A. Proofs

Proof of Theorem 1. Let $\boldsymbol{\mu}=(\boldsymbol{\mu}_k)_{k\in[K]}$ and $\mathcal{Q}=\{\boldsymbol{\mu}\mid\boldsymbol{\mu}_k\in\mathcal{Q}_k\ \forall k\in[K]\}$. We can re-express

$$\lambda^*=\sup_{\mathbb{P}\in\mathcal{F}}\mathbb{E}_{\mathbb{P}}\left[\boldsymbol{r}^\top(\tilde{s})\boldsymbol{G}_m(\tilde{s})\tilde{\boldsymbol{z}}+h_m(\tilde{s})\right]$$

by $\lambda^*=\sup_{(\boldsymbol{p},\boldsymbol{\mu})\in\mathcal{P}\times\mathcal{Q}}\lambda(\boldsymbol{p},\boldsymbol{\mu})$, where given $(\boldsymbol{p},\boldsymbol{\mu})\in\mathcal{P}\times\mathcal{Q}$, we define an ambiguity set

$$\mathcal{F}(\boldsymbol{p},\boldsymbol{\mu})=\left\{\mathbb{P}\in\mathcal{P}_0\left(\mathbb{R}^{I_z}\times[S]\right)\ \left|\ \begin{array}{ll}(\tilde{\boldsymbol{z}},\tilde{s})\sim\mathbb{P} & \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}\mid\tilde{s}\in\mathcal{E}_k]=\boldsymbol{\mu}_k & \forall k\in[K]\\ \mathbb{P}[\tilde{\boldsymbol{z}}\in\mathcal{Z}_s\mid\tilde{s}=s]=1 & \forall s\in[S]\\ \mathbb{P}[\tilde{s}=s]=p_s & \forall s\in[S]\end{array}\right.\right\}$$

and correspondingly the worst-case expectation

$$\lambda(\boldsymbol{p},\boldsymbol{\mu})=\sup_{\mathbb{P}\in\mathcal{F}(\boldsymbol{p},\boldsymbol{\mu})}\mathbb{E}_{\mathbb{P}}\left[\boldsymbol{r}^\top(\tilde{s})\boldsymbol{G}_m(\tilde{s})\tilde{\boldsymbol{z}}+h_m(\tilde{s})\right].$$

Using the law of total probability, we can construct the joint distribution $\mathbb{P}$ of $(\tilde{\boldsymbol{z}},\tilde{s})$ from the marginal distribution $\hat{\mathbb{P}}$ of $\tilde{s}$ supported on $[S]$ and the conditional distributions $\mathbb{P}_s$ of $\tilde{\boldsymbol{z}}$ given $\tilde{s}=s$, $s\in[S]$. In this way, we can reformulate $\lambda(\boldsymbol{p},\boldsymbol{\mu})$ as

$$\begin{aligned}\lambda(\boldsymbol{p},\boldsymbol{\mu})=\sup\ & \sum_{s\in[S]}p_s\mathbb{E}_{\mathbb{P}_s}\left[\boldsymbol{r}^\top(\tilde{s})\boldsymbol{G}_m(\tilde{s})\tilde{\boldsymbol{z}}+h_m(\tilde{s})\right]\\ \text{s.t.}\ & \sum_{s\in\mathcal{E}_k}p_s\mathbb{E}_{\mathbb{P}_s}[\tilde{\boldsymbol{z}}]=q_k\boldsymbol{\mu}_k & \forall k\in[K]\\ & \mathbb{P}_s[\tilde{\boldsymbol{z}}\in\mathcal{Z}_s]=1 & \forall s\in[S]\end{aligned}$$

with $q_k=\sum_{s\in\mathcal{E}_k}p_s$, $k\in[K]$. We can express the dual of $\lambda(\boldsymbol{p},\boldsymbol{\mu})$ as

$$\begin{aligned}\lambda_1(\boldsymbol{p},\boldsymbol{\mu})=\ \inf\ & \sum_{s\in[S]}\alpha_s+\sum_{k\in[K]}q_k\boldsymbol{\beta}_k^\top\boldsymbol{\mu}_k\\ \text{s.t.}\ & \alpha_s+p_s\sum_{k\in\mathcal{K}_s}\boldsymbol{\beta}_k^\top\boldsymbol{z}\geq p_s(\boldsymbol{r}^\top(s)\boldsymbol{G}_m(s)\boldsymbol{z}+h_m(s)) & \forall\boldsymbol{z}\in\mathcal{Z}_s,\ s\in[S]\\ & \boldsymbol{\alpha}\in\mathbb{R}^S,\ \boldsymbol{\beta}_k\in\mathbb{R}^{I_z} & \forall k\in[K]\\ =\ \inf\ & \boldsymbol{\alpha}^\top\boldsymbol{p}+\sum_{k\in[K]}q_k\boldsymbol{\beta}_k^\top\boldsymbol{\mu}_k\\ \text{s.t.}\ & \alpha_s+\sum_{k\in\mathcal{K}_s}\boldsymbol{\beta}_k^\top\boldsymbol{z}\geq\boldsymbol{r}^\top(s)\boldsymbol{G}_m(s)\boldsymbol{z}+h_m(s) & \forall\boldsymbol{z}\in\mathcal{Z}_s,\ s\in[S]\\ & \boldsymbol{\alpha}\in\mathbb{R}^S,\ \boldsymbol{\beta}_k\in\mathbb{R}^{I_z} & \forall k\in[K],\end{aligned}$$

where the second equality follows from for all $s\in[S]$, first changing variable from α_s to $p_s\alpha_s$ and then dividing both sides of the constraint by p_s, which is allowed since $\boldsymbol{p}\in\mathcal{P}$ is strictly positive.

By weak duality, $\lambda(\boldsymbol{p},\boldsymbol{\mu})\leq\lambda_1(\boldsymbol{p},\boldsymbol{\mu})$. By the general min-max theorem, we further observe that

$$\lambda_1^*=\sup_{(\boldsymbol{p},\boldsymbol{\mu})\in\mathcal{P}\times\mathcal{Q}}\lambda_1(\boldsymbol{p},\boldsymbol{\mu})\leq\lambda_2^*,$$

where

$$
\begin{aligned}
\lambda_2^* = \inf\ & \gamma \\
\text{s.t. }\ & \gamma \geq \boldsymbol{\alpha}^\top \boldsymbol{p} + \sum_{k\in[K]} q_k \boldsymbol{\beta}_k^\top \boldsymbol{\mu}_k && \forall \boldsymbol{p}\in\mathcal{P},\ \boldsymbol{\mu}_k \in \mathcal{Q}_k,\ k\in[K] \\
& \alpha_s + \sum_{k\in\mathcal{K}_s} \boldsymbol{\beta}_k^\top \boldsymbol{z} \geq \boldsymbol{r}^\top(s)\boldsymbol{G}_m(s)\boldsymbol{z} + h_m(s) && \forall \boldsymbol{z}\in\mathcal{Z}_s,\ s\in[S] \\
& \gamma\in\mathbb{R},\ \boldsymbol{\alpha}\in\mathbb{R}^S,\ \boldsymbol{\beta}_k\in\mathbb{R}^{I_z} && \forall k\in[K].
\end{aligned}
\tag{17}
$$

Due to the presence of products of uncertain variables (*e.g.*, $q_k\boldsymbol{\mu}_k$), problem (17) is nonconvex. Since $\boldsymbol{p} > 0$ (and hence $q_k > 0$), an equivalent convex representation can be obtained by changing variables in problem (17) from $q_k\boldsymbol{\mu}_k$ to $\boldsymbol{\mu}_k$ for all $k\in[K]$, which turns out to be problem (4).

Assuming the conic representation of the following system

$$
\begin{cases}
\dfrac{\sum_{s\in\mathcal{E}_k} \boldsymbol{\xi}_s}{\sum_{s\in\mathcal{E}_k} \tau_s} \in \mathcal{Q}_k & \forall k\in[K] \\
\dfrac{\boldsymbol{\xi}_s}{\tau_s} \in \mathcal{Z}_s & \forall s\in[S] \\
\boldsymbol{\tau}\in\mathcal{P}
\end{cases}
\tag{18}
$$

satisfies the Slater's condition (see Theorem 1.4.2 in Ben-Tal and Nemirovski 2001), one can establish strong duality, *i.e.*, $\lambda^* = \lambda_1^* = \lambda_2^*$ and show that problem (4) is solvable (Bertsimas et al. 2019b, Theorem 1). □

Proof of Theorem 2. We consider an ambiguity set without the auxiliary random variable $\tilde{v}$

$$
\bar{\mathcal{G}}_W(\theta) = \left\{ \mathbb{P}\in\mathcal{P}_0\left(\mathbb{R}^{I_u}\times[S]\right) \;\middle|\; \begin{array}{ll} (\tilde{\boldsymbol{u}},\tilde{s})\sim\mathbb{P} & \\ \mathbb{E}_{\mathbb{P}}\left[\rho(\tilde{\boldsymbol{u}},\hat{\boldsymbol{u}}_{\tilde{s}}) \mid \tilde{s}\in[S]\right] \leq \theta & \\ \mathbb{P}\left[\tilde{\boldsymbol{u}}\in\mathcal{U} \mid \tilde{s}=s\right] = 1 & \forall s\in[S] \\ \mathbb{P}\left[\tilde{s}=s\right] = \frac{1}{S} & \forall s\in[S] \end{array} \right\}.
\tag{19}
$$

Since this ambiguity set satisfies $\Pi_{(\tilde{\boldsymbol{u}},\tilde{s})}\mathcal{F}_W(\theta) = \bar{\mathcal{G}}_W(\theta)$ for all $\theta\geq 0$, thus it is sufficient to prove $\Pi_{\tilde{\boldsymbol{u}}}\bar{\mathcal{G}}_W(\theta) = \mathcal{G}_W(\theta)$ for all $\theta \geq 0$.

To this end, we first prove $\mathcal{G}_W(\theta)\subseteq\Pi_{\tilde{\boldsymbol{u}}}\bar{\mathcal{G}}_W(\theta)$. Consider $\tilde{\boldsymbol{u}}\sim\mathbb{P}$ for some $\mathbb{P}\in\mathcal{G}_W(\theta)$. By definition of the Wasserstein ambiguity set $\mathcal{G}_W(\theta)$, there exists a joint distribution $\mathbb{Q}\in\mathcal{P}(\mathbb{P},\hat{\mathbb{P}})$ of $(\tilde{\boldsymbol{u}},\tilde{\boldsymbol{u}}^\dagger)$ such that $\Pi_{\tilde{\boldsymbol{u}}}\mathbb{Q}=\mathbb{P}$, $\Pi_{\tilde{\boldsymbol{u}}^\dagger}\mathbb{Q}=\hat{\mathbb{P}}$, and $\mathbb{E}_{\mathbb{Q}}\left[\rho(\tilde{\boldsymbol{u}},\tilde{\boldsymbol{u}}^\dagger)\right]\leq\theta$. Since we can construct $\mathbb{Q}$ from the marginal distribution $\hat{\mathbb{P}}$ of $\tilde{\boldsymbol{u}}^\dagger$ supported on $\{\hat{\boldsymbol{u}}_1,\ldots,\hat{\boldsymbol{u}}_S\}$ and the conditional distributions $\mathbb{P}_s$ of $\tilde{\boldsymbol{u}}$, given the realization of $\tilde{\boldsymbol{u}}^\dagger$ is $\hat{\boldsymbol{u}}_s$, $s\in[S]$, we have $(\tilde{\boldsymbol{u}},\tilde{\boldsymbol{u}}^\dagger)\sim\frac{1}{S}\sum_{s\in[S]}\mathbb{P}_s\otimes\delta_{\hat{\boldsymbol{u}}_s}$. We can then construct a distribution $\mathbb{Q}'\in\mathcal{P}_0(\mathbb{R}^{I_u}\times[S])$ for the random variable $(\tilde{\boldsymbol{u}},\tilde{s})\sim\mathbb{Q}'$ via $\mathbb{Q}'=\frac{1}{S}\sum_{s\in[S]}\mathbb{P}_s\otimes\delta_s$. Observe that $\mathbb{Q}'\in\bar{\mathcal{G}}_W(\theta)$, hence $\mathcal{G}_W(\theta)\subseteq\Pi_{\tilde{\boldsymbol{u}}}\bar{\mathcal{G}}_W(\theta)$.

