# Online Appendices:
Supermodularity in Two-Stage Distributionally Robust Optimization

## A. Supplement

### A.1. Algorithm for checking supermodularity in Section 3

For any given matrices $\boldsymbol{U}, \boldsymbol{V}$, we provide the following algorithm to check explicitly whether the condition in Theorem 2 is met.

**Algorithm 2** algorithm for checking supermodularity

1: **Input:** $\boldsymbol{U} \in \Re^{r\times m}, \boldsymbol{V} \in \Re^{r\times n}$
2: **Initialization:** $r_0 = \mathrm{rank}(\boldsymbol{U}), s = 1$
3: **if** $r_0 < r$ **then**
4:     arbitrarily remove columns in $\boldsymbol{U}$, if any, until $\boldsymbol{U}$ has only $r_0$ linearly independent columns
5:     **for all** $\mathcal{I} \subseteq [r]$ with $|\mathcal{I}| = r_0$ and $\boldsymbol{U}_{\mathcal{I}}$ invertible, **do**
6:         **for** $i \in [r]\backslash\mathcal{I}$ **do**
7:             $\boldsymbol{d}_i^\top = \boldsymbol{v}_i^\top - \boldsymbol{u}_i^\top \boldsymbol{U}_{\mathcal{I}}^{-1} \boldsymbol{V}_{\mathcal{I}}$
8:             **if** there exist components $d_{ia}, d_{ib}$ such that $d_{ia}d_{ib} < 0$ **then**
9:                 $s = 0$, **go to** line 10
10: **return** $s$

**Theorem 5** *The condition in Theorem 2 is satisfied if and only if Algorithm 2 returns* $s = 1$.

We note that Algorithm 2 may take exponential number of steps. Specifically, the complexity is reflected in line 5, where we search for all row index sets subject to conditions on the number of rows and rank. The high complexity is essentially because this algorithm is for the necessary and sufficient condition. Indeed, if we aim only for necessary conditions, then it can be simplified by reducing the range of search. For example, only checking for submatrices containing consecutive rows of $\boldsymbol{U}$ and $\boldsymbol{V}$ can also be a necessary condition. If the condition is violated for any tested index set $\mathcal{I}$, then the function $g(\boldsymbol{x}, \boldsymbol{z})$ must not be supermodular for all $\boldsymbol{x}, \boldsymbol{b}, \boldsymbol{v}^0$. On the other hand, if we aim at sufficient conditions only, some matrices with simple structures can be easily shown to satisfy the conditions.

### A.2. A counterexample where segregated affine decision rules are suboptimal in Section 4

For simplicity, we drop the first-stage decision and consider the problem

$$\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[g(\tilde{z}_1,\tilde{z}_2)\right]$$

as an example, where we define

$$\mathcal{F}=\left\{\mathbb{P}\;\middle|\;\begin{array}{l}\mathbb{E}_{\mathbb{P}}[\tilde{z}_1]=3,\ \mathbb{E}_{\mathbb{P}}[\tilde{z}_2]=1\\ \mathbb{E}_{\mathbb{P}}[|\tilde{z}_1-3|]\le 1.2,\ \mathbb{E}_{\mathbb{P}}[|\tilde{z}_2-1|]\le 1.2\\ \mathbb{P}(\tilde{z}_1\in[0,4])=\mathbb{P}(\tilde{z}_2\in[0,4])=1\end{array}\right\},\qquad g(z_1,z_2)=\min\left\{y\;\middle|\;\begin{array}{l}y\ge z_1+z_2-4\\ y\ge -z_1-z_2+4\\ y\le 4\end{array}\right\}$$

for the problem setting. Easily we can check that $g(z_1,z_2)$ satisfies the condition in Theorem 2, hence is supermodular in $\boldsymbol{z}$. Applying Algorithm 1 we obtain a worst-case distribution $\mathbb{P}^\star$ with $\mathbb{P}^\star(\tilde{\boldsymbol{z}}=(0,0))=\mathbb{P}^\star(\tilde{\boldsymbol{z}}=(3,0))=\mathbb{P}^\star(\tilde{\boldsymbol{z}}=(4,0))=\mathbb{P}^\star(\tilde{\boldsymbol{z}}=(4,1))=\mathbb{P}^\star(\tilde{\boldsymbol{z}}=(4,4))=0.2$, hence the problem has an optimal value

$$\begin{aligned}\sup_{\mathbb{P}\in\mathcal{F}}\mathbb{E}_{\mathbb{P}}[g(\tilde{z}_1,\tilde{z}_2)]&=\mathbb{E}_{\mathbb{P}^\star}[g(\tilde{z}_1,\tilde{z}_2)]\\&=0.2\left(g(0,0)+g(3,0)+g(4,0)+g(4,1)+g(4,4)\right)\\&=0.2\left(4+1+0+1+4\right)\\&=2.\end{aligned}$$

On the other hand, we consider the decision following the segregated linear decision rule, i.e., $y(z_1,z_2)=\boldsymbol{\theta}^\top\left((3-z_1)^+,(1-z_2)^+,(z_1-3)^+,(z_2-1)^+\right)+\phi$. The problem with segregated linear decision rule (The detailed reformulation can be referred to Problem (26) in Appendix A.3) then becomes

$$\begin{aligned}\min\quad&\boldsymbol{\theta}^\top(0.6,0.6,0.6,0.6)+\phi\\ \text{s.t.}\quad&\boldsymbol{\theta}^\top\left((3-z_1)^+,(1-z_2)^+,(z_1-3)^+,(z_2-1)^+\right)\ge z_1+z_2-4,\qquad \forall(z_1,z_2)\in\{0,3,4\}\times\{0,1,4\}\\&\boldsymbol{\theta}^\top\left((3-z_1)^+,(1-z_2)^+,(z_1-3)^+,(z_2-1)^+\right)\ge -z_1-z_2+4,\quad \forall(z_1,z_2)\in\{0,3,4\}\times\{0,1,4\}\\&\boldsymbol{\theta}^\top\left((3-z_1)^+,(1-z_2)^+,(z_1-3)^+,(z_2-1)^+\right)\le 4,\qquad\qquad \forall(z_1,z_2)\in\{0,3,4\}\times\{0,1,4\}.\end{aligned}$$

This is a linear program with 27 constraints and 5 decision variables, hence can be handled by ordinary solvers. The optimal segregated linear decision rule turns out to be $y(z_1,z_2)=\frac{1}{3}(3-z_1)^+ + (1-z_2)^+ + (z_1-3)^+ + \frac{1}{3}(z_2-1)^+ + 2$, yielding the optimal value $0.6(\frac{1}{3}+1+1+\frac{1}{3})+2=3.6>2$. This implies that segregated linear decision rules can be suboptimal even if the second-stage cost is supermodular in $\boldsymbol{z}$.

### A.3. The CCG algorithm and segregated affine decision rules in Section 6

**CCG algorithm**

The CCG algorithm we use in this paper mostly resembles Saif and Delage (2021), who study a two-stage distributionally robust facility location problem. For Problem (2), the inner supremum problem $\sup_{\mathbb{P}\in\mathcal{F}}$ can be equivalently written as an infimum problem and the strong duality holds. Then, we reformulate Problem (2) as

$$\begin{aligned} \min\quad & \boldsymbol{a}^\top\boldsymbol{x}+\alpha+\boldsymbol{\mu}^\top\boldsymbol{\beta}+\boldsymbol{\delta}^\top\boldsymbol{\gamma} \\ \text{s.t.}\quad & \alpha+\boldsymbol{\beta}^\top\boldsymbol{z}+\boldsymbol{\gamma}^\top(|\boldsymbol{z}-\boldsymbol{\mu}|)\geq g(\boldsymbol{x},\boldsymbol{z}),\quad \forall\boldsymbol{z}\in[\underline{\boldsymbol{z}},\overline{\boldsymbol{z}}] \\ & \boldsymbol{\gamma}\geq\mathbf{0},\ \boldsymbol{x}\in\mathcal{X}. \end{aligned} \tag{20}$$

For the CCG algorithm, we assume that the relatively complete recourse holds and consider multiple iterations. In iteration $\tau$ ($\tau=0,1,2,\ldots$), we solve a relaxation of Problem (20) that imposes the first constraint to hold only on a subset $\{\boldsymbol{z}^0,\boldsymbol{z}^1,\ldots,\boldsymbol{z}^\tau\}\subseteq[\underline{\boldsymbol{z}},\overline{\boldsymbol{z}}]$. Formally, let

$$\begin{aligned} \hat{L}_\tau:=\min\quad & \boldsymbol{a}^\top\boldsymbol{x}+\alpha+\boldsymbol{\mu}^\top\boldsymbol{\beta}+\boldsymbol{\delta}^\top\boldsymbol{\gamma} \\ \text{s.t.}\quad & \alpha+\boldsymbol{\beta}^\top\boldsymbol{z}^i+\boldsymbol{\gamma}^\top(|\boldsymbol{z}^i-\boldsymbol{\mu}|)\geq\boldsymbol{b}^\top\boldsymbol{y}^i,\quad i\in[\tau] \\ & \boldsymbol{W}\boldsymbol{x}+\boldsymbol{U}\boldsymbol{y}^i\geq\boldsymbol{V}\boldsymbol{z}^i+\boldsymbol{v}^0,\qquad i\in[\tau] \\ & \boldsymbol{\gamma}\geq\mathbf{0},\ \boldsymbol{x}\in\mathcal{X}. \end{aligned} \tag{21}$$

This problem provides a lower bound to the optimal value of the original problem (2). Denoting by $\left(\hat{\boldsymbol{x}},\hat{\alpha},\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\gamma}},\{\hat{\boldsymbol{y}}^i\}_{i\in[\tau]}\right)$ a solution to Problem (21), we then examine the violation of the constraint in Problem (20) by evaluating

$$\begin{aligned} h\left(\hat{\boldsymbol{x}},\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\gamma}}\right) &= \max\left\{ g(\hat{\boldsymbol{x}},\boldsymbol{z})-\hat{\boldsymbol{\beta}}^\top\boldsymbol{z}-\hat{\boldsymbol{\gamma}}^\top(|\boldsymbol{z}-\boldsymbol{\mu}|) \;\middle|\; \boldsymbol{z}\in[\underline{\boldsymbol{z}},\overline{\boldsymbol{z}}] \right\} \\ &= \max\left\{ g(\hat{\boldsymbol{x}},\boldsymbol{z})-\hat{\boldsymbol{\beta}}^\top\boldsymbol{z}-\hat{\boldsymbol{\gamma}}^\top\boldsymbol{\theta} \;\middle|\; \begin{array}{l} \boldsymbol{\theta}\geq\boldsymbol{z}-\boldsymbol{\mu},\ \boldsymbol{\theta}\geq\boldsymbol{\mu}-\boldsymbol{z},\ \boldsymbol{\theta}\leq M_0\mathbf{1} \\ \boldsymbol{z}\in[\underline{\boldsymbol{z}},\overline{\boldsymbol{z}}] \end{array} \right\} \\ &= \max\left\{ (\boldsymbol{V}\boldsymbol{z}+\boldsymbol{v}^0-\boldsymbol{W}\hat{\boldsymbol{x}})^\top\boldsymbol{\eta}-\hat{\boldsymbol{\beta}}^\top\boldsymbol{z}-\hat{\boldsymbol{\gamma}}^\top\boldsymbol{\theta} \;\middle|\; \begin{array}{l} \boldsymbol{U}^\top\boldsymbol{\eta}=\boldsymbol{b},\ \boldsymbol{\eta}\geq\mathbf{0} \\ \boldsymbol{\theta}\geq\boldsymbol{z}-\boldsymbol{\mu},\ \boldsymbol{\theta}\geq\boldsymbol{\mu}-\boldsymbol{z},\ \boldsymbol{\theta}\leq M_0\mathbf{1} \\ \boldsymbol{z}\in[\underline{\boldsymbol{z}},\overline{\boldsymbol{z}}] \end{array} \right\} \\ &= \max_{\substack{\boldsymbol{\eta}:\boldsymbol{U}^\top\boldsymbol{\eta}=\boldsymbol{b} \\ \boldsymbol{\eta}\geq\mathbf{0}}} \left\{ (\boldsymbol{v}^0-\boldsymbol{W}\hat{\boldsymbol{x}})^\top\boldsymbol{\eta}+\max\left\{ (\boldsymbol{V}^\top\boldsymbol{\eta}-\hat{\boldsymbol{\beta}})^\top\boldsymbol{z}-(\hat{\gamma})^\top\boldsymbol{\theta} \;\middle|\; \begin{array}{l} \boldsymbol{z}-\boldsymbol{\theta}\leq\boldsymbol{\mu} \\ -\boldsymbol{z}-\boldsymbol{\theta}\leq-\boldsymbol{\mu} \\ \boldsymbol{\theta}\leq M_0\mathbf{1} \\ \boldsymbol{z}\leq\overline{\boldsymbol{z}} \\ -\boldsymbol{z}\leq-\underline{\boldsymbol{z}} \end{array} \right\} \right\}. \end{aligned} \tag{22}$$

Here the second equality follows from $\hat{\boldsymbol{\gamma}}\geq\mathbf{0}$, and the constraint $\boldsymbol{\theta}\leq M_0\mathbf{1}$ with $M_0$ being a sufficiently large real number guarantees the boundedness of the feasible set while keeping the optimal value unchanged. The third equality follows from the dual form of $g(\hat{\boldsymbol{x}},\boldsymbol{z})$, since the relatively complete course assumption guarantees that the strong duality holds. In the last line we seperate the maximization of $\boldsymbol{\eta}$ and $\boldsymbol{z},\boldsymbol{\theta}$. Further, replacing the constraints of the inner maximization by its KKT conditions, we get

$$
\begin{aligned}
h\left(\hat{\boldsymbol{x}},\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\gamma}}\right)=\max\quad & (\boldsymbol{v}^0-\boldsymbol{W}\hat{\boldsymbol{x}})^\top\boldsymbol{\eta}+\boldsymbol{\mu}^\top(\overline{\boldsymbol{\phi}}-\underline{\boldsymbol{\phi}})+M_0\mathbf{1}^\top\boldsymbol{\xi}+\overline{\boldsymbol{z}}^\top\overline{\boldsymbol{\rho}}-\underline{\boldsymbol{z}}^\top\underline{\boldsymbol{\rho}}\\
\text{s.t.}\quad & \boldsymbol{U}^\top\boldsymbol{\eta}=\boldsymbol{b}\\
& \boldsymbol{z}-\boldsymbol{\theta}\le\boldsymbol{\mu} && (\overline{\boldsymbol{\phi}})\\
& -\boldsymbol{z}-\boldsymbol{\theta}\le-\boldsymbol{\mu} && (\underline{\boldsymbol{\phi}})\\
& \boldsymbol{\theta}\le M_0\mathbf{1} && (\boldsymbol{\xi})\\
& \boldsymbol{z}\le\overline{\boldsymbol{z}} && (\overline{\boldsymbol{\rho}})\\
& -\boldsymbol{z}\le-\underline{\boldsymbol{z}} && (\underline{\boldsymbol{\rho}})\\
& \overline{\boldsymbol{\phi}}-\underline{\boldsymbol{\phi}}+\overline{\boldsymbol{\rho}}-\underline{\boldsymbol{\rho}}=\boldsymbol{V}^\top\boldsymbol{\eta}-\hat{\boldsymbol{\beta}} && (\boldsymbol{z})\\
& \overline{\boldsymbol{\phi}}-\underline{\boldsymbol{\phi}}-\boldsymbol{\xi}=\hat{\boldsymbol{\gamma}} && (\boldsymbol{\theta})\\
& \boldsymbol{\mu}-\boldsymbol{z}+\boldsymbol{\theta}\le M_0\boldsymbol{\lambda}^{\overline{\boldsymbol{\phi}}},\quad \overline{\boldsymbol{\phi}}\le M_0(\mathbf{1}-\boldsymbol{\lambda}^{\overline{\boldsymbol{\phi}}})\\
& -\boldsymbol{\mu}+\boldsymbol{z}+\boldsymbol{\theta}\le M_0\boldsymbol{\lambda}^{\underline{\boldsymbol{\phi}}},\quad \underline{\boldsymbol{\phi}}\le M_0(\mathbf{1}-\boldsymbol{\lambda}^{\underline{\boldsymbol{\phi}}})\\
& M_0\mathbf{1}-\boldsymbol{\theta}\le M_0\boldsymbol{\lambda}^{\boldsymbol{\xi}},\quad \boldsymbol{\xi}\le M_0(\mathbf{1}-\boldsymbol{\lambda}^{\boldsymbol{\xi}})\\
& \overline{\boldsymbol{z}}-\boldsymbol{z}\le M_0\boldsymbol{\lambda}^{\overline{\boldsymbol{\rho}}},\quad \overline{\boldsymbol{\rho}}\le M_0(\mathbf{1}-\boldsymbol{\lambda}^{\overline{\boldsymbol{\rho}}})\\
& \boldsymbol{z}-\underline{\boldsymbol{z}}\le M_0\boldsymbol{\lambda}^{\underline{\boldsymbol{\rho}}},\quad \underline{\boldsymbol{\rho}}\le M_0(\mathbf{1}-\boldsymbol{\lambda}^{\underline{\boldsymbol{\rho}}})\\
& \overline{\boldsymbol{\phi}},\underline{\boldsymbol{\phi}},\boldsymbol{\xi},\overline{\boldsymbol{\rho}},\underline{\boldsymbol{\rho}},\boldsymbol{\eta}\ge\mathbf{0}\\
& \boldsymbol{\lambda}^{\overline{\boldsymbol{\phi}}},\boldsymbol{\lambda}^{\underline{\boldsymbol{\phi}}},\boldsymbol{\lambda}^{\boldsymbol{\xi}},\boldsymbol{\lambda}^{\overline{\boldsymbol{\rho}}},\boldsymbol{\lambda}^{\underline{\boldsymbol{\rho}}}\in\{0,1\}^n.
\end{aligned}
\tag{23}
$$

Here the variables in parentheses specify the associated dual variable for each constraint, and the binary variables $\boldsymbol{\lambda}^{\overline{\boldsymbol{\phi}}},\boldsymbol{\lambda}^{\underline{\boldsymbol{\phi}}},\boldsymbol{\lambda}^{\boldsymbol{\xi}},\boldsymbol{\lambda}^{\overline{\boldsymbol{\rho}}},\boldsymbol{\lambda}^{\underline{\boldsymbol{\rho}}}$ are introduced such that $\boldsymbol{\mu}-\boldsymbol{z}+\boldsymbol{\theta}\le M\boldsymbol{\lambda}^{\overline{\boldsymbol{\phi}}},\overline{\boldsymbol{\phi}}\le M(\mathbf{1}-\boldsymbol{\lambda}^{\overline{\boldsymbol{\phi}}})$ is equivalent to the complementary slackness condition $(\mu_i-z_i+\theta_i)\overline{\boldsymbol{\phi}}_i=0$ $(i\in[n])$ for $\boldsymbol{\lambda}^{\overline{\boldsymbol{\phi}}}$ and so forth for the other binary variables. We hence reformulate the bilinear problem in the third line of Equation (22) into a mixed integer program.

By solving Problem (23), we derive the optimal solution of $\boldsymbol{z}$ and denote as $\boldsymbol{z}^{\tau+1}$. Further, we obtain an upper bound of Problem (2) as

$$
\hat{U}_\tau:=\boldsymbol{a}^\top\hat{\boldsymbol{x}}+h\left(\hat{\boldsymbol{x}},\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\gamma}}\right)+\boldsymbol{\mu}^\top\hat{\boldsymbol{\beta}}+\boldsymbol{\delta}^\top\hat{\boldsymbol{\gamma}}
\tag{24}
$$

This is an upper bound because the fixed $\left(\hat{\boldsymbol{x}},\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\gamma}}\right)$ and $\alpha=h\left(\hat{\boldsymbol{x}},\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\gamma}}\right)$ form a feasible solution to Problem (20).

