Supplementary Material

EC.1. Measures in $\mathcal{D}(\mathcal{S},\boldsymbol{\mu},\Psi)$ Are Less Dispersed

Lemma EC.1. *Given a distributional set $\mathcal{D}(\mathcal{S},\boldsymbol{\mu},\Psi)$ with $\mathcal{S}$ convex and a random vector $\boldsymbol{\xi}$ such that its distribution $F\in\mathcal{D}(\mathcal{S},\boldsymbol{\mu},\Psi)$, for all $0\le\alpha\le 1$ the random vector $\boldsymbol{\zeta}:=\alpha(\boldsymbol{\xi}-\boldsymbol{\mu})+\boldsymbol{\mu}$ also has a distribution that lies in $\mathcal{D}(\mathcal{S},\boldsymbol{\mu},\Psi)$.*

Proof: We simply need to verify systematically the conditions imposed on the members of $\mathcal{D}(\mathcal{S},\boldsymbol{\mu},\Psi)$. First,

$$\mathbb{P}_F(\boldsymbol{\zeta}\in\mathcal{S})=\mathbb{P}_F(\alpha\boldsymbol{\xi}+(1-\alpha)\boldsymbol{\mu}\in\mathcal{S})=1\,,$$

since $\mathbb{P}_F(\boldsymbol{\xi}\in\mathcal{S})=1$, $\boldsymbol{\mu}\in\mathcal{S}$, and the fact that $\mathcal{S}$ is a convex set. Second, it is easy to see that

$$\mathbb{E}_F[\boldsymbol{\zeta}]=\mathbb{E}_F[\alpha\boldsymbol{\xi}+(1-\alpha)\boldsymbol{\mu}]=\boldsymbol{\mu}\,.$$

Finally, for any $\psi(\cdot)\in\Psi$, we have that

$$\begin{aligned}\mathbb{E}_F[\psi(\alpha\boldsymbol{\xi}+(1-\alpha)\boldsymbol{\mu})]\le\mathbb{E}_F[\alpha\psi(\boldsymbol{\xi})+(1-\alpha)\boldsymbol{\mu}] &= \alpha\mathbb{E}_F[\psi(\boldsymbol{\xi})]+(1-\alpha)\psi(\boldsymbol{\mu})\\ \le\mathbb{E}_F[\psi(\boldsymbol{\xi})] &\le 0\,,\end{aligned}$$

where we applied Jensen's inequality for the two first upper bounding steps. □

EC.2. Concavity of $h(\boldsymbol{x},\xi)$ in ξ

Consider the second stage stochastic minimization problem

$$\begin{aligned}h(\boldsymbol{x},\xi):=\ &\text{minimize}_{\boldsymbol{y}}\ f(\boldsymbol{x},\boldsymbol{y},\xi)\\ &\text{s.t.}\qquad \boldsymbol{y}\in\mathcal{Y}(\boldsymbol{x}),\end{aligned}$$

where $\mathcal{Y}(\boldsymbol{x})$ is any given closed and bounded set function of $\boldsymbol{x}$. Here we make no assumption that set $\mathcal{Y}(\boldsymbol{x})$ is a polyhedral or even a convex set.

We assume that $f(\boldsymbol{x},\boldsymbol{y},\xi)$ is a concave function in the uncertain parameter-vector ξ. Note that this assumption has been made in most previous stochastic and robust optimization models (see Ben-Tal and Nemirovski (1998)).

Lemma EC.2. *The minimum value function $h(\boldsymbol{x},\xi)$ is a concave function in ξ.*

This lemma implies that our Proposition 1 is also applicable to most linear and nonlinear two-stage optimization problems. We omit including a detailed proof of this lemma as it is well known that the pointwise infimum of a set of affine functions is a concave function (see Boyd and Vandenberghe (2004) for more details).

EC.3. Proof of Proposition 2

We follow similar steps as followed in the proof of Proposition 1. We first underline the fact that implementing the MVP solution, a policy that does not adapt to the sequence of observable ξ, leads to a worst-case expected cost that is equal to the optimal value of the MVP problem, which we refer to as V_{MVP}.

$$\sup_{F\in\mathcal{D}(\boldsymbol{\mu}_{[1:T]})} \mathbb{E}_F\left[\sum_{t=1}^{T} \xi_t^\top C_t \bar{\boldsymbol{x}}_t\right] = \sup_{F\in\mathcal{D}(\boldsymbol{\mu}_{[1:T]})} \sum_{t=1}^{T} \mathbb{E}_F[\xi_t^\top] C_t \bar{\boldsymbol{x}}_t$$
$$= \sum_{t=1}^{T} \boldsymbol{\mu}_t^\top C_t \bar{\boldsymbol{x}}_t,$$

where $\mathcal{D}(\boldsymbol{\mu}_{[1:T]})$ is short for $\mathcal{D}(\mathcal{S},[\boldsymbol{\mu}_1,\boldsymbol{\mu}_2,...,\boldsymbol{\mu}_T],\Psi)$. Here, we used the fact that $\bar{x}_t$ does not adapt to the observed uncertain parameters and the fact that all the distributions in the set that is considered lead to the same expected value for the random vectors. We can therefore say that the optimal value of the distributionally robust multi-stage stochastic program must be smaller than V_{MVP}.