To prove $\Pi_{\tilde{\boldsymbol{u}}}\bar{\mathcal{G}}_W(\theta) \subseteq \mathcal{G}_W(\theta)$, we fix any $\mathbb{P} \in \bar{\mathcal{G}}_W(\theta)$ and we write its projection over $\tilde{\boldsymbol{u}}$ as $\Pi_{\tilde{\boldsymbol{u}}}\mathbb{P} = \frac{1}{S}\sum_{s\in[S]} \mathbb{P}_s$, where $\mathbb{P}_s$ is the conditional distribution of $\tilde{\boldsymbol{z}}$ given the outcome of the random scenario is s. We can then construct a joint distribution $\mathbb{Q} = \frac{1}{S}\sum_{s\in[S]} \mathbb{P}_s \otimes \delta_{\hat{\boldsymbol{u}}_s}$ of $(\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{u}}^\dagger)$ that satisfies

$$\mathbb{E}_{\mathbb{Q}}\left[\rho(\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{u}}^\dagger)\right] = \frac{1}{S}\sum_{s\in[S]} \mathbb{E}_{\mathbb{P}_s}[\rho(\tilde{\boldsymbol{u}}, \hat{\boldsymbol{u}}_s)] = \mathbb{E}_{\mathbb{P}}\left[\rho(\tilde{\boldsymbol{u}}, \hat{\boldsymbol{u}}_{\tilde{s}}) \mid \tilde{s} \in [S]\right] \le \theta.$$

Hence, $\Pi_{\tilde{\boldsymbol{u}}}\mathbb{P} \in \mathcal{G}_W(\theta)$, which gives $\Pi_{\tilde{\boldsymbol{u}}}\bar{\mathcal{G}}_W(\theta) \subseteq \mathcal{G}_W(\theta)$ to conclude $\mathcal{G}_W(\theta) = \Pi_{\tilde{\boldsymbol{u}}}\bar{\mathcal{G}}_W(\theta)$. □

Proof of Theorem 3. Previous derivations in the proof of Theorem 1 implies that *(i)* the worst-case expectation (9) is equivalent to the following problem

$$\begin{array}{lll} \inf & \gamma & \\ \text{s.t.} & \gamma \ge \boldsymbol{\alpha}^\top \boldsymbol{p} + \sum_{k\in[K]} \boldsymbol{\beta}_k^\top \boldsymbol{\mu}_k & \forall \boldsymbol{p} \in \mathcal{P},\ \frac{\boldsymbol{\mu}_k}{\sum_{s\in\mathcal{E}_k} p_s} \in \mathcal{Q}_k,\ k \in [K] \\ & \alpha_s + \sum_{k\in\mathcal{K}_s} \boldsymbol{\beta}_k^\top \boldsymbol{z} \ge \boldsymbol{r}^\top(s)\boldsymbol{G}_\ell(s)\boldsymbol{z} + h_\ell(s) & \forall \boldsymbol{z} \in \mathcal{Z}_s,\ s \in [S],\ \ell \in [L] \\ & \gamma \in \mathbb{R},\ \boldsymbol{\alpha} \in \mathbb{R}^S,\ \boldsymbol{\beta}_k \in \mathbb{R}^{I_z} & \forall k \in [K]; \end{array} \tag{20}$$

and *(ii)* problem (10) is equivalent to

$$\begin{array}{lll} \inf & \gamma & \\ \text{s.t.} & \gamma \ge \boldsymbol{\alpha}^\top \boldsymbol{p} + \sum_{k\in[K]} \boldsymbol{\beta}_k^\top \boldsymbol{\mu}_k & \forall \boldsymbol{p} \in \mathcal{P},\ \frac{\boldsymbol{\mu}_k}{\sum_{s\in\mathcal{E}_k} p_s} \in \mathcal{Q}_k,\ k \in [K] \\ & \alpha_s + \sum_{k\in\mathcal{K}_s} \boldsymbol{\beta}_k^\top \boldsymbol{z} \ge y(s, \boldsymbol{z}) & \forall \boldsymbol{z} \in \mathcal{Z}_s,\ s \in [S] \\ & y(s,\boldsymbol{z}) \ge \boldsymbol{r}^\top(s)\boldsymbol{G}_\ell(s)\boldsymbol{z} + h_\ell(s) & \forall \boldsymbol{z} \in \mathcal{Z}_s,\ s \in [S],\ \ell \in [L] \\ & y \in \bar{\mathcal{A}}(\bar{\mathcal{C}}, [I_z]) & \\ & \gamma \in \mathbb{R},\ \boldsymbol{\alpha} \in \mathbb{R}^S,\ \boldsymbol{\beta}_k \in \mathbb{R}^{I_z} & \forall k \in [K]. \end{array} \tag{21}$$

It is then sufficient to construct a feasible solution to problem (21) from a feasible solution to problem (20) such that the constructive solution yields the same objective. Indeed, given a feasible solution $(\gamma^\dagger, \boldsymbol{\alpha}^\dagger, (\boldsymbol{\beta}_k^\dagger)_{k\in[K]})$ to problem (20), we can construct such a desired solution via:

$$\gamma = \gamma^\dagger,\ \boldsymbol{\alpha} = \boldsymbol{\alpha}^\dagger,\ \boldsymbol{\beta}_k = \boldsymbol{\beta}_k^\dagger\ \forall k \in [K],\ y(s,\boldsymbol{z}) = \alpha_s^\dagger + \sum_{k\in\mathcal{K}_s} (\boldsymbol{\beta}_k^\dagger)^\top \boldsymbol{z}\ \forall s \in [S],$$

for which the recourse decision $y(\cdot,\cdot) \in \bar{\mathcal{A}}(\bar{\mathcal{C}}, [I_z])$. □

Proof of Theorem 4. Using the result of Popescu (2007), we first show that

$$\inf_{\mathbb{P}\in\mathcal{G}(\boldsymbol{w}^\top\boldsymbol{\mu}, \boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w})} \mathbb{E}_{\mathbb{P}}[U(\tilde{u})] = \sup_r \left\{ \inf_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[U(r\tilde{u} + \boldsymbol{w}^\top\boldsymbol{\mu})] \;\middle|\; r \ge \sqrt{\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}} \right\}.$$

By duality, we have

$$\begin{aligned} \inf_{\mathbb{P}\in\mathcal{G}(\boldsymbol{w}^\top\boldsymbol{\mu}, \boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w})} \mathbb{E}_{\mathbb{P}}[U(\tilde{u})] &= \inf_{\mathbb{P}\in\mathcal{G}(0, \boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w})} \mathbb{E}_{\mathbb{P}}[U(\tilde{u} + \boldsymbol{w}^\top\boldsymbol{\mu})] \\ &= \sup_{\alpha,\beta_1,\beta_2} \left\{ \alpha + \boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}\cdot\beta_2 \;\middle|\; \alpha + \beta_1 u + \beta_2 u^2 \le U(u + \boldsymbol{w}^\top\boldsymbol{\mu})\ \ \forall u \right\}. \end{aligned}$$

Note that it requires $\beta_2 \leq 0$ for the above problem to be feasible, as otherwise the constraint would be violated for some sufficiently large u. Hence, we can further rewrite this problem into

$$\begin{aligned}
&\sup_{\alpha,\beta_1,\beta_2,r} \left\{ \alpha + r^2\beta_2 \;\middle|\; \begin{array}{l} \alpha + \beta_1 u + \beta_2 u^2 \leq U(u + \boldsymbol{w}^\top \boldsymbol{\mu}) \quad \forall u \\ r \geq \sqrt{\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}} \end{array} \right\} \\
= &\sup_{r} \left\{ \inf_{\mathbb{P} \in \mathcal{G}(0,r^2)} \mathbb{E}_{\mathbb{P}}[U(\tilde{u} + \boldsymbol{w}^\top \boldsymbol{\mu})] \;\middle|\; r \geq \sqrt{\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}} \right\} \\
= &\sup_{r} \left\{ \inf_{\mathbb{P} \in \mathcal{G}(0,1)} \mathbb{E}_{\mathbb{P}}[U(r\tilde{u} + \boldsymbol{w}^\top \boldsymbol{\mu})] \;\middle|\; r \geq \sqrt{\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}} \right\} \\
= &\sup_{\alpha,\beta_1,\beta_2,r} \left\{ \alpha + \beta_2 \;\middle|\; \begin{array}{l} \alpha + \beta_1 u + \beta_2 u^2 \leq U(ru + \boldsymbol{w}^\top \boldsymbol{\mu}) \quad \forall u \\ r \geq \sqrt{\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}} \end{array} \right\} \\
= &\sup_{\alpha,\beta_1,\beta_2,r} \left\{ \alpha + \beta_2 \;\middle|\; \begin{array}{l} \alpha + \beta_1 u + \beta_2 v \leq U(ru + \boldsymbol{w}^\top \boldsymbol{\mu}) \quad \forall (u,v) : v \geq u^2 \\ \beta_2 \leq 0, \; r \geq \sqrt{\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}} \end{array} \right\} \\
= &\sup_{r} \left\{ \inf_{\mathbb{P} \in \mathcal{F}} \mathbb{E}_{\mathbb{P}}[U(r\tilde{u} + \boldsymbol{w}^\top \boldsymbol{\mu})] \;\middle|\; r \geq \sqrt{\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}} \right\}.
\end{aligned}$$

The result then follows by applying Theorem 3. □

Proof of Theorem 5. Observe that for any feasible recourse decision $\bar{\boldsymbol{r}}(\cdot)$ to problem (13), we have

$$\sup_{\mathbb{P} \in \mathcal{F}} \mathbb{E}_{\mathbb{P}}[\bar{\boldsymbol{r}}^\top(\tilde{s})\tilde{\boldsymbol{v}}] = \sup_{\mathbb{P} \in \mathcal{G}} \mathbb{E}_{\mathbb{P}}[\bar{\boldsymbol{r}}^\top(\tilde{s})\boldsymbol{\zeta}(\tilde{\boldsymbol{u}}, \tilde{s})].$$

In addition, the optimal $\bar{\boldsymbol{r}}^\star(\cdot)$ to problem (13) satisfies $\bar{r}^\star_\ell(s) = \xi_\ell(\boldsymbol{r}(s), s)$ for all $\ell \in [I_v]$ and $s \in [S]$. Therefore, our claim holds. □

B. Worst-Case Expectation of Quadratic Functions

The RSO framework can be used to provide a tight characterization of the worst-case expectation of some quadratic functions that are known in the literature and extend them to include discrete scenarios. Let $\mathbb{S}^I$ be the space of symmetric matrices in $\mathbb{R}^{I \times I}$. Given $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{S}^I$, we denote by $\boldsymbol{X} \succeq \boldsymbol{Y}$ (resp., $\boldsymbol{X} \succ \boldsymbol{Y}$) to represent $\boldsymbol{X} - \boldsymbol{Y}$ is positive semidefinite (resp., definite), and denote by $\boldsymbol{X} \bullet \boldsymbol{Y}$ as the trace inner product of $\boldsymbol{X}, \boldsymbol{Y}$. Special matrices and vectors of the appropriate dimension include $\boldsymbol{O}$, $\boldsymbol{I}$, and $\boldsymbol{0}$, which respectively correspond to the zero matrix, the identity matrix, and the zero vector.