We terminate the algorithm when the upper and lower bounds converge. If the algorithm does not stop in iteration $\tau$, we then add a decision variable $\boldsymbol{y}^{\tau+1}$ and the following constraints

$$
\begin{cases}
\alpha+\boldsymbol{\beta}^\top\boldsymbol{z}^{\tau+1}+\boldsymbol{\gamma}^\top(|\boldsymbol{z}^{\tau+1}-\boldsymbol{\mu}|)\ge\boldsymbol{b}^\top\boldsymbol{y}^{\tau+1},\\
\boldsymbol{W}\boldsymbol{x}+\boldsymbol{U}\boldsymbol{y}^{\tau+1}\ge\boldsymbol{V}\boldsymbol{z}^{\tau+1}+\boldsymbol{v}^0
\end{cases}
$$

to the lower bound problem (21) and solve for $\hat{L}_{\tau+1}$ in iteration $\tau+1$. The full procedure is formalized in Algorithm 3.

**Algorithm 3** A CCG algorithm for two-stage DRO

**Input:** two-stage formulation (2), second-stage formulation of $g(\boldsymbol{x},\boldsymbol{z})$, $\epsilon \geq 0$
**Initialization:** $LB=-\infty, UB=\infty$, $\tau=0$, $\boldsymbol{z}^0=\boldsymbol{\mu}$
**while** $UB-LB>\epsilon$ **do**
  solve Problem (21) with $\{\boldsymbol{z}^i\}_{i\in[\tau]}$
  denote by $(\hat{\boldsymbol{x}},\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\gamma}})$ the optimal solution, by $\hat{L}_\tau$ the optimal value
  update $LB=\hat{L}_\tau$
  solve Problem (23) with $(\hat{\boldsymbol{x}},\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\gamma}})$
  denote by $\boldsymbol{z}^{\tau+1}$ the optimal solution for $\boldsymbol{z}$, by $h(\hat{\boldsymbol{x}},\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\gamma}})$ the optimal value
  calculate $\hat{U}_\tau$ based on Equation (24) with $h(\hat{\boldsymbol{x}},\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\gamma}})$
  update $UB=\min\left\{UB,\hat{U}_\tau\right\}$
  update $\tau=\tau+1$
**Output:** optimal solution $\hat{\boldsymbol{x}}$

**Segregated affine decision rule**

For Problem (2), when we use the segregated decision rule method, the problem is reformulated as

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{\Theta},\boldsymbol{\phi}} \quad & \boldsymbol{a}^\top\boldsymbol{x}+\sup_{\mathbb{P}\in\mathcal{F}}\ \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{b}^\top\left(\boldsymbol{\Theta}\begin{bmatrix}(\boldsymbol{\mu}-\tilde{\boldsymbol{z}})^+\\(\tilde{\boldsymbol{z}}-\boldsymbol{\mu})^+\end{bmatrix}+\boldsymbol{\phi}\right)\right] \\ \text{s.t.} \quad & \boldsymbol{W}\boldsymbol{x}+\boldsymbol{U}\left(\boldsymbol{\Theta}\begin{bmatrix}(\boldsymbol{\mu}-\boldsymbol{z})^+\\(\boldsymbol{z}-\boldsymbol{\mu})^+\end{bmatrix}+\boldsymbol{\phi}\right)\geq\boldsymbol{V}\boldsymbol{z}+\boldsymbol{v}^0, \qquad \forall\boldsymbol{z}\in[\underline{\boldsymbol{z}},\overline{\boldsymbol{z}}] \\ & \boldsymbol{x}\in\mathcal{X}. \end{aligned}$$

By the property of segregated linearity, we can equivalently impose the constraints only on the breakpoints, hence obtain the following reformulation.

$$\begin{aligned} \min \quad & \boldsymbol{a}^\top\boldsymbol{x}+\sup_{\mathbb{P}\in\mathcal{F}}\mathbb{E}_{\mathbb{P}}\left[\boldsymbol{b}^\top\left(\boldsymbol{\Theta}\begin{bmatrix}(\boldsymbol{\mu}-\tilde{\boldsymbol{z}})^+\\(\tilde{\boldsymbol{z}}-\boldsymbol{\mu})^+\end{bmatrix}+\boldsymbol{\phi}\right)\right] \\ \text{s.t.} \quad & \boldsymbol{W}\boldsymbol{x}+\boldsymbol{U}\left(\boldsymbol{\Theta}\begin{bmatrix}(\boldsymbol{\mu}-\boldsymbol{z})^+\\(\boldsymbol{z}-\boldsymbol{\mu})^+\end{bmatrix}+\boldsymbol{\phi}\right)\geq\boldsymbol{V}\boldsymbol{z}+\boldsymbol{v}^0, \quad \forall\boldsymbol{z}\in\prod_{i\in[n]}\left\{\underline{z}_i,\mu_i,\overline{z}_i\right\}, \\ & \boldsymbol{x}\in\mathcal{X}. \end{aligned} \tag{25}$$

Similar to the proof of Proposition 1, we can derive a worst-case distribution $\mathbb{P}^\star$ and show that the MAD under $\mathbb{P}^\star$ is $\mathbb{E}_{\mathbb{P}^\star}\left[|\tilde{\boldsymbol{z}}-\boldsymbol{\mu}|\right]=\hat{\boldsymbol{\delta}}$. Observing that $\mathbb{E}_{\mathbb{P}^\star}\left[(\boldsymbol{\mu}-\tilde{\boldsymbol{z}})^+ + (\tilde{\boldsymbol{z}}-\boldsymbol{\mu})^+\right]=\mathbb{E}_{\mathbb{P}^\star}\left[|\tilde{\boldsymbol{z}}-\boldsymbol{\mu}|\right]=\hat{\boldsymbol{\delta}}$,

$\mathbb{E}_{\mathbb{P}^\star}[(\boldsymbol{\mu}-\tilde{\boldsymbol{z}})^+ - (\tilde{\boldsymbol{z}}-\boldsymbol{\mu})^+] = \mathbb{E}_{\mathbb{P}^\star}[\boldsymbol{\mu}-\tilde{\boldsymbol{z}}] = \boldsymbol{0}$, we have $\mathbb{E}_{\mathbb{P}^\star}[(\boldsymbol{\mu}-\tilde{\boldsymbol{z}})^+] = \mathbb{E}_{\mathbb{P}^\star}[(\tilde{\boldsymbol{z}}-\boldsymbol{\mu})^+] = \hat{\boldsymbol{\delta}}/2$. Hence, Problem (25) can be further reformulated as

$$\begin{aligned} \min \quad & \boldsymbol{a}^\top \boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{\Theta} \begin{bmatrix} \hat{\boldsymbol{\delta}}/2 \\ \hat{\boldsymbol{\delta}}/2 \end{bmatrix} + \boldsymbol{b}^\top \boldsymbol{\phi} \\ \text{s.t.} \quad & \boldsymbol{W}\boldsymbol{x} + \boldsymbol{U}\left(\boldsymbol{\Theta}\begin{bmatrix} (\boldsymbol{\mu}-\boldsymbol{z})^+ \\ (\boldsymbol{z}-\boldsymbol{\mu})^+ \end{bmatrix} + \boldsymbol{\phi}\right) \geq \boldsymbol{V}\boldsymbol{z} + \boldsymbol{v}^0, \quad \forall \boldsymbol{z} \in \prod_{i\in[n]} \{\underline{z}_i, \mu_i, \overline{z}_i\}, \\ & \boldsymbol{x} \in \mathcal{X}. \end{aligned} \tag{26}$$

This problem is essentially a linear program with an exponential size of constraints.

# B. Extensions

In this section, we introduce three possible extensions and show that, when the property of supermodularity holds, the exact tractable reformulation can be applied to more general settings.

## B.1. Left-hand-side uncertainties in the constraints

We consider that the matrix $\boldsymbol{W}$ on the left-hand side of the constraints is an affine function of the uncertain vector $\tilde{\boldsymbol{z}}$ as $\boldsymbol{W}(\tilde{\boldsymbol{z}}) = \boldsymbol{W}^0 + \sum_{i\in[n]} \boldsymbol{W}^i \tilde{z}_i$. In this case, the second-stage problem becomes

$$\begin{aligned} g^W(\boldsymbol{x},\boldsymbol{z}) = \min \quad & \boldsymbol{b}^\top \boldsymbol{y} \\ \text{s.t.} \quad & \left(\boldsymbol{W}^0 + \sum_{i\in[n]} \boldsymbol{W}^i z_i\right)\boldsymbol{x} + \boldsymbol{U}\boldsymbol{y} \geq \boldsymbol{V}\boldsymbol{z} + \boldsymbol{v}^0, \end{aligned}$$

where $\boldsymbol{W}^i, i \in \{0,1,\ldots,n\}$ are given constant matrices in $\Re^{r\times l}$. We next establish an equivalent condition for the supermodularity of $g^W$.

**Theorem 6** *$g^W(\boldsymbol{x},\boldsymbol{z})$ is supermodular in $\boldsymbol{z}$ for any $\boldsymbol{x}, \boldsymbol{b}$ and $\boldsymbol{v}^0$ if and only if $\boldsymbol{U} \in \Re^{r\times m}$, $\boldsymbol{V} \in \Re^{r\times n}$ and $\boldsymbol{W}_i \in \Re^{r\times l}, i \in [n]$ satisfy one of the following conditions:*

*1)* $\operatorname{rank}(\boldsymbol{U}) = r$,

*2) for all $\mathcal{I} \subseteq [r], \boldsymbol{\eta} \in \Re^{|\mathcal{I}|}$ with $|\mathcal{I}| = \operatorname{rank}(\boldsymbol{U}) + 1, \operatorname{rank}(\boldsymbol{U}_\mathcal{I}) = \operatorname{rank}(\boldsymbol{U})$ and $\boldsymbol{U}_\mathcal{I}^\top \boldsymbol{\eta} = \boldsymbol{0}$, we have*

*2a)* $(\boldsymbol{\eta}^\top(\boldsymbol{V}_\mathcal{I})_i)\cdot(\boldsymbol{\eta}^\top(\boldsymbol{V}_\mathcal{I})_j) \geq 0$, $(\boldsymbol{W}_\mathcal{I}^i)^\top \boldsymbol{\eta}\boldsymbol{\eta}^\top \boldsymbol{W}_\mathcal{I}^j$ *is positive semidefinite, for all $i,j \in [n]$;*

*2b)* $(\boldsymbol{\eta}^\top(\boldsymbol{V}_\mathcal{I})_i)\cdot\left(\boldsymbol{\eta}^\top \boldsymbol{W}_\mathcal{I}^j\right) = (\boldsymbol{\eta}^\top(\boldsymbol{V}_\mathcal{I})_j)\cdot(\boldsymbol{\eta}^\top \boldsymbol{W}_\mathcal{I}^i)$, *for all $i,j \in [n]$.*

For Condition 2) in Theorem 6, considering any concerned $\mathcal{I}$, i.e., $|\mathcal{I}| = \operatorname{rank}(\boldsymbol{U}) + 1$ and $\operatorname{rank}(\boldsymbol{U}_\mathcal{I}) = \operatorname{rank}(\boldsymbol{U})$, the null space of $\boldsymbol{U}_\mathcal{I}$ is of dimension 1. That is, there exists $\boldsymbol{\eta}^o$ such that for all $\boldsymbol{\eta}$ with $\boldsymbol{U}_\mathcal{I}^\top \boldsymbol{\eta} = \boldsymbol{0}$ we have $\boldsymbol{\eta} = k\boldsymbol{\eta}^o$ for some $k \in \Re$. We can easily observe that both Conditions 2a) and 2b) hold for all such $\boldsymbol{\eta}$ if and only if they hold for $\boldsymbol{\eta}^o$. Therefore, to verify whether Conditions 2a) and 2b) hold, it suffices to check for $\boldsymbol{\eta}^o$ only. Hence, as in Theorem 5, we can similarly build a corresponding algorithm to check the supermodularity of $g^W$.

### B.2. Non-linearity in the objective function

We extend our results by considering the objective as a more general function, which is nonlinear of the second-stage cost. For example, the objective can be either an expected disutility or a risk measure. Specifically, when the second-stage cost itself is supermodular in the uncertainty, the following lemma identifies mild conditions which are sufficient to preserve supermodularity. We subsequently show how our method can help us obtain tractable reformulations.

**Lemma 3** *Given any convex and non-decreasing function $u:\Re\rightarrow\Re$ and any monotone supermodular function $h:\Re^n\rightarrow\Re$, the function $\phi:\Re^n\rightarrow\Re$ defined as $\phi(\boldsymbol{z})=u(h(\boldsymbol{z}))$ is supermodular.*

This result can be applied when maximizing the decision maker's expected utility, or equivalently, minimizing the expected disutility. Consider the following problem

$$\min_{\boldsymbol{x}\in\mathcal{X}}\ \sup_{\mathbb{P}\in\mathcal{F}}\ \mathbb{E}_{\mathbb{P}}\left[u\left(\boldsymbol{a}^\top\boldsymbol{x}+g(\boldsymbol{x},\tilde{\boldsymbol{z}})\right)\right], \tag{27}$$

where $g(\boldsymbol{x},\boldsymbol{z})$ is the second-stage cost function defined by (1), and $u:\Re\rightarrow\Re$ is a piecewise linear convex and non-decreasing disutility function defined as

$$u(w)=\max_{j\in[J]}\ \{c_j w+d_j\},\quad \forall w\in\Re, \tag{28}$$

for some constants $c_j\geq 0$ and $d_j,\ j\in[J]$.

**Proposition 12** *If $g(\boldsymbol{x},\boldsymbol{z})$ is monotone and supermodular in $\boldsymbol{z}$ for all $\boldsymbol{x}\in\mathcal{X}$, then Problem (27) is equivalent to the following problem*

$$\begin{aligned}
\min\quad & \boldsymbol{\nu}^\top\boldsymbol{l} \\
\text{s.t.}\quad & \boldsymbol{R}_k^\top\boldsymbol{l}\geq\sum_{i\in[2n+1]}p_i^k f^{k,i}, && k\in[K]\\
& \boldsymbol{W}\boldsymbol{x}+\boldsymbol{U}\boldsymbol{y}^{k,i}\geq\boldsymbol{V}\boldsymbol{z}^{k,i}+\boldsymbol{v}^0, && k\in[K],i\in[2n+1]\\
& f^{k,i}\geq c_j\left(\boldsymbol{a}^\top\boldsymbol{x}+\boldsymbol{b}^\top\boldsymbol{y}^{k,i}\right)+d_j, && k\in[K],i\in[2n+1],j\in[J]\\
& \boldsymbol{l}\geq\boldsymbol{0},\ \boldsymbol{x}\in\mathcal{X},
\end{aligned} \tag{29}$$

*where $p_i^k,\boldsymbol{z}^{k,i},\ i\in[2n+1]$ are obtained from Algorithm 1 given the ambiguity sets $\mathcal{F}^k,k\in[K]$ defined by (4).*

We can also apply Lemma 3 when some risk measures are included in the objective. In particular, we study the case where the objective function is based on Optimized Certainty Equivalent (OCE) (Ben-Tal and Teboulle 1986). It is shown that the OCE models a broad range of risk measures

(Ben-Tal and Teboulle 2007), and includes the Conditional-Value-at-Risk (CVaR) as a special case. When evaluating the total cost by OCE, the two-stage problem is as follows,

$$\min_{\boldsymbol{x}\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{F}} \mathrm{OCE}_u\big(\boldsymbol{a}^\top\boldsymbol{x}+g(\boldsymbol{x},\tilde{\boldsymbol{z}})\big) = \min_{\boldsymbol{x}\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{F}} \inf_{\theta\in\Re} \{\theta+\mathbb{E}_{\mathbb{P}}\left[u\big(\boldsymbol{a}^\top\boldsymbol{x}+g(\boldsymbol{x},\tilde{\boldsymbol{z}})-\theta\big)\right]\}, \tag{30}$$

where $u(\cdot)$ is a piecewise linear convex and non-decreasing disutility function taken the form of (28). We now show that our method is applicable to Problem (30).

**Corollary 3** *If $g(\boldsymbol{x},\boldsymbol{z})$ is monotone and supermodular in $\boldsymbol{z}$ for all $\boldsymbol{x}\in\mathcal{X}$, then the OCE minimization problem (30) is equivalent to the following linear program*

$$\begin{array}{lll} \min & \theta+\boldsymbol{\nu}^\top\boldsymbol{l} & \\ \text{s.t.} & \boldsymbol{R}_k^\top\boldsymbol{l} \geq \sum_{i\in[2n+1]} p_i^k f^{k,i}, & k\in[K] \\ & \boldsymbol{W}\boldsymbol{x}+\boldsymbol{U}\boldsymbol{y}^{k,i} \geq \boldsymbol{V}\boldsymbol{z}^{k,i}+\boldsymbol{v}^0, & k\in[K], i\in[2n+1] \\ & f^{k,i} \geq c_j\left(\boldsymbol{a}^\top\boldsymbol{x}+\boldsymbol{b}^\top\boldsymbol{y}^{k,i}-\theta\right)+d_j, & k\in[K], i\in[2n+1], j\in[J] \\ & \boldsymbol{l}\geq\boldsymbol{0},\ \boldsymbol{x}\in\mathcal{X}, & \end{array} \tag{31}$$

*where $p_i^k, \boldsymbol{z}^{k,i},\ i\in[2n+1]$ are obtained from Algorithm 1 given the ambiguity sets $\mathcal{F}^k, k\in[K]$ defined by (4).*

### B.3. General ambiguity set

While the previous results are based on the ambiguity set that is constructed by mean, support and MAD in each scenario (see Equation (3)), now we extend that ambiguity set to a more general one and show that it is the most general case in which our results are still applicable. We define the ambiguity set based on piecewise linear convex functions, which are rather general and still maintain the linear structure in the reformulation. For notational simplicity, we do not incorporate the random scenario in this subsection. Specifically, we consider the ambiguity set defined as follows,

$$\mathcal{F}^G = \left\{ \mathbb{P} \;\middle|\; \begin{array}{l} \mathbb{P}\big(\underline{\boldsymbol{z}}\leq\tilde{\boldsymbol{z}}\leq\overline{\boldsymbol{z}}\big)=1 \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}]=\boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\left[h_i^j(\tilde{z}_i)\right]\leq\delta_i^j, \quad i\in[n], j\in[J_i] \end{array} \right\}, \tag{32}$$

where $J_i\geq 1$ is an integer, $h_i^j$ is a given piecewise linear convex function, $i\in[n],\ j\in[J_i]$. We assume $h_i^j$ has at least two pieces in $[\underline{z}_i,\overline{z}_i]$ to avoid the trivial case. The ambiguity set $\mathcal{F}^G$ generalizes $\mathcal{F}^k$, defined in Equation (4), as it replaces the MAD information in $\mathcal{F}^k$ by the expectations of several piecewise linear convex functions. Obviously, $\mathcal{F}^G$ includes $\mathcal{F}^k$ as a special case by choosing $J_i=1$ and $h_i^j(z)=|z-\mu_i|$ for all $z\in\Re$.

Unfortunately, as we will show later in this subsection, not all ambiguity sets $\mathcal{F}^G$ defined by Equation (32) can lead to a tractable reformulation using the procedures we discussed in Section 2.

Here we aim to identify the conditions for $\mathcal{F}^G$ such that the corresponding two-stage optimization problem, whenever the property of supermodularity holds for the second-stage cost function, can be solved with the methods in Section 2.

For any $i \in [n]$, we let $z_i^1 = \underline{z}_i, z_i^2 = \overline{z}_i$ and denote $z_i^3, z_i^4, \ldots, z_i^{S_i} \in (\underline{z}_i, \overline{z}_i)$ as the breakpoints of the piecewise linear functions $h_i^1, \ldots, h_i^{J_i}$. We now have the following result, which is essential for using the procedures in Section 2.