Secondly, after verifying that the Dirac measure $\delta_{\boldsymbol{\mu}_{[1:T]}}$ lies in the set $\mathcal{D}(\boldsymbol{\mu}_{[1:T]})$ (the argument being the same as in the proof of Proposition 1), one can show that V_{MVP} is actually also a lower bound for the same distributionally robust problem.

$$\min_{(\boldsymbol{x}_1,\boldsymbol{x}_2,...,\boldsymbol{x}_T)\in\mathcal{X} a.s.} \sup_{F\in\mathcal{D}(\boldsymbol{\mu}_{[1:T]})} \mathbb{E}_F\left[\sum_{t=1}^{T} \xi_t^\top C_t \boldsymbol{x}_t(\xi_{[1:t]})\right] \geq \min_{(\boldsymbol{x}_1,\boldsymbol{x}_2,...,\boldsymbol{x}_T)\in\mathcal{X} a.s.} \mathbb{E}_{\delta_{\boldsymbol{\mu}_{[1:T]}}}\left[\sum_{t=1}^{T} \xi_t^\top C_t \boldsymbol{x}_t(\xi_{[1:t]})\right]$$
$$= \min_{(\boldsymbol{x}_1,\boldsymbol{x}_2,...,\boldsymbol{x}_T)\in\mathcal{X} a.s.} \sum_{t=1}^{T} \boldsymbol{\mu}_t^\top C_t \boldsymbol{x}_t(\boldsymbol{\mu}_{[1:t]})$$
$$= V_{\text{MVP}}\,.$$

Hence, we conclude that the MVP solution is an optimal solution for the distributionally robust multi-stage stochastic program under $\mathcal{D}(\boldsymbol{\mu}_{[1:T]})$. □

EC.4. Proof of Proposition 4

The proof consists of showing that the sequence of distributions $F_1, F_2, ...$ with

$$F_k(\boldsymbol{\xi}) = (1-(1+\beta_k)^{-1})\delta_{\boldsymbol{\mu}}(\boldsymbol{\xi}) + (1+\beta_k)^{-1} G_k(\boldsymbol{\xi})\,,$$

satisfies $\mathbb{E}_{F_k}[\psi(\boldsymbol{\xi})] = 0\,,\ \forall\,\psi\in\Psi, \forall k$ and that given any $\epsilon>0$, one can choose k large enough so that F_k satisfies:

$$\mathbb{P}_{F_k}(\boldsymbol{\xi}\in\mathcal{S}) \geq 1-\epsilon$$
$$\|\mathbb{E}_{F_k}[\boldsymbol{\xi}] - \boldsymbol{\mu}\| \leq \epsilon\,.$$

Based on Condition 1, we can easily show the first part:

$$\begin{aligned}\mathbb{E}_{F_k}[\psi(\boldsymbol{\xi})] &= (1-(1+\beta_k)^{-1})\mathbb{E}_{\delta_{\boldsymbol{\mu}}}[\psi(\boldsymbol{\xi})] + (1+\beta_k)^{-1}\mathbb{E}_{G_k}[\psi(\boldsymbol{\xi})] \\ &= (1-(1+\beta_k)^{-1})\psi(\boldsymbol{\mu}) + (1+\beta_k)^{-1}\beta_k = 0\,.\end{aligned}$$

Based on Condition 2, we also have the property that $\mathbb{E}_{G_k}[\|\boldsymbol{\xi}\|] = O(\beta_k^{1/\gamma})$, for some $\gamma > 1$. This is due to the fact that there exists an $a > 0$, $b \in \mathbb{R}$, and $\gamma > 1$:

$$\begin{aligned}\mathbb{E}_{G_k}[\|\boldsymbol{\xi}\|]^{\gamma} \leq \mathbb{E}_{G_k}[\|\boldsymbol{\xi}\|^{\gamma}] \;&=\; (1/a)(\mathbb{E}_{G_k}[b + a\|\boldsymbol{\xi}\|^{\gamma}] - b) \\ &\leq (1/a)(\mathbb{E}_{G_k}[\psi_0(\boldsymbol{\xi})] - b) \;=\; (1/a)(\beta_k - b)\,,\end{aligned}$$

where we used Jensen's inequality and the fact that $\psi_0(\boldsymbol{\xi}) = \sum_i \theta_i \psi_i(\boldsymbol{\xi})$ for some conical combination of $\psi_i \in \Psi$ allowing us to derive

$$\mathbb{E}_{G_k}[\psi_0(\boldsymbol{\xi})] = \mathbb{E}_{G_k}\left[\sum_i \theta_i \psi_i(\boldsymbol{\xi})\right] = \sum_i \theta_i \beta_k = \beta_k\,.$$

Hence, we have that $\mathbb{E}_{G_k}[\|\boldsymbol{\xi}\|] = \mathbb{E}_{G_k}[\|\boldsymbol{\xi}\|]^{\gamma/\gamma} \leq ((1/a)(\beta_k - b))^{1/\gamma} = O(\beta_k^{1/\gamma})$.