Bi-Convex-Quadratic Function

We explore the following bi-convex-quadratic function as an extension of Ben-Tal and Nemirovski (1998) to include discrete scenarios:

$$g(\boldsymbol{r}(s), \boldsymbol{u}, s) \triangleq \boldsymbol{u}^\top \boldsymbol{A}^\top(\boldsymbol{r}(s), s)\boldsymbol{A}(\boldsymbol{r}(s), s)\boldsymbol{u} + 2\boldsymbol{u}^\top \boldsymbol{b}(\boldsymbol{r}(s), s) + c(\boldsymbol{r}(s), s),$$

where given a scenario s, $\boldsymbol{A}(\boldsymbol{r}(s),s), \boldsymbol{b}(\boldsymbol{r}(s),s), c(\boldsymbol{r}(s),s)$ are affine mappings of $\boldsymbol{r}(s)$. The event-wise ambiguity set is given by

$$\mathcal{G} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^{I_u} \times [S]\right) \;\middle|\; \begin{array}{ll} (\tilde{\boldsymbol{u}}, \tilde{s}) \sim \mathbb{P} & \\ \mathbb{E}_{\mathbb{P}}\left[\binom{1}{\tilde{\boldsymbol{u}}}\binom{1}{\tilde{\boldsymbol{u}}}^\top \;\middle|\; \tilde{s} \in \mathcal{E}_k\right] \in \mathcal{Q}_k & \forall k \in [K] \\ \mathbb{P}\left[\binom{1}{\tilde{\boldsymbol{u}}}\binom{1}{\tilde{\boldsymbol{u}}}^\top \in \mathcal{U}_s \;\middle|\; \tilde{s} = s\right] = 1 & \forall s \in [S] \\ \mathbb{P}[\tilde{s} = s] = p_s & \forall s \in [S] \\ \text{for some } \boldsymbol{p} \in \mathcal{P} & \end{array} \right\}.$$

The support set $\mathcal{U}_s$ is general enough to capture the ubiquitous uncertainty set $\{\boldsymbol{u} \mid \boldsymbol{u}^\top \boldsymbol{\Lambda}_s \boldsymbol{u} \leq 1\}$ parameterized by some $\boldsymbol{\Lambda}_s \succ \boldsymbol{0}$, for which we only need to define

$$\mathcal{U}_s = \left\{ \boldsymbol{U} \in \mathbb{S}^{I_u+1} \;\middle|\; \boldsymbol{U} \bullet \begin{pmatrix} -1 & \boldsymbol{0}^\top \\ \boldsymbol{0} & \boldsymbol{\Lambda}_s \end{pmatrix} \leq 0 \right\}. \tag{22}$$

THEOREM 6. *The worst-case expectation*

$$\sup_{\mathbb{P} \in \mathcal{G}} \mathbb{E}_{\mathbb{P}}\left[g(\boldsymbol{r}(\tilde{s}), \tilde{\boldsymbol{u}}, \tilde{s})\right] \tag{23}$$

is bounded from above by

$$\begin{array}{lll} \min & \sup_{\mathbb{P} \in \mathcal{F}} \mathbb{E}_{\mathbb{P}}[\boldsymbol{R}(\tilde{s}) \bullet \tilde{\boldsymbol{Z}}] & \\ \text{s.t.} & \begin{pmatrix} \boldsymbol{R}(s) - \begin{pmatrix} 1 & \boldsymbol{b}^\top(\boldsymbol{r}(s),s) \\ \boldsymbol{b}(\boldsymbol{r}(s),s) & \boldsymbol{O} \end{pmatrix} & \begin{pmatrix} 0 & \boldsymbol{0}^\top \\ \boldsymbol{0} & \boldsymbol{A}^\top(\boldsymbol{r}(s),s) \end{pmatrix} \\ \begin{pmatrix} 0 & \boldsymbol{0}^\top \\ \boldsymbol{0} & \boldsymbol{A}(\boldsymbol{r}(s),s) \end{pmatrix} & \boldsymbol{I} \end{pmatrix} \succeq \boldsymbol{O} & \forall s \in [S] \\ & \boldsymbol{R}(s) \in \mathbb{S}^{I_u+1} & \forall s \in [S], \end{array} \tag{24}$$

where the lifted event-wise ambiguity set

$$\mathcal{F} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{S}^{I_u+1} \times [S]\right) \;\middle|\; \begin{array}{ll} (\tilde{\boldsymbol{Z}}, \tilde{s}) \sim \mathbb{P} & \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{Z}} \mid \tilde{s} \in \mathcal{E}_k] \in \mathcal{Q}_k & \forall k \in [K] \\ \mathbb{P}[\tilde{\boldsymbol{Z}} \in \mathcal{Z}_s \mid \tilde{s} = s] = 1 & \forall s \in [S] \\ \mathbb{P}[\tilde{s} = s] = p_s & \forall s \in [S] \\ \textit{for some } \boldsymbol{p} \in \mathcal{P} & \end{array} \right\}$$

takes lifted support sets $\mathcal{Z}_s = \{\boldsymbol{Z} \in \mathcal{U}_s \mid \boldsymbol{Z} \succeq \boldsymbol{O},\ [\boldsymbol{Z}]_{1,1} = 1\}$, $s \in [S]$. *Moreover, the bound is tight for ellipsoidal support sets defined in (22).*

Proof of Theorem 6. We note that

$$g(\boldsymbol{r}(s), \boldsymbol{u}, s) = \begin{pmatrix} 1 & \boldsymbol{b}^\top(\boldsymbol{r}(s),s) \\ \boldsymbol{b}(\boldsymbol{r}(s),s) & \boldsymbol{A}^\top(\boldsymbol{r}(s),s)\boldsymbol{A}(\boldsymbol{r}(s),s) \end{pmatrix} \bullet \left(\begin{pmatrix} 1 \\ \boldsymbol{u} \end{pmatrix} \begin{pmatrix} 1 \\ \boldsymbol{u} \end{pmatrix}^\top \right).$$

By Schur complement, each positive semidefinite constraint of problem (24) is equivalent to

$$\boldsymbol{R}(s) \succeq \begin{pmatrix} 1 & \boldsymbol{b}^\top(\boldsymbol{r}(s),s) \\ \boldsymbol{b}(\boldsymbol{r}(s),s) & \boldsymbol{A}^\top(\boldsymbol{r}(s),s)\boldsymbol{A}(\boldsymbol{r}(s),s) \end{pmatrix}.$$

Since $\boldsymbol{Z} \in \mathcal{Z}_s$ is positive semidefinite, an optimal $\boldsymbol{R}(s)$ would be

$$\boldsymbol{R}(s) = \begin{pmatrix} 1 & \boldsymbol{b}^\top(\boldsymbol{r}(s),s) \\ \boldsymbol{b}(\boldsymbol{r}(s),s) & \boldsymbol{A}^\top(\boldsymbol{r}(s),s)\boldsymbol{A}(\boldsymbol{r}(s),s) \end{pmatrix}.$$

Observe that the ambiguity set $\mathcal{F}$ coincides with $\mathcal{G}$ if every support set $\mathcal{Z}_s$ is replaced by $\bar{\mathcal{Z}}_s = \{\boldsymbol{Z} \in \mathcal{U}_s \mid \boldsymbol{Z} \succeq \boldsymbol{O},\ [\boldsymbol{Z}]_{1,1} = 1,\ \mathrm{rank}(\boldsymbol{Z}) = 1\}$, which however, would lead to a harder problem to solve due to the rank constraint. Since $\bar{\mathcal{Z}}_s \subseteq \mathcal{Z}_s$, we obtain the conservative upper bound.

We next show that the bound is tight for ellipsoidal uncertainty sets defined in (22). After using Theorem 1 to reformulate problem (23), we need to deal with the following robust counterpart.

$$\alpha_s \geq \boldsymbol{\Phi}_s \bullet \left(\begin{pmatrix} 1 \\ \boldsymbol{u} \end{pmatrix} \begin{pmatrix} 1 \\ \boldsymbol{u} \end{pmatrix}^\top \right) \quad \forall \boldsymbol{u}^\top \boldsymbol{\Lambda}_s \boldsymbol{u} \leq 1$$

for some $\alpha_s \in \mathbb{R}, \boldsymbol{\Phi}_s \in \mathbb{S}^{I_u+1}$, which by S-lemma, is equivalent to

$$\begin{pmatrix} \alpha_s & \boldsymbol{0}^\top \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} + \delta_s \begin{pmatrix} -1 & \boldsymbol{0}^\top \\ \boldsymbol{0} & \boldsymbol{\Lambda}_s \end{pmatrix} \succeq \boldsymbol{\Phi}_s,$$

for some $\delta_s \geq 0$. On the other hand, the robust counterpart in the reformulation of problem (24)

$$\alpha_s \geq \boldsymbol{\Phi}_s \bullet \boldsymbol{Z} \quad \forall \boldsymbol{Z} \bullet \begin{pmatrix} -1 & \boldsymbol{0}^\top \\ \boldsymbol{0} & \boldsymbol{\Lambda}_s \end{pmatrix} \leq 0,\ [\boldsymbol{Z}]_{1,1} = 1,\ \boldsymbol{Z} \succeq \boldsymbol{O}$$

is equivalent to

$$\begin{pmatrix} \tau_s & \boldsymbol{0}^\top \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} + \delta_s \begin{pmatrix} -1 & \boldsymbol{0}^\top \\ \boldsymbol{0} & \boldsymbol{\Lambda}_s \end{pmatrix} \succeq \boldsymbol{\Phi}_s,$$

for some $\tau_s \leq \alpha_s$ and $\delta_s \geq 0$, for which we can replace τ_s with α_s without affecting its feasibility. This establishes the desired tight bound for ellipsoidal uncertainty sets. □

Bi-Conic-Quadratic Function

We can also extend the bi-conic-quadratic function considered in Ben-Tal and Nemirovski (1998) to include discrete scenarios as follows:

$$h(\boldsymbol{r}(s), \boldsymbol{u}, s) \triangleq \|\boldsymbol{A}(\boldsymbol{r}(s),s)\boldsymbol{u} + \boldsymbol{b}(\boldsymbol{r}(s),s)\|_2,$$

where given s, $\boldsymbol{A}(\boldsymbol{r}(s),s), \boldsymbol{b}(\boldsymbol{r}(s),s)$ are affine mappings of $\boldsymbol{r}(s)$. The event-wise ambiguity set takes

$$\mathcal{G} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^{I_u} \times [S]\right) \;\middle|\; \begin{array}{ll} (\tilde{\boldsymbol{u}}, \tilde{s}) \sim \mathbb{P} & \\ \mathbb{P}\left[\binom{1}{\tilde{\boldsymbol{u}}}\binom{1}{\tilde{\boldsymbol{u}}}^\top \in \mathcal{U}_s \;\middle|\; \tilde{s} = s\right] = 1 & \forall s \in [S] \\ \mathbb{P}[\tilde{s} = s] = p_s & \forall s \in [S] \\ \text{for some } \boldsymbol{p} \in \mathcal{P} & \end{array} \right\}.$$