**Theorem 7** *The following two statements are equivalent.*

1. *Given any $\delta_i^j,\ i \in [n],\ j \in [J_i]$ satisfying $\mathcal{F}^G \neq \emptyset$, there exists $\boldsymbol{p}_i = (p_{i1}, \ldots, p_{iS_i}) \in \Re_+^{S_i}, i \in [n]$ such that for all convex function $f : \Re^n \to \Re$, we have $\mathbb{P}^* \in \arg\sup_{\mathbb{P} \in \mathcal{F}^G} \mathbb{E}_{\mathbb{P}} [f(\tilde{\boldsymbol{z}})]$ and for any $i \in [n]$,*
$$\mathbb{P}^*(\tilde{z}_i = w) = \begin{cases} p_{is} & \text{if } w = z_i^s,\ s \in [S_i], \\ 0 & \text{otherwise.} \end{cases}$$
2. *For all $i \in [n],\ j \in [J_i]$, $h_i^j$ has exactly two pieces on $[\underline{z}_i, \overline{z}_i]$.*

We observe that the worst-case distribution $\mathbb{P}^*$ provided in Theorem 7 has the same structure as that in Proposition 1. Essentially, we can characterize their marginal distributions for both settings. Moreover, the marginal distribution depends only on the ambiguity set itself, and is independent of the objective function $f$ (in Theorem 7) or the first-stage decision $\boldsymbol{x}$ (in Proposition 1). Therefore, if $\mathcal{F}^G$ satisfies the condition in Theorem 7, we can adopt a similar procedure to that in Section 2 to solve the two-stage optimization problem. In particular, we first obtain the marginal distribution, and then find the worst-case distribution based on the chained support, after which we can reformulate the two-stage problem as a linear program with low dimension. By contrast, if $\mathcal{F}^G$ violates the condition in Theorem 7, there are two-stage problems such that the worst-case distribution would depend on the first-stage decision $\boldsymbol{x}$, and hence our method cannot work.

# C. Proofs

### C.1. Proof of Proposition 1

For a one-dimensional random variable, when its MAD is known exactly, Ben-Tal and Hochman (1972, Theorem 3) have derived the worst-case distribution. Here since in $\mathcal{F}^k$, it involves multiple dimensions and only the upper bound of MADs is known, we need to prove differently as follows. For notational brevity, we drop the superscript $k$.

We now consider any $i \in [n]$ such that $\overline{z}_i > \underline{z}_i$. For the $i$-th marginal, given the support $[\underline{z}_i, \overline{z}_i]$ and mean $\mu_i$, the maximum possible value of MAD is $\frac{2(\overline{z}_i-\mu_i)(\mu_i-\underline{z}_i)}{\overline{z}_i-\underline{z}_i}$ (see Ben-Tal and Hochman (1972, Lemma 1)). Hence, we let $\hat{\delta}_i = \min\left\{\delta_i, \frac{2(\overline{z}_i-\mu_i)(\mu_i-\underline{z}_i)}{\overline{z}_i-\underline{z}_i}\right\}$. Then, the worst-case expectation of $g(\boldsymbol{x},\tilde{\boldsymbol{z}})$ under the $\mathcal{F}^k$ defined in Equation (4) can be reformulated as

$$\begin{aligned}\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[g(\boldsymbol{x},\tilde{\boldsymbol{z}})\right] &= \sup\left\{\mathbb{E}_{\mathbb{P}}\left[g(\boldsymbol{x},\tilde{\boldsymbol{z}})\right] \;\middle|\; \begin{array}{l}\mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}]=\boldsymbol{\mu},\\ \mathbb{E}_{\mathbb{P}}[|\tilde{\boldsymbol{z}}-\boldsymbol{\mu}|]\le\boldsymbol{\delta},\\ \mathbb{P}(\underline{\boldsymbol{z}}\le\boldsymbol{z}\le\bar{\boldsymbol{z}})=1\end{array}\right\} = \sup\left\{\mathbb{E}_{\mathbb{P}}\left[g(\boldsymbol{x},\tilde{\boldsymbol{z}})\right] \;\middle|\; \begin{array}{l}\mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}]=\boldsymbol{\mu},\\ \mathbb{E}_{\mathbb{P}}[|\tilde{\boldsymbol{z}}-\boldsymbol{\mu}|]\le\hat{\boldsymbol{\delta}},\\ \mathbb{P}(\underline{\boldsymbol{z}}\le\boldsymbol{z}\le\bar{\boldsymbol{z}})=1\end{array}\right\}\\ &= \sup_{\boldsymbol{0}\le\boldsymbol{d}\le\hat{\boldsymbol{\delta}}} V(\boldsymbol{d}),\end{aligned}$$

where

$$V(\boldsymbol{d}) = \sup\left\{\mathbb{E}_{\mathbb{P}}\left[g(\boldsymbol{x},\tilde{\boldsymbol{z}})\right] \;\middle|\; \begin{array}{l}\mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}]=\boldsymbol{\mu},\\ \mathbb{E}_{\mathbb{P}}[|\tilde{\boldsymbol{z}}-\boldsymbol{\mu}|]=\boldsymbol{d},\\ \mathbb{P}(\underline{\boldsymbol{z}}\le\boldsymbol{z}\le\bar{\boldsymbol{z}})=1\end{array}\right\}.$$

We prove our proposition by two steps.

**Step 1.** Considering any given $\boldsymbol{d} \in [\boldsymbol{0}, \hat{\boldsymbol{\delta}}]$, we will show that there must exist an optimal probability distribution, $\mathbb{P}^*$, for the problem in defining $V(\boldsymbol{d})$ such that the marginal distribution is as follows,

$$\mathbb{P}^*(\tilde{z}_i = w) = \begin{cases} \frac{d_i}{2(\mu_i - \underline{z}_i)} & \text{if } w = \underline{z}_i \\ 1 - \frac{d_i(\overline{z}_i - \underline{z}_i)}{2(\overline{z}_i - \mu_i)(\mu_i - \underline{z}_i)} & \text{if } w = \mu_i \\ \frac{d_i}{2(\overline{z}_i - \mu_i)} & \text{if } w = \overline{z}_i \\ 0 & \text{otherwise.} \end{cases} \tag{33}$$

We prove this by discussing two scenarios, depending on whether $V(\boldsymbol{d})$ is finite or not.

Consider the first case where $V(\boldsymbol{d}) = \infty$. In this case, there exists $\mathbb{P}'$ such that $\mathbb{E}_{\mathbb{P}'}[g(\boldsymbol{x}', \tilde{\boldsymbol{z}})] = \infty, \mathbb{E}_{\mathbb{P}'}[\tilde{\boldsymbol{z}}] = \boldsymbol{\mu}, \mathbb{E}_{\mathbb{P}'}[|\tilde{\boldsymbol{z}} - \boldsymbol{\mu}|] = \boldsymbol{d}$ and $\mathbb{P}'(\underline{\boldsymbol{z}} \leq \boldsymbol{z} \leq \bar{\boldsymbol{z}}) = 1$. We denote by $\text{supp}(\mathbb{P})$ the support of any probability distribution $\mathbb{P}$. Observing that the feasible set of $V(\boldsymbol{d})$ is nonempty (any distribution with marginal distribution as in (33) is feasible), $\mathbb{E}_{\mathbb{P}'}[g(\boldsymbol{x}', \tilde{\boldsymbol{z}})] = \infty$ implies that there must exist $\boldsymbol{z}' \in \text{supp}(\mathbb{P}') \subseteq [\underline{\boldsymbol{z}}, \overline{\boldsymbol{z}}]$ such that $g(\boldsymbol{x}, \boldsymbol{z}') = \infty$. We let $\mathbb{P}''$ be any probability distribution with marginal distribution as defined in (33), then $\text{supp}(\mathbb{P}'') = \prod_{i \in [n]} S_i$ where $S_i = \{\mu_i\}$ if $d_i = 0$, $S_i = \{\underline{z}_i, \mu_i, \overline{z}_i\}$ if $d_i \in (0, \hat{\delta}_i)$ and $S_i = \{\underline{z}_i, \overline{z}_i\}$ if $d_i = \hat{\delta}_i$, $i \in [n]$. Now, consider any $i \in [n]$. If $d_i = 0$, we must have $\mathbb{P}'(\tilde{z}_i = \mu_i) = 1$; hence, $z'_i = \mu_i \in \text{conv}(S_i)$; if $d_i > 0$, $z'_i \in \text{conv}(S_i) = [\underline{z}_i, \overline{z}_i]$ since $\boldsymbol{z} \in [\underline{\boldsymbol{z}}, \overline{\boldsymbol{z}}]$. Hence, in any case, $z'_i \in \text{conv}(S_i)$. Consequently, we have $\boldsymbol{z}' \in \text{conv}(\text{supp}(\mathbb{P}''))$. Since function $g(\boldsymbol{x}, \boldsymbol{z})$ is convex in $\boldsymbol{z}$ (see Theorem 2, Section 3.1 in Birge and Louveaux (2011)), there must exist $\boldsymbol{z}'' \in \text{supp}(\mathbb{P}'')$ such that $g(\boldsymbol{x}', \boldsymbol{z}'') = \infty$. Hence, $\mathbb{P}''$ is also a worst-case distribution.

For the second case where $V(\boldsymbol{d})$ is finite, by strong duality (e.g., Shapiro 2001),

$$V(\boldsymbol{d}) = \min\left\{ s + \boldsymbol{\mu}^\top \boldsymbol{t} + \boldsymbol{d}^\top \boldsymbol{r} \;\middle|\; s + \boldsymbol{z}^\top \boldsymbol{t} + (|\boldsymbol{z} - \boldsymbol{\mu}|)^\top \boldsymbol{r} \geq g(\boldsymbol{x}, \boldsymbol{z}),\ \forall \underline{\boldsymbol{z}} \leq \boldsymbol{z} \leq \overline{\boldsymbol{z}} \right\}. \tag{34}$$

For any given $\boldsymbol{t}, \boldsymbol{\mu}, \boldsymbol{r}$, since $g(\boldsymbol{x}, \boldsymbol{z})$ is convex in $\boldsymbol{z}$, the function $g(\boldsymbol{x}, \boldsymbol{z}) - \boldsymbol{z}^\top \boldsymbol{t} - (|\boldsymbol{z} - \boldsymbol{\mu}|)^\top \boldsymbol{r}$ is convex in $\boldsymbol{z}$ if $\boldsymbol{z} \in [\boldsymbol{a}, \boldsymbol{b}]$ where for all $i \in [n]$, $(a_i, b_i)$ takes value of $(\underline{z}_i, \mu_i)$ or $(\mu_i, \overline{z}_i)$. Hence, the constraint in (34) is equivalent to

$$s \geq g(\boldsymbol{x}, \boldsymbol{z}) - \boldsymbol{z}^\top \boldsymbol{t} - (|\boldsymbol{z} - \boldsymbol{\mu}|)^\top \boldsymbol{r}, \quad \forall \boldsymbol{z} \in \prod_{i \in [n]} \{\underline{z}_i, \mu_i, \overline{z}_i\}. \tag{35}$$

Substituting the constraints in Problem (34) by (35) and writing its dual form again, we obtain

$$V(\boldsymbol{d}) = \sup\left\{ \sum_{\tau=1}^{3^n} p_\tau g(\boldsymbol{x}, \boldsymbol{z}^\tau) \;\middle|\; \begin{array}{ll} \sum_{\tau=1}^{3^n} p_\tau z_i^\tau = \mu_i, & i \in [n] \\ \sum_{\tau=1}^{3^n} p_\tau |z_i^\tau - \mu_i| = d_i, & i \in [n] \\ \sum_{\tau=1}^{3^n} p_\tau = 1 & \\ p_\tau \geq 0, & \tau \in [3^n] \end{array} \right\}, \tag{36}$$

where $\boldsymbol{z}^1, \ldots, \boldsymbol{z}^{3^n}$ represent all $\boldsymbol{z} \in \prod_{i \in [n]} \{\underline{z}_i, \mu_i, \overline{z}_i\}$, and $p_1, \ldots, p_{3^n}$ are the associated decision variables. Therefore, we can find a distribution $\mathbb{P}^*$ which is optimal to $V(\boldsymbol{d})$, with its support being $\boldsymbol{z}^1,\ldots,\boldsymbol{z}^{3^n}$. This implies $\mathbb{P}^*(\tilde{z}_i=w)=0$ whenever $w\notin\{\underline{z}_i,\mu_i,\overline{z}_i\}$, for all $i\in[n]$. Given any three-point support $\{\underline{z}_i,\mu_i,\overline{z}_i\}$, mean $\mu_i$ and MAD value $d_i\in[0,\hat{\delta}_i]$, we observe that a distribution which places $\frac{d_i}{2(\mu_i-\underline{z}_i)}$ amount of mass at $\underline{z}_i$, $1-\frac{d_i(\overline{z}_i-\underline{z}_i)}{2(\overline{z}_i-\mu_i)(\mu_i-\underline{z}_i)}$ at $\mu_i$, and $\frac{d_i}{2(\overline{z}_i-\mu_i)}$ at $\overline{z}_i$, is uniquely determined. Hence, $\mathbb{P}^*$ must have a marginal distribution as in (33).

**Step 2.** We next show that function $V(\boldsymbol{d})$ is non-decreasing in $\boldsymbol{d}$.

Consider any $\boldsymbol{0}\le\boldsymbol{d}'\le\boldsymbol{d}''\le\hat{\boldsymbol{\delta}}$ with $\boldsymbol{d}''-\boldsymbol{d}'=\theta\boldsymbol{e}_{i^\circ}$ for some $\theta>0, i^\circ\in[n]$, and the probability distribution $\mathbb{P}'$ with $\mathbb{P}'(\tilde{\boldsymbol{z}}=\boldsymbol{z}^\tau)=p'_\tau,\tau\in[3^n]$ such that $p'_1,\ldots,p'_{3^n}$ is the worst-case distribution in Problem (36) when $\boldsymbol{d}=\boldsymbol{d}'$. WLOG, we let $i^\circ=1$. We define another distribution $\mathbb{P}''$ with $\mathbb{P}''(\tilde{\boldsymbol{z}}=\boldsymbol{z}^\tau)=p''_\tau,\tau\in[3^n]$ as

$$\mathbb{P}''(\tilde{\boldsymbol{z}}=\boldsymbol{z})=\begin{cases}\mathbb{P}'(\tilde{\boldsymbol{z}}=\boldsymbol{z})+\frac{\epsilon(\boldsymbol{z})}{\mu_1-\underline{z}_1} & \text{if } z_1=\underline{z}_1,\\ \mathbb{P}'(\tilde{\boldsymbol{z}}=\boldsymbol{z})-\frac{\epsilon(\boldsymbol{z})}{\mu_1-\underline{z}_1}-\frac{\epsilon(\boldsymbol{z})}{\overline{z}_1-\mu_1} & \text{if } z_1=\mu_1,\\ \mathbb{P}'(\tilde{\boldsymbol{z}}=\boldsymbol{z})+\frac{\epsilon(\boldsymbol{z})}{\overline{z}_1-\mu_1} & \text{if } z_1=\overline{z}_1,\end{cases}$$

where $\epsilon:\prod_{i\in[n]}\{\underline{z}_i,\mu_i,\overline{z}_i\}\to[0,1]$ is a mapping defined by

$$\epsilon(\boldsymbol{z})=\epsilon(\boldsymbol{z}+(\underline{z}_1-\mu_1)\boldsymbol{e}_1)=\epsilon(\boldsymbol{z}+(\overline{z}_1-\mu_1)\boldsymbol{e}_1)=\frac{\theta/2}{1-\frac{d'_1}{2}\left(\frac{1}{\mu_1-\underline{z}_1}+\frac{1}{\overline{z}_1-\mu_1}\right)}\mathbb{P}'(\tilde{\boldsymbol{z}}=\boldsymbol{z}) \tag{37}$$

for all $\boldsymbol{z}\in\prod_{i\in[n]}\{\underline{z}_i,\mu_i,\overline{z}_i\}$ with $z_1=\mu_1$. We next verify that $p''_1,\ldots,p''_{3^n}$ satisfy the constraints in (36) when replacing $\boldsymbol{d}$ by $\boldsymbol{d}''$.

From the definition of $\theta$, we observe that for all $\boldsymbol{z}$ such that $z_1=\mu_1$,

$$\frac{\epsilon(\boldsymbol{z})}{\mu_1-\underline{z}_1}+\frac{\epsilon(\boldsymbol{z})}{\overline{z}_1-\mu_1}=\frac{\frac{d''_1-d'_1}{2}\left(\frac{1}{\mu_1-\underline{z}_1}+\frac{1}{\overline{z}_1-\mu_1}\right)}{1-\frac{d'_1}{2}\left(\frac{1}{\mu_1-\underline{z}_1}+\frac{1}{\overline{z}_1-\mu_1}\right)}\mathbb{P}'(\tilde{\boldsymbol{z}}=\boldsymbol{z})=\left(1-\frac{1-\frac{d''_1}{2}\left(\frac{1}{\mu_1-\underline{z}_1}+\frac{1}{\overline{z}_1-\mu_1}\right)}{1-\frac{d'_1}{2}\left(\frac{1}{\mu_1-\underline{z}_1}+\frac{1}{\overline{z}_1-\mu_1}\right)}\right)\mathbb{P}'(\tilde{\boldsymbol{z}}=\boldsymbol{z}).$$

Because $0<d'_1<d''_1\le\hat{\delta}_1$ and the three-point distribution is uniquely determined, we have $1-\frac{d'_1}{2}\left(\frac{1}{\mu_1-\underline{z}_1}+\frac{1}{\overline{z}_1-\mu_1}\right)\ge1-\frac{d''_1}{2}\left(\frac{1}{\mu_1-\underline{z}_1}+\frac{1}{\overline{z}_1-\mu_1}\right)\ge0$. Hence $\mathbb{P}''(\tilde{\boldsymbol{z}}=\boldsymbol{z})\in[0,\mathbb{P}'(\tilde{\boldsymbol{z}}=\boldsymbol{z})]$ for all $\boldsymbol{z}$ with $z_1=\mu_1$. By the definition of $\mathbb{P}''$ we notice that $\sum_{\boldsymbol{z}\in\prod_{i\in[n]}\{\underline{z}_i,\mu_i,\overline{z}_i\}}\mathbb{P}''(\tilde{\boldsymbol{z}}=\boldsymbol{z})=\sum_{\boldsymbol{z}\in\prod_{i\in[n]}\{\underline{z}_i,\mu_i,\overline{z}_i\}}\mathbb{P}'(\tilde{\boldsymbol{z}}=\boldsymbol{z})=1$, we have $\mathbb{P}''(\tilde{\boldsymbol{z}}=\boldsymbol{z})\in[0,1]$ for all $\boldsymbol{z}\in\prod_{i\in[n]}\{\underline{z}_i,\mu_i,\overline{z}_i\}$.

To see $\mathbb{E}_{\mathbb{P}''}[\tilde{\boldsymbol{z}}]=\boldsymbol{\mu}$, we observe that for the first dimension,

$$\begin{aligned}\sum_{\tau=1}^{3^n}p''_\tau z_1^\tau&=\sum_{\boldsymbol{z}:z_1=\mu_1}\mathbb{P}''(\tilde{\boldsymbol{z}}=\boldsymbol{z})\mu_1+\sum_{\boldsymbol{z}:z_1=\underline{z}_1}\mathbb{P}''(\tilde{\boldsymbol{z}}=\boldsymbol{z})\underline{z}_1+\sum_{\boldsymbol{z}:z_1=\overline{z}_1}\mathbb{P}''(\tilde{\boldsymbol{z}}=\boldsymbol{z})\overline{z}_1\\&=\sum_{\boldsymbol{z}:z_1=\mu_1}\left(\mathbb{P}''(\tilde{\boldsymbol{z}}=\boldsymbol{z})\mu_1+\mathbb{P}''(\tilde{\boldsymbol{z}}=\boldsymbol{z}+(\underline{z}_1-\mu_1)\boldsymbol{e}_1)\underline{z}_1+\mathbb{P}''(\tilde{\boldsymbol{z}}=\boldsymbol{z}+(\overline{z}_1-\mu_1)\boldsymbol{e}_1)\overline{z}_1\right)\\&=\sum_{\tau=1}^{3^n}p'_\tau z_1^\tau+\sum_{\boldsymbol{z}:z_1=\mu_1}\left(\left(-\frac{\epsilon(\boldsymbol{z})}{\mu_1-\underline{z}_1}-\frac{\epsilon(\boldsymbol{z})}{\overline{z}_1-\mu_1}\right)\mu_1+\frac{\epsilon(\boldsymbol{z})}{\mu_1-\underline{z}_1}\underline{z}_1+\frac{\epsilon(\boldsymbol{z})}{\overline{z}_1-\mu_1}\overline{z}_1\right)\\&=\mu_1+\sum_{\boldsymbol{z}:z_1=\mu_1}\left(\frac{\epsilon(\boldsymbol{z})}{\mu_1-\underline{z}_1}(\underline{z}_1-\mu_1)+\frac{\epsilon(\boldsymbol{z})}{\overline{z}_1-\mu_1}(\overline{z}_1-\mu_1)\right)\\&=\mu_1,\end{aligned}$$

where the third equality is due to the property of $\epsilon$ stated in (37). For any other dimension $i$ with $i \neq 1$, by the construction of $\mathbb{P}''$ we can observe that the marginal masses on the $i$-th dimension remain identical with $\mathbb{P}'$. Hence, $\sum_{\tau=1}^{3^n} p''_\tau \boldsymbol{z}^\tau = \boldsymbol{\mu}$.