Thus, we can demonstrate that both $\mathbb{P}_{F_k}(\boldsymbol{\xi} \in \mathcal{S})$ and $\mathbb{E}_{F_k}[\boldsymbol{\xi}]$ will become feasible for k large enough:

$$\mathbb{P}_{F_k}(\boldsymbol{\xi} \in \mathcal{S}) = (1-(1+\beta_k)^{-1})\mathbb{P}_{\delta_\mu}(\boldsymbol{\xi} \in \mathcal{S}) + (1+\beta_k)^{-1})\mathbb{P}_{G_k}(\boldsymbol{\xi} \in \mathcal{S}) \geq (1-(1+\beta_k)^{-1})\mathbb{P}_{\delta_\mu}(\boldsymbol{\xi} \in \mathcal{S}) = 1-(1+\beta_k)$$

$$\begin{aligned}\|\mathbb{E}_{F_k}[\boldsymbol{\xi}] - \boldsymbol{\mu}\| &\leq (1-(1+\beta_k)^{-1})\|\mathbb{E}_{\delta_{\boldsymbol{\mu}}}[\boldsymbol{\xi}] - \boldsymbol{\mu}\| + (1+\beta_k)^{-1}\|\mathbb{E}_{G_k}[\boldsymbol{\xi}] - \boldsymbol{\mu}\| \\ &\leq (1+\beta_k)^{-1}(\|\mathbb{E}_{G_k}[\boldsymbol{\xi}]\| + \|\boldsymbol{\mu}\|) \leq (1+\beta_k)^{-1}(\mathbb{E}_{G_k}[\|\boldsymbol{\xi}\|] + \|\boldsymbol{\mu}\|) = O(\beta_k^{-(1-1/\gamma)})\,.\end{aligned}$$

Ultimately, based on the Lipschitz property of $h(x,\cdot)$ demonstrated in Lemma 1, we can verify that

$$\begin{aligned}\sup_{F \in \bar{\mathcal{D}}(\mathcal{S},\boldsymbol{\mu},\Psi,\epsilon)} \mathbb{E}_F[h(\boldsymbol{x},\boldsymbol{\xi})] \;&\geq\; \sup_{\{k | F_k \in \bar{\mathcal{D}}(\mathcal{S},\boldsymbol{\mu},\Psi,\epsilon)\}} \mathbb{E}_{F_k}[h(\boldsymbol{x},\boldsymbol{\xi})] \\ &= \sup_{\{k | F_k \in \bar{\mathcal{D}}(\mathcal{S},\boldsymbol{\mu},\Psi,\epsilon)\}} (1-(1+\beta_k)^{-1})h(\boldsymbol{x},\boldsymbol{\mu}) + (1+\beta_k)^{-1}\mathbb{E}_{G_k}[h(\boldsymbol{x},\boldsymbol{\xi})] \\ &\geq \sup_{\{k | F_k \in \bar{\mathcal{D}}(\mathcal{S},\boldsymbol{\mu},\Psi,\epsilon)\}} (1-(1+\beta_k)^{-1})h(\boldsymbol{x},\boldsymbol{\mu}) + (1+\beta_k)^{-1}\mathbb{E}_{G_k}[h(\boldsymbol{x},\boldsymbol{\mu}) - R\|\boldsymbol{C}_2\|\|\boldsymbol{\xi} - \boldsymbol{\mu}\|] \\ &\geq \sup_{\{k | F_k \in \bar{\mathcal{D}}(\mathcal{S},\boldsymbol{\mu},\Psi,\epsilon)\}} h(\boldsymbol{x},\boldsymbol{\mu}) - (1+\beta_k)^{-1}R\|\boldsymbol{C}_2\|\mathbb{E}_{G_k}[\|\boldsymbol{\xi}\| + \|\boldsymbol{\mu}\|] \\ &= \sup_{\{k | F_k \in \bar{\mathcal{D}}(\mathcal{S},\boldsymbol{\mu},\Psi,\epsilon)\}} h(\boldsymbol{x},\boldsymbol{\mu}) - (1+\beta_k)^{-1}R\|\boldsymbol{C}_2\|\|\boldsymbol{\mu}\| - (1+\beta_k)^{-1}R\|\boldsymbol{C}_2\|\,O(\beta_k^{1/\gamma}) \\ &= h(\boldsymbol{x},\boldsymbol{\mu})\,.\end{aligned}$$

We can then resolve an approximate upper bound using Proposition 1 and the Lipschitz property of $h(\boldsymbol{x},\cdot)$:

$$\begin{aligned}\sup_{F\in\bar{\mathcal{D}}(\mathcal{S},\boldsymbol{\mu},\Psi,\epsilon)} \mathbb{E}_F[h(\boldsymbol{x},\boldsymbol{\xi})] &\le \sup_{\{\boldsymbol{z}\,|\,\|\boldsymbol{z}-\boldsymbol{\mu}\|\le\epsilon\}} \sup_{F\in\mathcal{D}(\mathbb{R}^d,\boldsymbol{z},\Psi)} \mathbb{E}_F[h(\boldsymbol{x},\boldsymbol{\xi})] \\ &\le \sup_{\{\boldsymbol{z}\,|\,\|\boldsymbol{z}-\boldsymbol{\mu}\|\le\epsilon\}} h(\boldsymbol{x},\boldsymbol{z}) \le h(\boldsymbol{x},\boldsymbol{\mu})+O(\epsilon)\,.\end{aligned}$$

We are left with demonstrating near-optimality of the MVP solution. This is easily done using the following argument, where for conciseness we let $\boldsymbol{x}_{\text{MVP}}$ refer to the optimal solution of the MVP problem, $g(\boldsymbol{x})$ be equal to $\sup_{F\in\bar{\mathcal{D}}(\mathcal{S},\boldsymbol{\mu},\Psi,\epsilon)} \boldsymbol{c}_1^\mathsf{T}\boldsymbol{x}+\mathbb{E}_F[h(\boldsymbol{x},\boldsymbol{\xi})]$, and $\boldsymbol{x}_g$ be the member of $\mathcal{X}$ that minimizes $g(\boldsymbol{x})$.