Theorem 7. *The worst-case expectation*

$$\sup_{\mathbb{P}\in\mathcal{G}} \mathbb{E}_{\mathbb{P}}\left[h(\boldsymbol{r}(\tilde{s}),\tilde{\boldsymbol{u}},\tilde{s})\right]$$

is bounded from above by

$$\begin{array}{lll} \min & \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[x(\tilde{s})] & \\ \text{s.t.} & x(s) \geq \boldsymbol{R}(s)\bullet\boldsymbol{Z} & \forall \boldsymbol{Z}\in\mathcal{Z}_s,\ s\in[S] \\ & \begin{pmatrix} \boldsymbol{R}(s) & \begin{pmatrix} \boldsymbol{b}^\top(\boldsymbol{r}(s),s) \\ \boldsymbol{A}^\top(\boldsymbol{r}(s),s)\end{pmatrix} \\ (\boldsymbol{b}(\boldsymbol{r}(s),s)\ \ \boldsymbol{A}(\boldsymbol{r}(s),s)) & x(s)\boldsymbol{I}\end{pmatrix} \succeq \boldsymbol{O} & \forall s\in[S] \\ & \boldsymbol{R}(s)\in\mathbb{S}^{I_u+1} & \forall s\in[S] \\ & x\in\mathcal{A}(\bar{\mathcal{C}}), & \end{array} \tag{25}$$

where $\bar{\mathcal{C}} \triangleq \{\{s\}\mid s\in[S]\}$ *and where the lifted event-wise ambiguity set*

$$\mathcal{F} = \left\{ \mathbb{P}\in\mathcal{P}_0\left(\mathbb{S}^{I_u+1}\times[S]\right) \left| \begin{array}{ll} (\tilde{\boldsymbol{Z}},\tilde{s})\sim\mathbb{P} & \\ \mathbb{P}[\tilde{\boldsymbol{Z}}\in\mathcal{Z}_s\mid\tilde{s}=s]=1 & \forall s\in[S] \\ \mathbb{P}[\tilde{s}=s]=p_s & \forall s\in[S] \\ \text{for some } \boldsymbol{p}\in\mathcal{P} & \end{array}\right.\right\}$$

takes lifted support sets $\mathcal{Z}_s = \{\boldsymbol{Z}\in\mathcal{U}_s \mid \boldsymbol{Z}\succeq\boldsymbol{O},\ [\boldsymbol{Z}]_{1,1}=1\}$, $s\in[S]$*. Moreover, the bound is tight for ellipsoidal uncertainty sets defined in* (22)*.*

Proof of Theorem 7. Since the ambiguity set does not contain any expectation constraint, we can obtain a tractable reformulation by replacing $h(\boldsymbol{r}(s),\boldsymbol{u},s)$ with a recourse variable $x(s)$ and imposing the following constraint (see reformulation in Theorem 1):

$$x^2(s)\geq h^2(\boldsymbol{r}(s),\boldsymbol{u},s) \quad \forall \begin{pmatrix}1\\ \boldsymbol{u}\end{pmatrix}\begin{pmatrix}1\\ \boldsymbol{u}\end{pmatrix}^\top \in\mathcal{U}_s, s\in[S].$$

We next discuss how such a constraint can be specified in problem (25). Observe that

$$h^2(\boldsymbol{r}(s),\boldsymbol{u},s) = \left(\begin{pmatrix}\boldsymbol{b}^\top(\boldsymbol{r}(s),s)\\ \boldsymbol{A}^\top(\boldsymbol{r}(s),s)\end{pmatrix}(\boldsymbol{b}(\boldsymbol{r}(s),s)\ \ \boldsymbol{A}(\boldsymbol{r}(s),s))\right)\bullet\left(\begin{pmatrix}1\\ \boldsymbol{u}\end{pmatrix}\begin{pmatrix}1\\ \boldsymbol{u}\end{pmatrix}^\top\right).$$

By Schur complement, each positive semidefinite constraint of problem (24) is equivalent to

$$x(s)\boldsymbol{R}(s)\succeq\begin{pmatrix}\boldsymbol{b}^\top(\boldsymbol{r}(s),s)\\ \boldsymbol{A}^\top(\boldsymbol{r}(s),s)\end{pmatrix}(\boldsymbol{b}(\boldsymbol{r}(s),s)\ \ \boldsymbol{A}(\boldsymbol{r}(s),s)).$$

Since $\boldsymbol{Z}\in\mathcal{Z}_s$ is positive semidefinite and $x(s)\geq 0$, an optimal $\boldsymbol{R}(s)$ would be

$$x(s)\boldsymbol{R}(s)=\begin{pmatrix}\boldsymbol{b}^\top(\boldsymbol{r}(s),s)\\ \boldsymbol{A}^\top(\boldsymbol{r}(s),s)\end{pmatrix}(\boldsymbol{b}(\boldsymbol{r}(s),s)\ \ \boldsymbol{A}(\boldsymbol{r}(s),s)).$$

The rest of the proof follows similarly as in the proof of Theorem 6. □

Affine-Quadractic Function

As an extension of Tütüncü and Koenig (2004), we consider a saddle function that is convex quadratic with respect to the decision variable and that is affine with respect to $\boldsymbol{z}$:

$$g(\boldsymbol{r}(s), \boldsymbol{z}, s) \triangleq \boldsymbol{r}^\top(s)\boldsymbol{H}(s,\boldsymbol{z})\boldsymbol{r}(s) + \boldsymbol{r}^\top(s)\boldsymbol{G}(s)\boldsymbol{z} + h(s), \tag{26}$$

where given a scenario s, $\boldsymbol{H}(s,\boldsymbol{z})$ is an affine mapping of $\boldsymbol{z}$ and $\mathcal{Z}_s \subseteq \{\boldsymbol{z} \mid \boldsymbol{H}(\boldsymbol{z}, s) \succeq \boldsymbol{O}\}$. Introducing auxiliary variables $\boldsymbol{R}(s) \in \mathbb{S}^{I_v}$, $s \in [S]$ and using the Schur complement, the robust expectation $\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[g(\boldsymbol{r}(\tilde{s}), \tilde{\boldsymbol{z}}, \tilde{s})\right]$ is the same as

$$\begin{array}{lll}
\min & \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{R}(\tilde{s}) \bullet \boldsymbol{H}(\tilde{s},\tilde{\boldsymbol{z}}) + \boldsymbol{r}^\top(\tilde{s})\boldsymbol{G}(\tilde{s})\tilde{\boldsymbol{z}} + h(\tilde{s})\right] & \\
\text{s.t.} & \begin{pmatrix} 1 & \boldsymbol{r}^\top(s) \\ \boldsymbol{r}(s) & \boldsymbol{R}(s) \end{pmatrix} \succeq \boldsymbol{0} & \forall s \in [S] \\
& \boldsymbol{R}(s) \in \mathbb{S}^{I_v} & \forall s \in [S],
\end{array}$$

which falls within the RSO framework.

C. Representation of Wasserstein Ambiguity Sets

For $p \in [1, \infty)$, the type-p Wasserstein metric between two distributions $\mathbb{P}$ and $\hat{\mathbb{P}}$ for a given distance metric ρ is defined as

$$d_W^p(\mathbb{P}, \hat{\mathbb{P}}) \triangleq \inf_{\mathbb{Q}\in\mathcal{Q}(\mathbb{P},\hat{\mathbb{P}})} \left(\mathbb{E}_{\mathbb{Q}}\left[\rho^p(\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{u}}^\dagger)\right]\right)^{\frac{1}{p}}.$$

Correspondingly, the type-p Wasserstein ambiguity set is defined by

$$\mathcal{G}_W^p(\theta) = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathcal{U}) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{u}} \sim \mathbb{P} \\ d_W^p(\mathbb{P}, \hat{\mathbb{P}}) \leq \theta \end{array} \right\}.$$

Consider another distance metric ρ^p and the corresponding type-1 Wasserstein metric $\bar{d}_W(\mathbb{P}, \hat{\mathbb{P}})$ between $\mathbb{P}$ and $\hat{\mathbb{P}}$ which is determined by

$$\bar{d}_W(\mathbb{P}, \hat{\mathbb{P}}) \triangleq \inf_{\mathbb{Q}\in\mathcal{Q}(\mathbb{P},\hat{\mathbb{P}})} \mathbb{E}_{\mathbb{Q}}\left[\rho^p(\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{u}}^\dagger)\right].$$

We then have

$$\mathcal{G}_W^p(\theta) = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathcal{U}) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{u}} \sim \mathbb{P} \\ \bar{d}_W(\mathbb{P}, \hat{\mathbb{P}}) \leq \theta^p \end{array} \right\}.$$

Equivalently, for $p \in [1, \infty)$, the type-p Wasserstein ambiguity set of radius θ can be re-interpreted as a type-1 Wasserstein ambiguity set of radius θ^p where the type-1 Wasserstein metric between $\mathbb{P}$ and $\hat{\mathbb{P}}$ is $\bar{d}_W(\mathbb{P}, \hat{\mathbb{P}})$. From this perspective, we can directly use Theorem 2 to represent the type-p Wasserstein ambiguity set $\mathcal{G}_W^p(\theta)$ in the format of an event-wise ambiguity set.

For $p=\infty$, the type-∞ Wasserstein metric between two distributions $\mathbb{P}$ and $\hat{\mathbb{P}}$ is defined as

$$d_W^{\infty}(\mathbb{P},\hat{\mathbb{P}}) \triangleq \inf_{\mathbb{Q}\in\mathcal{Q}(\mathbb{P},\hat{\mathbb{P}})} \mathbb{Q}\text{-ess}\sup_{\mathcal{U}\times\mathcal{U}} \rho(\tilde{\boldsymbol{u}},\tilde{\boldsymbol{u}}^{\dagger}),$$

where the essential supremum of the joint distribution $\mathbb{Q}$ is defined by

$$\mathbb{Q}\text{-ess}\sup_{\mathcal{U}\times\mathcal{U}} \rho(\tilde{\boldsymbol{u}},\tilde{\boldsymbol{u}}^{\dagger}) = \inf\{M:\ \mathbb{Q}\left[\rho(\tilde{\boldsymbol{u}}-\tilde{\boldsymbol{u}}^{\dagger})>M\right]=0\}.$$

Bertsimas et al. (2018) show that a distribution $\mathbb{P}$ in the type-∞ Wasserstein ambiguity set

$$\mathcal{G}_W^{\infty}(\theta)=\left\{\mathbb{P}\in\mathcal{P}_0(\mathcal{U}) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{u}}\sim\mathbb{P} \\ d_W^{\infty}(\mathbb{P},\hat{\mathbb{P}})\le\theta \end{array}\right\}$$

is indeed a mixture distribution $\mathbb{P}=\frac{1}{S}\sum_{s\in[S]}\mathbb{P}_s$ consisting of ambiguous components such that for every $s\in[S]$, $\mathbb{P}_s\in\mathcal{P}_0(\mathcal{U})$ and $\mathbb{P}_s[\rho(\tilde{\boldsymbol{u}},\hat{\boldsymbol{u}}_s)\le\theta]=1$. Therefore, one can represent the type-∞ Wasserstein ambiguity set using the following mixture-distribution ambiguity set

$$\mathcal{F}_W^{\infty}(\theta)=\left\{\mathbb{P}\in\mathcal{P}_0\left(\mathbb{R}^{I_u}\times[S]\right) \;\middle|\; \begin{array}{ll} (\tilde{\boldsymbol{u}},\tilde{s})\sim\mathbb{P} & \\ \mathbb{P}\left[\tilde{\boldsymbol{u}}\in\mathcal{U},\rho(\tilde{\boldsymbol{u}},\hat{\boldsymbol{u}}_s)\le\theta \mid \tilde{s}=s\right]=1 & \forall s\in[S] \\ \mathbb{P}\left[\tilde{s}=s\right]=\frac{1}{S} & \forall s\in[S] \end{array}\right\},$$

which is an event-wise ambiguity set satisfying $\mathcal{G}_W^{\infty}(\theta)=\Pi_{\tilde{\boldsymbol{u}}}\mathcal{F}_W^{\infty}(\theta)$ for all $\theta\ge 0$.