For the MAD information, we start from the first dimension and notice that

$$\begin{aligned}
\sum_{\boldsymbol{z}:z_1=\mu_1} \mathbb{P}''(\tilde{\boldsymbol{z}}=\boldsymbol{z}) &= \sum_{\boldsymbol{z}:z_1=\mu_1} \mathbb{P}'(\tilde{\boldsymbol{z}}=\boldsymbol{z}) - \left(\frac{1}{\mu_1-\underline{z}_1}+\frac{1}{\overline{z}_1-\mu_1}\right)\sum_{\boldsymbol{z}:z_1=\mu_1}\epsilon(\boldsymbol{z}) \\
&= \left(1-\left(\frac{1}{\mu_1-\underline{z}_1}+\frac{1}{\overline{z}_1-\mu_1}\right)\frac{\theta/2}{1-\frac{d'_1}{2}\left(\frac{1}{\mu_1-\underline{z}_1}+\frac{1}{\overline{z}_1-\mu_1}\right)}\right)\sum_{\boldsymbol{z}:z_1=\mu_1}\mathbb{P}'(\tilde{\boldsymbol{z}}=\boldsymbol{z}) \\
&= \left(1-\left(\frac{1}{\mu_1-\underline{z}_1}+\frac{1}{\overline{z}_1-\mu_1}\right)\frac{\theta/2}{1-\frac{d'_1}{2}\left(\frac{1}{\mu_1-\underline{z}_1}+\frac{1}{\overline{z}_1-\mu_1}\right)}\right)\left(1-\frac{d'_1}{2}\left(\frac{1}{\mu_1-\underline{z}_1}+\frac{1}{\overline{z}_1-\mu_1}\right)\right) \\
&= 1-\frac{d''_1}{2}\left(\frac{1}{\mu_1-\underline{z}_1}+\frac{1}{\overline{z}_1-\mu_1}\right),
\end{aligned}$$

where the second equality is due to (37), the third equality holds since the marginal distribution is uniquely determined in $\mathbb{P}'$. Similarly, we have $\sum_{\boldsymbol{z}:z_1=\underline{z}_1}\mathbb{P}''(\tilde{\boldsymbol{z}}=\boldsymbol{z}) = \frac{d''_1}{2(\mu_1-\underline{z}_1)}$ and $\sum_{\boldsymbol{z}:z_1=\overline{z}_1}\mathbb{P}''(\tilde{\boldsymbol{z}}=\boldsymbol{z}) = \frac{d''_1}{2(\overline{z}_1-\mu_1)}$. It then follows that $\sum_{\tau=1}^{3^n} p''_\tau |z_1^\tau-\mu_1| = d''_1$. Since the marginal masses at all remaining dimensions are unchanged, we have $\sum_{\tau=1}^{3^n} p''_\tau|\boldsymbol{z}^\tau-\boldsymbol{\mu}| = \boldsymbol{d}''$. Hence $\mathbb{P}''$ is a feasible solution to the set (36) when we replace $\boldsymbol{d}$ by $\boldsymbol{d}''$.

Consequently,

$$\begin{aligned}
V(\boldsymbol{d}'') &\geq \sum_{\tau=1}^{3^n} p''_\tau g(\boldsymbol{x},\boldsymbol{z}^\tau) \\
&= \sum_{\tau=1}^{3^n} p'_\tau g(\boldsymbol{x},\boldsymbol{z}^\tau) + \sum_{\boldsymbol{z}:z_1=\mu_1}\left(-\frac{\epsilon(\boldsymbol{z})}{\mu_1-\underline{z}_1}-\frac{\epsilon(\boldsymbol{z})}{\overline{z}_1-\mu_1}\right)g(\boldsymbol{x},\boldsymbol{z}) + \sum_{\boldsymbol{z}:z_1=\underline{z}_1}\frac{\epsilon(\boldsymbol{z})}{\mu_1-\underline{z}_1}g(\boldsymbol{x},\boldsymbol{z}) + \sum_{\boldsymbol{z}:z_1=\overline{z}_1}\frac{\epsilon(\boldsymbol{z})}{\overline{z}_1-\mu_1}g(\boldsymbol{x},\boldsymbol{z}) \\
&= \sum_{\tau=1}^{3^n} p'_\tau g(\boldsymbol{x},\boldsymbol{z}^\tau) + \sum_{\boldsymbol{z}:z_1=\mu_1}\Bigg(\left(-\frac{\epsilon(\boldsymbol{z})}{\mu_1-\underline{z}_1}-\frac{\epsilon(\boldsymbol{z})}{\overline{z}_1-\mu_1}\right)g(\boldsymbol{x},\boldsymbol{z}) \\
&\qquad + \frac{\epsilon(\boldsymbol{z})}{\mu_1-\underline{z}_1}g(\boldsymbol{x},\boldsymbol{z}+(\underline{z}_1-\mu_1)\boldsymbol{e}_1) + \frac{\epsilon(\boldsymbol{z})}{\overline{z}_1-\mu_1}g(\boldsymbol{x},\boldsymbol{z}+(\overline{z}_1-\mu_1)\boldsymbol{e}_1)\Bigg) \\
&= \sum_{\tau=1}^{3^n} p'_\tau g(\boldsymbol{x},\boldsymbol{z}^\tau) + \sum_{\boldsymbol{z}:z_1=\mu_1}\epsilon(\boldsymbol{z})\Bigg(\frac{1}{\mu_1-\underline{z}_1}\left(g(\boldsymbol{x},\boldsymbol{z}+(\underline{z}_1-\mu_1)\boldsymbol{e}_1)-g(\boldsymbol{x},\boldsymbol{z})\right) \\
&\qquad + \frac{1}{\overline{z}_1-\mu_1}\left(g(\boldsymbol{x},\boldsymbol{z}+(\overline{z}_1-\mu_1)\boldsymbol{e}_1)-g(\boldsymbol{x},\boldsymbol{z})\right)\Bigg) \\
&\geq \sum_{\tau=1}^{3^n} p'_\tau g(\boldsymbol{x},\boldsymbol{z}^\tau) = V(\boldsymbol{d}'),
\end{aligned}$$

where the first inequality holds because $p''_1,\ldots,p''_{3^n}$ is a feasible solution, the second equality is based on (37), and the second inequality follows from the convexity of $g$. Hence, $V(\boldsymbol{d})$ is non-decreasing on $[\mathbf{0},\hat{\boldsymbol{\delta}}]$. Therefore, $\sup_{\mathbb{P}\in\mathcal{F}}\mathbb{E}_{\mathbb{P}}[g(\boldsymbol{x},\tilde{\boldsymbol{z}})]=\sup_{\mathbf{0}\le\boldsymbol{d}\le\hat{\boldsymbol{\delta}}}V(\boldsymbol{d})=V(\hat{\boldsymbol{\delta}})$. The worst-case is with the form of (33) when $\boldsymbol{d}=\hat{\boldsymbol{\delta}}$, which is as proposed in our Proposition. □

### C.2. Proof of Proposition 2

1) $\Longrightarrow$ 2). Consider any $\mathbb{P}\in\mathcal{P}$ such that there exists an unordered pair $\boldsymbol{w}',\boldsymbol{w}''$ with $p'=\mathbb{P}(\tilde{\boldsymbol{w}}=\boldsymbol{w}')>0$, $p''=\mathbb{P}(\tilde{\boldsymbol{w}}=\boldsymbol{w}'')>0$. WLOG, assume $p'\le p''$. We construct a new probability distribution $\mathbb{P}^o$, such that

$$\mathbb{P}^o(\tilde{\boldsymbol{w}}=\boldsymbol{w})=\begin{cases}0 & \text{if } \boldsymbol{w}=\boldsymbol{w}'\\ p''-p' & \text{if } \boldsymbol{w}=\boldsymbol{w}''\\ \mathbb{P}(\tilde{\boldsymbol{w}}=\boldsymbol{w}'\wedge\boldsymbol{w}'')+p' & \text{if } \boldsymbol{w}=\boldsymbol{w}'\wedge\boldsymbol{w}''\\ \mathbb{P}(\tilde{\boldsymbol{w}}=\boldsymbol{w}'\vee\boldsymbol{w}'')+p' & \text{if } \boldsymbol{w}=\boldsymbol{w}'\vee\boldsymbol{w}''\\ \mathbb{P}(\tilde{\boldsymbol{w}}=\boldsymbol{w}) & \text{otherwise.}\end{cases}$$

In particular, based on $\mathbb{P}$, we move the probability mass $p'$ from the realization of $\boldsymbol{w}',\boldsymbol{w}''$ to $\boldsymbol{w}'\wedge\boldsymbol{w}'',\boldsymbol{w}'\vee\boldsymbol{w}''$. That does not change the marginal distribution and hence $\mathbb{P}^o\in\mathcal{P}$. Moreover, compared with the support of $\mathbb{P}$, that of $\mathbb{P}^o$ has one less unordered pair. We also observe that

$$\mathbb{E}_{\mathbb{P}^o}[f(\tilde{\boldsymbol{w}})]-\mathbb{E}_{\mathbb{P}}[f(\tilde{\boldsymbol{w}})]=p'\left(f(\boldsymbol{w}'\wedge\boldsymbol{w}'')+f(\boldsymbol{w}'\vee\boldsymbol{w}'')-f(\boldsymbol{w}')-f(\boldsymbol{w}'')\right)\ge 0,$$

where the last inequality is due to the supermodularity of $f$. Therefore, we can always reduce the number of unordered pairs (if there is any) in the support while the value of expectation on $f(\tilde{\boldsymbol{w}})$ either increases or remains unchanged. Since any $\mathbb{P}\in\mathcal{P}$ has nonzero probability mass only at a finite number of discrete points (by the definition of $\mathcal{P}$), the number of unordered pairs must be finite and hence will be decreased to zero after a finite number of such steps. Therefore, finally we obtain $\mathbb{P}^*\in\arg\sup_{\mathbb{P}\in\mathcal{P}}\mathbb{E}_{\mathbb{P}}[f(\tilde{\boldsymbol{w}})]$ such that the support of $\mathbb{P}^*$ has no unordered pair. Since there are $m_i$ points in the support of the $i$-th marginal, moving along the chain in ascending order from $(x_{11},\ldots,x_{n1})$ to $(x_{1m_1},\ldots,x_{nm_n})$ takes $m_i-1$ steps on the $i$-th dimension. Hence, the chain has its maximum length being 1+the total number of steps, i.e., $\sum_{i\in[n]}(m_i-1)+1$.

2) $\Longrightarrow$ 1). Assuming the contrary of 1), i.e., $f$ is not supermodular, then there exists a pair of unordered $\boldsymbol{w}',\boldsymbol{w}''\in\Re^n$ such that $f(\boldsymbol{w}')+f(\boldsymbol{w}'')>f(\boldsymbol{w}'\wedge\boldsymbol{w}'')+f(\boldsymbol{w}'\vee\boldsymbol{w}'')$. We denote $\mathcal{I}'=\{i\in[n]\mid w_i'<w_i''\}$, $\mathcal{I}''=\{i\in[n]\mid w_i'>w_i''\}$ and $\mathcal{I}_e=\{i\in[n]\mid w_i'=w_i''\}$. As $\boldsymbol{w}',\boldsymbol{w}''$ are unordered, we know $\mathcal{I}'$, $\mathcal{I}''$ are both nonempty. For all $i\in\mathcal{I}'\cup\mathcal{I}''$, we let $m_i=2$, $x_{i1}=w_i'\wedge w_i''$, $x_{i2}=w_i'\vee w_i''$, $p_{i1}=p_{i2}=0.5$; for all $i\in\mathcal{I}_e$, we let $m_i=1$, $x_{i1}=w_i'=w_i''$ and $p_{i1}=1$. Correspondingly, $\mathcal{P}=\{\mathbb{P}\mid\mathbb{P}(\tilde{w}_i=x_{ij})=p_{ij},j\in[m_i],i\in[n]\}$, and any $\mathbb{P}\in\mathcal{P}$ must has its support in $\mathcal{W}=\Pi_{i\in\mathcal{I}'\cup\mathcal{I}''}\{x_{i1},x_{x2}\}\times\Pi_{i\in\mathcal{I}_e}\{x_{i1}\}$. Consider any $\mathbb{P}^o\in\mathcal{P}$ such that its support $\mathcal{W}_{\mathbb{P}^o}=\{\boldsymbol{w}\in\Re^n\mid\mathbb{P}^o(\tilde{\boldsymbol{w}}=\boldsymbol{w})>0\}$ forms a chain. We now show that $\mathbb{P}^o\notin\arg\sup_{\mathbb{P}\in\mathcal{P}}\mathbb{E}_{\mathbb{P}}[f(\tilde{\boldsymbol{w}})]$ and hence statement 2) in the proposition is false, then the proof can be completed. To this end, we notice since $\mathcal{W}_{\mathbb{P}^o}$ forms a chain, we can label the elements in $\mathcal{W}_{\mathbb{P}^o}$ in ascending order, i.e., $\boldsymbol{w}^1\le\boldsymbol{w}^2\le\ldots$.

We first show $\boldsymbol{w}^1 = \boldsymbol{w}' \wedge \boldsymbol{w}''$. Consider any $i \in [n]$. If $w_i^1 < x_{i1}$, then $w_i^1 \notin \{x_{ij} \mid j \in [m_i]\}$, contradicts with $\mathbb{P}^o \in \mathcal{P}$. If $w_i^1 > x_{i1}$, then $w_i > x_{i1}$ for all $\boldsymbol{w} \in \mathcal{W}_{\mathbb{P}^o}$, $\mathbb{P}^o(\tilde{w}_i = x_{i1}) = 0$, which also contradicts with $\mathbb{P}^o \in \mathcal{P}$. Therefore, we must have $w_i^1 = x_{i1}$ for all $i \in [n]$, i.e., $\boldsymbol{w}^1 = (x_{11}, \ldots, x_{n1}) = \boldsymbol{w}' \wedge \boldsymbol{w}''$.

We next show $\boldsymbol{w}^2 = \boldsymbol{w}' \vee \boldsymbol{w}''$. Assume that there exists $i \in \mathcal{I}' \cup \mathcal{I}''$ with $w_i^2 = w_i^1$. Since $\boldsymbol{w}^2 \geq \boldsymbol{w}^1$ and $\boldsymbol{w}^2 \neq \boldsymbol{w}^1$, we know that there exists $j \in \mathcal{I}' \cup \mathcal{I}''$ with $w_j^2 > w_j^1$. By $w_i^2 = w_i^1 = x_{i1}$, $\mathbb{P}^o(\tilde{\boldsymbol{w}} = \boldsymbol{w}^1) + \mathbb{P}^o(\tilde{\boldsymbol{w}} = \boldsymbol{w}^2) \leq \mathbb{P}^o(\tilde{w}_i = x_{i1}) = p_{i1} = 0.5$, we know $\mathbb{P}^o(\tilde{\boldsymbol{w}} = \boldsymbol{w}^1) < 0.5$; by $w_j^2 > w_j^1 = x_{j1}$, we know $\mathbb{P}^o(\tilde{\boldsymbol{w}} = \boldsymbol{w}^1) = p_{j1} = 0.5$. Hence, we have contradiction. It follows that $w_i^2 > w_i^1$ for all $i \in \mathcal{I}' \cup \mathcal{I}''$, i.e., $\boldsymbol{w}^2 = \boldsymbol{w}' \vee \boldsymbol{w}''$. Moreover, we have $|\mathcal{W}| = 2$ since $\boldsymbol{w}^2$ is the maximum element of $\mathcal{W}$.

Therefore, $\mathbb{P}^o$ is such that $\mathbb{P}^o(\tilde{\boldsymbol{w}} = \boldsymbol{w}' \wedge \boldsymbol{w}'') = \mathbb{P}^o(\tilde{\boldsymbol{w}} = \boldsymbol{w}' \vee \boldsymbol{w}'') = 0.5$. Consider another distribution $\mathbb{P}^*$ such that $\mathbb{P}^*(\tilde{\boldsymbol{w}} = \boldsymbol{w}') = \mathbb{P}^*(\tilde{\boldsymbol{w}} = \boldsymbol{w}'') = 0.5$. We can easily have $\mathbb{P}^* \in \mathcal{P}$, and

$$\mathbb{E}_{\mathbb{P}^o}\left[f(\tilde{\boldsymbol{w}})\right] = 0.5 \times \left(f(\boldsymbol{w}' \wedge \boldsymbol{w}'') + f(\boldsymbol{w}' \vee \boldsymbol{w}'')\right) < 0.5 \times \left(f(\boldsymbol{w}') + f(\boldsymbol{w}'')\right) = \mathbb{E}_{\mathbb{P}^*}\left[f(\tilde{\boldsymbol{w}})\right].$$

□

### C.3. Proof of Proposition 3

For notational simplicity, we drop the superscript $k$ which represents the scenario $k$; we also assume $\underline{z}_i = -1, \mu_i = 0, \bar{z}_i = 1$ for all $i \in [n]$, since the general case can be proved in the same way.

During the progress of this algorithm, for each $j \in [2n+1]$, we define $\mathbf{mp}^{j,i}$, which stands for the remaining marginal probability for iteration $j$ at dimension $i$, as

$$\mathbf{mp}^{j,i} = \begin{cases} q_i^j & \text{if } z_i^j = 1 \quad (\mathbf{mp}^{j,i} \in \Re \text{ in this case}) \\ (q_i^j,\ \mathbb{P}^*(\tilde{z}_i = 1)) & \text{if } z_i^j = 0 \quad (\mathbf{mp}^{j,i} \in \Re^2 \text{ in this case}) \\ (q_i^j,\ \mathbb{P}^*(\tilde{z}_i = 0),\ \mathbb{P}^*(\tilde{z}_i = 1)) & \text{if } z_i^j = -1 \ (\mathbf{mp}^{j,i} \in \Re^3 \text{ in this case}) \end{cases}.$$

We also define $c_j = \mathbf{1}^\top \mathbf{mp}^{j,1}$ which represents the remaining total probability mass. Correspondingly, we denote the set of information $\mathcal{I}^j = \{\boldsymbol{z}^j, \mathbf{mp}^{j,1}, \ldots, \mathbf{mp}^{j,n}, c_j\}$.

Given a set of information $\mathcal{I}^j$, we say it is valid if it satisfies the following four conditions: 1) $\boldsymbol{z}^j \in \{-1, 0, 1\}^n$; 2) $\mathbf{mp}^{j,i} \in [0,1]^{2 - z_i^j}$ for all $i \in [n]$; 3) $\mathrm{mp}_{\text{end}}^{j,i} > 0$ for all $i \in [n]$, where we denote $\mathrm{mp}_{\text{end}}^{j,i}$ as the last element of the vector $\mathbf{mp}^{j,i}$; and 4) $\mathbf{1}^\top \mathbf{mp}^{j,i} = c_j$ for all $i \in [n]$.

By induction, we now show that $\mathcal{I}^j$ is valid for all $j \in [2n+1]$.

First, when $j = 1$, the conditions 1), 2) and 3) are obviously satisfied. The condition 4) is also satisfied since $\mathbf{1}^\top \mathbf{mp}^{1,i} = \mathbb{P}^*(\tilde{z}_i = -1) + \mathbb{P}^*(\tilde{z}_i = 0) + \mathbb{P}^*(\tilde{z}_i = 1) = 1$ for all $i \in [n]$, and $c_1 = 1$.

