) = 0 \). To obtain an explicit formulation of the corresponding adversarial impact measure, we use the result of Ben-Tal and Teboulle (2007) that relates \( \phi \)-divergence to a class of convex risk measures termed as optimized certainty equivalent.

\[
\mu_{OCE}(\hat{v}) = \inf \{ \eta + \mathbb{E}_P [\phi^*(\hat{v} - \eta)] \} = \sup_{\mathcal{P}_0} \left\{ \mathbb{E}_P [\hat{v}] - D_{\phi}(P||\hat{P}) \right\},
\]

where \( \phi^* \) is the convex conjugate of \( \phi \).

**Proposition 3 (\( \phi \)-divergence-based adversarial impact measure).** The \( \phi \)-divergence-based adversarial impact measure,

\[
\rho_{\phi}(\hat{v}) := \inf \left\{ k \geq 0 \mid \mathbb{E}_P [\hat{v}] \leq kD_{\phi}(P||\hat{P}), \ \forall P \in \mathcal{P}_0 \right\},
\]

is equivalent to

\[
\rho_{\phi}(\hat{v}) := \inf \{ k > 0 \mid \exists \eta \in \mathbb{R} : \eta + k\mathbb{E}_P [\phi^*((\hat{v} - \eta)/k)] \leq 0 \}.
\]

A commonly used \( \phi \)-divergence is the *total variation distance* with \( \phi(t) = |t - 1| \). The corresponding adversarial impact measure is

\[
\rho_{TV}(\hat{v}) := \inf \{ k > 0 \mid \exists \eta \in \mathbb{R} : \eta + k\mathbb{E}_P [\phi_{TV}^*((\hat{v} - \eta)/k)] \leq 0 \},
\]

where

\[
\phi_{TV}^*(s) = \begin{cases} 
-1 & \text{if } s < -1 \\
\eta & \text{if } s \in [-1, 1] \\
\infty & \text{if } s > 1.
\end{cases}
\]

We refer interested reader to Ben-Tal et al. (2013) for more choices of function \( \phi \) and their convex conjugate.
Kullback-Leibler divergence and riskiness index

Kullback-Leibler divergence is a particular $\phi$-divergence with $\phi(t) = t \log(t) - t + 1$ and can be explicitly written as:

$$D_{KL}(P \| \hat{P}) := \begin{cases} \mathbb{E}_P \left[ \log \left( \frac{dP}{d\hat{P}} \right) \right] & \text{if } P \ll \hat{P} \\ \infty & \text{otherwise.} \end{cases}$$

Observe the dual presentation of entropic convex risk measure and Kullback-Leibler divergence (Foellmer and Schied 2002), i.e.,

$$\sup_{P \in \mathcal{P}_0} \left\{ \mathbb{E}_P [\tilde{v}] - D_{KL}(P \| \hat{P}) \right\} = \log \mathbb{E}_{\hat{P}} [\exp (\tilde{v})],$$

which we can easily derive from Equation (11) with the corresponding convex conjugate being $\phi^*(r) = \exp(r) - 1$. Hence, we can express the corresponding adversarial impact measure as

$$\rho_{KL}(\tilde{v}) := \inf \left\{ k > 0 \mid k \log \mathbb{E}_{\hat{P}} [\exp (\tilde{v} / k)] \leq 0 \right\},$$

which is the same as the riskiness index of Aumann and Serrano (2008). The riskiness index has recently been used in lieu of maximizing success probability, and it has the benefits of having greater computational tractability and generating superior solutions under stochastic uncertainty with known distribution (see, e.g., Chen et al. 2015, Hall et al. 2015, Jaillet et al. 2016, Chen and Tang 2019). The connection between the riskiness index and adversarial impact measure via the Kullback-Leibler divergence also implies that the solutions that minimize the index are also robust to distribution ambiguity.

Convex risk measure and penalty function

A convex risk measure, which has been well studied in the literature, has a dual representation (see, e.g., Heath and Ku 2004, Foellmer and Schied 2004),

$$\mu(\tilde{v}) = \sup_{P \in \mathcal{P}_0} \left\{ \mathbb{E}_P [\tilde{v}] - \alpha(P) \right\},$$

where $\alpha(P)$ is the penalty function and given a convex risk measure, $\mu$, the penalty function can be determined as

$$\alpha(P) = \sup_{\tilde{v} \in \mathcal{L}} \left\{ \mathbb{E}_P [\tilde{v}] - \mu(\tilde{v}) \right\}.$$

The penalty function plays a similar role as the probability distance metric. As we have presented above, while some convex risk measures can be associated with $\phi$-divergence, others may not. One of the prominent class of convex risk measures is the shortfall risk measures defined as

$$\mu_S(\tilde{v}) = \inf \left\{ a \mid \mathbb{E}_P [u(\tilde{v} - a)] \leq 0 \right\},$$
for a given convex and non-decreasing disutility function \( u : \mathbb{R} \to \mathbb{R} \) such that \( u(0) = 0 \) and \( 1 \in \partial u(0) \). The penalty function associated with the shortfall risk measure is given by (Foellmer and Schied 2002, Theorem 10)

\[
\alpha(P) = \begin{cases} 
\inf_{\lambda > 0} \frac{1}{\lambda} \mathbb{E}_{\tilde{P}} \left[ u^* \left( \lambda \frac{d\tilde{P}}{dP} \right) \right] & \text{if } P \ll \tilde{P}, \\
+\infty & \text{otherwise},
\end{cases}
\]

where \( u^* \) is the convex conjugate of \( u \).

**Proposition 4 (Shortfall-based Adversarial Impact Measure).** The shortfall-based adversarial impact measure,

\[
\rho_S(\tilde{v}) := \inf \left\{ k > 0 \mid \mathbb{E}_{\tilde{P}}[\tilde{v}] \leq k \inf_{\lambda > 0} \frac{1}{\lambda} \mathbb{E}_{\tilde{P}} \left[ u^* \left( \lambda \frac{d\tilde{P}}{dP} \right) \right] \right\} \forall P \in \mathcal{P}_0, P \ll \tilde{P}
\]

is equivalent to

\[
\rho_S(\tilde{v}) := \inf \{ k > 0 \mid k \mathbb{E}_{\tilde{P}}[u(\tilde{v}/k)] \leq 0 \}. \tag{12}
\]

It is interesting to note that the riskiness index of Aumann and Serrano (2008) is also a special case of shortfall-based adversarial impact measure in which the disutility function is given by \( u(v) = \exp(v) - 1 \). The simplest case of shortfall-based adversarial impact measure has been proposed in Zhang et al. (2019) for which the disutility function is given by \( u(v) = \max\{v, -1\} \). Given the simplicity, we call this the essential shortfall-based adversarial impact measure as follows:

\[
\rho_E(\tilde{v}) := \min \{ k > 0 \mid \mathbb{E}_{\tilde{P}}[\max\{\tilde{v}, -k\}] \leq 0 \}.
\]

**Optimal-transport metric**

We can also extend the probability distance metric to Wasserstein metric, also known as the Kantorovich-Rubinstein metric, which has been widely used in data-driven distributionally robust optimization (see, e.g., Gao and Kleywegt 2016, Gao et al. 2017, Mohajerin Esfahani and Kuhn 2018). Unlike \( \phi \)-divergence, the optimal transport metric does not require absolute continuity. The random variable at the model’s constraint, \( v(\tilde{z}) \) is now a measurable function of a multivariate random variable, \( \tilde{z} \), with distribution \( \tilde{z} \sim \tilde{P}, \tilde{P} \in \mathcal{P}_0(\mathcal{Z}) \). For two distributions \( P, \tilde{P} \in \mathcal{P}_0(\mathcal{Z}) \), the Wasserstein metric (of type-1) is defined as:

\[
W(P, \tilde{P}) := \inf_{Q \in \mathcal{P}_0(\mathcal{Z} \times \mathcal{Z})} \left\{ \mathbb{E}_Q[d(\tilde{z}, \tilde{u})] \left| (\tilde{z}, \tilde{u}) \sim Q, \tilde{z} \sim P, \tilde{u} \sim \tilde{P} \right. \right\},
\]

where \( d \) is an appropriate distance metric, e.g., Euclidean distance.