In recent independent works, based on a generalized 'primal-worst equals dual-best' duality scheme, Kuhn et al. (2019) and Zhen et al. (2019) provide convex reformulations for many distributionally robust optimization problems with type-p Wasserstein metric for $p\in[1,\infty]$.

D. Computational Experiments with Wasserstein Ambiguity Sets

We focus on two-stage linear optimization problems with the data-driven Wasserstein ambiguity set of type-1 in the form (7), given some past observations $\hat{\boldsymbol{u}}_1,\ldots,\hat{\boldsymbol{u}}_S$ of the uncertainty.

Multi-Item Newsvendor Problem

We consider a multi-item newsvendor problem with I_u different items. For each item i ($i\in[I_u]$), its unit selling price and ordering cost are denoted by p_i and c_i, respectively. Under a total budget d, the decision maker decides the ordering quantity w_i of each item before its random demand $\tilde{u}_i$ is observed. Once the demand realizes, the selling quantity of each item is decided as $\min\{w_i,u_i\}$. The decision maker maximizes the worst-case expected operating revenue by solving

$$\begin{array}{rl} \max & \displaystyle\inf_{\mathbb{P}\in\mathcal{F}_W(\theta)} \mathbb{E}_{\mathbb{P}}\left[\sum_{i\in[I_u]} p_i\min\{w_i,\tilde{u}_i\}\right] \\ \text{s.t.} & \boldsymbol{c}^{\top}\boldsymbol{w}=d,\ \boldsymbol{w}\ge\boldsymbol{0}, \end{array}$$

which can be recast as a minimization problem,

$$\begin{aligned} \min\ & -\boldsymbol{p}^\top \boldsymbol{w} + \sup_{\mathbb{P}\in\mathcal{F}_W(\theta)} \mathbb{E}_{\mathbb{P}}\left[\sum_{i\in[I_u]} p_i(w_i - \tilde{u}_i)^+\right] \\ \text{s.t.}\ & \boldsymbol{c}^\top \boldsymbol{w} = d,\ \boldsymbol{w} \geq \boldsymbol{0}. \end{aligned} \tag{27}$$

In the objective function, $\sum_{i\in[I_u]} p_i(w_i - u_i)^+ = \max_{\mathcal{J}\subseteq[I_u]} \sum_{j\in\mathcal{J}} p_j(w_j - u_j)$ is convex and piecewise affine involving 2^{I_u} pieces. Thus by Theorem 3, problem (27) can be exactly solved by

$$\begin{aligned} \lambda^\star = \min\ & -\boldsymbol{p}^\top \boldsymbol{w} + \sup_{\mathbb{P}\in\mathcal{F}_W(\theta)} \mathbb{E}_{\mathbb{P}}\left[y(\tilde{s}, \tilde{\boldsymbol{z}})\right] \\ \text{s.t.}\ & y(s, \boldsymbol{z}) \geq \sum_{j\in\mathcal{J}} p_j(w_j - u_j) \qquad \forall \boldsymbol{z}\in\mathcal{Z}_s,\ s\in[S],\ \mathcal{J}\subseteq[I_u] \\ & \boldsymbol{c}^\top \boldsymbol{w} = d,\ \boldsymbol{w}\geq \boldsymbol{0} \\ & y \in \bar{\mathcal{A}}(\bar{\mathcal{C}}, [I_u + 1]), \end{aligned} \tag{28}$$

where we introduce a recourse variable $y(\cdot,\cdot)$ following the event-wise affine adaptation with the collection $\bar{\mathcal{C}} \triangleq \{\{s\} \mid s\in[S]\}$. Problem size of this exact approach however, increases exponentially in the number of items. Alternatively, we can obtain an upper bound by solving an RSO problem:

$$\begin{aligned} \lambda = \min\ & -\boldsymbol{p}^\top \boldsymbol{x} + \sup_{\mathbb{P}\in\mathcal{F}_W(\theta)} \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{p}^\top \boldsymbol{y}(\tilde{s}, \tilde{\boldsymbol{z}})\right] \\ \text{s.t.}\ & \boldsymbol{y}(s, \boldsymbol{z}) \geq \boldsymbol{0} && \forall \boldsymbol{z}\in\mathcal{Z}_s,\ s\in[S] \\ & \boldsymbol{y}(s, \boldsymbol{z}) \geq \boldsymbol{w} - \boldsymbol{u} && \forall \boldsymbol{z}\in\mathcal{Z}_s,\ s\in[S] \\ & \boldsymbol{c}^\top \boldsymbol{w} = d,\ \boldsymbol{w}\geq \boldsymbol{0} \\ & y_i \in \bar{\mathcal{A}}(\mathcal{C}, \mathcal{I}) && \forall i\in[I_u], \end{aligned} \tag{29}$$

where we control how the recourse decision $\boldsymbol{y}(\cdot,\cdot)$ adapts to $(\tilde{\boldsymbol{u}}, \tilde{v})$ and $\tilde{s}$ through choosing the collection $\mathcal{C}$ of MECE events and the information index set $\mathcal{I}$.

We consider $I_u \in \{5, 7\}$ and $S \in \{5, 10, 20, 50\}$. The random demand belongs to a support set $\mathcal{U} = [\boldsymbol{0}, \bar{\boldsymbol{u}}]$, and we use the Euclidean norm $\|\cdot\|_2$ as the distance metric. In each instance, we randomly generate the upper bound $\bar{\boldsymbol{u}}$ from a uniform distribution on $[0, 100]^{I_u}$. Subsequently, past observations are randomly generated from the uniform distribution on $[\boldsymbol{0}, \bar{\boldsymbol{u}}]$. We set $c_i = 1$, $i\in[I_u]$ and $b = 50 I_u$, and we generate $\boldsymbol{p}$ from a uniform distribution on $[0, 5]^{I_u}$. For different choices of θ, we run 100 random instances and compare the performance of different cases of event-wise recourse adaptation against the exact approach:

- case 1: $\boldsymbol{y}(\cdot,\cdot)$ adapts on $\tilde{\boldsymbol{u}}, \tilde{v}, \tilde{s}$, *i.e.*, $\mathcal{C} = \{\{s\} \mid s\in[S]\}$ and $\mathcal{I} = [I_u + 1]$;
- case 2: $\boldsymbol{y}(\cdot,\cdot)$ adapts only on $\tilde{\boldsymbol{u}}, \tilde{v}$, *i.e.*, $\mathcal{C} = \{\{1, \ldots, S\}\}$ and $\mathcal{I} = [I_u + 1]$;
- case 3: $\boldsymbol{y}(\cdot,\cdot)$ adapts only on $\tilde{\boldsymbol{u}}$, *i.e.*, $\mathcal{C} = \{\{1, \ldots, S\}\}$ and $\mathcal{I} = [I_u]$.

Case 1 corresponds to the full event-wise affine adaptation. We turn off the event-wise adaptation in case 2, while in case 3 we further deprive the recourse decision $\boldsymbol{y}(\cdot,\cdot)$ of the affine adaptation on

		θ														
		1			2			5			10			20		
S	5	<0.1	2.0	8.7	<0.1	3.4	8.8	0.1	5.4	9.2	0.2	7.0	9.9	0.5	9.4	10.5
	10	<0.1	4.5	11.6	<0.1	6.4	11.9	0.2	8.6	12.3	0.3	11.0	12.7	1.0	12.6	13.0
	20	<0.1	4.8	5.7	<0.1	5.3	5.8	<0.1	5.6	6.1	0.1	6.4	6.5	0.3	7.9	7.9
	50	<0.1	6.8	7.9	<0.1	7.4	8.0	<0.1	8.1	8.4	0.1	9.2	9.2	0.3	11.3	11.3

Table 1 5 items: 90-th percentile optimality gaps (in %) of case 1 (left), case 2 (middle), and case 3 (right).

		θ														
		1			2			5			10			20		
S	5	<0.1	1.3	8.9	0.1	2.3	9.0	0.2	3.8	9.3	0.3	5.6	9.8	0.6	8.2	10.5
	10	<0.1	2.5	6.2	0.1	3.5	6.2	0.1	5.0	6.4	0.2	6.0	6.6	0.5	7.6	7.7
	20	<0.1	4.8	6.7	0.1	5.7	6.8	0.1	6.7	6.9	0.2	7.3	7.3	0.6	7.9	7.9
	50	0.1	5.5	6.1	0.1	5.8	6.1	0.2	6.3	6.3	0.2	6.3	6.6	0.5	7.6	7.6

Table 2 7 items: 90-th percentile optimality gaps (in %) of case 1 (left), case 2 (middle), and case 3 (right).

the auxiliary random variable $\tilde{v}$. For each case, we consider the following relative gap between the objective value using the event-wise recourse adaptation and the exact optimal objective value:

$$\frac{\lambda^\star - \lambda}{\lambda^\star} \times 100\%.$$

Results for $I_u = 5$ and $I_u = 7$ are summarized in Table 1 and Table 2, respectively. With *(i)* the notion of event-wise adaptation and *(ii)* the inclusion of auxiliary random variable $\tilde{v}$, the full event-wise affine adaptation could provide a reasonably good conservative approximation to the exact approach; while excluding either *(i)* or *(ii)* may lead to a more conservative approximation.

We evaluate the scalability of the full event-wise affine adaptation (case 1), the affine adaptation without event-wise dependence (case 2), and the exact approach, by comparing their computation times and limits for different pairs of problem sizes. For the exact approach, the computer runs out of memory when the number of items exceeds 10 and the number of samples exceeds 5 (see Table 3), which is not practically favorable. In contrast, we are able to obtain a conservative solution via the full event-wise affine adaptation with modest computational effort. Quite interestingly, the event-wise adaptation that plays the key role in delivering the less conservative approximation seems to require only a little extra computational effort (see Table 4).

(S, I_u)						
(5, 5)	(10, 5)	(20, 5)	(50, 5)	(100, 5)	(200, 5)	(5, 10)
0.1	1.1	0.2	0.5	1.3	9.4	9.8

Table 3 Computation times of the exact approach. We report only those (S, I_u) pairs for which the exact approach were solved.