$$\begin{aligned}g(\boldsymbol{x}_{\text{MVP}})-g(\boldsymbol{x}_g) &\le \boldsymbol{c}_1^\mathsf{T}\boldsymbol{x}_{\text{MVP}}+h(\boldsymbol{x}_{\text{MVP}},\boldsymbol{\mu})+O(\epsilon)-\boldsymbol{c}_1^\mathsf{T}\boldsymbol{x}_g-h(\boldsymbol{x}_g,\boldsymbol{\mu}) \\ &= \min_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{c}_1^\mathsf{T}\boldsymbol{x}+h(\boldsymbol{x},\boldsymbol{\mu})-(\boldsymbol{c}_1^\mathsf{T}\boldsymbol{x}_g+h(\boldsymbol{x}_g,\boldsymbol{\mu}))+O(\epsilon) \le O(\epsilon)\,,\end{aligned}$$

so that $g(\boldsymbol{x}_{\text{MVP}})\le g(\boldsymbol{x}_g)+O(\epsilon)$. $\square$

EC.5. Proof of Lemma 2

We first re-parametrize as $z(\Delta,\delta):=2\boldsymbol{\mu}+\delta\Delta$. We therefore need to show that for any $\Delta\in\mathbb{R}^d$

$$g(\delta):=\|\boldsymbol{\mu}+\delta\Delta\|_\alpha^\gamma-\frac{1}{2^{\gamma-1}}\|2\boldsymbol{\mu}+\delta\Delta\|_\alpha^\gamma+\|\boldsymbol{\mu}\|_\alpha^\gamma\ge 0\,,\ \forall\,\delta\in\mathbb{R}\,.$$

This is done by showing that $g(\delta)$ is decreasing for negative δ's, achieves the value of zero at $\delta=0$ and is increasing for positive δ's. The function $g(\delta)$ therefore achieves its minimum value of zero at $\delta=0$ for any Δ. The simplest step is to show that $g(0)=0$.

$$g(0)=\|\boldsymbol{\mu}\|_\alpha^\gamma-\frac{1}{2^{\gamma-1}}\|2\boldsymbol{\mu}\|_\alpha^\gamma+\|\boldsymbol{\mu}\|_\alpha^\gamma=0\,.$$

The second step requires us to realize that $g(\delta)$ can be represented in terms of the convex function $g_2(\delta)=\|\mu+\delta\Delta\|_\alpha^\gamma$:

$$g(\delta)=g_2(\delta)-2g_2(\delta/2)+\|\mu\|_\alpha^\gamma\,.$$

One can verify that $g_2(\delta)$ is convex using the fact that it is the composition of a function y^γ, which is convex and increasing over $y\ge 0$ for $\gamma\ge 1$, and of a convex function $\|\mu+\delta\Delta\|_\alpha$. The convexity of $g_2(\delta)$ tells us that $g_2'(\delta)\ge g_2'(\delta/2)$ for $\delta\ge 0$ and that $g_2'(\delta)\le g_2'(\delta/2)$ for $\delta\le 0$. Thus, we can easily verify the properties of the derivative of $g(\delta)$. While for $\delta\ge 0$, we have $g'(\delta)=g_2'(\delta)-g_2'(\delta/2)\ge 0$, we can also easily show that, for $\delta\le 0$, we have $g'(\delta)=g_2'(\delta)-g_2'(\delta/2)\le 0$. This completes our proof. $\square$

EC.6. Proof of Lemma 3

To make this demonstration, we use the fact that $h(x_2,\boldsymbol{\xi}) = x_2 h(1,\boldsymbol{\xi})$. First, in the case where $x_2 = 0$, we already verified that $h(0,\boldsymbol{\xi}) = 0$. When $x_2 > 0$, one can easily show that $h(x_2,\boldsymbol{\xi}) = x_2 h(1,\boldsymbol{\xi})$ for all $\boldsymbol{\xi} \in \mathbb{R}^d$ by replacing the inner variable $\boldsymbol{y}$ by $\boldsymbol{z} = \boldsymbol{y}/x_2$. Thus, for all $x_2 \geq 0$, the objective of problem (6) reduces to

$$\begin{aligned}\sup_{F\in\mathcal{D}(\mathbb{R}^d,\mathbf{0}_d,\mathbf{I})} \{\mathbb{E}_F[-cx_2 - h(x_2,\boldsymbol{\xi})]\} &= \sup_{F\in\mathcal{D}(\mathbb{R}^d,\mathbf{0}_d,\mathbf{I})} \{\mathbb{E}_F[(-c-h(1,\boldsymbol{\xi}))x_2]\} \\ &= \sup_{F\in\mathcal{D}(\mathbb{R}^d,\mathbf{0}_d,\mathbf{I})} \{\mathbb{E}_F[(-c-h(1,\boldsymbol{\xi}))]\}\, x_2 . \quad \square\end{aligned}$$