**Proposition 5 (Wasserstein-based adversarial impact measure).** The Wasserstein-based adversarial impact measure,

\[
\rho_W(v(\tilde{z})) := \inf \{ k \geq 0 \mid \mathbb{E}_P[v(\tilde{z})] \leq kW(P, \tilde{P}), \forall P \in \mathcal{P}_0(\mathcal{Z}) \}
\]
is equivalent to
\[ \rho_W(v(\tilde{z})) := \inf \left\{ k \geq 0 \ \middle| \ \sup_{Q \in \mathcal{Q}} \mathbb{E}_Q [v(\tilde{z}) - kd(\tilde{z}, \tilde{u})] \leq 0 \right\}, \]
where
\[ \mathcal{Q} = \left\{ Q \in \mathcal{P}_0(\mathcal{Z} \times \mathcal{Z}) \ \middle| \ (\tilde{z}, \tilde{u}) \sim Q, \ \tilde{u} \sim \hat{P} \right\}. \]

4. Target-based Choices

The adversarial impact measure is a target-based decision criterion or satisficing measure proposed in Brown and Sim (2009). Apart from establishing the normative properties of the adversarial impact measure, we also explore the descriptive aspect, i.e., in relation to how human subjects would make those choices. In behavioral decision making, probability distribution may be distorted by the decision maker and the choices she makes may not be consistent with the paradigm of expected utility (Kahneman and Tversky 1979). Incidentally, such distortion and its impact on choices may well be reflected via the adversarial impact measure.

We consider a lottery \( \tilde{r} \) with random discrete outcomes, \( \hat{P}[\tilde{r} = r_i] = q_i, i \in [N] \) and \( r_1 \leq \cdots \leq r_N \). In the target-based choice, the decision maker’s preference of the lotteries is determined by her desired target, \( \tau \in [r_1, r_N] \) that she would like to attain. Specifically, the preference is determined by the adversarial impact measure on the target shortfall \( \rho(\tau - \tilde{r}) \), for which the baseline probability is \( \hat{P} \). The lower the value, the more robust and hence more preferred the random lottery would be for the desired target.

Observe that whenever \( \mathbb{E}_\hat{P} [\tilde{r}] < \tau \), \( \rho(\tau - \tilde{r}) = \infty \) and hence, the adversarial impact measure is unable to differentiate among lotteries for which the expected returns are below the target. While this is reasonable in normative decision making to avoid over risky positions, for descriptive purpose, we can also extend the adversarial impact measure to elicit preferences over fragile positions for which the target return cannot be attained in expectation. Specifically, whenever \( \mathbb{E}_\hat{P} [\tilde{r}] < \tau \), \( \rho(\tilde{r} - \tau) \) evaluates the fragility of the lottery in meeting the desired target; the smaller the value, the higher the fragility and consequently, the lower the preference.

To incorporate preferences of both robust and fragile positions in a single measure, we follow Brown et al. (2012) to define an aspiration measure, \( \alpha_\rho : \mathcal{L} \mapsto [-\infty, \infty] \),
\[
\alpha_\rho(\tilde{v}) := \begin{cases} 
\frac{1}{\rho(\tilde{v})} & \text{if } \rho(\tilde{v}) < \infty \\
-\frac{1}{\rho(-\tilde{v})} & \text{otherwise}.
\end{cases}
\]
The preference over the lotteries would then be consistent with how high the value of the aspiration measure, \( \alpha_\rho(\tau - \tilde{r}) \) could attain.
Hence, the target plays a similar role as the reference point in prospect theory. One tends to be risk averse in making choices among risky positions with high returns relative to one’s target or reference point, and the converse is true among those with returns that are expected to perform poorly. Empirical evidences have shown that targets are critical in decision making (see, e.g., Payne et al. 1980, 1981) and is usually set according to some simple heuristics (see, e.g., Merchant and Manzoni 1989, Köszegi and Rabin 2006).

Allais Paradox

We further illustrate our choice over lotteries using the paradox from Allais (1953), which is a well studied example in descriptive decision theory.

Consider the following four lotteries:

- Lottery A: Wins $1M for sure.
- Lottery B: 1% chance of nothing, 10% of winning $5M, and 89% chance of winning $1M.
- Lottery C: 89% chance of nothing, 11% chance of winning $1M.
- Lottery D: 90% chance of nothing, 10% chance of winning $5M.

Under the expected utility paradigm, the possible preferences are either A>B and C>D, or B>A and D>C (here A>B means lottery A is preferred over lottery B and A~B implies indifference), and it is not possible to induce a preference of either A>B and D>C or B>A and C>D. To see this, we note that for any utility function, u, A>B implies \( u(1) > 0.01u(0) + 0.10u(5) + 0.89u(1) \) or \( 0.11u(1) > 0.01u(0) + 0.10u(5) \), while D>C demands \( 0.90u(0) + 0.10u(5) > 0.89u(0) + 0.11u(1) \) or \( 0.01u(0) + 0.10u(5) > 0.11u(1) \), which is a contradiction. While it may well be reasonable to rule out the incoherent preference B>A and C>D, behavioral studies reflect a dominant preference of A>B and D>C, which is inconsistent with the paradigm of expected utility. In fact, this preference cannot also be reconciled by other popular normative preferences such as value-at-risk (VaR) and conditional value-at-risk (CVaR) (Rockafellar and Uryasev 2000) defined respectively as

\[
\text{VaR}_{\hat{P}, \epsilon}(\hat{r}) := \inf_w \left\{ w \mid \mathbb{P}[\hat{r} \leq w] \geq 1 - \epsilon \right\},
\]
\[
\text{CVaR}_{\hat{P}, \epsilon}(\hat{r}) := \inf_w \left\{ w + \frac{1}{\epsilon} \mathbb{E}_{\hat{P}} \left[ (\hat{r} - w)^+ \right] \right\},
\]

for \( \epsilon \in (0, 1] \), and also the mean-variance criterion, defined as \( \mathbb{E}_{\hat{P}}[\hat{r}] - \lambda \cdot \text{var}_{\hat{P}}(\hat{r}) \) for \( \lambda \in \mathbb{R} \), where \( \text{var}_{\hat{P}}(\hat{r}) \) is the variance of \( \hat{r} \) under \( \hat{P} \). The choices in Allais paradox according to these three criteria are presented in Tables 1-2.

We now explore choices over lotteries using the aspirational measures that are based on the following adversarial impact measures:

1. Riskiness index of Aumann and Serrano (2008), \( \rho_{KL} \).
2. Essential shortfall-based adversarial impact measure, \( \rho_E \).
3. $\phi$-divergence-based adversarial impact measure with total variation distance, $\rho_{TV}$. Note that the last two adversarial impact measures can be solved via linear optimization.

We present the preferences of lotteries in Table 3. Note that, except for the incoherent preference where $B\succ A$ and $C\succ D$, all other preferences are possible. In fact, that there is a broad range of targets $\tau \in (0.11M, 1M)$ reflects a dominant preference of $A\succ B$ and $D\succ C$, which is consistent with behavioral studies.