		I_u					
		5	10	15	20	25	30
S	5	< 0.1	< 0.1	0.2	0.5	0.7	1.1
	10	< 0.1	0.1	0.3	0.6	2.4	3.7
	20	< 0.1	0.3	0.7	1.3	2.2	3.8
	50	0.2	0.8	1.8	3.7	6.8	10.9
	100	0.4	1.8	5.2	10.3	17.8	31.0
	200	0.5	2.0	5.5	11.0	19.1	32.8
	300	0.8	3.1	7.8	16.5	32.3	49.6
	500	1.4	5.7	16.6	36.6	—	—
	800	2.5	14.8	—	—	—	—

		I_u					
		5	10	15	20	25	30
S	5	< 0.1	0.2	0.2	0.2	0.4	0.5
	10	< 0.1	< 0.1	0.2	0.3	0.7	2.3
	20	< 0.1	0.1	0.3	0.6	1.3	1.9
	50	0.1	0.4	0.8	1.9	3.0	4.4
	100	0.2	1.1	2.7	5.1	9.2	10.5
	200	0.4	1.7	5.2	15.8	30.7	25.3
	300	0.8	3.2	8.6	14.2	36.1	48.4
	500	1.1	5.3	13.8	—	—	—
	800	4.3	13.8	—	—	—	—

Table 4 Computation times and limits of the affine approximation without event-wise adaptation (left) and the full event-wise affine adaptation (right). The symbol "—" indicates "out of memory".

The following code segment shows how to implement the full event-wise affine adaptation for the multi-item newsvendor problem with Wasserstein ambiguity sets in RSOME.

```
% I: number of items
% S: number of past observations
% theta: radius
% Gamma: total budget
% cost (price): cost (price) parameters
% ubar: upper bound of demand
% U = (u_1, ..., u_S): past realizations

% Create RSOME model
model = rsome('newsvendor');

% Define random variables
u = model.random;                    % random demand
v = model.random;                    % auxiliary random variable
P = model.ambiguity(S);              % create ambiguity set
```

```
% Define support sets for scenarios
for s = 1:S
    P(s).suppset(0 <= u, u <= ubar, norm(u - U(:,s)) <= v);
end

% Define probabilities for scenarios
pr = P.prob;
P.probset(pr == 1/S);

% Define event-wise expectation
P.exptset(expect(u) <= theta);

% Declare Warsserstein ambiguity set
model.with(P);

% Define decision variables
w = model.decision(I,1);
y = model.decision(I,1);

% Define event-wise adaptation
for s = 1:S
    y.evtadapt(s);
end

% Define affine adaptation
y.affadapt(u);
y.affadapt(v);

% Define objective function
model.min(-price'*w - expect(price'*y));

% Define constraints
model.append(y >= 0);
model.append(y >= w - u);
model.append(w >= 0);
model.append(cost'*w == Gamma);

% Solution
model.solve;
```

Experiment of Hanasusanto and Kuhn (2018)

We benchmark the RSO model against a state-of-art approximation scheme proposed by Hanasusanto and Kuhn (2018). Particularly, we repeat their experiment using the same set-ups.

Consider the second-stage problem of the form

$$f(\boldsymbol{u}) = \min\{\boldsymbol{e}^\top \boldsymbol{y} \mid \boldsymbol{y} \geq \boldsymbol{0}, \boldsymbol{y} \geq \boldsymbol{A}\boldsymbol{u} - \boldsymbol{b}\}, \tag{30}$$

where $\boldsymbol{e}$ is a vector of ones. The problem does not have any here-and-now decision $\boldsymbol{w}$ and assumes that the random variable $\tilde{\boldsymbol{u}}$ resides in a box $\mathcal{U} = [0,1]^{I_u}$. Under the distance metric $\rho(\boldsymbol{u}, \boldsymbol{u}^\dagger) =$

$\|\boldsymbol{u}-\boldsymbol{u}^\dagger\|_2^2$, Hanasusanto and Kuhn (2018) have shown that the worst-case expectation

$$\sup_{\mathbb{P}\in\mathcal{F}_W(\theta)} \mathbb{E}_{\mathbb{P}}\left[f(\tilde{\boldsymbol{u}})\right] \tag{31}$$

amounts exactly to the optimal value of the following copositive program.

$$\begin{array}{rll}
\inf & \displaystyle\frac{1}{S}\sum_{s\in[S]}\left(\alpha_s+\overline{\boldsymbol{q}}^\top\boldsymbol{\psi}_s-\beta\|\hat{\boldsymbol{u}}_s\|_2^2+\sum_{\ell\in[I_u+J_y]}\phi_{s\ell}\overline{q}_\ell^2\right)+\beta\theta^2 & \\
\text{s.t.} & \begin{pmatrix} \beta\boldsymbol{I}+\overline{\boldsymbol{Q}}^\top\mathrm{diag}(\boldsymbol{\phi}_s)\overline{\boldsymbol{Q}} & -\frac{1}{2}\overline{\boldsymbol{T}}^\top-\overline{\boldsymbol{Q}}^\top\mathrm{diag}(\boldsymbol{\phi}_s)\overline{\boldsymbol{W}}^\top & -\beta\hat{\boldsymbol{u}}_s-\frac{1}{2}\overline{\boldsymbol{Q}}^\top\boldsymbol{\psi}_s \\ -\frac{1}{2}\overline{\boldsymbol{T}}-\overline{\boldsymbol{W}}\mathrm{diag}(\boldsymbol{\phi}_s)\overline{\boldsymbol{Q}} & \overline{\boldsymbol{W}}\mathrm{diag}(\boldsymbol{\phi}_s)\overline{\boldsymbol{W}}^\top & \frac{1}{2}(\overline{\boldsymbol{W}}\boldsymbol{\phi}_s-\overline{\boldsymbol{h}}) \\ -\beta\hat{\boldsymbol{u}}_s^\top-\frac{1}{2}\boldsymbol{\psi}_s^\top\overline{\boldsymbol{Q}} & \frac{1}{2}(\overline{\boldsymbol{W}}\boldsymbol{\phi}_s-\overline{\boldsymbol{h}})^\top & \alpha_s \end{pmatrix}\succeq_{K_{\mathrm{cop}}}\boldsymbol{O} & \forall s\in[S] \\
& \boldsymbol{\alpha}\in\mathbb{R}^S,\ \beta\in\mathbb{R}_+,\ \boldsymbol{\psi}_s,\boldsymbol{\phi}_s\in\mathbb{R}^{I_u+J_y} & \forall s\in[S].
\end{array} \tag{32}$$

Here, $K_{\mathrm{cop}}=\{\boldsymbol{M}\in\mathbb{S}^K \mid \boldsymbol{x}^\top\boldsymbol{M}\boldsymbol{x}\geq 0\ \forall\boldsymbol{x}\geq\boldsymbol{0}\}$ is the copositive cone,

$$\overline{\boldsymbol{Q}}=\begin{pmatrix}\boldsymbol{O}\\ \boldsymbol{I}\end{pmatrix},\ \overline{\boldsymbol{q}}=\begin{pmatrix}\boldsymbol{e}\\ -\boldsymbol{e}\end{pmatrix},\ \overline{\boldsymbol{T}}=\begin{pmatrix}\boldsymbol{A}\\ \boldsymbol{O}\end{pmatrix},\ \overline{\boldsymbol{h}}=\begin{pmatrix}-\boldsymbol{b}\\ \boldsymbol{0}\end{pmatrix},\ \overline{\boldsymbol{W}}=\begin{pmatrix}\boldsymbol{W} & \boldsymbol{O}\\ \boldsymbol{O} & -\boldsymbol{I}\end{pmatrix} \text{ with } \boldsymbol{W}=\begin{pmatrix}\boldsymbol{I}\\ \boldsymbol{I}\end{pmatrix},$$

and $\boldsymbol{O}$, $\boldsymbol{I}$, $\boldsymbol{0}$ and $\boldsymbol{e}_i$ respectively correspond to the zero matrix, the identity matrix, the zero vector and the i-th standard unit basis, all of which are of the appropriate dimension. Because the copostive program (32) is generally intractable, Hanasusanto and Kuhn (2018) adopt a conservative K_0-approximation by replacing the copostive cone K_{cop} with

$$K_0=\{\boldsymbol{M}\in\mathbb{S}^K \mid \boldsymbol{M}=\boldsymbol{P}+\boldsymbol{N},\boldsymbol{P}\succeq\boldsymbol{O},\boldsymbol{N}\geq\boldsymbol{O}\}\subseteq K_{\mathrm{cop}},$$

which leads problem (32) to a semidefinite program.

We run numerical tests for different pairs of the uncertainty dimension I_u and the sample size S, and for each pair, we use the same set-ups as in Hanasusanto and Kuhn (2018) to generate 100 random instances. The Wasserstein radius is set to $\theta=1/S$. The dimension J_y of the recourse decision is sampled uniformly at random from $\{1,2,\ldots,\lceil\log(I_u+1)\rceil\}$, $\boldsymbol{A}$ is sampled uniformly from $[0,1]^{J_y\times I_u}$, and $\boldsymbol{b}$ is sampled uniformly from $[0,\boldsymbol{e}^\top\boldsymbol{A}_{1:}]\times\cdots\times[0,\boldsymbol{e}^\top\boldsymbol{A}_{J_y:}]$. Here, $\boldsymbol{A}_{1:}$ stands for the first row of $\boldsymbol{A}$ and so forth. Lastly, we generate independent training samples from the uniform distribution on $[0,1]^{I_u}$. We evaluate the worst-case expectation (31) approximately by using *(i)* the K_0-approximation and *(ii)* the following full event-wise affine adaptation:

$$\begin{array}{rll}
\min & \displaystyle\sup_{\mathbb{P}\in\mathcal{F}_W(\theta)}\mathbb{E}_{\mathbb{P}}\left[\boldsymbol{e}^\top\boldsymbol{y}(\tilde{s},\tilde{\boldsymbol{z}})\right] & \\
\text{s.t.} & \boldsymbol{y}(s,\boldsymbol{z})\geq\boldsymbol{0} & \forall\boldsymbol{z}\in\mathcal{Z}_s,\ s\in[S] \\
& \boldsymbol{y}(s,\boldsymbol{z})\geq\boldsymbol{A}\boldsymbol{u}-\boldsymbol{b} & \forall\boldsymbol{z}\in\mathcal{Z}_s,\ s\in[S] \\
& y_j\in\bar{\mathcal{A}}(\{\{1\},\ldots,\{S\}\},[I_u+1]) & \forall j\in[J_y],
\end{array}$$

		I_u									
		1		2		4		8		16	
S	5	0.3	<0.1	0.3	<0.1	0.3	<0.1	0.5	<0.1	2.3	<0.1
	10	0.4	<0.1	0.4	<0.1	0.5	<0.1	0.9	<0.1	5.1	<0.1
	20	0.6	<0.1	0.7	<0.1	0.8	<0.1	1.8	<0.1	11.6	<0.1
	40	1.1	<0.1	1.3	<0.1	1.6	<0.1	3.5	<0.1	23.1	0.1
	80	2.3	<0.1	2.5	<0.1	3.2	<0.1	7.8	0.1	51.1	0.1
	160	4.5	<0.1	5.1	<0.1	7.0	0.1	18.0	0.2	118.0	0.3
	320	9.2	0.1	10.8	0.1	15.5	0.2	45.4	0.3	281.5	0.6
	640	19.7	0.1	26.9	0.2	43.9	0.3	141.5	1.0	684.3	2.3

Table 5 Computation times (in seconds) of K_0-approximation (left) and event-wise affine adaptation (right).

where for each $s \in [S]$, $\mathcal{Z}_s = \{(\boldsymbol{u}, v) \mid \boldsymbol{u} \in [0,1]^{I_u}, v \geq \|\boldsymbol{u} - \hat{\boldsymbol{u}}_s\|_2^2\}$.