EC.7. Proof of Lemma 4

Based on Lemma 1 of Delage and Ye (2010), considering that $h(1,\boldsymbol{\xi})$ is F-integrable for all $F \in \mathcal{D}(\mathbb{R}^d,\mathbf{0}_d,\mathbf{I})$ since

$$|\mathbb{E}_F[h(1,\boldsymbol{\xi})]| \leq \mathbb{E}_F[\|\boldsymbol{\xi}\|_1] \leq \sqrt{d}\mathbb{E}_F[\|\boldsymbol{\xi}\|_2] \leq \sqrt{d}\sqrt{\mathbb{E}_F[\|\boldsymbol{\xi}\|_2^2]} = d\,,$$

by duality theory we can say that evaluating the supremum of such an expression is equivalent to finding the optimal value of

$$\begin{aligned}\underset{\boldsymbol{Q},\boldsymbol{q},t}{\text{minimize}} \quad & t + \mathbf{I} \bullet \boldsymbol{Q} \\ \text{subject to} \quad & t \geq \max_{\boldsymbol{y}\in\mathcal{Y}(1)} -c - \boldsymbol{\xi}^\mathsf{T}\boldsymbol{y} - \boldsymbol{\xi}^\mathsf{T}\boldsymbol{Q}\boldsymbol{\xi} - \boldsymbol{q}^\mathsf{T}\boldsymbol{\xi}\,, \forall \boldsymbol{\xi} \in \mathbb{R}^d\,,\end{aligned}$$

where $\mathcal{Y}(1) = \{\boldsymbol{y} \in \mathbb{R}^d | \boldsymbol{a}^\mathsf{T}\boldsymbol{y} = 0 \;\&\; -1 \leq y_i \leq 1\,, \forall i\}$. This is necessarily a convex optimization problem with linear objective function since each constraint is jointly convex in $\boldsymbol{Q}$, $\boldsymbol{q}$, and t. We now show that the separation problem associated with the only constraint of this problem is NP-hard. Thus, by the equivalence of optimization and separation (see Grötschel et al. (1981)) finding the optimal value of this problem can be shown to be NP-hard.

Consider separating the solution $\boldsymbol{Q} = (1/4)\mathbf{I}$, $\boldsymbol{q} = \mathbf{0}_m$, and $t = d - c$ from the feasible set. In this case, we must be able to verify its feasibility with respect to the only constraint. This can be shown to reduce to verifying if

$$\sup_{\boldsymbol{\xi}\in\mathbb{R}^d,\, \boldsymbol{y}\in\mathcal{Y}(1)} -\boldsymbol{\xi}^\mathsf{T}\boldsymbol{y} - (1/4)\boldsymbol{\xi}^\mathsf{T}\boldsymbol{\xi} \leq d\,.$$

After solving in terms of $\boldsymbol{\xi}$, we get that we need to verify if the optimal value of the problem

$$\begin{aligned}\underset{\boldsymbol{y}}{\text{maximize}} \quad & \boldsymbol{y}^\mathsf{T}\boldsymbol{y} \\ \text{subject to} \quad & -1 \leq y_i \leq 1\,, \forall i \in \{1,2,...,d\} \\ & \boldsymbol{a}^\mathsf{T}\boldsymbol{y} = 0\;,\end{aligned}$$

is greater or equal to d or not. This is equivalent to showing if the set defined by

$$\Upsilon(\boldsymbol{a}) = \{\, \boldsymbol{y} \in \{-1,1\}^d \mid \boldsymbol{a}^\top \boldsymbol{y} = 0 \,\}\,,$$

for any $\boldsymbol{a} \in \mathbb{R}^d$, is empty or not since the extreme point of the unit box are the only ones for which $\|\boldsymbol{y}\|^2 \geq d$. Finally, one can easily confirm that the NP-complete Partition problem can be reduced to verifying whether some $\Upsilon(\boldsymbol{a})$ is empty or not.

Partition Problem: Given a set of positive integers $\{s_i\}_{i=1}^d$ with index set $\mathcal{A} = \{1,2,...,d\}$ is there a partition $(\mathcal{B}_1, \mathcal{B}_2)$ of $\mathcal{A}$ such that $\sum_{i\in\mathcal{B}_1} s_i = \sum_{i\in\mathcal{B}_2} s_i$?

The reduction is obtained by casting $a_i := s_i$ for all i, and considering that a feasible solution $\boldsymbol{y} \in \Upsilon(\boldsymbol{a})$ only exists if such a partition exists and identifies the partition through $\mathcal{B}_1 = \{i \in \mathcal{A} | y_i = 1\}$. This completes our proof. □

EC.8. Proof of Proposition 6

We first represent $\mathcal{D}(\mathcal{S}, \boldsymbol{\mu}, \mathbf{I})$ in the form proposed in Example 1, i.e., using

$$\Psi = \{\psi : \mathbb{R}^d \to \mathbb{R} \mid \exists \boldsymbol{Q} \succeq 0\,,\ \psi(\boldsymbol{\xi}) = \boldsymbol{Q} \bullet ((\boldsymbol{\xi} - \boldsymbol{\mu})(\boldsymbol{\xi} - \boldsymbol{\mu})^\top - \mathbf{I})\}$$

It is indeed the case that this set Ψ satisfies Assumption 3 since the positive semi-definite cone is one for which it is relatively easy to verify feasibility using singular value decomposition. In formulating problem (8), we get that constraint (8b) takes the shape:

$$t \geq w_i^k + \xi v_i^k - \alpha_i^k - \beta_i^k(\xi - \xi_i^k) - q_i(\xi - \mu_i) - \boldsymbol{Q} \bullet ((\xi - \mu_i)^2 \boldsymbol{e}_i \boldsymbol{e}_i^\top - \mathbf{I})\,, \begin{cases} \forall \xi \in [\nu_i - \tau_i, \nu_i + \tau_i] \\ \forall i \in \{1,2,...,d\} \\ \forall k \in \{1,2,...,K\} \end{cases}$$