This study also elucidates the usefulness of target-based decision criteria, which has well been acknowledged in behavioral decision making, though it is rarely used as a decision criterion for optimization under uncertainty. Charnes and Cooper (1963) introduce the P-model that aims to maximize the probability of target attainment. However, as an expected step utility criterion, it is neither amiable to optimization nor descriptively useful for decision making under uncertainty (see Diecidue and Van De Ven 2008).

5. Practicable Robustness Optimization Models

We present and compare the robustness optimization models that are analogous to some of the common robust optimization models found in the literature. We also show how the framework can be used to improve the robustness of dynamic optimization problems.
Linear optimization problem
Following Bertsimas and Sim (2004), we consider the following baseline linear optimization problem,

$$Z_0 = \max c^\top x$$
$$\text{s.t. } a_i(\hat{z})^\top x \leq b_i, \ \forall i \in [I]$$
$$x \in \mathcal{X}.$$  

The coefficients affinely depend on some exogenous parameter $z \in \mathcal{Z}$, $\mathcal{Z} = [-1, 1]^N$:

$$a_i(z) = a_{i,0} + \sum_{n \in [N]} a_{i,n} z_n, \ \forall i \in [I],$$

with baseline parameter $\hat{z} = 0$. Bertsimas and Sim (2004) propose the following robust optimization model:

$$\begin{align*}
\max c^\top x \\
\text{s.t. } a_{i,0}^\top x + \sum_{n \in [N]} a_{i,n}^\top x z_n \leq b_i, \ \forall i \in [I], \ \forall z \in \mathcal{U}(r_i) \tag{14}
\end{align*}$$

where $\mathcal{U}(r_i) = \{z \in \mathcal{Z} \mid \|z\|_1 \leq r_i\}$, $i \in [I]$ are the budgeted uncertainty sets. Here, the modeler specifies the parameters $r_i$, $i \in [I]$ depending on the relative importance of individual constraints. Problem (14) can be reformulated as the following linear optimization problem:

$$\begin{align*}
\max c^\top x \\
\text{s.t. } a_{i,0}^\top x + \sum_{n \in [N]} s_{i,n} + r_i t_i \leq b_i, \ \forall i \in [I] \\
t_i + s_{i,n} \geq a_{i,n}^\top x, \quad \forall i \in [I], n \in [N] \\
t_i + s_{i,n} \geq -a_{i,n}^\top x, \quad \forall i \in [I], n \in [N] \\
t \geq 0, \ s \geq 0 \\
x \in \mathcal{X}.
\end{align*}$$

The robustness linear optimization problem analogous the Bertsimas and Sim (2004) robust optimization model is as follows

$$\begin{align*}
\min w^\top \kappa \\
\text{s.t. } a_{i,0}^\top x + \sum_{n \in [N]} a_{i,n}^\top x z_n - b_i \leq \kappa_i \|z\|_1, \ \forall i \in [I], \ \forall z \in \mathcal{Z} \tag{15}
\end{align*}$$

where $w_i$, $i \in [I]$ are weights to reflect the relative importance of individual constraints. A reasonable choice would be to set $w_i = 1/|b_i|$, if $|b_i| > 0$, so that the effective adversarial impact is the weighted sum of adversarial impacts normalized by the right-hand-side values of the constraints.
By reformulating the robust counterpart to its dual formulation, Problem (15) admits the following linear formulation:

\[
\begin{align*}
\min & \quad w^\top \kappa \\
\text{s.t.} & \quad a_{i,0}^\top x + \sum_{n \in [N]} s_{i,n} \leq b_i, \quad \forall i \in [I] \\
& \quad \kappa_i + s_{i,n} \geq a_{i,n}^\top x, \quad \forall i \in [I], n \in [N] \\
& \quad \kappa_i + s_{i,n} \geq -a_{i,n}^\top x, \quad \forall i \in [I], n \in [N] \\
& \quad \kappa \geq 0, s \geq 0 \\
& \quad c^\top x \geq \tau \\
& \quad x \in \mathcal{X},
\end{align*}
\]

which is almost similar in complexity as the Bertsimas and Sim (2004) robust optimization model.

**Combinatorial optimization problem**

Following Bertsimas and Sim (2003), we consider the following baseline combinatorial optimization problem:

\[
Z_0 = \min_c c(\hat{z})^\top x \\
\text{s.t.} \quad x \in \mathcal{X},
\]

where \(\mathcal{X} \subseteq \{0,1\}^N\) and the cost parameter \(c(z)\) depends on some exogenous parameter \(z \in \mathcal{Z}\), \(\mathcal{Z} = [0,1]^N\). More specifically, for any \(n \in [N]\):

\[
c_n(z) = c_n + d_n z_n,
\]

with \(c, d \geq 0\) and baseline parameter \(\hat{z} = 0\). Bertsimas and Sim (2003) present the following robust optimization model:

\[
\begin{align*}
\min & \quad c^\top x + \max_{z \in \mathcal{U}(r)} \sum_{n \in [N]} d_n x_n z_n \\
\text{s.t.} & \quad x \in \mathcal{X},
\end{align*}
\]

where the budgeted uncertainty set, \(\mathcal{U}(r)\), is defined as:

\[
\mathcal{U}(r) = \{z \in \mathcal{Z} \mid \|z\|_1 \leq r\}.
\]

They show that the optimal solution can be obtained by solving \(N+1\) of the following combinatorial optimization problems with different linear objective functions:

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

for \(k \in \{0, d_1, \ldots, d_N\}\). Hence, the robust model (16) is polynomial-time solvable if the baseline combinatorial problem is also polynomial-time solvable (Bertsimas and Sim 2003, Theorem 3).
We present the robustness optimization model analogous to Bertsimas and Sim (2003) robust optimization model as follows:

\[
\begin{align*}
\min & \quad k \\
\text{s.t.} & \quad c^\top x + \sum_{n \in [N]} d_n x_n z_n - \tau \leq k \|z\|_1, \forall z \in \mathcal{Z} \\
& \quad x \in \mathcal{X}, k \geq 0,
\end{align*}
\]  \tag{17}

for some cost budget \( \tau \geq Z_0 \).

**Theorem 2.** The combinatorial robustness optimization problem (17) is equivalent to the following combinatorial optimization problem:

\[
\begin{align*}
\min & \quad k \\
\text{s.t.} & \quad \min_{x \in \mathcal{X}} \left\{ c^\top x + \sum_{n=1}^{N} (d_n - k)^+ x_n \right\} \leq \tau \\
& \quad k \geq 0.
\end{align*}
\]

By performing bisection search on \( k \) we can obtain the optimal solution up to \( \epsilon > 0 \) accuracy by solving at most \( \lceil \log_2(\tilde{d}/\epsilon) \rceil \), \( \tilde{d} = \max \{d_1, \ldots, d_N\} \) combinatorial optimization problems with different linear objective functions.

Similar to Bertsimas and Sim (2003) robust optimization model, Theorem 2 implies that the robustness optimization model (17) is polynomial-time solvable if the baseline combinatorial problem is also polynomial-time solvable, though the number of baseline problems to solve does not grow with \( N \). As a simple comparison, consider a shortest path problem with \( N = 1000 \) arcs and the largest deviation of travel time along an arc is \( \tilde{d} = 10 \). The robust model requires solving 1001 baseline combinatorial optimization problems, while the robustness model only requires solving 20 baseline problems to achieve an accuracy of \( \epsilon = 10^{-5} \).