Suppose $\mathcal{I}^j$ is valid for some $j \in [2n]$. Based on the algorithm, the elements in $\mathcal{I}^{j+1}$ are obtained as follows. First, $p_j = \min\{\mathrm{mp}_1^{j,1}, \ldots, \mathrm{mp}_1^{j,n}\}$, $r_j = \min\{i \in [n] \mid \mathrm{mp}_1^{j,i} = p_j\}$. After that, $\boldsymbol{z}^{j+1} = \boldsymbol{z}^j + \boldsymbol{e}_{r_j}$. We now prove that $z_{r_j}^j \neq 1$ by contradiction. Assume to the contrary, i.e., $z_{r_j}^j = 1$, then $\mathbf{mp}^{j,r_j} \in \Re$, we have $c_j = \mathbf{1}^\top \mathbf{mp}^{j,r_j} = \mathrm{mp}_1^{j,r_j} = p_j$. For any $i \in [n] \setminus \{r_j\}$, we observe i) $\mathrm{mp}_1^{j,i} \geq p_j = c_j$ (the inequality is because of our choice of $p_j$); ii) $\mathrm{mp}_{\text{end}}^{j,i} > 0$; and iii) $\mathbf{1}^\top \mathbf{mp}^{j,i} = c_j$ and $\mathbf{mp}^{j,i} \geq \mathbf{0}$.

The last two observations are because $\mathcal{I}^j$ is valid and hence satisfies conditions 2), 3) and 4). Hence, we have $\mathbf{mp}^{j,i} \in \Re$ and then $z_i^j = 1$. That implies $\boldsymbol{z}^j = \mathbf{1}$. We notice that for any $t \in [j-1]$, $\boldsymbol{z}^{t+1} = \boldsymbol{z}^t + \boldsymbol{e}_i$ for some $i \in [n]$. So moving from $\boldsymbol{z}^1 = -\mathbf{1}$ to $\boldsymbol{z}^j = \mathbf{1}$ requires $2n$ steps, i.e., $j = 2n+1$, which contradicts $j \in [2n]$. Hence, $z_{r_j}^j = 1$ is false, and we must have $z_{r_j}^j \in \{-1, 0\}$. We can conclude that $\boldsymbol{z}^{j+1} = \boldsymbol{z}^j + \boldsymbol{e}_{r_j} \in \{-1,0,1\}^n$, the condition 1) is satisfied for $\mathcal{I}^{j+1}$. As a result, condition 2) is obviously satisfied by the way $\mathbf{mp}^{j,i}$ is calculated.

With the algorithm, we know $\mathbf{mp}^{j+1,r_j}$ can be obtained from the vector of $\mathbf{mp}^{j,r_j}$ by removing the first component. Therefore, $\mathrm{mp}_{\mathrm{end}}^{j+1,r_j} = \mathrm{mp}_{\mathrm{end}}^{j,r_j} > 0$, the condition 3) is satisfied when $i = r_j$. Moreover, $\mathbf{1}^\top \mathbf{mp}^{j+1,r_j} = \mathbf{1}^\top \mathbf{mp}^{j,r_j} - \mathrm{mp}_1^{j,r_j} = c_j - p_j$. We also observe $c_j - p_j = \mathbf{1}^\top \mathbf{mp}^{j+1,r_j} \geq \mathrm{mp}_{\mathrm{end}}^{j+1,r_j} > 0$ and hence $c_j > p_j$.

For any $i \in [n] \setminus \{r_j\}$, since $z_i^{j+1} = z_i^j$, $\mathbf{mp}^{j+1,i}$ and $\mathbf{mp}^{j,i}$ are both of dimension $(2 - z^{j+1,i})$, they differ only at the first dimension; in particular,

$$\mathrm{mp}_s^{j+1,i} = \begin{cases} \mathrm{mp}_1^{j,i} - p_j & \text{if } s = 1 \\ \mathrm{mp}_s^{j,i} & \text{if } z_i^{j+1} \in \{-1,0\} \text{ and } s \neq 1 \end{cases}. \tag{38}$$

We note that if $z_i^{j+1} = z_i^j = 1$, then $\mathbf{mp}^{j,i}, \mathbf{mp}^{j+1,i} \in \Re_+$, and $\mathrm{mp}_1^{j,i} = \mathbf{1}^\top \mathbf{mp}^{j,i} = c_j > p_j$, $\mathrm{mp}_{\mathrm{end}}^{j+1,i} = \mathrm{mp}_1^{j+1,i} = \mathrm{mp}_1^{j,i} - p_j > 0$. If $z_i^{j+1} = z_i^j \in \{-1,0\}$, obviously $\mathrm{mp}_{\mathrm{end}}^{j+1,i} = \mathrm{mp}_{\mathrm{end}}^{j,i} > 0$. Therefore, condition 3) is satisfied for $i$. Moreover, by Equation (38) we also know $\mathbf{1}^\top \mathbf{mp}^{j+1,i} = \mathbf{1}^\top \mathbf{mp}^{j,i} - p_j = c_j - p_j$. Since we have previously obtained $\mathbf{1}^\top \mathbf{mp}^{j+1,r_j} = c_j - p_j$, condition 4) is also satisfied. We conclude $\mathcal{I}^{j+1}$ is also valid and it finishes the induction, i.e., $\mathcal{I}^j$ is valid for all $j \in [2n+1]$.

Now, for any $j \in [2n+1]$, we define $\mathcal{Q}^j$ as the set of all mass functions with the marginal mass given by $\mathbf{mp}^{j,1}, \ldots, \mathbf{mp}^{j,n}$ and the possible realizations forming a chain. More specifically, define $\boldsymbol{w}^{j,i} \in \{-1,0,1\}^{2-z_i^j}$ by

$$\boldsymbol{w}^{j,i} = \begin{cases} (-1,0,1) & \text{if } z_i^j = -1 \\ (0,1) & \text{if } z_i^j = 0 \\ 1 & \text{if } z_i^j = 1 \end{cases},$$

which is the vector of all possible realizations at dimension $i$, $i \in [n]$, and $\mathcal{W}^j = \{\boldsymbol{z} \mid \boldsymbol{z}^j \leq \boldsymbol{z} \leq \mathbf{1}\} \cap \{-1,0,1\}^n$ which is the set of all possible realizations of vector $\boldsymbol{z}$; then

$$\mathcal{Q}^j = \left\{ f^j : \mathcal{W}^j \to [0,1] \;\middle|\; \begin{array}{l} \sum_{\boldsymbol{z} \in \mathcal{W}^j : z_i = w_s^{j,i}} f^j(\boldsymbol{z}) = \mathrm{mp}_s^{j,i}, \quad i \in [n], s \in [2 - z_i^j] \\ \{\boldsymbol{z} \mid f^j(\boldsymbol{z}) > 0\} \text{ forms a chain} \end{array} \right\}.$$

Noticing that $\mathcal{W}^{j+1} = \left\{ \boldsymbol{z} \in \mathcal{W}^j \;\middle|\; z_{r_j} \neq z_{r_j}^j \right\}$, we define another set $\hat{\mathcal{Q}}^j$ by

$$\hat{\mathcal{Q}}^j = \left\{ f^j : \mathcal{W}^j \to [0,1] \;\middle|\; \begin{array}{ll} f^j(\boldsymbol{z}^j) = p_j & \\ f^j(\boldsymbol{z}) = 0, & \forall \boldsymbol{z} \in \mathcal{W}^j \text{ such that } z_{r_j} = z_{r_j}^j, \boldsymbol{z} \neq \boldsymbol{z}^j \\ f^j(\boldsymbol{z}) = f^{j+1}(\boldsymbol{z}), & \forall \boldsymbol{z} \in \mathcal{W}^{j+1} \\ f^{j+1} \in \mathcal{Q}^{j+1} & \end{array} \right\}.$$

We next prove $\mathcal{Q}^j = \hat{\mathcal{Q}}^j$.

First, consider any $f^j \in \mathcal{Q}^j$. Suppose there exist $\boldsymbol{z}^o \in \mathcal{W}^j$ with $z^o_{r_j} = z^j_{r_j}$ and $\boldsymbol{z}^o \neq \boldsymbol{z}^j$ such that $f^j(\boldsymbol{z}^o) > 0$. That implies the existence of $s \in [n] \setminus \{r_j\}$ such that $z^o_s > z^j_s$. Hence,

$$\sum_{\substack{\boldsymbol{z}\in\mathcal{W}^j: z_{r_j}=z^j_{r_j} \\ z_s = z^j_s}} f^j(\boldsymbol{z}) = \sum_{\boldsymbol{z}\in\mathcal{W}^j: z_{r_j}=z^j_{r_j}} f^j(\boldsymbol{z}) - \sum_{\substack{\boldsymbol{z}\in\mathcal{W}^j: z_{r_j}=z^j_{r_j} \\ z_s \neq z^j_s}} f^j(\boldsymbol{z}) = p_j - \sum_{\substack{\boldsymbol{z}\in\mathcal{W}^j: z_{r_j}=z^j_{r_j} \\ z_s \neq z^j_s}} f^j(\boldsymbol{z}) \le p_j - f^j(\boldsymbol{z}^o) < p_j,$$

$$\sum_{\substack{\boldsymbol{z}\in\mathcal{W}^j: z_{r_j}>z^j_{r_j} \\ z_s = z^j_s}} f^j(\boldsymbol{z}) = \sum_{\boldsymbol{z}\in\mathcal{W}^j: z_s=z^j_s} f^j(\boldsymbol{z}) - \sum_{\substack{\boldsymbol{z}\in\mathcal{W}^j: z_{r_j}=z^j_{r_j} \\ z_s = z^j_s}} f^j(\boldsymbol{z}) \ge p_j - \sum_{\substack{\boldsymbol{z}\in\mathcal{W}^j: z_{r_j}=z^j_{r_j} \\ z_s = z^j_s}} f^j(\boldsymbol{z}) > 0.$$

Therefore, we have $\boldsymbol{z}^* \in \mathcal{W}^j$ such that $z^*_{r_j} > z^j_{r_j} = z^o_{r_j}$, $z^*_s = z^j_s < z^0_s$ and $f^j(\boldsymbol{z}^*) > 0$, contradicting that $\{\boldsymbol{z} \mid f^j(\boldsymbol{z}) > 0\}$ forms a chain. Therefore, $f^j(\boldsymbol{z}) = 0$ whenever $\boldsymbol{z} \in \mathcal{W}^j$ has $z_{r_j} = z^j_{r_j}, \boldsymbol{z} \neq \boldsymbol{z}^j$, and $f^j(\boldsymbol{z}^j) = \mathrm{mp}^{j,r_j}_1 - \sum_{\boldsymbol{z}\in\mathcal{W}^j, z_{r_j}=z^j_{r_j}, \boldsymbol{z}\neq\boldsymbol{z}^j} f^j(\boldsymbol{z}) = p_j - 0 = p_j$. Therefore, $f^j$ satisfies the first two conditions in $\hat{Q}^j$. The corresponding $f^{j+1}$ is in $\mathcal{Q}^{j+1}$ can be easily verified by showing the chain structure and checking the equality constraints on the marginal mass. Hence, we have $f^j \in \hat{\mathcal{Q}}^j$.

We now prove the reverse. Consider any $f^j \in \hat{\mathcal{Q}}^j$ and we check whether it satisfies the two conditions in $\mathcal{Q}^j$. The first condition, which is on the marginal mass, can be verified by standard algebra. The second condition, which is on the chain structure, is straightforward. Therefore, we have $f^j \in \mathcal{Q}^j$. We can conclude that $\mathcal{Q}^j = \hat{\mathcal{Q}}^j$ for all $j \in [2n+1]$.

Finally, by representing $\mathcal{Q}^j$ in the form of $\hat{Q}^j$, with recursion we can easily get

$$\mathcal{Q}^1 = \left\{ f : \mathcal{W}^j \to [0,1] \;\middle|\; \begin{array}{ll} f(\boldsymbol{z}^i) = p_i, & i \in [2n] \\ f(\boldsymbol{z}) = 0, & \forall \boldsymbol{z} \in \mathcal{W}^1 \setminus \{\boldsymbol{z}^i, i \in [2n]\} \setminus \mathcal{W}^{2n+1} \\ f(\boldsymbol{z}) = \hat{f}(\boldsymbol{z}), & \forall \boldsymbol{z} \in \mathcal{W}^{2n+1} \\ \hat{f} \in \mathcal{Q}^{2n+1} & \end{array} \right\} \tag{39}$$

We note that since $\boldsymbol{z}^j \in \{-1,0,1\}^n$, $\boldsymbol{z}^1 = -\mathbf{1}$, and any time the movement from $\boldsymbol{z}^j$ to $\boldsymbol{z}^{j+1}$ is to increase one dimension by 1 while maintaining other dimensions unchanged, and hence we have $\boldsymbol{z}^{2n+1} = \mathbf{1}$. Therefore, $\mathcal{W}^{2n+1} = \{\boldsymbol{z}^{2n+1}\}$. Then by Equation (39), we have

$$\mathcal{Q}^1 = \left\{ f : \mathcal{W}^j \to [0,1] \;\middle|\; \begin{array}{ll} f(\boldsymbol{z}^i) = p_i, \; i \in [2n+1] \\ f(\boldsymbol{z}) = 0, \quad \forall \boldsymbol{z} \in \mathcal{W}^1 \setminus \{\boldsymbol{z}^i, i \in [2n+1]\} \end{array} \right\}$$

Hence, the result is proved. □

### C.4. Proof of Theorem 1

We define the function $f(\boldsymbol{x})$ as

$$\begin{aligned} f(\boldsymbol{x}) = \min \quad & \boldsymbol{\nu}^\top \boldsymbol{l} \\ \text{s.t.} \quad & \boldsymbol{R}_k^\top \boldsymbol{l} \ge \sum_{i\in[2n+1]} p^k_i \boldsymbol{b}^\top \boldsymbol{y}^{k,i}, \qquad k \in [K] \\ & \boldsymbol{W}\boldsymbol{x} + \boldsymbol{U}\boldsymbol{y}^{k,i} \ge \boldsymbol{V}\boldsymbol{z}^{k,i} + \boldsymbol{v}^0, \quad k \in [K], i \in [2n+1] \\ & \boldsymbol{l} \ge \mathbf{0}, \end{aligned}$$

then Problem (6) is equivalent with $\min_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x})$.

We further denote $\mathcal{X}_{fea} = \{\boldsymbol{x}\in\mathcal{X} \mid \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[g(\boldsymbol{x},\tilde{\boldsymbol{z}})] < \infty\}$. Recall that we assume Problem (2) has finite optimal value, so $\mathcal{X}_{fea}\neq\emptyset$.

Consider any $\boldsymbol{x}\in\mathcal{X}\setminus\mathcal{X}_{fea}$, we have

$$\infty = \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[g(\boldsymbol{x},\tilde{\boldsymbol{z}})] = \max\left\{ \sum_{k\in[K]} q_k \sup_{\mathbb{P}^k\in\mathcal{F}^k} \mathbb{E}_{\mathbb{P}^k}[g(\boldsymbol{x},\tilde{\boldsymbol{z}})] \;\middle|\; \boldsymbol{q}\in\mathcal{Q} \right\}.$$

Since any feasible $\boldsymbol{q}\in\mathcal{Q}\subseteq\{\boldsymbol{q}\in\Re^K_+ \mid \sum_{k\in[K]} q_k = 1\}$ is bounded, there must be $k\in[K]$ such that

$$\infty = \sup_{\mathbb{P}^k\in\mathcal{F}^k} \mathbb{E}_{\mathbb{P}^k}[g(\boldsymbol{x},\tilde{\boldsymbol{z}})] = \sum_{i\in[2n+1]} p_i^k g(\boldsymbol{x},\boldsymbol{z}^{k,i}),$$

where the last equality follows from Proposition 3. Again, since $p_i^k\in[0,1]$ for all $i\in[2n+1]$, there exists a specific $i\in[2n+1]$ such that $g(\boldsymbol{x},\boldsymbol{z}^{k,i}) = \infty$. It is equivalent to the infeasibility of the constraint $\boldsymbol{W}\boldsymbol{x}+\boldsymbol{U}\boldsymbol{y}^{k,i}\geq\boldsymbol{V}\boldsymbol{z}^{k,i}+\boldsymbol{v}^0$, which is involved in the problem defining $f(\boldsymbol{x})$. Hence, $f(\boldsymbol{x})=\infty$.

Therefore, Problem (6) is equivalent with $\min_{\boldsymbol{x}\in\mathcal{X}_{fea}} f(\boldsymbol{x})$. We notice that Problem (2) is equivalent to $\min_{\boldsymbol{x}\in\mathcal{X}_{fea}} \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[g(\boldsymbol{x},\tilde{\boldsymbol{z}})]$. Hence, for proving this theorem, now it suffices to show that for all $\boldsymbol{x}\in\mathcal{X}_{fea}$, we have $\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[g(\boldsymbol{x},\tilde{\boldsymbol{z}})] = f(\boldsymbol{x})$. To this end, consider any $\boldsymbol{x}\in\mathcal{X}_{fea}$, we then know $\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[g(\boldsymbol{x},\tilde{\boldsymbol{z}})]$ is finite. Notice that 1) $\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[g(\boldsymbol{x},\tilde{\boldsymbol{z}})] = \max\left\{\sum_{k\in[K]} q_k \sup_{\mathbb{P}^k\in\mathcal{F}^k} \mathbb{E}_{\mathbb{P}^k}[g(\boldsymbol{x},\tilde{\boldsymbol{z}})] \;\middle|\; \boldsymbol{q}\in\mathcal{Q}\right\}$ and 2) by the assumption on $\mathcal{Q}$, for any $k\in[K]$ there exists $\boldsymbol{q}\in\mathcal{Q}$ with $q_k>0$. Hence, for all $k\in[K]$, $\sup_{\mathbb{P}^k\in\mathcal{F}^k} \mathbb{E}_{\mathbb{P}^k}[g(\boldsymbol{x},\tilde{\boldsymbol{z}})]$ must be finite. It implies that $g(\boldsymbol{x},\boldsymbol{z})$ is finite for all $\boldsymbol{z}\in[\underline{\boldsymbol{z}}^k,\overline{\boldsymbol{z}}^k]$. Moreover,

$$\begin{aligned}
\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[g(\boldsymbol{x},\tilde{\boldsymbol{z}})] &= \max\left\{ \sum_{k\in[K]} q_k \sup_{\mathbb{P}^k\in\mathcal{F}^k} \mathbb{E}_{\mathbb{P}^k}[g(\boldsymbol{x},\tilde{\boldsymbol{z}})] \;\middle|\; \boldsymbol{R}\boldsymbol{q}\leq\boldsymbol{\nu}, \boldsymbol{q}\geq\boldsymbol{0} \right\} \\
&= \min\left\{ \boldsymbol{\nu}^\top\boldsymbol{l} \;\middle|\; \begin{array}{l} \boldsymbol{R}_k^\top\boldsymbol{l}\geq \sup_{\mathbb{P}^k\in\mathcal{F}^k} \mathbb{E}_{\mathbb{P}^k}[g(\boldsymbol{x},\tilde{\boldsymbol{z}})],\ k\in[K] \\ \boldsymbol{l}\geq\boldsymbol{0} \end{array} \right\} \\
&= f(\boldsymbol{x}),
\end{aligned}$$

where the second equality is due to strong duality. □

### C.5. Proof of Corollary 1

It has been proved in the proof for Theorem 1. □

### C.6. Proof of Proposition 4

We first prove the "if" part. Suppose $\mathcal{S}(\boldsymbol{x})$ is an inverse additive lattice, then given any $\boldsymbol{z}',\boldsymbol{z}'',\boldsymbol{p},\boldsymbol{q}$ with $(\boldsymbol{p},\boldsymbol{z}'\wedge\boldsymbol{z}''),(\boldsymbol{q},\boldsymbol{z}'\vee\boldsymbol{z}'')\in\mathcal{S}(\boldsymbol{x})$, there exist $\boldsymbol{y}',\boldsymbol{y}''$ such that $(\boldsymbol{y}',\boldsymbol{z}'),(\boldsymbol{y}'',\boldsymbol{z}'')\in\mathcal{S}(\boldsymbol{x})$ and $\boldsymbol{y}'+\boldsymbol{y}''=\boldsymbol{p}+\boldsymbol{q}$. We then have

$$g(\boldsymbol{x},\boldsymbol{z}')+g(\boldsymbol{x},\boldsymbol{z}'')\leq\boldsymbol{b}^\top\boldsymbol{y}'+\boldsymbol{b}^\top\boldsymbol{y}''=\boldsymbol{b}^\top\boldsymbol{p}+\boldsymbol{b}^\top\boldsymbol{q}.$$

Taking the minimum on the right-hand-side over all $\boldsymbol{p},\boldsymbol{q}$ with $(\boldsymbol{p},\boldsymbol{z}'\wedge\boldsymbol{z}''),(\boldsymbol{q},\boldsymbol{z}'\vee\boldsymbol{z}'')\in\mathcal{S}(\boldsymbol{x})$, we obtain $g(\boldsymbol{x},\boldsymbol{z}')+g(\boldsymbol{x},\boldsymbol{z}'')\le g(\boldsymbol{x},\boldsymbol{z}'\wedge\boldsymbol{z}'')+g(\boldsymbol{x},\boldsymbol{z}'\vee\boldsymbol{z}'')$.