A simple replacement of t for $t + \mathbf{I} \bullet \boldsymbol{Q}$ leads to the equivalent formulation of problem (8)

$$\begin{aligned} \underset{t,\boldsymbol{q},\boldsymbol{Q},\{\boldsymbol{w}^k,\boldsymbol{v}^k\}_{k=1}^K}{\text{minimize}} \quad & t + \mathbf{I} \bullet \boldsymbol{Q} \\ \text{subject to} \quad & t \geq w_i^k + \xi v_i^k - \alpha_i^k - \beta_i^k(\xi - \xi_i^k) - q_i(\xi - \mu_i) - (\xi - \mu_i)^2 Q_{i,i}\,, \begin{cases} \forall \xi \in [\nu_i - \tau_i, \nu_i + \tau_i] \\ \forall i \in \{1,2,...,d\} \\ \forall k \in \{1,2,...,K\} \end{cases} \\ & w_i^k + v_i^k \xi_i^m \geq \boldsymbol{c}_1^\top \boldsymbol{x}_1 + h(\boldsymbol{x}_1, \boldsymbol{\mu} + (\xi_i^m - \mu_i)\boldsymbol{e}_i)\,, \begin{cases} \forall i \in \{1,2,...,d\} \\ \forall m,\, k \in \{1,2,...,K\} \end{cases} \\ & \boldsymbol{Q} \succeq 0\,. \end{aligned}$$

Since only the diagonal terms of $\boldsymbol{Q}$ are involved in both the objective function and the constraints, one can arbitrarily set all off diagonal terms of $\boldsymbol{Q}$ to zero. Thus, we are left with

$$\underset{t,\boldsymbol{q},\boldsymbol{r},\{\boldsymbol{w}^k,\boldsymbol{v}^k\}_{k=1}^K}{\text{minimize}} \quad t + \boldsymbol{e}^\top \boldsymbol{r}$$

$$\text{subject to} \quad t \geq w_i^k + \xi v_i^k - \alpha_i^k - \beta_i^k(\xi - \xi_i^k) - q_i(\xi - \mu_i) - (\xi_i - \mu_i)^2 r_i\,, \begin{cases} \forall \xi \in [\nu_i - \tau_i, \nu_i + \tau_i] \\ \forall i \in \{1,2,...,d\} \\ \forall k \in \{1,2,...,K\} \end{cases}$$

$$w_i^k + v_i^k \xi_i^m \geq \boldsymbol{c}_1^\mathsf{T}\boldsymbol{x}_1 + h(\boldsymbol{x}_1, \boldsymbol{\mu} + (\xi_i^m - \mu_i)\boldsymbol{e}_i)\,, \begin{cases} \forall i \in \{1,2,...,d\} \\ \forall m,\, k \in \{1,2,...,K\} \end{cases}$$

$$\boldsymbol{r} \geq 0\,.$$

We finally apply the S-Lemma (cf., Theorem 2.2 in Pólik and Terlaky (2007)) to replace, for each fixed i and k, the set of constraints indexed over the interval $[\nu_i - \tau_i, \nu_i + \tau_i]$ by an equivalent linear matrix inequality:

$$\begin{bmatrix} r_i & \frac{\beta_i^k - v_i^k + q_i - 2\mu_i r_i}{2} \\ \frac{\beta_i^k - v_i^k + q_i - 2\mu_i r_i}{2} & t - w_i^k + \alpha_i^k - \beta_i^k \xi_i^k - \mu_i q_i + \mu_i^2 r_i \end{bmatrix} \succeq -s_i^k \begin{bmatrix} 1 & -\nu_i \\ -\nu_i & \tau_i^2 \end{bmatrix} \quad , \quad s_i^k \geq 0\,.$$

Now regarding the complexity of solving this semi-definite programming problem, it is well known that in the standard form

$$\begin{aligned} &\underset{\boldsymbol{x}\in\mathbb{R}^{\tilde{n}}}{\text{minimize}} \quad \boldsymbol{c}^\mathsf{T}\boldsymbol{x} \\ &\text{subject to} \quad \boldsymbol{A}_i(\boldsymbol{x}) \succeq 0 \quad \forall i = 1,2,...,\tilde{K} \end{aligned}$$

the problem can be solved in $O\left(\left(\sum_{i=1}^{\tilde{K}} \tilde{m}_i\right)^{0.5}\left(\tilde{n}^2 \sum_{i=1}^{\tilde{K}} \tilde{m}_i^2 + \tilde{n}\sum_{i=1}^{\tilde{K}} \tilde{m}_i^3\right)\right)$, where $\tilde{m}_i$ stands for the dimension of the positive semi-definite cone (*i.e.*, $\boldsymbol{A}_i(\boldsymbol{x}) \in \mathbb{R}^{\tilde{m}_i \times \tilde{m}_i}$) (see Nesterov and Nemirovski (1994)). In the SDP that interest us, one can show that $\tilde{n} = 1 + (3K+2)d$ and that both problems can be solved in $O(K^5 d^{3.5})$ operations, with K being the number of scenarios for each random variable that composes $\boldsymbol{\xi}$. In calculating the total complexity of evaluating $\mathcal{LB}(\boldsymbol{x}_1, \mathcal{X}_2, \{\boldsymbol{\xi}_i^k\})$ under such a Ψ, we also need to account for $O(KdT_{\text{MVP}})$ operations for evaluating the required α_i^k and β_i^k. This completes our proof. □

EC.9. Proof of Proposition 7

We first present without proof a Lemma that describes how to approximate a concave function on the real line to any level of accuracy by containing it between an outer and an inner piecewise linear concave functions.