**Data-driven utility maximization problem**

We consider the following baseline data-driven utility maximization problem:

\[
Z_0 = \max \mathbb{E}_{\hat{\mathbb{P}}} \left[ u(x^\top \hat{z}) \right] \\
\text{s.t.} \quad x \in \mathcal{X},
\]

for some empirical distribution \( \hat{\mathbb{P}} \) and a concave piecewise affine increasing utility, \( u(v) = \min_{i \in [I]} \{a_i v + b_i\} \). Mohajerin Esfahani and Kuhn (2018) propose the following data-driven distributionally robust optimization model

\[
\max \inf_{x \in \mathcal{X}} \mathbb{E}_{\hat{\mathbb{P}}} \left[ u(x^\top \hat{z}) \right]
\]
where the ambiguity set
\[ B_W(r) = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathcal{Z}) \bigg| \hat{z} \sim \mathbb{P}, W(\mathbb{P}, \hat{\mathbb{P}}) \leq r \right\} \]
is a Wasserstein ball of radius \( r \) around the empirical distribution \( \hat{\mathbb{P}} \), constructed from historical data \( (\hat{z}_1, \ldots, \hat{z}_M) \). For the distance function, \( d(z, u) = \|z - u\| \) and support set \( \mathcal{Z} = \{ z \in \mathbb{R}^N \mid Cz \leq h \} \), they provide the following reformulation:

\[
\begin{align*}
\max \ -\lambda r + \frac{1}{M} \sum_{m \in [M]} y_m \\
\text{s.t. } y_m &\leq a_i x^\top \hat{z}_m + b_i + \eta_{im}^\top h - \eta_{im}^\top C \hat{z}_m, \quad \forall i \in [I], \ m \in [M] \\
\lambda &\geq \|C^\top \eta_{im} - a_i x\|_*, \quad \forall i \in [I], \ m \in [M] \\
\eta_{im} &\leq 0, \quad \forall i \in [I], \ m \in [M] \\
x &\in \mathcal{X}.
\end{align*}
\]

Using the Wasserstein-based adversarial impact measure and following Proposition 5, we formulate the data-driven robustness optimization problem for given target utility \( \tau \leq Z_0 \) as follows:

\[
\begin{align*}
\min k \\
\text{s.t. } \mathbb{E}_Q \left[ \tau - u(x^\top \hat{z}) - k \|\hat{z} - \hat{u}\| \right] &\leq 0 \ \forall Q \in \mathcal{Q} \\
x &\in \mathcal{X}, \ k \geq 0,
\end{align*}
\]

where
\[ \mathcal{Q} = \left\{ Q \in \mathcal{P}_0(\mathbb{R}^N \times \mathbb{R}^N) \mid (\hat{z}, \hat{u}) \sim Q, \mathbb{Q} [\hat{u} = \hat{z}_m] = 1/M, \ \forall m \in [M] \right\}. \]

**Theorem 3.** Problem (18) is equivalent to the following convex optimization problem:

\[
\begin{align*}
\min \ k \\
\text{s.t. } \frac{1}{M} \sum_{m \in [M]} y_m &\geq \tau \\
y_m &\leq a_i x^\top \hat{z}_m + b_i + \eta_{im}^\top h - \eta_{im}^\top C \hat{z}_m, \quad \forall i \in [I], \ m \in [M] \\
k &\geq \|C^\top \eta_{im} - a_i x\|_*, \quad \forall i \in [I], \ m \in [M] \\
\eta_{im} &\leq 0, \quad \forall i \in [I], \ m \in [M] \\
x &\in \mathcal{X}.
\end{align*}
\]

**Data-driven portfolio optimization problem**

We first illustrate the robustness optimization framework in the context of portfolio selection under ambiguity. For the baseline problem, we consider a risk averse investor investing in \( N \) risky assets using the empirical distribution \( \hat{\mathbb{P}} \) constructed from historical data \( (\hat{z}_1, \ldots, \hat{z}_M) \) on the returns of
these risky assets. The investor minimizes the empirical CVaR defined in (13) in the following optimization problem,
\[
Z_0 = \min w + \frac{1}{\epsilon} \mathbb{E}_{\hat{\mathcal{P}}^k} \left[ (-x^T \bar{z} - w)^+ \right] \\
\text{s.t. } 1^T x = 1 \\
x \in \mathbb{R}^N, \; w \in \mathbb{R}.
\]

To guard against ambiguity, we adopt a robustness optimization approach using the Wasserstein-based adversarial impact measure with $l_1$-norm. Following from Proposition 5, we formulate the data-driven robustness optimization problem for a given target on CVaR, $\tau \geq Z_0$, as follows:
\[
\begin{align*}
\min k \\
\text{s.t. } w + \frac{1}{\epsilon} \mathbb{E}_Q \left[ (-x^T \bar{z} - w)^+ \right] - \tau \leq k \mathbb{E}_Q \left[ || \bar{z} - \bar{u} ||_1 \right] \forall Q \in \mathcal{Q} \\
1^T x = 1 \\
x \in \mathbb{R}^N, \; w \in \mathbb{R}, \; k \geq 0,
\end{align*}
\]
\[\text{where } \mathcal{Q} = \left\{ Q \in \mathcal{P}_0(\mathbb{R}^N \times \mathbb{R}^N) \left| \begin{array}{l}
(\bar{z}, \bar{u}) \sim Q \\
Q[\bar{u} = \bar{z}_m] = 1/M, \forall m \in [M]
\end{array} \right. \right\}. \tag{20}\]

**Theorem 4.** The optimal portfolio in Problem (20) is equivalent to the optimal portfolio in the following linear optimization problem:
\[
\begin{align*}
\min k \\
\text{s.t. } \text{CVaR}_{1/\epsilon}(x^T \bar{z}) \leq \tau \\
x^T 1 = 1 \\
k \geq ||x||_\infty \\
x \in \mathbb{R}^N.
\end{align*}
\]

It is clear the robustness optimization approach improves the robustness by diversifying as much as possible, while trying to attain the target level of risk, $\tau$ over the empirical distribution.

**Corollary 1.** When $\tau \geq \text{CVaR}_{1/\epsilon}(1^T \bar{z}/N)$, the optimal portfolio is the equal-weighted $1/N$ portfolio.

To provide a simple illustration of the above proposition, consider a risk-neutral case i.e., $\epsilon = 1$ and CVaR$_{1/\epsilon}(x^T \bar{z}) = -\mu^T x$, where $\mu = \mathbb{E}_{\hat{\mathcal{P}}} [\bar{z}]$. Whenever the target return $-\tau$ is less than the average return of an equal weighted portfolio, i.e., $-\tau \leq 1^T \mu/N$, then the optimum portfolio would also be the equal-weighted portfolio. As the target return increases, the decision maker commits to less ambiguity averse portfolio and diversify less in order to achieve a more ambitious target.

Despite the simplicity, it has well been known from the empirical study of DeMiguel et al. (2009) that equal-weighted portfolio outperforms many risk models in practice. In particular, when they
compare the equal-weighted portfolio against 14 models across seven empirical datasets, none of the model is consistently better than the equal-weighted portfolio in out-of-sample performance. When there are 25 assets, they show that mean-variance strategy and its extensions require an estimation window of over 3000 months in order to outperform the equal-weighted portfolio, alluding to their impractically in addressing actual portfolio selection problems.

The optimality of equal weighted portfolios has also been established in distributionally robust portfolio optimization models (see, e.g., Pflug et al. 2012, Mohajerin Esfahani and Kuhn 2018) as the Wasserstein ball increases in size within the ambiguity set. In the robustness optimization model, the optimal portfolio depends on the target of the decision maker, which we believe is more interpretable for the decision maker to specify than the radius of the Wasserstein ball.