Quite surprisingly, for all pairs of problem sizes, the solutions of both approximation approaches coincide for all 100 randomly generated instances. Unfortunately, we are not able to give a formal proof for this observation. Nevertheless, this observation supports that our proposed event-wise affine adaptation delivers solutions with competitive approximation quality as the state-of-the-art approximation scheme by Hanasusanto and Kuhn (2018). We report in Table 5 the average computation times of both approaches. In terms of computation efficiency, the event-wise affine adaptation outperforms because it leads to a second-order cone approximation to problem (32).

We note that the K_0-approximation by Hanasusanto and Kuhn (2018) also works when the cost vector of the second-stage problem (30) is affinely affected by the uncertainty, while our event-wise affine adaptation does not. On the other hand, the event-wise affine adaptation works with more general distance metrics and more general support sets that are not necessarily polyhedral (in the current experiment, the support set is a box $[0,1]^{I_u}$), while the K_0-approximation does not.

E. Multi-Stage Stochastic Financial Planning Problem

We adopt a financial planning problem from Birge and Louveaux (2011) to illustrate how to incorporate the scenario tree approach in the RSO framework.

At the beginning of the first stage in this multi-stage problem, the decision maker allocates the wealth d into two possible investment types, stocks (S) and bonds (B). Eight possible scenarios may occur, which corresponds to independent and equal likelihoods of having high returns of 1.25 for stocks and 1.14 for bonds, or low returns of 1.06 for stocks and 1.12 for bonds over subsequent

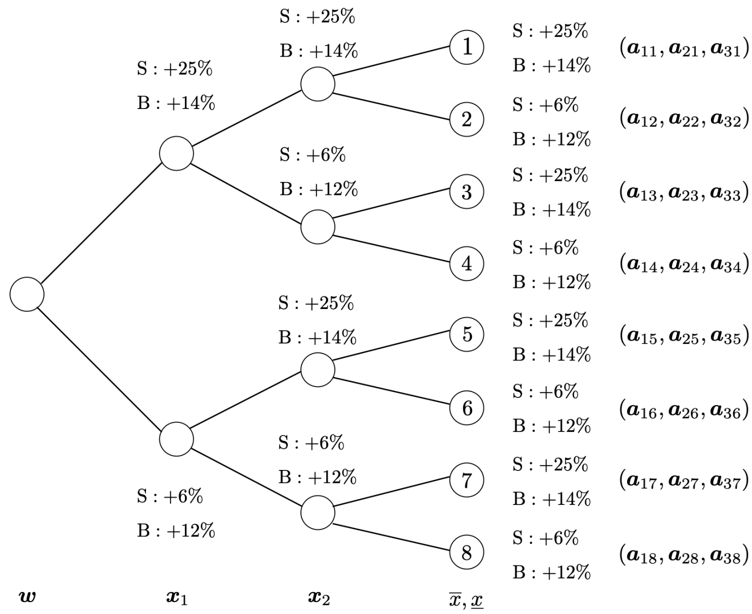

Figure 5 Scenario tree of the financial planning problem.

stages (see Figure 5). Hence we can construct the following singleton ambiguity set of the known discrete distribution of uncertain returns over all stages.

$$\mathcal{F} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2 \times [S]\right) \left| \begin{array}{ll} (\tilde{\boldsymbol{z}}_1, \tilde{\boldsymbol{z}}_2, \tilde{\boldsymbol{z}}_3, \tilde{s}) \sim \mathbb{P} & \\ \mathbb{P}\left[(\tilde{\boldsymbol{z}}_1, \tilde{\boldsymbol{z}}_2, \tilde{\boldsymbol{z}}_3) \in \mathcal{Z}_s \mid \tilde{s} = s\right] = 1 & \forall s \in [S] \\ \mathbb{P}\left[\tilde{s} = s\right] = \frac{1}{S} & \forall s \in [S] \end{array} \right. \right\},$$

where $S = 8$ and the singleton support sets $\mathcal{Z}_s = \{(\boldsymbol{a}_{1s}, \boldsymbol{a}_{2s}, \boldsymbol{a}_{3s})\}, s \in [S]$ are determined by $\boldsymbol{a}_{1s} = (1.25, 1.14)$ for $s \in \{1,2,3,4\}$, $\boldsymbol{a}_{2s} = (1.25, 1.14)$ for $s \in \{1,2,5,6\}$, $\boldsymbol{a}_{3s} = (1.25, 1.14)$ for $s \in \{1,3,5,7\}$, and $\boldsymbol{a}_{is} = (1.06, 1.12)$ otherwise.

The decision maker evaluates the difference between the final return r and a prescribed target τ based on a concave and piecewise affine utility function that takes $U(r-\tau) = r - \tau$ if $r \geq \tau$ and $U(r-\tau) = 4(r-\tau)$ otherwise. The initial investment decisions $\boldsymbol{w}$, made before the first stage returns of stocks and bonds realize, must be indifferent among all eight scenarios. The rebalanced investment decison $\boldsymbol{x}_1$, made after the first stage returns realize but before the second stage returns realize, shall be indifferent among scenarios $\{1,2,3,4\}$ and indifferent among scenarios $\{5,6,7,8\}$. Similarly, $\boldsymbol{x}_2$ is indifferent between scenarios $\{1,2\}$ as well as between scenarios $\{3,4\}$, $\{5,6\}$, and $\{7,8\}$. Finally, the nonnegative auxiliary recourse decisions $\overline{x}$ and $\underline{x}$, respectively standing for the excess above or shortfall below the target, are adaptive to revealed uncertainties and thus can be different across the eight scenarios. In all, we can formulate the RSO model as follows:

$$
\begin{aligned}
\max\ & \inf_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[\overline{x}(\tilde{s}) - 4\underline{x}(\tilde{s})\right] \\
\text{s.t.}\ & w_1, w_2 \geq 0,\ w_1 + w_2 = d \\
& x_{11}(s) + x_{12}(s) - \boldsymbol{z}_1^\top \boldsymbol{w} = 0 && \forall \boldsymbol{z} \in \mathcal{Z}_s, s\in[S] \\
& x_{21}(s) + x_{22}(s) - \boldsymbol{z}_2^\top \boldsymbol{x}_1(s) = 0 && \forall \boldsymbol{z} \in \mathcal{Z}_s, s\in[S] \\
& \boldsymbol{z}_3^\top \boldsymbol{x}_2(s) - \overline{x}(s) + \underline{x}(s) = \tau && \forall \boldsymbol{z} \in \mathcal{Z}_s, s\in[S] \\
& x_{11}(s), x_{12}(s), x_{21}(s), x_{22}(s), \overline{x}(s), \underline{x}(s) \geq 0 && \forall s\in[S] \\
& x_{11}, x_{12} \in \mathcal{A}(\{\{1,2,3,4\},\{5,6,7,8\}\}) \\
& x_{21}, x_{22} \in \mathcal{A}(\{\{1,2\},\{3,4\},\{5,6\},\{7,8\}\}) \\
& \overline{x}, \underline{x} \in \mathcal{A}(\{\{1\},\{2\},\{3\},\{4\},\{5\},\{6\},\{7\},\{8\}\}).
\end{aligned}
$$

F. Portfolio Management with K-means Adaptive Rebalancing

We consider a three-period portfolio allocation and rebalancing problem to minimize the investment risk at the last period taking into account of transaction costs. At the beginning of the first period, we decide the number of shares $w_i \geq 0$ of stock $i \in [I_u]$ to invest at price a_i, incurring a transaction cost $b_i w_i$. The price of stock i in the second period is $\tilde{a}_i^1 \triangleq a_i(\tilde{u}_i^1 + 1)$, where $\tilde{u}_i^1$ is the corresponding return. Subsequently, for each stock i, we rebalance its shares to $x_i \geq 0$, which incurs a transaction cost $b_i|x_i - w_i|$. In the last period, the price of stock i is $\tilde{a}_i^2 \triangleq a_i(\tilde{u}_i^2 + 1)$, where $\tilde{u}_i^2$ is the third period return with respect to the first period price. The effective portfolio return at the last period, taking into account of the total transaction costs, amounts to

$$
\boldsymbol{w}^\top \tilde{\boldsymbol{a}}^1 - \boldsymbol{w}^\top \boldsymbol{a} + \boldsymbol{x}^\top \tilde{\boldsymbol{a}}^2 - \boldsymbol{x}^\top \tilde{\boldsymbol{a}}^1 - \boldsymbol{b}^\top(\boldsymbol{w} + |\boldsymbol{x} - \boldsymbol{w}|) = \boldsymbol{w}^\top \boldsymbol{A}\tilde{\boldsymbol{u}}^1 + \boldsymbol{x}^\top \boldsymbol{A}(\tilde{\boldsymbol{u}}^2 - \tilde{\boldsymbol{u}}^1) - \boldsymbol{b}^\top(\boldsymbol{w} + |\boldsymbol{x} - \boldsymbol{w}|),
$$

where $\boldsymbol{A} = \mathrm{diag}(\boldsymbol{a})$ and the operator $|\cdot|$ takes the absolute value component-wise.

Ideally, the rebalancing decision $\boldsymbol{x}$ should only depend on the realization of $\tilde{\boldsymbol{u}}^1$. However, this would lead to an intractable problem. Instead, we propose an alternative K-means adaptive approach, where the recourse decision $\boldsymbol{x}(\tilde{s})$ depends on the random scenario $\tilde{s}$ that is associated with the realization of $\tilde{\boldsymbol{u}}^1$. In particular, using the available historical returns $\{(\hat{\boldsymbol{u}}_n^1, \hat{\boldsymbol{u}}_n^2)\}_{n\in[N]}$, we construct a two-layer K-means ambiguity set by first partitioning $\{\hat{\boldsymbol{u}}_n^1\}_{n\in[N]}$ into K_1 clusters, each of which we then further partition into K_2 clusters based on a subset of $\{\hat{\boldsymbol{u}}_n^2\}_{n\in[N]}$ that are affiliated with this specific first-layer cluster. As a result, we obtain a total number of $S = K_1 K_2$ scenarios, each of which corresponds to a unique cluster determined by the first and second layers; see an illustration in Figure 2. For each of the first-layer cluster $k \in [K_1]$, we denote by $\mathcal{E}_k \subseteq [S]$ as the set of scenarios associated with that cluster. Correspondingly, we define $\kappa(s) \in [K_1]$ as the specific

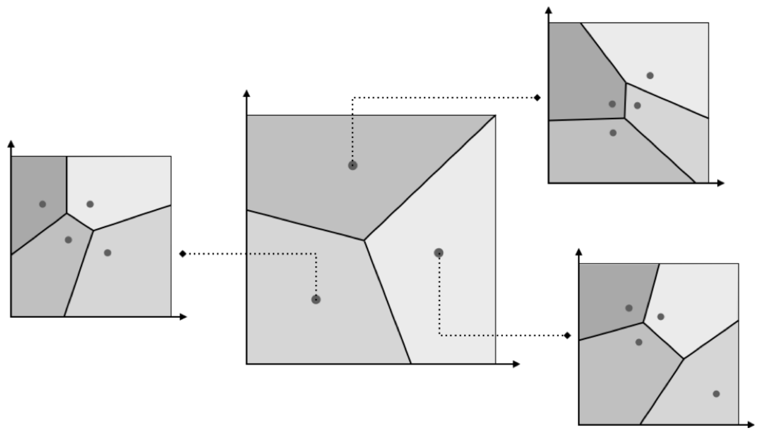

Figure 6 Two-layer K-means clustering. There are 3 clusters for $\{\hat{\boldsymbol{u}}_n^1\}_{n\in[N]}$ in the first layer, and a subset of $\{\hat{\boldsymbol{u}}_n^2\}_{n\in[N]}$ affiliated with each of these clusters is further partitioned into 4 clusters in the second layer. In total, we have 12 distinctive clusters.