Lemma EC.3. *Given any set of points $\{z_i\}_{k=1}^K$, a concave function $g:\mathbb{R}\to\mathbb{R}$ is contained between the two piecewise linear concave functions. Specifically, $g_{inner}(z) \leq g(z) \leq g_{outer}(z)$, where*

$$g_{outer}(z) := \min_{k\in\{1,2,\ldots,K\}} g(z_k) + (z - z_k)g'(z_k)\,,$$

with $g'(z_k) \in \mathbb{R}$ as any super-gradient of $g(z)$ at z_k, and

$$\begin{aligned} g_{inner}(z) := \quad &\min_{w,v} \quad w + zv \\ &\text{subject to} \quad w + z_k v \geq g(z_k)\,,\ \forall k \in \{1,2,...,K\}\,. \end{aligned}$$

We then reduce the intractable task of evaluating MVSM($\boldsymbol{x}_1$) to the search for a distribution in the more manageable set $\mathcal{D}(\mathcal{S}_0,\boldsymbol{\mu},\Psi)$. The analysis is also limited to the potential regret of having committed to $\boldsymbol{x}_1$ instead of $\hat{\boldsymbol{x}}_2$. This naturally leads to a lower bound for the *MVSM* which considers distributions in the larger set $\mathcal{D}(\mathcal{S},\boldsymbol{\mu},\Psi)$ and any alternative decision in $\mathcal{X}_2$

$$\mathrm{MVSM}(\boldsymbol{x}_1)\geq \sup_{F\in\mathcal{D}(\mathcal{S}_0,\boldsymbol{\mu},\Psi)} \mathbb{E}_F[\boldsymbol{c}_1^\top(\boldsymbol{x}_1-\hat{\boldsymbol{x}}_2)+h(\boldsymbol{x}_1,\boldsymbol{\xi})-h(\hat{\boldsymbol{x}}_2,\boldsymbol{\xi})]$$

We then apply duality theory to the $\sup_{F\in\mathcal{D}(\mathcal{S}_0,\boldsymbol{\mu},\Psi)}$ operation. Considering the primal problem as a semi-infinite linear conic problem

$$\begin{aligned}
&\sup_{F\in\mathcal{M}} && \mathrm{E}_F[\boldsymbol{c}_1^\top(\boldsymbol{x}_1-\hat{\boldsymbol{x}}_2)+h(\boldsymbol{x}_1,\boldsymbol{\xi})-h(\hat{\boldsymbol{x}}_2,\boldsymbol{\xi})] && \text{(EC.1a)}\\
&\text{subject to} && \mathbb{E}_F[\mathbb{1}\{\boldsymbol{\xi}\in\mathcal{S}_0\}]=1 && \text{(EC.1b)}\\
& && \mathbb{E}_F[\boldsymbol{\xi}]=\boldsymbol{\mu} && \text{(EC.1c)}\\
& && \mathbb{E}_F[\boldsymbol{r}^\top\boldsymbol{\psi}(\boldsymbol{\xi})]\leq 0\,,\ \forall\boldsymbol{r}\in\mathcal{K}\,, && \text{(EC.1d)}
\end{aligned}$$

the dual form takes the shape

$$\begin{aligned}
&\underset{t,\boldsymbol{q},\boldsymbol{r}}{\text{minimize}} && t && \text{(EC.2a)}\\
&\text{subject to} && t\geq \boldsymbol{c}_1^\top(\boldsymbol{x}_1-\hat{\boldsymbol{x}}_2)+h(\boldsymbol{x}_1,\boldsymbol{\xi})-h(\hat{\boldsymbol{x}}_2,\boldsymbol{\xi})-(\boldsymbol{\xi}-\boldsymbol{\mu})^\top\boldsymbol{q}-\boldsymbol{r}^\top\boldsymbol{\psi}(\boldsymbol{\xi})\,,\ \forall\boldsymbol{\xi}\in\mathcal{S}_0 && \text{(EC.2b)}\\
& && \boldsymbol{r}\in\mathcal{K}\,, && \text{(EC.2c)}
\end{aligned}$$

where t, $\boldsymbol{q}$, and $\boldsymbol{r}$ are the dual variables associated respectively with constraints (EC.1b), (EC.1c), and (EC.1d). One can verify that there is no duality gap between primal and dual problems using the weaker version of Proposition 3.4 in Shapiro (2001) and the fact that the Dirac measure $\delta_{\boldsymbol{\mu}}$ lies in the relative interior of $\mathcal{D}(\mathcal{S}_0,\boldsymbol{\mu},\Psi)$.