**Dynamic optimization problem**

Robust dynamic optimization problems have been well studied in the literature (see, e.g., Hansen and Sargent 2001, Iyengar 2005, Lim and Shanthikumar 2007, Xu and Mannor 2012, Wiesemann et al. 2013). Here we revisit this from the perspective of robustness optimization. For simplicity, we focus on the case of discrete time but the results can be extended to continuous time. We consider the dynamic system

\[ s_{t+1} = f_t(s_t, a_t, z_t), \quad t \in [T], \]

where for all \( t \in [T] \), \( s_t \in \mathbb{R}^{N_s}, a_t \in \mathbb{R}^{N_a}, z_t \in \mathbb{R}^{N_z} \) are the state of the system, the control, and the realization of uncertainty \( \tilde{z}_t \) in period \( t \), respectively. The control \( a_t \) is constrained to be in a set \( A_t(s_t) \) which is specified by the system state \( s_t \). We assume that with the baseline distribution \( \tilde{P} \), is such that \( \tilde{z}_1, \ldots, \tilde{z}_t \) are independently distributed. Given an initial state \( s_1 \), we want to find a policy \( \pi = \{x_1, \ldots, x_T\} \), where \( x_t : \mathbb{R}^{N_z} \to \mathbb{R}^{N_x} \) maps the system state to a control. Denote \( \Pi \) as the set of all admissible policies, i.e., the set of all \( \pi = \{x_1, \ldots, x_T\} \) with \( x_t(s_t) \in A_t(s_t) \) for all \( s_t \in \mathbb{R}^{N_s}, \ t \in [T] \). The problem of minimizing expected loss can be formulated as

\[
Z_0 = \min_{\pi \in \Pi} \mathbb{E}_{\tilde{P}} \left[ \sum_{t \in [T]} l_t(s_t, x_t(s_t), \tilde{z}_t) \right]
\]

where the function \( l_t \) maps to the loss in period \( t \). Typically, this problem is solved via Bellman equations.

Based on the shortfall-based adversarial impact measure as defined in Equation (12), we reformulate the robustness optimization problem as

\[
R^* = \min_{k} \quad k \\
\text{s.t.} \quad \inf_{\pi \in \Pi} \left\{ k \mathbb{E}_{\tilde{P}} \left[ u \left( \frac{\sum_{t \in [T]} l_t(s_t, x_t(s_t), \tilde{z}_t) - \tau}{k} \right) \right] \right\} \leq 0 
\]

(22)
for some convex and non-decreasing disutility, \( u \) normalized by \( u(0) = 0 \) and \( 1 \in \partial u(0) \).

A similar problem that minimizes a target oriented objective function under risk aversion has also been studied in Chen et al. (2015), while our approach aims to minimize the adversarial impact under distributional ambiguity. Chen et al. (2015) show that given any \( k > 0 \), the constraint is feasible if and only if \( k \geq R^* \). Hence, we can perform bisection search on \( k \) to find an \( \epsilon \)-optimal policy by solving at most \( \lceil \log_2(R/\epsilon) \rceil \) dynamic optimization problems under expected disutility as follows,

\[
\min_{\pi \in \Pi} k \mathbb{E}_{\hat{\mathbb{P}}} \left[ u \left( \frac{\sum_{t \in [T]} l_t(s_t, x_t(s_t), \tilde{z}_t)}{k} - \tau \right) \right].
\]

(23)

While this problem can be solved via standard dynamic optimization procedure, in general, it increases the dimension of state space when searching for the optimal policy (see, for instance, Chen et al. 2015). The only exception is the case where \( u \) is chosen as the exponential disutility function, \( u(v) = \exp(v) - 1 \). With this disutility function, Problem (23) can be equivalently formulated as

\[
G^* = \min_{\pi \in \Pi} k \log \mathbb{E}_{\hat{\mathbb{P}}} \left[ \exp \left( \frac{\sum_{t \in [T]} l_t(s_t, x_t(s_t), \tilde{z}_t)}{k} \right) \right].
\]

(24)

We remark that with \( u \) being the exponential disutility, the target \( \tau \) does not affect the optimal solution in Problem (23), and hence \( \tau \) is omitted in its reformulation. Nevertheless, \( \tau \) affects the optimal value in Problem (23). Therefore, it affects both the optimal value and the optimal solution in our robustness optimization problem (22).

Problem (24), which optimizes the certainty equivalent for exponential disutility, has been discussed by Jacobson (1973). Specifically, the optimal policy can be obtained via the following recursion,

\[
G_t(s_t) = \min_{a_t \in A_t(s_t)} k \log \mathbb{E}_{\hat{\mathbb{P}}} \left[ \exp \left( \frac{l_t(s_t, a_t, \tilde{z}_t) + G_{t+1} ( f_t(s_t, a_t, \tilde{z}_t) )}{k} \right) \right], \quad \forall s_t \in \mathbb{R}^{N_s}, \ t \in [T]
\]

(25)

where we let \( G_{T+1}(s) = 0 \) for all \( s \); we then have \( G^* = G_1(s_1) \).

We remark the philosophical difference from Jacobson (1973) expected exponential disutility model for which the decision maker specifies her risk tolerance parameter \( k \) under the subjective distribution \( \hat{\mathbb{P}} \). In the robustness optimization model, the decision maker specifies a target \( \tau > Z_0 \) and minimizes the adversarial impact to best protect objective from exceeding the target due to potential misspecification of the subjective probability, which could occur, for instance, when the independence assumption does not hold.

Our method can be applied in a broad class of practical dynamic decision making problems. With the dynamic recursion equation (25), the optimal policy from robustness optimization model usually preserves the same structure as the model with given distribution. For example, consider
the classical multi-period inventory control and pricing problem. With the recursion (25) and the technique in Chen et al. (2007), we can show that the optimal policy has the structure of \((s, S, A, p)\), preserving the same structure as the baseline dynamic optimization problem of Chen and Simchi-Levi (2004).

6. Numerical Study: Robust vs Robustness

In our numerical study, we consider the following baseline knapsack problem proposed in Bertsimas and Sim (2004),

\[
\begin{align*}
\max_{i \in [N]} & \quad c_i x_i \\
\text{s.t.} & \quad \sum_{i \in [N]} w_i(\hat{z}_i) x_i \leq b \\
& \quad x_i \in \{0, 1\} \quad \forall i \in [N],
\end{align*}
\]

where the uncertain weights are affinely dependent on the random variable, \(z \in \mathcal{Z}, \mathcal{Z} = [-1, 1]^N\):

\[
w_i(z_i) = \hat{w}_i + \delta_i z_i, \quad \forall i \in [I],
\]

with the baseline parameter \(\hat{z} = 0\).

In the experiments, we have \(N = 50, b = 2000\), each unit profit \(c_i\) taken randomly from \(\{10, 12, 14, 16, 18\}\), each baseline weight \(\hat{w}_i\) drawn randomly from \(\{20, 22, 24, \ldots, 80\}\) and \(\delta = 0.2 \hat{w}\). Hence, the uncertain weights will take values in \([0.8 \hat{w}, 1.2 \hat{w}]\), though the decision maker does not know the actual distribution. We compare our robustness optimization model to the robust optimization model of Bertsimas and Sim (2004). Specifically, we vary the budget of uncertainty \(r\) and for each \(r\), we solve the corresponding robust optimization problem to obtain the optimal robust solution at the profit of \(Z_r\). Subsequently, we obtain the solutions to the robustness optimization with the target profit at \(\tau = Z_r\). Hence, all comparisons of solutions between the two approaches are made on the same attainable profit. In our instance, the baseline profit is \(Z_0 = 592\), and the worst-case profit is \(Z_\infty = 534\).