Next we prove the "only if" part by contradiction. Suppose $\mathcal{S}(\boldsymbol{x})$ is not an inverse additive lattice, then there exist $\boldsymbol{z}',\boldsymbol{z}'',\boldsymbol{p},\boldsymbol{q}$ with $(\boldsymbol{p},\boldsymbol{z}'\wedge\boldsymbol{z}''),(\boldsymbol{q},\boldsymbol{z}'\vee\boldsymbol{z}'')\in\mathcal{S}(\boldsymbol{x})$ but $\boldsymbol{p}+\boldsymbol{q}\notin\mathcal{W}$, where the set $\mathcal{W}$ is defined as $\mathcal{W}=\{\boldsymbol{r}+\boldsymbol{s}\mid(\boldsymbol{r},\boldsymbol{z}'),(\boldsymbol{s},\boldsymbol{z}'')\in\mathcal{S}(\boldsymbol{x})\}$. According to the definition of $\mathcal{S}(\boldsymbol{x})$, we can easily see that $\mathcal{W}$ is convex and closed. By the Hyperplane Separation Theorem, there exist a vector $\boldsymbol{\eta}$ and a real number $\lambda$ such that,

$$\boldsymbol{\eta}^\top(\boldsymbol{p}+\boldsymbol{q})<\lambda<\boldsymbol{\eta}^\top\boldsymbol{w}\qquad\forall\boldsymbol{w}\in\mathcal{W}.$$

Consider the second-stage cost function $g(\boldsymbol{x},\boldsymbol{z})$ (defined in Equation (1)) with coefficient $\boldsymbol{b}=\boldsymbol{\eta}$. We have

$$\begin{aligned}
g(\boldsymbol{x},\boldsymbol{z}')+g(\boldsymbol{x},\boldsymbol{z}'')&=\min\left\{\boldsymbol{\eta}^\top\boldsymbol{y}\;\middle|\;(\boldsymbol{y},\boldsymbol{z}')\in\mathcal{S}(\boldsymbol{x})\right\}+\min\left\{\boldsymbol{\eta}^\top\boldsymbol{y}\;\middle|\;(\boldsymbol{y},\boldsymbol{z}'')\in\mathcal{S}(\boldsymbol{x})\right\}\\
&=\min\left\{\boldsymbol{\eta}^\top(\boldsymbol{r}+\boldsymbol{s})\;\middle|\;(\boldsymbol{r},\boldsymbol{z}'),(\boldsymbol{s},\boldsymbol{z}'')\in\mathcal{S}(\boldsymbol{x})\right\}\\
&=\min\left\{\boldsymbol{\eta}^\top\boldsymbol{w}\;\middle|\;\boldsymbol{w}\in\mathcal{W}\right\}>\lambda,\\
g(\boldsymbol{x},\boldsymbol{z}'\wedge\boldsymbol{z}'')+g(\boldsymbol{x},\boldsymbol{z}'\vee\boldsymbol{z}'')&=\min\left\{\boldsymbol{\eta}^\top\boldsymbol{y}\;\middle|\;(\boldsymbol{y},\boldsymbol{z}'\wedge\boldsymbol{z}'')\in\mathcal{S}(\boldsymbol{x})\right\}+\min\left\{\boldsymbol{\eta}^\top\boldsymbol{y}\;\middle|\;(\boldsymbol{y},\boldsymbol{z}''\vee\boldsymbol{z}'')\in\mathcal{S}(\boldsymbol{x})\right\}\\
&\le\boldsymbol{\eta}^\top(\boldsymbol{p}+\boldsymbol{q})<\lambda.
\end{aligned}$$

Therefore, $g(\boldsymbol{x},\boldsymbol{z}')+g(\boldsymbol{x},\boldsymbol{z}'')>g(\boldsymbol{x},\boldsymbol{z}'\wedge\boldsymbol{z}'')+g(\boldsymbol{x},\boldsymbol{z}'\vee\boldsymbol{z}'')$, which contradicts the supermodularity. The "only if" part is completed. ☐

### C.7. Proof of Theorem 2

Based on Proposition 4, the above theorem is equivalent to this statement: $\mathcal{S}(\boldsymbol{x})$ is an additive inverse lattice for all $\boldsymbol{x}$ and $\boldsymbol{v}^0$ if and only if $\boldsymbol{U}$ and $\boldsymbol{V}$ satisfy one of the two conditions in the above theorem. We prove the equivalent statement as follows.

First we prove the "if" direction by contradiction. Suppose there exist $\boldsymbol{x}$ and $\boldsymbol{v}^0$ such that $\mathcal{S}(\boldsymbol{x})$ is not an additive inverse lattice, i.e., we have $\boldsymbol{z}',\boldsymbol{z}'',\boldsymbol{p},\boldsymbol{q}$ with $\boldsymbol{z}^\wedge=\boldsymbol{z}'\wedge\boldsymbol{z}'',\boldsymbol{z}^\vee=\boldsymbol{z}'\vee\boldsymbol{z}''$ and $(\boldsymbol{p},\boldsymbol{z}^\wedge),(\boldsymbol{q},\boldsymbol{z}^\vee)\in\mathcal{S}(\boldsymbol{x})$, such that $\boldsymbol{y}'+\boldsymbol{y}''\neq\boldsymbol{p}+\boldsymbol{q}$ holds for all $\boldsymbol{y}',\boldsymbol{y}''$ with $(\boldsymbol{y}',\boldsymbol{z}'),(\boldsymbol{y}'',\boldsymbol{z}'')\in\mathcal{S}(\boldsymbol{x})$.

We denote $\boldsymbol{c}=-\boldsymbol{W}\boldsymbol{x}+\boldsymbol{v}^0$, $\boldsymbol{t}^1=\boldsymbol{U}\boldsymbol{p}-\boldsymbol{V}\boldsymbol{z}^\wedge\ge\boldsymbol{c}$, $\boldsymbol{t}^2=\boldsymbol{U}\boldsymbol{q}-\boldsymbol{V}\boldsymbol{z}^\vee\ge\boldsymbol{c}$. Here the two inequalities are due to $(\boldsymbol{p},\boldsymbol{z}^\wedge),(\boldsymbol{q},\boldsymbol{z}^\vee)\in\mathcal{S}(\boldsymbol{x})$ and the definition of $\mathcal{S}(\boldsymbol{x})$. We define a set $\mathcal{W}$ as

$$\mathcal{W}=\left\{\boldsymbol{y}\in\Re^m\;\middle|\;(\boldsymbol{t}^1\wedge\boldsymbol{t}^2)+\boldsymbol{V}\boldsymbol{z}'\le\boldsymbol{U}\boldsymbol{y}\le(\boldsymbol{t}^1\vee\boldsymbol{t}^2)+\boldsymbol{V}\boldsymbol{z}'\right\}.$$

Note that $\mathcal{W}$ should be an empty set, otherwise there exists a $\boldsymbol{y}^0\in\mathcal{W}$ and hence

$$\begin{gathered}
\boldsymbol{U}\boldsymbol{y}^0-\boldsymbol{V}\boldsymbol{z}'\ge(\boldsymbol{t}^1\wedge\boldsymbol{t}^2)\ge\boldsymbol{c},\\
\boldsymbol{U}(\boldsymbol{p}+\boldsymbol{q}-\boldsymbol{y}^0)-\boldsymbol{V}\boldsymbol{z}''=\boldsymbol{U}\boldsymbol{p}-\boldsymbol{V}\boldsymbol{z}^\wedge+\boldsymbol{U}\boldsymbol{q}-\boldsymbol{V}\boldsymbol{z}^\vee-(\boldsymbol{U}\boldsymbol{y}^0-\boldsymbol{V}\boldsymbol{z}')\ge\boldsymbol{t}^1+\boldsymbol{t}^2-(\boldsymbol{t}^1\vee\boldsymbol{t}^2)=\boldsymbol{t}^1\wedge\boldsymbol{t}^2\ge\boldsymbol{c},
\end{gathered}$$

which implies both $(\boldsymbol{y}^0, \boldsymbol{z}'), (\boldsymbol{p}+\boldsymbol{q}-\boldsymbol{y}^0, \boldsymbol{z}'') \in \mathcal{S}(\boldsymbol{x})$, and contradicts the previous statement on $\boldsymbol{y}', \boldsymbol{y}''$ resulting from the assumption.

We now show that the first part of the condition in our theorem is not true. If $\text{rank}(\boldsymbol{U}) = r$, we can solve $\boldsymbol{y}$ with $\boldsymbol{U}\boldsymbol{y} = (\boldsymbol{t}^1 \wedge \boldsymbol{t}^2) + \boldsymbol{V}\boldsymbol{z}' \le (\boldsymbol{t}^1 \vee \boldsymbol{t}^2) + \boldsymbol{V}\boldsymbol{z}'$, which contradicts the emptiness of $\mathcal{W}$. Therefore, $\text{rank}(\boldsymbol{U}) < r$.

We then focus on the second part of the condition in our theorem. The emptiness of $\mathcal{W}$ leads to the infeasibility of the following optimization problem:

$$\begin{aligned} \max \quad & 0 \\ \text{s.t.} \quad & \begin{bmatrix} \boldsymbol{U} \\ -\boldsymbol{U} \end{bmatrix} \boldsymbol{y} \le \begin{bmatrix} (\boldsymbol{t}^1 \vee \boldsymbol{t}^2) + \boldsymbol{V}\boldsymbol{z}' \\ -(\boldsymbol{t}^1 \wedge \boldsymbol{t}^2) - \boldsymbol{V}\boldsymbol{z}' \end{bmatrix}. \end{aligned}$$

Furthermore, by Lemma 4 we know that there exists $\mathcal{I} \subseteq [r], |\mathcal{I}| = \text{rank}(\boldsymbol{U}) + 1$ with $\text{rank}(\boldsymbol{U}_{\mathcal{I}}) = \text{rank}(\boldsymbol{U})$ such that the problem

$$\begin{aligned} \max \quad & 0 \\ \text{s.t.} \quad & \begin{bmatrix} \boldsymbol{U}_{\mathcal{I}} \\ -\boldsymbol{U}_{\mathcal{I}} \end{bmatrix} \boldsymbol{y} \le \begin{bmatrix} (\boldsymbol{t}^1_{\mathcal{I}} \vee \boldsymbol{t}^2_{\mathcal{I}}) + \boldsymbol{V}_{\mathcal{I}}\boldsymbol{z}' \\ -(\boldsymbol{t}^1_{\mathcal{I}} \wedge \boldsymbol{t}^2_{\mathcal{I}}) - \boldsymbol{V}_{\mathcal{I}}\boldsymbol{z}' \end{bmatrix} \end{aligned} \tag{40}$$

is also infeasible. We write the dual of (40) as follows,

$$\begin{aligned} \min \quad & \boldsymbol{r}^\top \left( (\boldsymbol{t}^1_{\mathcal{I}} \vee \boldsymbol{t}^2_{\mathcal{I}}) + \boldsymbol{V}_{\mathcal{I}}\boldsymbol{z}' \right) - \boldsymbol{s}^\top \left( (\boldsymbol{t}^1_{\mathcal{I}} \wedge \boldsymbol{t}^2_{\mathcal{I}}) + \boldsymbol{V}_{\mathcal{I}}\boldsymbol{z}' \right) \\ \text{s.t.} \quad & \boldsymbol{U}_{\mathcal{I}}^\top (\boldsymbol{r} - \boldsymbol{s}) = \boldsymbol{0} \\ & \boldsymbol{r}, \boldsymbol{s} \ge \boldsymbol{0}. \end{aligned} \tag{41}$$

Observing that $\boldsymbol{r} = \boldsymbol{s} = \boldsymbol{0}$ gives a feasible solution of (41), the infeasibility of the primal problem implies the unboundedness of the above dual problem. Therefore, there exist $\boldsymbol{r}, \boldsymbol{s} \ge \boldsymbol{0}$ with $\boldsymbol{U}_{\mathcal{I}}^\top (\boldsymbol{r} - \boldsymbol{s}) = \boldsymbol{0}$ such that the following inequalities holds,

$$\begin{aligned} 0 &> \boldsymbol{r}^\top \left( (\boldsymbol{t}^1_{\mathcal{I}} \vee \boldsymbol{t}^2_{\mathcal{I}}) + \boldsymbol{V}_{\mathcal{I}}\boldsymbol{z}' \right) - \boldsymbol{s}^\top \left( (\boldsymbol{t}^1_{\mathcal{I}} \wedge \boldsymbol{t}^2_{\mathcal{I}}) + \boldsymbol{V}_{\mathcal{I}}\boldsymbol{z}' \right) \\ &= \boldsymbol{r}^\top \left( (\boldsymbol{t}^1_{\mathcal{I}} \vee \boldsymbol{t}^2_{\mathcal{I}}) + \boldsymbol{V}_{\mathcal{I}}\boldsymbol{z}' - \boldsymbol{U}_{\mathcal{I}}\boldsymbol{q} \right) - \boldsymbol{s}^\top \left( (\boldsymbol{t}^1_{\mathcal{I}} \wedge \boldsymbol{t}^2_{\mathcal{I}}) + \boldsymbol{V}_{\mathcal{I}}\boldsymbol{z}' - \boldsymbol{U}_{\mathcal{I}}\boldsymbol{q} \right) \\ &\ge \boldsymbol{r}^\top \left( \boldsymbol{t}^2_{\mathcal{I}} + \boldsymbol{V}_{\mathcal{I}}\boldsymbol{z}' - \boldsymbol{U}_{\mathcal{I}}\boldsymbol{q} \right) - \boldsymbol{s}^\top \left( \boldsymbol{t}^2_{\mathcal{I}} + \boldsymbol{V}_{\mathcal{I}}\boldsymbol{z}' - \boldsymbol{U}_{\mathcal{I}}\boldsymbol{q} \right) \\ &= (\boldsymbol{r} - \boldsymbol{s})^\top \boldsymbol{V}_{\mathcal{I}} (\boldsymbol{z}' - \boldsymbol{z}^\vee), \end{aligned}$$

where the first inequality is obtained from the unboundedness of (41), the first equality is due to $\boldsymbol{U}_{\mathcal{I}}^\top (\boldsymbol{r} - \boldsymbol{s}) = \boldsymbol{0}$, the second inequality follows from $\boldsymbol{t}^1_{\mathcal{I}} \wedge \boldsymbol{t}^2_{\mathcal{I}} \le \boldsymbol{t}^1_{\mathcal{I}} \le \boldsymbol{t}^1_{\mathcal{I}} \vee \boldsymbol{t}^2_{\mathcal{I}}$, and the second equality comes from $\boldsymbol{t}^2_{\mathcal{I}} = \boldsymbol{U}_{\mathcal{I}}\boldsymbol{q} - \boldsymbol{V}_{\mathcal{I}}\boldsymbol{z}^\vee$. We remark that in the above equation, if we use $\boldsymbol{U}_{\mathcal{I}}\boldsymbol{p}$ instead of $\boldsymbol{U}_{\mathcal{I}}\boldsymbol{q}$ in the first equality, and $\boldsymbol{t}^1_{\mathcal{I}}$ instead of $\boldsymbol{t}^2_{\mathcal{I}}$ in the second inequality, then $0 > (\boldsymbol{r} - \boldsymbol{s})^\top \boldsymbol{V}_{\mathcal{I}} (\boldsymbol{z}' - \boldsymbol{z}^\wedge)$ can be obtained similarly.

We define $\Delta_1 = (\boldsymbol{r} - \boldsymbol{s})^\top \boldsymbol{V}_{\mathcal{I}} (\boldsymbol{z}' - \boldsymbol{z}^\vee)$, $\Delta_2 = (\boldsymbol{r} - \boldsymbol{s})^\top \boldsymbol{V}_{\mathcal{I}} (\boldsymbol{z}' - \boldsymbol{z}^\wedge)$, and $\boldsymbol{\beta} = \dfrac{\boldsymbol{z}' - \boldsymbol{z}^\vee}{\Delta_1} - \dfrac{\boldsymbol{z}' - \boldsymbol{z}^\wedge}{\Delta_2}$.

We have three observations on $\boldsymbol{\beta}$. First, $\boldsymbol{\beta} \ge \boldsymbol{0}$ since $\Delta_1, \Delta_2 < 0$ and $\boldsymbol{z}^\wedge \le \boldsymbol{z}' \le \boldsymbol{z}^\vee$.

Second, $\boldsymbol{V}_{\mathcal{I}}\boldsymbol{\beta} \in \operatorname{span}(\boldsymbol{U}_{\mathcal{I}})$. To see this, recall that for any matrix, its column space is the orthogonal complement of the null space of its transpose; therefore, we can equivalently show that $\operatorname{null}(\boldsymbol{U}_{\mathcal{I}}^\top) \subseteq \operatorname{null}((\boldsymbol{V}_{\mathcal{I}}\boldsymbol{\beta})^\top)$, where $\operatorname{null}(\cdot)$ is the null space of a given matrix. Since $|\mathcal{I}| = \operatorname{rank}(\boldsymbol{U}) + 1 = \operatorname{rank}(\boldsymbol{U}_{\mathcal{I}}) + 1 = \operatorname{rank}(\boldsymbol{U}_{\mathcal{I}}^\top) + 1$, $\operatorname{null}(\boldsymbol{U}_{\mathcal{I}}^\top)$ is of dimension 1. That implies for any $\boldsymbol{w} \in \operatorname{null}(\boldsymbol{U}_{\mathcal{I}}^\top)$, we have $\boldsymbol{w} = k(\boldsymbol{r} - \boldsymbol{s})$ for some $k \in \Re$. Therefore,

$$(\boldsymbol{V}_{\mathcal{I}}\boldsymbol{\beta})^\top \boldsymbol{w} = k(\boldsymbol{r}-\boldsymbol{s})^\top \boldsymbol{V}_{\mathcal{I}}\boldsymbol{\beta} = k\left(\frac{(\boldsymbol{r}-\boldsymbol{s})^\top \boldsymbol{V}_{\mathcal{I}}(\boldsymbol{z}'-\boldsymbol{z}^\vee)}{\Delta_1} - \frac{(\boldsymbol{r}-\boldsymbol{s})^\top \boldsymbol{V}_{\mathcal{I}}(\boldsymbol{z}'-\boldsymbol{z}^\wedge)}{\Delta_2}\right) = k\,(1-1) = 0.$$

That is, $\boldsymbol{w} \in \operatorname{null}((\boldsymbol{V}_{\mathcal{I}}\boldsymbol{\beta})^\top)$. Hence, $\operatorname{null}(\boldsymbol{U}_{\mathcal{I}}^\top) \subseteq \operatorname{null}((\boldsymbol{V}_{\mathcal{I}}\boldsymbol{\beta})^\top)$ and then we have $\boldsymbol{V}_{\mathcal{I}}\boldsymbol{\beta} \in \operatorname{span}(\boldsymbol{U}_{\mathcal{I}})$.