Exploiting the structure of $\mathcal{S}_0\subseteq\bigcup_{i=1}^d\{\boldsymbol{\xi}\,|\,\exists\,\xi\in[\nu_i-\tau_i,\nu_i+\tau_i]\,,\ \boldsymbol{\xi}=\boldsymbol{\mu}+(\xi-\mu_i)\boldsymbol{e}_i\}$, we can decompose constraint (EC.2b) into d simpler constraints

$$t\geq \max_{\xi\in[\nu_i-\tau_i,\nu_i+\tau_i]} g_i(\boldsymbol{x}_1,\xi)-g_i(\hat{\boldsymbol{x}}_2,\xi)-q_i(\xi-\mu_i)-\boldsymbol{r}^\top\boldsymbol{\psi}(\boldsymbol{\mu}+(\xi-\mu_i)\boldsymbol{e}_i)\ \forall i\in\{1,2,...,d\} \quad \text{(EC.3)}$$

where $g_i(\boldsymbol{x},\xi)=\boldsymbol{c}_1^\top\boldsymbol{x}+h(\boldsymbol{x},\boldsymbol{\mu}+(\xi-\mu_i)\boldsymbol{e}_i)$. Note that if $S_0=S_0\cap\mathcal{B}_\infty(\boldsymbol{\nu},\boldsymbol{\tau})$, then this set of constraints is entirely equivalent to the original one.

For any fixed i, we now go through a sequence of relaxation steps for constraint (EC.3).

$$\begin{aligned}
t&\geq \max_{\xi\in[\nu_i-\tau_i,\nu_i+\tau_i]} g_i(\boldsymbol{x}_1,\xi)-g_i(\hat{\boldsymbol{x}}_2,\xi)-q_i(\xi-\mu_i)-\boldsymbol{r}^\top\boldsymbol{\psi}(\boldsymbol{\mu}+(\xi-\mu_i)\boldsymbol{e}_i)\\
&\geq \max_{\xi\in[\nu_i-\tau_i,\nu_i+\tau_i]} g_i(\boldsymbol{x}_1,\xi)-\min_k\{\alpha_i^k+(\xi-\xi_i^k)\beta_i^k\}-q_i(\xi-\mu_i)-\boldsymbol{r}^\top\boldsymbol{\psi}(\boldsymbol{\mu}+(\xi-\mu_i)\boldsymbol{e}_i)\\
&=\max_k \max_{\xi\in[\nu_i-\tau_i,\nu_i+\tau_i]} g_i(\boldsymbol{x}_1,\xi)-\alpha_i^k-(\xi-\xi_i^k)\beta_i^k-q_i(\xi-\mu_i)-\boldsymbol{r}^\top\boldsymbol{\psi}(\boldsymbol{\mu}+(\xi-\mu_i)\boldsymbol{e}_i)\\
&\geq \max_k \max_{\xi\in[\nu_i-\tau_i,\nu_i+\tau_i]} \min_{(w,v)\in\mathcal{W}_i}\{w+\xi v\}-\alpha_i^k-\beta_i^k(\xi-\xi_i^k)-q_i(\xi-\mu_i)-\boldsymbol{r}^\top\boldsymbol{\psi}(\boldsymbol{\mu}+(\xi-\mu_i)\boldsymbol{e}_i)\\
&=\max_k \min_{(w,v)\in\mathcal{W}_i} \max_{\xi\in[\nu_i-\tau_i,\nu_i+\tau_i]} w+\xi v-\alpha_i^k-\beta_i^k(\xi-\xi_i^k)-q_i(\xi-\mu_i)-\boldsymbol{r}^\top\boldsymbol{\psi}(\boldsymbol{\mu}+(\xi-\mu_i)\boldsymbol{e}_i)\,,
\end{aligned}$$

where $\mathcal{W}_i$ is short for $\{(w,v) \in \mathbb{R}^2 \mid w + \xi_m v \geq g_i(\boldsymbol{x}_1, \xi_m) \forall m \in \{1,2,...,K\}\}$. The two relaxation steps are obtained through the application of Lemma EC.3 to first replace $g(\boldsymbol{x}_2, \boldsymbol{\xi})$ by its outer approximation, and then $g(x_1, \boldsymbol{\xi})$ by its inner approximation. The last equality is obtained by inversing the order of $\max_{\xi \in [\nu_i - \tau_i, \nu_i + \tau_i]}$ and $\min_{(w,v) \in \mathcal{W}_i}$ using Sion's minimax theorem (see Sion (1958)). Thus, we obtain the optimization problem presented in Definition 1. $\square$

References

Ben-Tal, A., A. Nemirovski. 1998. Robust convex optimization. *Mathematics of Operations Research* **23**(4) 769–805.

Boyd, S., L. Vandenberghe. 2004. *Convex Optimization*. 1st ed. Cambridge University Press.

Delage, E., Y. Ye. 2010. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research* **58**(3) 595–612.

Grötschel, M., L. Lovász, A. Schrijver. 1981. The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica* **1** 169–197.

Nesterov, Y., A. Nemirovski. 1994. *Interior-point polynomial methods in convex programming*, vol. 13. Studies in Applied Mathematics, Philadelphia.

Pólik, I., T. Terlaky. 2007. A survey of the S-Lemma. *SIAM Review* **49**(3) 371–418.

Shapiro, A. 2001. On duality theory of conic linear problems. M. A. Goberna, M. A. López, eds., *Semi-Infinite Programming: Recent Advances*. Kluwer Academic Publishers, Dordrecht, 135–165.

Sion, M. 1958. On general minimax theorems. *Pacific Journal of Mathematics* **8**(1) 171–176.