In the first simulation experiment, we assume that \(\tilde{z}\) is uniformly distributed on the support \(\mathcal{Z}\), and we generate 20,000 samples beforehand to compare the average performance of the two models. We plot in Figure 1(a) the probability of infeasibility of the robust optimization as the solutions vary with \(r\). We also compare this with the solutions of the robustness optimization, setting the target profit to be the same as profit attainable by the robust optimization problem. In Figure 1(b), we replicate the same plot against the target profit, \(\tau\). Similarly, in Figure 1(c) and 1(d), we compare the average violations of the capacity constraints for the solutions of the robust optimization and the robustness optimization problems, as they vary with \(r\) and \(\tau = Z_r\), respectively. In the second simulation experiment, we assume that \(\tilde{z}\) is now uniformly distributed
in $[-0.5, 1]^N$ instead of the full support $\mathcal{Z}$. Other setups remain the same and we summarize the comparison in Figure 2.

We observe from these figures that for the same attainable profit, the solutions obtained by robustness optimization would generally dominate those obtained by robust optimization. While we observe monotonicity in the performance of solutions as we vary $\tau$ in the robustness optimization model, this is not the case as we vary $r$ in the robust optimization problem. We also observe that for a modest reduction of profit from the baseline knapsack problem, we can significantly improve the robustness of the solutions. We note that it is generally difficult to determine $r$ to match the desired performance for the robust optimization model. Even if the assumption of ambiguous distributions is correct, the probability bound of Bertsimas and Sim (2004) may also be too weak to be useful as a guideline for the choice of $r$. In sum, for the same price of robustness, robustness optimization consistently yields more robust solutions than those obtained by robust optimization, and doing so for the same computational effort.
Figure 2  Robust knapsack experiment 2 – asymmetrical deviation.

References


A. Proof of Results

Proof of Proposition 1. Consider the $x$ given as in the proposition, and denote $\bar{f} = \sup_{z \in Z} f(x, z) < \infty$. The case of $\bar{f} \leq 0$ is trivial and we just focus on the case where $\bar{f} > 0$. Since $U$ contains $\tilde{z}$ in its interior, there exists $\epsilon > 0$ such that for all $z$ with $\|z - \tilde{z}\| \leq \epsilon$, we have $z \in U$. Choose $k = \bar{f}/\epsilon > 0$. For any $z \in Z$ with $\|z - \tilde{z}\| \leq \epsilon$, we know $z \in U$ and hence $f(x, z) \leq 0 \leq k\|z - \tilde{z}\|$. For any $z \in Z$ with $\|z - \tilde{z}\| > \epsilon$, we know $f(x, z) \leq \bar{f} = k\epsilon < k\|z - \tilde{z}\|$. Therefore, $f(x, z) \leq k\|z - \tilde{z}\|$ holds for all $z \in Z$. 

Proof of Theorem 1. We first prove the “only if” direction. Given an adversarial impact measure $\rho$ defined by Equation (8), we show that it has all the five properties in this theorem and is lower semi-continuous, i.e., $\{\tilde{v} \mid \rho(\tilde{v}) \leq a\}$ is a closed set for any $a > 0$. For convenience, we define

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

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

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

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

$$ \rho(\lambda \tilde{v}) = \inf \left\{ k > 0 \mid \mathbb{E}_\mathbb{P}[\lambda \tilde{v}] \leq k\Delta(\mathbb{P}, \tilde{\mathbb{P}}), \ \forall \mathbb{P} \in \mathcal{P}_0 \right\} $$

$$ = \inf \left\{ k > 0 \mid \mathbb{E}_\mathbb{P}[\tilde{v}] \leq \frac{k}{\lambda}\Delta(\mathbb{P}, \tilde{\mathbb{P}}), \ \forall \mathbb{P} \in \mathcal{P}_0 \right\} $$

$$ = \lambda \inf \left\{ \beta > 0 \mid \mathbb{E}_\mathbb{P}[\tilde{v}] \leq \beta\Delta(\mathbb{P}, \tilde{\mathbb{P}}), \ \forall \mathbb{P} \in \mathcal{P}_0 \right\} $$

$$ = \lambda \rho(\tilde{v}). $$

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

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

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

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

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

The lower semi-continuity of $\rho$ can be shown as follows. Consider any converging sequence of random variable $\tilde{v}_1, \ldots, \tilde{v}_n$ such that $\tilde{v}_n \to \tilde{v}$ as $n \to +\infty$. For any fixed value $a > 0$, we need to show that $\rho(\tilde{v}) \leq a$ if $\rho(\tilde{v}_n) \leq a$ for all $n > 0$. This is true because the expectation is a continuous measure, i.e., $\lim_{n \to +\infty} \mathbb{E}_\mathbb{P}[\tilde{v}_n] = \mathbb{E}_\mathbb{P}[\tilde{v}]$. 

Long, Sim, and Zhou: The Dao of Robustness
We now prove the “if” direction. Consider any given lower semi-continuous function $\rho$ which satisfies all the five properties in this theorem. Note that $\rho$ is also convex, because positive homogeneity and subadditivity imply convexity. We first show that the function $\Delta$ defined by Equation (9) is a probability distance function. We note that since $\rho(0) = 0 < 1$ by Absolute Robustness, $\Delta(\mathbb{P}, \mathbb{P}) \geq \mathbb{E}_{\mathbb{P}} [0] = 0$ for all $\mathbb{P} \in \mathcal{P}_0$. Moreover, due to the property of Baseline Infeasibility, $\rho(\mathbb{v}) = \infty > 1$ for all $\mathbb{v}$ with $\mathbb{E}_{\mathbb{P}} [\mathbb{v}] > 0$. Therefore, by Equation (9),

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

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

With $\Delta$ defined as in Equation (9), given any $\mathbb{v} \in \mathcal{L}$, we define the set $\mathcal{K}(\mathbb{v})$ as in Equation (26). Now it remains to prove $\inf \mathcal{K}(\mathbb{v}) = \rho(\mathbb{v})$. To see this, we start from the case where $\rho(\mathbb{v}) \in (0, \infty)$.

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

We then show $\inf \mathcal{K}(\mathbb{v}) \geq \rho(\mathbb{v})$. Consider any $0 < k < \rho(\mathbb{v})$, and hence $\rho(\mathbb{v}/k) = \rho(\mathbb{v})/k > 1$. Denote a set $\mathcal{W} = \{ \mathbb{w} \in \mathcal{L} \mid \rho(\mathbb{w}) \leq 1 \}$. Then by convexity and lower semi-continuity of $\rho$, $\mathcal{W}$ is a closed convex set and $\mathbb{v}/k \notin \mathcal{W}$. Therefore, by Hahn-Banach separation theorem, there exists a linear functional $l$ with

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

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

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

\[\Box\]

Proof of Proposition 2. This follows by noticing there exists a dual representation of $\mu$:

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

The proof follows largely from existing proofs, see, e.g., Hall et al. (2015). Therefore, we omit it here. \[\Box\]
Proof of Proposition 3. Given any \( k > 0 \), we define a function \( \hat{\phi}(t) = k\phi(t) \) \( \forall t \geq 0 \). Its conjugate is
\[
\hat{\phi}^*(s) = \sup_{t \geq 0} \left\{ st - \hat{\phi}(t) \right\} = k \sup_{t \geq 0} \left\{ \frac{st}{k} - \phi(t) \right\} = k\phi^*(\frac{s}{k}).
\]
Therefore, in the definition of \( \rho_\phi \), the constraint on \( k \) is equivalent to
\[
0 \geq \sup_{P \in \mathcal{P}_0} \left\{ \mathbb{E}_P [\hat{v}] - kD_\phi(P||\hat{P}) \right\} = \sup_{P \in \mathcal{P}_0} \left\{ \mathbb{E}_P [\hat{v}] - D_\phi(P||\hat{P}) \right\} = \inf_n \left\{ \eta + \mathbb{E}_P \left[ \hat{\phi}^*(\hat{v} - \eta) \right] \right\} = \inf_n \left\{ \eta + k\mathbb{E}_P \left[ \phi^*((\hat{v} - \eta)/k) \right] \right\},
\]
where the second equality follows from Equation (11). \( \square \)