The third observation is that there exists some $i \in [n]$ such that $(\boldsymbol{V}_{\mathcal{I}})_i\beta_i \notin \operatorname{span}(\boldsymbol{U}_{\mathcal{I}})$. To show this, we denote $\mathcal{H} = \{i \in [n] \mid z_i' \leq z_i''\}$. We then have for every $i \in \mathcal{H}$, $z_i^\wedge = z_i'$, $z_i^\vee = z_i''$ and hence $\beta_i = \dfrac{z_i' - z_i^\vee}{\Delta_1}$. In addition, since for every $i \in [n] \setminus \mathcal{H}$, $z_i' > z_i''$, $\dfrac{z_i' - z_i^\vee}{\Delta_1} = \dfrac{z_i' - z_i'}{\Delta_1} = 0$, we have

$$(\boldsymbol{r}-\boldsymbol{s})^\top \sum_{i\in\mathcal{H}} \beta_i (\boldsymbol{V}_{\mathcal{I}})_i = (\boldsymbol{r}-\boldsymbol{s})^\top \left(\sum_{i\in\mathcal{H}} \beta_i(\boldsymbol{V}_{\mathcal{I}})_i + \sum_{i\in[n]\setminus\mathcal{H}} 0\cdot(\boldsymbol{V}_{\mathcal{I}})_i\right) = (\boldsymbol{r}-\boldsymbol{s})^\top \boldsymbol{V}_{\mathcal{I}} \frac{\boldsymbol{z}'-\boldsymbol{z}^\vee}{\Delta_1} = 1.$$

Hence, $(\boldsymbol{r}-\boldsymbol{s}) \notin \operatorname{null}\left(\left(\sum_{i\in\mathcal{H}}\beta_i(\boldsymbol{V}_{\mathcal{I}})_i\right)^\top\right)$, which implies that $\operatorname{null}(\boldsymbol{U}_{\mathcal{I}}^\top)$ is not a subset of $\operatorname{null}\left(\left(\sum_{i\in\mathcal{H}}\beta_i(\boldsymbol{V}_{\mathcal{I}})_i\right)^\top\right)$. Consequently we have $\sum_{i\in\mathcal{H}}\beta_i(\boldsymbol{V}_{\mathcal{I}})_i \notin \operatorname{span}(\boldsymbol{U}_{\mathcal{I}})$, implying that there exists some $i \in \mathcal{H}$ such that $(\boldsymbol{V}_{\mathcal{I}})_i\beta_i \notin \operatorname{span}(\boldsymbol{U}_{\mathcal{I}})$.

With the three observations, we have a contradiction of the second condition in Theorem 2.

We next prove the "only if" direction by contradiction. Assume the condition in the theorem is not satisfied. That is, $\operatorname{rank}(\boldsymbol{U}) < r$ and there exist some $\mathcal{I} \subseteq [r]$, $\boldsymbol{\beta} \in \Re^n_+$ satisfying $|\mathcal{I}| = \operatorname{rank}(\boldsymbol{U}) + 1$, $\operatorname{rank}(\boldsymbol{U}_{\mathcal{I}}) = \operatorname{rank}(\boldsymbol{U})$ and $\boldsymbol{V}_{\mathcal{I}}\boldsymbol{\beta} \in \operatorname{span}(\boldsymbol{U}_{\mathcal{I}})$, such that $\beta_i(\boldsymbol{V}_{\mathcal{I}})_i \notin \operatorname{span}(\boldsymbol{U}_{\mathcal{I}})$ for some $i \in [n]$. Note that in this case, we can find a vector $\boldsymbol{\alpha} \in \Re^m$ such that $\boldsymbol{U}_{\mathcal{I}}\boldsymbol{\alpha} = \boldsymbol{V}_{\mathcal{I}}\boldsymbol{\beta}$.

We arbitrarily choose $\boldsymbol{z}^\wedge \in \Re^n, \boldsymbol{p} \in \Re^m$ and let $\boldsymbol{z}^\vee = \boldsymbol{z}^\wedge + \boldsymbol{\beta} \geq \boldsymbol{z}^\wedge, \boldsymbol{q} = \boldsymbol{p} + \boldsymbol{\alpha}$, then $\boldsymbol{U}_{\mathcal{I}}\boldsymbol{p} - \boldsymbol{V}_{\mathcal{I}}\boldsymbol{z}^\wedge = \boldsymbol{U}_{\mathcal{I}}(\boldsymbol{q}-\boldsymbol{\alpha}) - \boldsymbol{V}_{\mathcal{I}}(\boldsymbol{z}^\vee - \boldsymbol{\beta}) = \boldsymbol{U}_{\mathcal{I}}\boldsymbol{q} - \boldsymbol{V}_{\mathcal{I}}\boldsymbol{z}^\vee$. We also arbitrarily choose $\boldsymbol{x}$, and then choose $\boldsymbol{v}^0$ such that $\boldsymbol{c} = -\boldsymbol{W}\boldsymbol{x} + \boldsymbol{v}^0$ is with $\boldsymbol{c}_{\mathcal{I}} = \boldsymbol{U}_{\mathcal{I}}\boldsymbol{p} - \boldsymbol{V}_{\mathcal{I}}\boldsymbol{z}^\wedge$ and $c_j$ being sufficiently small for all $j \notin \mathcal{I}$ . Then we have $(\boldsymbol{p}, \boldsymbol{z}^\wedge), (\boldsymbol{q}, \boldsymbol{z}^\vee) \in \mathcal{S}(\boldsymbol{x})$. We further define $\boldsymbol{z}' = \boldsymbol{z}^\wedge + \beta_i\boldsymbol{e}_i, \boldsymbol{z}'' = \boldsymbol{z}^\vee - \beta_i\boldsymbol{e}_i$ so that $\boldsymbol{z}' \wedge \boldsymbol{z}'' = \boldsymbol{z}^\wedge, \boldsymbol{z}' \vee \boldsymbol{z}'' = \boldsymbol{z}^\vee$. Then we have

$$\boldsymbol{c}_{\mathcal{I}} + \boldsymbol{V}_{\mathcal{I}}\boldsymbol{z}' = \boldsymbol{c}_{\mathcal{I}} + \boldsymbol{V}_{\mathcal{I}}(\boldsymbol{z}^\wedge + \beta_i\boldsymbol{e}_i) = \boldsymbol{U}_{\mathcal{I}}\boldsymbol{p} - \boldsymbol{V}_{\mathcal{I}}\boldsymbol{z}^\wedge + \boldsymbol{V}_{\mathcal{I}}(\boldsymbol{z}^\wedge + \beta_i\boldsymbol{e}_i) = \boldsymbol{U}_{\mathcal{I}}\boldsymbol{p} + \beta_i(\boldsymbol{V}_{\mathcal{I}})_i \notin \operatorname{span}(\boldsymbol{U}_{\mathcal{I}}),$$

where the last relationship holds since $\boldsymbol{U}_{\mathcal{I}}\boldsymbol{p} \in \operatorname{span}(\boldsymbol{U}_{\mathcal{I}})$ but $\beta_i(\boldsymbol{V}_{\mathcal{I}})_i \notin \operatorname{span}(\boldsymbol{U}_{\mathcal{I}})$.

Hence, $\{\boldsymbol{y} \in \Re^m \mid \boldsymbol{U}_{\mathcal{I}}\boldsymbol{y} = \boldsymbol{c}_{\mathcal{I}} + \boldsymbol{V}_{\mathcal{I}}\boldsymbol{z}'\} = \emptyset$, i.e. for any $\boldsymbol{y}'$ satisfying $\boldsymbol{U}\boldsymbol{y}' - \boldsymbol{V}\boldsymbol{z}' \geq \boldsymbol{c}$, there exists $j \in \mathcal{I}$ such that $\boldsymbol{u}_j^\top\boldsymbol{y}' - \boldsymbol{v}_j^\top\boldsymbol{z}' > c_j$. If there exists some $\boldsymbol{y}''$ with $\boldsymbol{U}\boldsymbol{y}'' - \boldsymbol{V}\boldsymbol{z}'' \geq \boldsymbol{c}$ satisfies $\boldsymbol{y}' + \boldsymbol{y}'' = \boldsymbol{p} + \boldsymbol{q}$,

$$\begin{aligned}
\boldsymbol{U}_{\mathcal{I}}\boldsymbol{y}'' - \boldsymbol{V}_{\mathcal{I}}\boldsymbol{z}'' &= \boldsymbol{U}_{\mathcal{I}}(\boldsymbol{p}+\boldsymbol{q}-\boldsymbol{y}') - \boldsymbol{V}_{\mathcal{I}}(\boldsymbol{z}^\wedge + \boldsymbol{z}^\vee - \boldsymbol{z}') \\
&= \boldsymbol{U}_{\mathcal{I}}\boldsymbol{p} - \boldsymbol{V}_{\mathcal{I}}\boldsymbol{z}^\wedge + \boldsymbol{U}_{\mathcal{I}}\boldsymbol{q} - \boldsymbol{V}_{\mathcal{I}}\boldsymbol{z}^\vee - (\boldsymbol{U}_{\mathcal{I}}\boldsymbol{y}' - \boldsymbol{V}_{\mathcal{I}}\boldsymbol{z}') \\
&= 2\boldsymbol{c}_{\mathcal{I}} - (\boldsymbol{U}_{\mathcal{I}}\boldsymbol{y}' - \boldsymbol{V}_{\mathcal{I}}\boldsymbol{z}'),
\end{aligned}$$

then we should have $2c_j - (\boldsymbol{u}_j^\top \boldsymbol{y}' - \boldsymbol{v}_j^\top \boldsymbol{z}') < c_j$ for the above mentioned $j$, which contradicts the assumption $(\boldsymbol{y}'', \boldsymbol{z}'') \in \mathcal{S}(\boldsymbol{x})$. Hence we prove the necessity of the conditions on $\boldsymbol{U}, \boldsymbol{V}$. $\square$

**Lemma 4 (Chen et al. 2021)** *Consider any matrix* $\boldsymbol{U} \in \Re^{r\times m}$ *with* $\text{rank}(\boldsymbol{U}) < r$. *Suppose that system* $\begin{cases} \boldsymbol{U}\boldsymbol{x} \le \overline{\boldsymbol{c}} \\ -\boldsymbol{U}\boldsymbol{x} \le -\underline{\boldsymbol{c}} \end{cases}$ *is infeasible. Then there exists* $\mathcal{I} \subseteq [r]$ *with* $|\mathcal{I}| = \text{rank}(\boldsymbol{U}) + 1$ *and* $\text{rank}(\boldsymbol{U}_\mathcal{I}) = \text{rank}(\boldsymbol{U})$ *such that system* $\begin{cases} \boldsymbol{U}_\mathcal{I}\boldsymbol{x} \le \overline{\boldsymbol{c}}_\mathcal{I} \\ -\boldsymbol{U}_\mathcal{I}\boldsymbol{x} \le -\underline{\boldsymbol{c}}_\mathcal{I} \end{cases}$ *is also infeasible.*

### C.8. Proof of Proposition 5

"$\Leftarrow$" Assume there exists a $2\times 3$ submatrix of $\boldsymbol{U}^\circ$ such that any pair of columns in it are linearly independent. WLOG, let $\boldsymbol{U}^\circ_{\{1,2\},\{1,2,3\}}$ be such matrix and we denote it by $\boldsymbol{A} = [\boldsymbol{A}_1\ \boldsymbol{A}_2\ \boldsymbol{A}_3] \in \Re^{2\times 3}$. WLOG, assume $\boldsymbol{A}_3 = t_1\boldsymbol{A}_1 + t_2\boldsymbol{A}_2$ with $t_1, t_2 > 0$. Choose $\boldsymbol{V}^1 = \boldsymbol{I}_{m\times m}$, $\mathcal{I} = [m+2]\setminus\{3\}$, $\boldsymbol{\beta} = t_1\boldsymbol{e}_1 + t_2\boldsymbol{e}_2 \ge \boldsymbol{0}$, $\boldsymbol{\alpha} = t_1\boldsymbol{e}_1 + t_2\boldsymbol{e}_2 - \boldsymbol{e}_3$. We then have $\boldsymbol{V}_\mathcal{I}\boldsymbol{\beta} = t_1\boldsymbol{e}_1 + t_2\boldsymbol{e}_2$; at the same time, $\boldsymbol{U}_\mathcal{I}\boldsymbol{\alpha} = t_1\boldsymbol{e}_1 + t_2\boldsymbol{e}_2$ since $\boldsymbol{A}_3 = t_1\boldsymbol{A}_1 + t_2\boldsymbol{A}_2$. Hence, $\boldsymbol{V}_\mathcal{I}\boldsymbol{\beta} = \boldsymbol{U}_\mathcal{I}\boldsymbol{\alpha} \in \text{span}(\boldsymbol{U}_\mathcal{I})$. However, $(\boldsymbol{V}_\mathcal{I})_1\beta_1 = \beta_1\boldsymbol{e}_1 \notin \text{span}(\boldsymbol{U}_\mathcal{I})$. Therefore, the second condition in Theorem 2 is violated, there exists an instance of $g(\boldsymbol{x}, \boldsymbol{z})$ which is not supermodular in $\boldsymbol{z}$.

"$\Rightarrow$" Assume that every $2\times 3$ submatrix of $\boldsymbol{U}^\circ$ contains at least one pair of column vectors which are linearly dependent. We prove the result by showing the second condition in Theorem 2 is always satisfied. To see this, consider any $\mathcal{I} \subseteq [r]$ such that $|\mathcal{I}| = m+1$, $\text{rank}(\boldsymbol{U}_\mathcal{I}) = m$. Let $\mathcal{I}_1 = \mathcal{I} \cap [m]$ and $\mathcal{I}_0 = \mathcal{I} \cap \{m+1, \ldots, r\}$ be a partition of $\mathcal{I}$, hence the submatrix $\boldsymbol{U}_{\mathcal{I}_1}$ is extracted from $\boldsymbol{I}_{m\times m}$ and $\boldsymbol{U}_{\mathcal{I}_0}$ is from $\boldsymbol{U}^\circ$. We further let $\mathcal{J}_1$, $\mathcal{J}_0$ be a partition of $[m]$ such that $U_{\mathcal{I}_1,\mathcal{J}_1}$ contains all nonzero columns in $U_{\mathcal{I}_1}$ and hence $U_{\mathcal{I}_1,\mathcal{J}_0} = \boldsymbol{0}$. Noting that $U_{\mathcal{I}_1}$ contains rows extracted from $\boldsymbol{I}_{m\times m}$, we know $\mathcal{I}_1 = \mathcal{J}_1$. Hence, $|\mathcal{I}_0| = m+1-|\mathcal{I}_1| = m+1-|\mathcal{J}_1| = m+1-(m-|\mathcal{J}_0|) = |\mathcal{J}_0|+1$. We illustrate the partition of $U_\mathcal{I}$ as follows,

$$\boldsymbol{U}_\mathcal{I} = \begin{bmatrix} \boldsymbol{U}_{\mathcal{I}_1} \\ \boldsymbol{U}_{\mathcal{I}_0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_{|\mathcal{I}_1|\times|\mathcal{I}_1|} & \boldsymbol{0}_{|\mathcal{I}_1|\times|\mathcal{J}_0|} \\ \boldsymbol{U}_{\mathcal{I}_0,\mathcal{J}_1} & \boldsymbol{U}_{\mathcal{I}_0,\mathcal{J}_0} \end{bmatrix}.$$

We first prove that there exists a unit vector $\boldsymbol{p} \in \Re^{|\mathcal{I}_0|}$, such that it is orthogonal to $\text{span}(\boldsymbol{U}_{\mathcal{I}_0,\mathcal{J}_0})$ and $\text{span}[\boldsymbol{U}_{\mathcal{I}_0,\mathcal{J}_0}\ \boldsymbol{p}] = \Re^{|\mathcal{I}_0|}$. Notice that $\boldsymbol{U}_\mathcal{I}$ is of full column rank, and hence so does its submatrix $\boldsymbol{U}_{\mathcal{I},\mathcal{J}_0} = \begin{bmatrix} \boldsymbol{0}_{|\mathcal{I}_1|\times|\mathcal{J}_0|} \\ \boldsymbol{U}_{\mathcal{I}_0,\mathcal{J}_0} \end{bmatrix}$, which implies $\boldsymbol{U}_{\mathcal{I}_0,\mathcal{J}_0} \in \Re^{|\mathcal{I}_0|\times|\mathcal{J}_0|}$ is also of full column rank. Therefore, $\text{span}(\boldsymbol{U}_{\mathcal{I}_0,\mathcal{J}_0})$ is of dimension $|\mathcal{J}_0| = |\mathcal{I}_0| - 1$, the existence of $\boldsymbol{p}$ can be proved.

We now show that the orthogonal unit vector $\boldsymbol{p}$ can be chosen such that for all $i \in \mathcal{J}_1$, there exist some $\boldsymbol{s}_i \in \text{span}(\boldsymbol{U}_{\mathcal{I}_0,\mathcal{J}_0})$ and $\gamma_i \ge 0$ such that $(\boldsymbol{U}_{\mathcal{I}_0})_i = \boldsymbol{s}_i + \gamma_i\boldsymbol{p}$. For those $i \in \mathcal{J}_1$ with $(\boldsymbol{U}_{\mathcal{I}_0})_i \in \text{span}(\boldsymbol{U}_{\mathcal{I}_0,\mathcal{J}_0})$, we always have $\gamma_i = 0$ regardless of the choice of orthogonal vector $\boldsymbol{p}$. Now we consider any given $j \in \mathcal{J}_1$ with $(\boldsymbol{U}_{\mathcal{I}_0})_j \notin \text{span}(\boldsymbol{U}_{\mathcal{I}_0,\mathcal{J}_0})$. Since $\text{span}[\boldsymbol{U}_{\mathcal{I}_0,\mathcal{J}_0}\ \boldsymbol{p}] = \Re^{|\mathcal{I}_0|}$, we can surely represent $(\boldsymbol{U}_{\mathcal{I}_0})_j = \boldsymbol{s}_j + \gamma_j\boldsymbol{p}$ for some $\boldsymbol{s}_j \in \text{span}(\boldsymbol{U}_{\mathcal{I}_0,\mathcal{J}_0})$ and $\gamma_j \ne 0$. Moreover, the unit vector $\boldsymbol{p}$ can be chosen (as $-\boldsymbol{p}$, if necessary) to make $\gamma_j > 0$. Consider any $k \in \mathcal{J}_1\setminus\{j\}$ with $(\boldsymbol{U}_{\mathcal{I}_0})_j \notin \text{span}(\boldsymbol{U}_{\mathcal{I}_0,\mathcal{J}_0})$. Denote

$\boldsymbol{Q} = [(\boldsymbol{U}_{\mathcal{I}_0})_j \ (\boldsymbol{U}_{\mathcal{I}_0})_k \ \boldsymbol{U}_{\mathcal{I}_0,\mathcal{J}_0}]$. Notice that every $2\times 3$ submatrix of $\boldsymbol{U}^\circ$, and hence that of $\boldsymbol{Q}$, contains at least one pair of column vectors which are linearly dependent. By Lemma 5, there are at least one pair of columns in $\boldsymbol{Q}$ which are linearly dependent. Since $\boldsymbol{U}_{\mathcal{I}_0,\mathcal{J}_0}$ is of full column rank and $(\boldsymbol{U}_{\mathcal{I}_0})_j, (\boldsymbol{U}_{\mathcal{I}_0})_k \notin \operatorname{span}(\boldsymbol{U}_{\mathcal{I}_0,\mathcal{J}_0})$, the two linearly dependent columns can only be $(\boldsymbol{U}_{\mathcal{I}_0})_j, (\boldsymbol{U}_{\mathcal{I}_0})_k$, i.e., $(\boldsymbol{U}_{\mathcal{I}_0})_k = \zeta(\boldsymbol{U}_{\mathcal{I}_0})_j$ for some $\zeta \neq 0$ (recall that both $(\boldsymbol{U}_{\mathcal{I}_0})_k$ and $(\boldsymbol{U}_{\mathcal{I}_0})_j$ are nonzero vector since they are not in $\operatorname{span}(\boldsymbol{U}_{\mathcal{I}_0,\mathcal{J}_0})$). As all components in the same row of $\boldsymbol{U}$ are with the same sign, we know $\zeta > 0$. Therefore, $(\boldsymbol{U}_{\mathcal{I}_0})_k = \zeta\,(\boldsymbol{s}_j + \gamma_j \boldsymbol{p}) = \zeta \boldsymbol{s}_j + \zeta\gamma_j \boldsymbol{p}$ where $\zeta\gamma_j > 0$.