first-layer cluster that the scenario s affiliates with. Observe that $\mathcal{C} = \{\mathcal{E}_k \mid k \in [K_1]\}$ is a collection of MECE events. In this way, we obtain the two-layer K-means ambiguity set

$$\mathcal{F} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^{2I_u+2I_v} \times [S]\right) \left| \begin{array}{ll} ((\tilde{\boldsymbol{u}}^1, \tilde{\boldsymbol{u}}^2, \tilde{\boldsymbol{v}}^1, \tilde{\boldsymbol{v}}^2), \tilde{s}) \sim \mathbb{P} & \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{u}}^1 \mid \tilde{s} \in \mathcal{E}_k\right] = \hat{\boldsymbol{\mu}}_k^1 & \forall k \in [K_1] \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{v}}^1 \mid \tilde{s} \in \mathcal{E}_k\right] \leq \hat{\boldsymbol{\sigma}}_k^1 & \forall k \in [K_1] \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{u}}^2 \mid \tilde{s} = s\right] = \hat{\boldsymbol{\mu}}_s^2 & \forall s \in [S] \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{v}}^2 \mid \tilde{s} = s\right] \leq \hat{\boldsymbol{\sigma}}_s^2 & \forall s \in [S] \\ \mathbb{P}\left[(\tilde{\boldsymbol{u}}^1, \tilde{\boldsymbol{u}}^2, \tilde{\boldsymbol{v}}^1, \tilde{\boldsymbol{v}}^2) \in \mathcal{Z}_s \mid \tilde{s} = s\right] = 1 & \forall s \in [S] \\ \mathbb{P}\left[\tilde{s} = s\right] = p_s & \forall s \in [S] \end{array} \right. \right\},$$

where for each $s \in [S]$, the cluster-wise support set is determined by

$$\mathcal{Z}_s = \{(\boldsymbol{u}^1, \boldsymbol{u}^2, \boldsymbol{v}^1, \boldsymbol{v}^2) \mid \boldsymbol{u}^1 \in \mathcal{U}^1_{\kappa(s)}, \boldsymbol{u}^2 \in \mathcal{U}^2_s, \boldsymbol{v}^1 \geq \boldsymbol{\phi}(\boldsymbol{u}^1), \boldsymbol{v}^2 \geq \boldsymbol{\phi}(\boldsymbol{u}^2)\}.$$

The objective function evaluates the worst-case conditional value-at-risk (CVaR) of the final return at a pre-specified risk threshold $\varepsilon \in (0, 1)$.

$$\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{P}\text{-CVaR}_{\varepsilon}\left(\boldsymbol{w}^\top \boldsymbol{A}\tilde{\boldsymbol{u}}^1 + \boldsymbol{x}^\top(\tilde{s})\boldsymbol{A}(\tilde{\boldsymbol{u}}^2 - \tilde{\boldsymbol{u}}^1) - \boldsymbol{b}^\top(\boldsymbol{w} + |\boldsymbol{x}(\tilde{s}) - \boldsymbol{w}|)\right),$$

which using now standard techniques, can be rewritten as

$$\min_{\delta} \ \delta + \frac{1}{\varepsilon}\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[\left(\boldsymbol{x}^\top(\tilde{s})\boldsymbol{A}(\tilde{\boldsymbol{u}}^1 - \tilde{\boldsymbol{u}}^2) - \boldsymbol{w}^\top \boldsymbol{A}\tilde{\boldsymbol{u}}^1 + \boldsymbol{b}^\top(\boldsymbol{w} + |\boldsymbol{x}(\tilde{s}) - \boldsymbol{w}|) - \delta\right)^+\right].$$

Using Theorem 3, we formulate the RSO model for this portfolio optimization problem as follows:

$$\begin{array}{lll}
\min & \delta + \dfrac{1}{\varepsilon} \sup\limits_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[y(\tilde{s}, \tilde{\boldsymbol{z}})\right] & \\
\text{s.t.} & y(s, \boldsymbol{z}) \geq 0 & \forall \boldsymbol{z} \in \mathcal{Z}_s,\ s \in [S] \\
& y(s, \boldsymbol{z}) \geq \boldsymbol{x}^\top(s)\boldsymbol{A}(\boldsymbol{u}^1 - \boldsymbol{u}^2) - \boldsymbol{w}^\top\boldsymbol{A}\boldsymbol{u}^1 + \boldsymbol{b}^\top(\boldsymbol{w} + \bar{\boldsymbol{x}}(s)) - \delta & \forall \boldsymbol{z} \in \mathcal{Z}_s,\ s \in [S] \\
& \bar{\boldsymbol{x}}(s) \geq \boldsymbol{x}(s) - \boldsymbol{w} & \forall s \in [S] \\
& \bar{\boldsymbol{x}}(s) \geq \boldsymbol{w} - \boldsymbol{x}(s) & \forall s \in [S] \\
& \boldsymbol{b}^\top \bar{\boldsymbol{x}}(s) \leq \eta & \forall s \in [S] \\
& \boldsymbol{x}(s) \geq \boldsymbol{0} & \forall s \in [S] \\
& \boldsymbol{a}^\top \boldsymbol{w} = d & \\
& \boldsymbol{w} \geq \boldsymbol{0} & \\
& x_i, \bar{x}_i \in \mathcal{A}(\mathcal{C}) & \forall i \in [I_u] \\
& y \in \bar{\mathcal{A}}(\bar{\mathcal{C}}, [2I_u + 2I_v]) &
\end{array}$$

where $\tilde{\boldsymbol{z}} \triangleq (\tilde{\boldsymbol{u}}^1, \tilde{\boldsymbol{u}}^2, \tilde{\boldsymbol{v}}^1, \tilde{\boldsymbol{v}}^2)$ and $\bar{\mathcal{C}} \triangleq \{\{s\} \mid s \in [S]\}$ consists of singleton MECE events. For simplicity, we impose a limit η on the transaction cost to prohibit over rebalancing in the second period, and d is the initial allocation budget. We next provide the sample code in RSOME to elucidate the intuitive implementation of the RSO model via an algebraic modeling language.

Sample Code for K-means Adaptive Rebalancing

We assume $\mathcal{E}_k = \{(k-1)K_2 + 1, \ldots, kK_2\} \subseteq [S]$ for all $k \in [K_1]$. Correspondingly, $\kappa(s) = \lceil \frac{s}{K_2} \rceil$ for all $s \in [S]$. We take a convex function $\boldsymbol{\phi}$ that specifies the mean absolute deviation of each random return within a particular cluster. Hence for each $s \in [S]$, the cluster-wise support set is given by

$$\mathcal{Z}_s = \{(\boldsymbol{u}^1, \boldsymbol{u}^2, \boldsymbol{v}^1, \boldsymbol{v}^2) \mid \boldsymbol{D}^1_{\kappa(s)}\boldsymbol{u}^1 \leq \boldsymbol{f}^1_{\kappa(s)}, \boldsymbol{D}^2_s\boldsymbol{u}^2 \leq \boldsymbol{f}^2_s, \boldsymbol{v}^1 \geq |\boldsymbol{u}^1 - \hat{\boldsymbol{\mu}}^1_{\kappa(s)}|, \boldsymbol{v}^2 \geq |\boldsymbol{u}^2 - \hat{\boldsymbol{\mu}}^2_s|\},$$

where each cluster is in fact a polyhedron and where $|\cdot|$ applies component-wise. The estimates $\{\boldsymbol{D}^1_k\}_{k\in[K_1]}$, $\{\boldsymbol{f}^1_k\}_{k\in[K_1]}$, $\{\hat{\boldsymbol{\mu}}^1_k\}_{k\in[K_1]}$, $\{\hat{\boldsymbol{\sigma}}^1_k\}_{k\in[K_1]}$ are contained in MATLAB cells D1, f1, mu1, sigma1, and similarly, $\{\boldsymbol{D}^2_s\}_{s\in[S]}$, $\{\boldsymbol{f}^2_s\}_{s\in[S]}$, $\{\hat{\boldsymbol{\mu}}^2_s\}_{s\in[S]}$, $\{\hat{\boldsymbol{\sigma}}^2_s\}_{s\in[S]}$ are contained in D2, f2, mu2, sigma2.

```
% I: number of stocks
% K1: number of first-layer clusters
% K2: number of second-layer clusters
% ps: probabilities of clusters
% epsilon: risk threshold
% a,b,d,eta: parameters

% Create RSOME model
model = rsome('portfolio');

% Define random variables
```

```
u = model.random(I,2);                  % random demand
v = model.random(I,2);                  % auxiliary random variable
P = model.ambiguity(K1*K2);             % create ambiguity set

% Define support sets for scenarios
for s = 1:K1*K2
    P(s).suppset(D1{ceil(s/K2)}*u(:,1) <= f1{ceil(s/K2)}, ...
                 D2{s}*u(:,2) <= f2{s}, ...
                 v(:,1) >= abs(u(:,1) - mu1{ceil(s/K2)}), ...
                 v(:,2) >= abs(u(:,2) - mu2{s}));
end

% Define probabilities for scenarios
pr = P.prob;
P.probset(pr == ps);

% Define event-wise expectation
for k = 1:K1
    P((k-1)*K2+1:k*K2).exptset(expect(u(:,1)) == mu1{k}, ...
                               expect(v(:,1)) <= sigma1{k});
end
for s = 1:K1*K2
    P(s).exptset(expect(u(:,2)) == mu2{s}, ...
                 expect(v(:,2)) <= sigma2{s});
end

% Declare K-means ambiguity set
model.with(P);

% Define decision variables
w = model.decision(I,1);
x = model.decision(I,1);
xbar = model.decision(I,1);
y = model.decision;
delta = model.decision;

% Define event-wise adaptation
for k = 1:K1
    x.evtadapt((k-1)*K2+1:k*K2);
    xbar.evtadapt((k-1)*K2+1:k*K2);
end
for s = 1:K1*K2
    y.evtadapt(s);
end

% Define affine adaptation
y.affadapt(u);
y.affadapt(v);

% Define objective function
model.min(delta + expect((1/epsilon)*y));
```

```
% Define constraints
model.append(y >= 0);
model.append(y >= x'*diag(a)*(u(:, 1)- u(:, 2)) ...
             - w'*diag(a)*u(:, 1) + b'*(w + xbar) - delta);
model.append(xbar >= abs(x - w));
model.append(b'*xbar <= eta);
model.append(x >= 0);
model.append(a'*w == d);
model.append(w >= 0);

% Solution
model.solve;
```

G. Endnote

All mathematical programs in numerical experiments are solved using MOSEK on an Intel Core (TM) @ 3.40 GHz with 8GB RAM. The semidefinite program related to the K_0-approximation is implemented using the CVX interface (Grant and Boyd 2008), while remaining models are implemented using our developped algebraic modeling package RSOME (available at `https://www.rsomerso.com/`).