Proof of Proposition 4. By the monotonicity of \( u \) and the definition of \( \mu_S \), \( \mathbb{E}_P [u(\hat{v}/k)] \leq 0 \) if and only if \( \mu_S(\hat{v}/k) \leq 0 \). By the dual representation of \( \mu_S \), \( \mu_S(\hat{v}/k) \leq 0 \) can be reformulated as
\[
\sup_{P \in \mathcal{P}_0, \mathcal{P} \ll \hat{P}} \left\{ \mathbb{E}_P [\hat{v}/k] - \inf_{\lambda > 0} \frac{1}{\lambda} \mathbb{E}_P \left[ u^* \left( \frac{\lambda dP}{d\hat{P}} \right) \right] \right\} \leq 0,
\]
which is equivalent to \( \mathbb{E}_P [\hat{v}] \leq k \inf_{\lambda > 0} \frac{1}{\lambda} \mathbb{E}_P \left[ u^* \left( \frac{\lambda dP}{d\hat{P}} \right) \right] \), \( \forall P \in \mathcal{P}_0 \) with \( P \ll \hat{P} \). \( \square \)

Proof of Proposition 5. The equivalent reformulation can be achieved by directly applying the definition of Wasserstein metric. \( \square \)

Proof of Theorem 2. First, the robustness counterpart can be written as:
\[
\max_{n=1}^N d_n x_n z_n - k \| z \|_1 \iff \max_{n=1}^N (d_n x_n - k) z_n \\
\text{s.t. } z \in \mathcal{Z} \quad \text{s.t. } z \geq 0, z \leq 1.
\]
It follows from the strong duality that the formulation on the right is equivalent to:
\[
Y = \min_{n=1}^N y_n \\
\text{s.t. } y_n \geq d_n x_n - k, \forall n \in [N] \\
y_n \geq 0, \forall n \in [N].
\]
Because \( x_n \) is binary for any \( n \in [N] \) and \( k \geq 0 \), the above optimization problem has a closed-form solution:
\[
Y = \sum_{n=1}^N \max\{d_n - k, 0\} x_n
\]
Therefore, the original robustness optimization problem (17) can be written as:

\[
K^* = \min k \\
\text{s.t. } c^\top x + \sum_{n=1}^{N} \max\{d_n - k, 0\} x_n \leq \tau \\
x \in X, k \geq 0.
\]

For the second part, since the objective function is \(k\), we can do a bisecton search on \(k\). Note that \(K^* \leq \tilde{d}\) because we choose \(\tau \geq Z_0\), for practicality. For any fixed \(k \leq \tilde{d}\), we solve a feasibility problem that has the same complexity as the baseline combinatorial optimization problem. Starting from a feasible region, \(0, \tilde{d}\), each bisection search iteration reduces the searching region by half. Therefore, to achieve any given accuracy \(\epsilon\), we solve at most \(\lceil \log_2(\tilde{d}/\epsilon) \rceil\) baseline problems with different linear objective functions.

\[\square\]

**Proof of Theorem 3.** Recall the robustness optimization model is given by (18), and \(u(x^\top z) = \min_{i \in [I]} \{a_i x^\top z + b_i\}\). First, \(\inf_{Q \in Q} \mathbb{E}_Q [u(x^\top z) + k\|z - \tilde{u}\|]\) can be reformulated as:

\[
\inf_{z_m \in Z, m \in [M]} \frac{1}{M} \sum_{m \in [M]} \left( \min_{i \in [I]} \{a_i x^\top z_m + b_i\} + k\|z_m - \hat{z}_m\| \right),
\]

which is then equivalent to:

\[
\frac{1}{M} \sum_{m \in [M]} \min_{i \in [I]} \{a_i x^\top z + b_i + k\|z - \hat{z}_m\|\}.
\]

Now, we focus on the inner minimization, for any \(i \in [I], m \in [M]\), the inner minimization problem can be reformulated as:

\[
\inf_{z \in Z} \{a_i x^\top z + b_i + k\|z - \hat{z}_m\|\} = \inf_{z \in Z} \left\{a_i x^\top z + b_i + \max_{\gamma \in \gamma_{im}} \gamma^\top (z - \hat{z}_m) \right\} = \max_{\|\gamma\| \leq k} \inf_{z \in Z} \left\{a_i x^\top z + \gamma_{im}^\top z \right\} + b_i - \gamma_{im}^\top \hat{z}_m,
\]

where the interchange of maximization and minimization follows because the function is affine in both variables. Now, focus on the inner minimization. For any \(i \in [I], m \in [M]\), let \(Z_{im} = \inf_{z \in Z} \{a_i x^\top z + \gamma_{im}^\top z\}\). By strong duality, we have:

\[
Z_{im} = \sup \eta_{im} h \\
\text{s.t. } C^\top \eta_{im} = a_i x + \gamma_{im} \\
\eta_{im} \leq 0
\]
Finally, by above reformulations, the minimization problem (27) can be written as:

$$\max \frac{1}{M} \sum_{m \in [M]} y_m$$

s.t. $$y_m \leq a_i x^T \tilde{z}_m + b_i - \eta_{im}^T C \tilde{z}_m + \eta_{im}^T h_i, \ \forall m \in [M], \ i \in [I]$$

$$k \geq \|C^T \eta_{im} - a_i x\|_1, \ \forall m \in [M], \ i \in [I]$$

$$\eta_{im} \leq 0, \ \forall m \in [M], \ i \in [I].$$

By substituting in the robustness counterpart, the final robustness optimization model becomes:

$$\min k$$

s.t. $$\frac{1}{M} \sum_{m \in [M]} y_m \geq \tau$$

$$y_m \leq a_i x^T \tilde{z}_m + b_i - \eta_{im}^T C \tilde{z}_m + \eta_{im}^T h_i, \ \forall m \in [M], \ i \in [I]$$

$$k \geq \|C^T \eta_{im} - a_i x\|_1, \ \forall m \in [M], \ i \in [I]$$

$$\eta_{im} \leq 0, \ \forall m \in [M], \ i \in [I]$$

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

which gives the final formulation in Theorem 3. \hfill \square

Proof of Theorem 4. The robustness constraint in Problem (20) can be written as:

$$\mathbb{E}_Q \left[ -\tau + \frac{w}{\epsilon}(-x^T \tilde{z} - w)^+ - k\|\tilde{z} - \tilde{u}\|_1 \right] \leq 0, \ \forall Q \in \mathcal{Q}.$$ 

Now, notice that

$$w + \frac{1}{\epsilon}(-x^T z - w)^+ = \max \{(\epsilon - 1)w/\epsilon - x^T z/\epsilon, w\} = -u(x^T z),$$

where $u$ is piecewise affine utility function, $u(v) = \min_{i \in [2]} \{a_i v + b_i\}$, where $a_1 = 1/\epsilon, a_2 = 0$ and $b_1 = -(\epsilon - 1)w/\epsilon, b_2 = -w$. Therefore, the robustness constraint can be written as:

$$\mathbb{E}_Q \left[ -\tau - u(x^T \tilde{z}) - k\|\tilde{z} - \tilde{u}\|_1 \right] \leq 0, \ \forall Q \in \mathcal{Q},$$

which has the same structure as the robustness constraint in Problem (18). By Theorem 3, the final reformulation for the robustness optimization problem (20) becomes:

$$\min k$$

s.t. $$\frac{1}{M} \sum_{m \in [M]} s_m \geq -\tau$$

$$s_m \leq a_i x^T \tilde{z}_m + b_i \ \forall m \in [M], \ i \in [2]$$

$$k \geq -a_i x_n \ \forall n \in [N], \ i \in [2]$$

$$k \geq a_i x_n \ \forall n \in [N], \ i \in [2]$$

$$1^T x = 1.$$
By a change of variable, \( y_m = -s_m \), and notice that
\[
\max_{i \in [2]} \{-a_i x^\top z_m - b_i\} = \omega + \frac{1}{\epsilon} (-x^\top \hat{z}_m - \omega)^+,
\]
then the above problem can be equivalently written as:
\[
\begin{align*}
\min & \quad k \\
\text{s.t.} & \quad \omega + \frac{1}{\epsilon} \mathbb{E}_{\hat{P}} [(-x^\top \hat{z} - \omega)^+] \leq \tau \\
& \quad k \geq -x_n/\epsilon & \forall n \in [N], \\
& \quad k \geq x_n/\epsilon & \forall n \in [N], \\
& \quad 1^\top x = 1, w \in \mathbb{R},
\end{align*}
\]
then, following the definition of CVaR and let \( \kappa = \epsilon k \), the above problem becomes:
\[
\begin{align*}
\min & \quad \kappa/\epsilon \\
\text{s.t.} & \quad \text{CVaR}_{\hat{P}, \epsilon}(x^\top \hat{z}) \leq \tau \\
& \quad \kappa \geq \|x\|_{\infty} \\
& \quad 1^\top x = 1.
\end{align*}
\]
Since \( \epsilon \) is a constant and the objective function \( \kappa/\epsilon \) is monotone in \( \kappa \), the optimal solution to the above problem is the same as the optimal solution to Problem (21).
\[\square\]

B. Distributionally Robust Baseline Optimization Problem

In the most general framework, we consider a distributionally robust baseline optimization problem, which is associated with a random variable \( \hat{z} \) with distribution \( \hat{P} \in \mathcal{B} \), \( \hat{z} \sim \hat{P} \) being ambiguous over the ambiguity set \( \mathcal{B} \) as follows:
\[
Z_0 = \max \quad c^\top x \\
\text{s.t.} \quad \mathbb{E}_{\hat{P}} [f(x, \hat{z})] \leq 0, \forall \hat{P} \in \mathcal{B} \tag{28}
\]
\[
x \in \mathcal{X}.
\]

The above distributionally robust problem is usually constructed in a data-driven way by using moment information and empirical distribution from historical data (see, e.g., Gao et al. 2017, Zhao and Jiang 2018). Since the baseline problem is already a distributionally robust optimization problem, we present the generalized robustness optimization as follows
\[
\begin{align*}
\min & \quad k \\
\text{s.t.} & \quad \mathbb{E}_{\hat{P}} [f(x, \hat{z})] \leq k \Delta(\mathcal{P}, \hat{P}), \forall \hat{P} \in \mathcal{P}_0(\mathcal{Z}), \hat{P} \in \mathcal{B} \\
& \quad c^\top x \geq \tau \\
& \quad x \in \mathcal{X}, k \geq 0. \tag{29}
\end{align*}
\]
Observe that since \( \Delta(\hat{\mathbb{P}}, \tilde{\mathbb{P}}) = 0 \), any feasible solution to the robustness optimization problem (29) would also be feasible in the baseline distributionally robust optimization problem (28). In addition, in contrast to Problem (28), the adversarial impact is accounted for in Problem (29) for distributions outside the baseline ambiguity set \( \mathcal{B} \). Most of the results in the main text can directly extend to robustness optimization with ambiguous baseline uncertainty. For an illustration, we provide the following tractable example to show how Proposition 5 can be applied.

**Distributional robustness optimization with Wasserstein metric**

Now, we consider the distributionally robust problem (28) as the baseline problem, recall that the robustness optimization model is given by Problem (29). By Proposition 5, we can write Problem (29) as:

\[
\begin{align*}
\min k \\
\text{s.t.} \quad \sup_{Q \in \mathcal{Q}} \mathbb{E}_Q[f(x, \tilde{z}) - kd(\tilde{z}, \tilde{u})] & \leq 0 \\
& c^\top x \geq \tau \\
x & \in \mathcal{X},
\end{align*}
\]

where

\[
\mathcal{Q} = \left\{ Q \in \mathcal{P}_0(\mathcal{Z} \times \mathcal{Z}) \mid (\tilde{z}, \tilde{u}) \sim Q, \begin{array}{l}
\tilde{u} \sim \tilde{\mathbb{P}}, \text{ for some } \tilde{\mathbb{P}} \in \mathcal{B}
\end{array} \right\}.
\]

Suppose the baseline ambiguity set, \( \mathcal{B} \), is given by a moment-based ambiguity set defined as follows:

\[
\mathcal{B} = \left\{ \tilde{\mathbb{P}} \in \mathcal{P}_0(\mathcal{Z}) \mid \begin{array}{l}
\tilde{z} \sim \tilde{\mathbb{P}}, \\
\mathbb{E}_\tilde{\mathbb{P}}[g_i(\tilde{z})] \leq \mu_i, \forall i \in [I] \\
\tilde{\mathbb{P}}[\tilde{z} \in \mathcal{Z}] = 1
\end{array} \right\},
\]

where \( g_i(\tilde{z}) \), for \( i \in [I] \), represent some generalized moments. This baseline ambiguity set is general enough to capture a large class of distributions (Chen et al. 2019). Then, the joint ambiguity set, \( \mathcal{Q} \), can be written compactly as:

\[
\mathcal{Q} = \left\{ Q \in \mathcal{P}_0(\mathcal{Z} \times \mathcal{Z}) \mid (\tilde{z}, \tilde{u}) \sim Q, \begin{array}{l}
\mathbb{E}_Q[g_i(\tilde{u})] \leq \mu_i, \forall i \in [I] \\
Q[\tilde{z} \in \mathcal{Z}, \tilde{u} \in \mathcal{Z}] = 1
\end{array} \right\}.
\]

By further lifting the ambiguity set (see, e.g., Wiesemann et al. 2014, Bertsimas et al. 2019), one can get the following reformulation:

\[
\begin{align*}
\min k \\
\text{s.t.} \quad \sup_{Q \in \mathcal{Q}'} \mathbb{E}_Q[\tilde{v} - k\tilde{r}] & \leq 0 \\
& c^\top x \geq \tau \\
x & \in \mathcal{X},
\end{align*}
\]

(30)
where the lifted joint ambiguity set is
\[
Q' = \left\{ Q \in \mathcal{P}_0(\mathcal{Z} \times \mathcal{Z} \times \mathbb{R}^I \times \mathbb{R}) \mid \begin{array}{l}
(\hat{z}, \hat{u}, \hat{m}, \hat{v}, \hat{r}) \sim Q \\
E_Q[\bar{m}_i] = \mu_i, \ \forall i \in [I] \\
Q[(\hat{z}, \hat{u}, \hat{m}, \hat{v}, \hat{r})] = 1
\end{array} \right\},
\]
and the lifted support set is defined as
\[
\bar{Z} = \{ (z, u, m, v, r) \in \mathcal{Z} \times \mathcal{Z} \times \mathbb{R}^I \times \mathbb{R} \times \mathbb{R} \mid m \geq g(u), v \geq f(x, z), r \geq d(z, u) \}.
\]