We are now ready to prove the second condition in Theorem 2 holds. Consider any $\boldsymbol{\beta} \geq \boldsymbol{0}$ and $\boldsymbol{\alpha}$ such that $\boldsymbol{V}_{\mathcal{I}}\boldsymbol{\beta} = \boldsymbol{U}_{\mathcal{I}}\boldsymbol{\alpha}$. Observing the first block, characterized by $\mathcal{I}_1$, we have $\boldsymbol{V}_{\mathcal{I}_1}\boldsymbol{\beta} = \boldsymbol{U}_{\mathcal{I}_1}\boldsymbol{\alpha} = (\boldsymbol{I}_{m\times m})_{\mathcal{I}_1}\boldsymbol{\alpha} = \boldsymbol{\alpha}_{\mathcal{I}_1}$; since $\boldsymbol{V}, \boldsymbol{\beta}$ are both nonnegative, we have $\boldsymbol{\alpha}_{\mathcal{I}_1} \geq \boldsymbol{0}$. Observing the second block, characterized by $\mathcal{I}_0$, by $\boldsymbol{V}_{\mathcal{I}_0} = \boldsymbol{0}$, we have

$$\boldsymbol{0} = \boldsymbol{V}_{\mathcal{I}_0}\boldsymbol{\beta} = \boldsymbol{U}_{\mathcal{I}_0}\boldsymbol{\alpha} = [\boldsymbol{U}_{\mathcal{I}_0,\mathcal{J}_1} \ \boldsymbol{U}_{\mathcal{I}_0,\mathcal{J}_0}] \begin{bmatrix} \boldsymbol{\alpha}_{\mathcal{J}_1} \\ \boldsymbol{\alpha}_{\mathcal{J}_0} \end{bmatrix} = \boldsymbol{U}_{\mathcal{I}_0,\mathcal{J}_1}\boldsymbol{\alpha}_{\mathcal{J}_1} + \boldsymbol{U}_{\mathcal{I}_0,\mathcal{J}_0}\boldsymbol{\alpha}_{\mathcal{J}_0} = \boldsymbol{s} + \boldsymbol{p}\sum_{i\in\mathcal{N}} \alpha_i\gamma_i \tag{42}$$

for some $\boldsymbol{s} \in \operatorname{span}(\boldsymbol{U}_{\mathcal{I}_0,\mathcal{J}_0})$. Here we denote the index set $\mathcal{N} = \{i \in \mathcal{J}_1 \mid (\boldsymbol{U}_{\mathcal{I}_0})_i \notin \operatorname{span}(\boldsymbol{U}_{\mathcal{I}_0,\mathcal{J}_0})\}$ and hence the last equality above holds due to the argument proved in the last paragraph. Moreover, since $\boldsymbol{s}$ and $\boldsymbol{p}$ are orthogonal, by (42) we have $\sum_{i\in\mathcal{N}} \alpha_i\gamma_i = 0$, which implies $\alpha_i = 0$ for all $i \in \mathcal{N}$, as we have already known $\gamma_i > 0, \alpha_i \geq 0$ holds for all $i \in \mathcal{N}$ (recall that $\mathcal{N} \subseteq \mathcal{J}_1 = \mathcal{I}_1$, and $\boldsymbol{\alpha}_{\mathcal{I}_1} \geq \boldsymbol{0}$). Therefore, the equation $0 = \alpha_i = \boldsymbol{u}_i^\top \boldsymbol{\alpha} = \boldsymbol{v}_i^\top \boldsymbol{\beta}$ holds for any $i \in \mathcal{N}$, where the last equality is due to $\boldsymbol{V}_{\mathcal{I}}\boldsymbol{\beta} = \boldsymbol{U}_{\mathcal{I}}\boldsymbol{\alpha}$. As $\boldsymbol{V}, \boldsymbol{\beta} \geq \boldsymbol{0}$ for all $i \in \mathcal{N}$, $\boldsymbol{v}_i^\top\boldsymbol{\beta} = 0$ implies $v_{ik}\beta_k = 0$ for all $k \in [m]$. We now consider any $j \in [m]$ and it remains to show $(\boldsymbol{V}_{\mathcal{I}})_j\beta_j = \boldsymbol{U}_{\mathcal{I}}\boldsymbol{\eta}$ for some $\boldsymbol{\eta} \in \Re^m$. To this end, we choose $\boldsymbol{\eta} \in \Re^m$ with $\eta_i = v_{ij}\beta_j$ for all $i \in \mathcal{J}_1 = \mathcal{I}_1$ and we determine $\boldsymbol{\eta}_{\mathcal{J}_0}$ later. Then $\boldsymbol{u}_i^\top\boldsymbol{\eta} = \eta_i = v_{ij}\beta_j$ for all $i \in \mathcal{J}_1 = \mathcal{I}_1$. We additionally observe that $\eta_i = 0$ for all $i \in \mathcal{N}$, following from $v_{ij}\beta_j = 0$. We now move on to $\mathcal{I}_0$, and have

$$\boldsymbol{U}_{\mathcal{I}_0}\boldsymbol{\eta} = [\boldsymbol{U}_{\mathcal{I}_0,\mathcal{J}_1} \ \boldsymbol{U}_{\mathcal{I}_0,\mathcal{J}_0}] \begin{bmatrix} \boldsymbol{\eta}_{\mathcal{J}_1} \\ \boldsymbol{\eta}_{\mathcal{J}_0} \end{bmatrix} = \boldsymbol{U}_{\mathcal{I}_0,\mathcal{J}_1}\boldsymbol{\eta}_{\mathcal{J}_1} + \boldsymbol{U}_{\mathcal{I}_0,\mathcal{J}_0}\boldsymbol{\eta}_{\mathcal{J}_0} = \sum_{i\in\mathcal{J}_1\setminus\mathcal{N}} \boldsymbol{s}_i\eta_i + \boldsymbol{U}_{\mathcal{I}_0,\mathcal{J}_0}\boldsymbol{\eta}_{\mathcal{J}_0},$$

where the last equality is due to that when $i \in \mathcal{N}$, $\eta_i = 0$ and when $j \in \mathcal{J}_1 \setminus \mathcal{N}$, $(\boldsymbol{U}_{\mathcal{I}_0})_j = \boldsymbol{s}_j + \gamma_j \boldsymbol{u}$ with $\gamma_j = 0$. Since $\boldsymbol{s}_i \in \operatorname{span}(\boldsymbol{U}_{\mathcal{I}_0,\mathcal{J}_0})$, we can choose $\boldsymbol{\eta}_{\mathcal{J}_0}$ such that $\sum_{i\in\mathcal{J}_1\setminus\mathcal{N}} \boldsymbol{s}_i\eta_i + \boldsymbol{U}_{\mathcal{I}_0,\mathcal{J}_0}\boldsymbol{\eta}_{\mathcal{J}_0} = \boldsymbol{0}$. In this case, $\boldsymbol{U}_{\mathcal{I}_0}\boldsymbol{\eta} = \boldsymbol{0} = (\boldsymbol{V}_{\mathcal{I}_0})_j\beta_j$. Hence, we conclude $(\boldsymbol{V}_{\mathcal{I}})_j\beta_j = \boldsymbol{U}_{\mathcal{I}}\boldsymbol{\eta} \in \operatorname{span}(\boldsymbol{U}_{\mathcal{I}})$. □

**Lemma 5 (Chen et al. 2021)** *Consider any matrix* $\boldsymbol{Q} \in \Re^{s\times(s+1)}$ *with* $\operatorname{rank}(\boldsymbol{Q}) = s, s \geq 2$. *If every* $2\times 3$ *submatrix of* $\boldsymbol{Q}$ *contains at least one pair of column vectors which are linearly dependent, then* $\boldsymbol{Q}$ *has at least one pair of column vectors which are linearly dependent.*

### C.9. Proof of Lemma 2

For notational simplicity, we remove the superscript $k$ throughout this proof. To see $\boldsymbol{\zeta}^1,\ldots,\boldsymbol{\zeta}^{2n+1}$ are vertices of a $2n$-simplex, it suffices to show these $2n+1$ points are affinely independent. That is, we need to prove that $\boldsymbol{\zeta}^2-\boldsymbol{\zeta}^1,\ldots,\boldsymbol{\zeta}^{2n+1}-\boldsymbol{\zeta}^1$ are linearly independent. First, we scale each elements in $\boldsymbol{\omega}^i,\boldsymbol{v}^i,\boldsymbol{\zeta}^i$ such that all nonzero elements become 1 and denote the corresponding vectors as $\hat{\boldsymbol{\omega}}^i,\hat{\boldsymbol{v}}^i,\hat{\boldsymbol{\zeta}}^i$. Notice that we still have $\hat{\boldsymbol{\zeta}}^i=\begin{bmatrix}\hat{\boldsymbol{\omega}}^i\\ \hat{\boldsymbol{v}}^i\end{bmatrix}$. In this case, $\hat{\boldsymbol{\omega}}^1=\mathbf{1},\hat{\boldsymbol{v}}^1=\mathbf{0}$ since $\boldsymbol{z}^1=\underline{\boldsymbol{z}}$, $\hat{\boldsymbol{\omega}}^{2n+1}=\mathbf{0},\hat{\boldsymbol{v}}^{2n+1}=\mathbf{1}$ since $\boldsymbol{z}^{2n+1}=\overline{\boldsymbol{z}}$. Moreover, we have

$$\{\hat{\boldsymbol{\omega}}^i-\hat{\boldsymbol{\omega}}^{i+1},\hat{\boldsymbol{v}}^{i+1}-\hat{\boldsymbol{v}}^i\}=\{\mathbf{0},\boldsymbol{e}_{\kappa_i}\}$$

for some $\kappa_i\in[n]$, $i\in[2n]$. This follows from that $\boldsymbol{z}^{i+1}-\boldsymbol{z}^i$ has exactly one nonzero entry, the index of which is denoted as $\kappa_i$. Specifically, for the $\kappa_i$-th entry where $\boldsymbol{z}^i$ moves to $\boldsymbol{z}^{i+1}$, 1) if the move is from the lower bound to the mean, then $\hat{\boldsymbol{\omega}}^{i+1}=\hat{\boldsymbol{\omega}}^i-\boldsymbol{e}_{\kappa_i},\hat{\boldsymbol{v}}^{i+1}=\hat{\boldsymbol{v}}^i$ and hence $\hat{\boldsymbol{\zeta}}^{i+1}-\hat{\boldsymbol{\zeta}}^i=\begin{bmatrix}-\boldsymbol{e}_{\kappa_i}\\ \mathbf{0}\end{bmatrix}$; 2) if the move is from the mean to the upper bound, then $\hat{\boldsymbol{\omega}}^{i+1}=\hat{\boldsymbol{\omega}}^i,\hat{\boldsymbol{v}}^{i+1}=\hat{\boldsymbol{v}}^i+\boldsymbol{e}_{\kappa_i}$ and hence $\hat{\boldsymbol{\zeta}}^{i+1}-\hat{\boldsymbol{\zeta}}^i=\begin{bmatrix}\mathbf{0}\\ \boldsymbol{e}_{\kappa_i}\end{bmatrix}$. We also notice that for each dimension, there is exactly one move from the lower bound to the mean, and one from the mean to the upper bound. Therefore, the matrix $\begin{bmatrix}\hat{\boldsymbol{\zeta}}^2-\hat{\boldsymbol{\zeta}}^1 & \cdots & \hat{\boldsymbol{\zeta}}^{2n+1}-\hat{\boldsymbol{\zeta}}^{2n}\end{bmatrix}=\begin{bmatrix}-\boldsymbol{I}_{n\times n} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n\times n}\end{bmatrix}\boldsymbol{P}$ for a $2n\times 2n$ permutation matrix $\boldsymbol{P}$. Notice that $\boldsymbol{\zeta}^i=\begin{bmatrix}((\mu_j-\underline{z}_j)\cdot\hat{\omega}^i_j)_{j\in[n]}\\ ((\overline{z}_j-\mu_j)\cdot\hat{v}^i_j)_{j\in[2n]}\end{bmatrix}=\begin{bmatrix}\operatorname{diag}(\boldsymbol{\mu}-\underline{\boldsymbol{z}}) & \mathbf{0}\\ \mathbf{0} & \operatorname{diag}(\overline{\boldsymbol{z}}-\boldsymbol{\mu})\end{bmatrix}\hat{\boldsymbol{\zeta}}^i$ for all $i\in[2n+1]$. We then have

$$\begin{aligned}
[\boldsymbol{\zeta}^2-\boldsymbol{\zeta}^1\ \cdots\ \boldsymbol{\zeta}^{2n+1}-\boldsymbol{\zeta}^{2n}] &= \begin{bmatrix}\operatorname{diag}(\boldsymbol{\mu}-\underline{\boldsymbol{z}}) & \mathbf{0}\\ \mathbf{0} & \operatorname{diag}(\overline{\boldsymbol{z}}-\boldsymbol{\mu})\end{bmatrix}\begin{bmatrix}\hat{\boldsymbol{\zeta}}^2-\hat{\boldsymbol{\zeta}}^1 & \cdots & \hat{\boldsymbol{\zeta}}^{2n+1}-\hat{\boldsymbol{\zeta}}^{2n}\end{bmatrix}\\
&= \begin{bmatrix}\operatorname{diag}(\boldsymbol{\mu}-\underline{\boldsymbol{z}}) & \mathbf{0}\\ \mathbf{0} & \operatorname{diag}(\overline{\boldsymbol{z}}-\boldsymbol{\mu})\end{bmatrix}\begin{bmatrix}-\boldsymbol{I}_{n\times n} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n\times n}\end{bmatrix}\boldsymbol{P}\\
&= \begin{bmatrix}\operatorname{diag}(\underline{\boldsymbol{z}}-\boldsymbol{\mu}) & \mathbf{0}\\ \mathbf{0} & \operatorname{diag}(\overline{\boldsymbol{z}}-\boldsymbol{\mu})\end{bmatrix}\boldsymbol{P}.
\end{aligned}$$

This implies that the matrix $[\boldsymbol{\zeta}^2-\boldsymbol{\zeta}^1\ \cdots\ \boldsymbol{\zeta}^{2n+1}-\boldsymbol{\zeta}^1]$ are also invertible. □

### C.10. Proof of Proposition 6

We first let $V_{adapt}$ and $V_{ldr}$ represent the optimal values for Problems (9) and (11), respectively. Our aim is to show that $V_{adapt}=V_{ldr}$.

We first prove $V_{adapt}\le V_{ldr}$. To show this, we define a new problem by relaxing Problem (9) such that the constraints of second-stage problem apply only to the realizations $\boldsymbol{z}^{k,i},k\in[K],i\in[2n+1]$ and denote the optimal value as $V_{relax}$, i.e.,

$$\begin{aligned}
V_{relax}=\min\ \ & \boldsymbol{a}^\top\boldsymbol{x}+\sup_{\mathbb{P}\in\mathcal{F}}\ \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{b}^\top\boldsymbol{y}(\tilde{k},\tilde{\boldsymbol{z}})\right]\\
\text{s.t.}\ \ & \boldsymbol{W}\boldsymbol{x}+\boldsymbol{U}\boldsymbol{y}(k,\boldsymbol{z}^{k,i})\ge\boldsymbol{V}\boldsymbol{z}^{k,i}+\boldsymbol{v}^0,\quad k\in[K],i\in[2n+1]\\
& \boldsymbol{x}\in\mathcal{X}.
\end{aligned}\tag{43}$$

By the minimax theorem (von Neumann 1928), we can interchange "$\sup_{\mathbb{P}}$" and "$\min_{\boldsymbol{y}(k,\boldsymbol{z})}$" equivalently. Omitting the dependency between $\boldsymbol{y}$ and the uncertainty realizations, we rewrite $V_{relax} = \min_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{a}^\top \boldsymbol{x} + \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[g'(\boldsymbol{x},\boldsymbol{z})]$, where

$$g'(\boldsymbol{x},\boldsymbol{z}) = \begin{cases} \min\left\{\boldsymbol{b}^\top \boldsymbol{y} \;\middle|\; \boldsymbol{W}\boldsymbol{x} + \boldsymbol{U}\boldsymbol{y} \ge \boldsymbol{V}\boldsymbol{z} + \boldsymbol{v}^0\right\} & \text{if } \boldsymbol{z} \in \bigcup_{k\in[K]}\{\boldsymbol{z}^{k,1},\ldots,\boldsymbol{z}^{k,2n+1}\}, \\ -\infty & \text{otherwise.} \end{cases}$$

Fixing any $\boldsymbol{x}\in\mathcal{X}$, we recall that $p_i^k, \boldsymbol{z}^{k,i}, k\in[K], i\in[2n+1]$ returned by Algorithm 1 gives a worst-case distribution to Problem (9), and, at the same time, is an admissible probability distribution to Problem (11) because the two problems share the same ambiguity set. It follows that

$$\begin{aligned} V_{adapt} &= \min_{\boldsymbol{x}\in\mathcal{X}} \; \boldsymbol{a}^\top\boldsymbol{x} + \sup_{\mathbb{P}\in\mathcal{F}} \; \mathbb{E}_{\mathbb{P}}[g(\boldsymbol{x},\boldsymbol{z})] \\ &= \min_{\boldsymbol{x}\in\mathcal{X}} \; \boldsymbol{a}^\top\boldsymbol{x} + \max_{\boldsymbol{q}\in\mathcal{Q}} \sum_{k\in[K],i\in[2n+1]} q_k p_i^k g(\boldsymbol{x},\boldsymbol{z}^{k,i}) \\ &= \min_{\boldsymbol{x}\in\mathcal{X}} \; \boldsymbol{a}^\top\boldsymbol{x} + \max_{\boldsymbol{q}\in\mathcal{Q}} \sum_{k\in[K],i\in[2n+1]} q_k p_i^k g'(\boldsymbol{x},\boldsymbol{z}^{k,i}) \\ &\le \min_{\boldsymbol{x}\in\mathcal{X}} \; \boldsymbol{a}^\top\boldsymbol{x} + \sup_{\mathbb{P}\in\mathcal{F}} \; \mathbb{E}_{\mathbb{P}}[g'(\boldsymbol{x},\boldsymbol{z})] = V_{relax}, \end{aligned}$$

where the second equality follows from Proposition 3, the third equality holds because $g$ and $g'$ have the same value whenever $\boldsymbol{z}\in\bigcup_{k\in[K]}\{\boldsymbol{z}^{k,1},\ldots,\boldsymbol{z}^{k,2n+1}\}$, and the inequality follows from the feasibility of the distribution characterized by $p_i^k, \boldsymbol{z}^{k,i}, i\in[2n+1]$.

Further, we observe that Problem (11) can be directly obtained from Problem (43) by imposing a restriction of linearity structure on $\boldsymbol{y}(k,\boldsymbol{z})$. This implies any feasible $\boldsymbol{\Theta}^k, \boldsymbol{\phi}^k$ to Problem (11) determines a function $\boldsymbol{y}(k,\boldsymbol{z})$ that is feasible to Problem (43). Hence, $V_{relax} \le V_{ldr}$. We then conclude that $V_{adapt} \le V_{relax} \le V_{ldr}$.

We next show $V_{adapt} \ge V_{ldr}$. To this end, we construct a recourse decision rule that is feasible to Problem (11) and returns the optimal value of Problem (9).

We first consider the case of fixed scenario; for brevity, we remove the notation $k$ (or $\tilde{k}$) that denotes realized (or random) scenarios. The construction is similar to the proof of Bertsimas and Goyal (2012, Theorem 1). Define auxiliary uncertain factors $\tilde{\boldsymbol{\omega}} = (\boldsymbol{\mu}-\tilde{\boldsymbol{z}})^+, \tilde{\boldsymbol{v}} = (\tilde{\boldsymbol{z}}-\boldsymbol{\mu})^+, \tilde{\boldsymbol{\zeta}} = (\tilde{\boldsymbol{\omega}},\tilde{\boldsymbol{v}})$, and let $\boldsymbol{\omega},\boldsymbol{v},\boldsymbol{\zeta}$ be the counterpart when $\tilde{\boldsymbol{z}}$ is realized as $\boldsymbol{z}$. Then $\tilde{\boldsymbol{z}} = \boldsymbol{\mu} - \tilde{\boldsymbol{\omega}} + \tilde{\boldsymbol{v}} = \boldsymbol{\mu} + [-\boldsymbol{I}_{n\times n}\ \boldsymbol{I}_{n\times n}]\,\tilde{\boldsymbol{\zeta}}$, $|\tilde{\boldsymbol{z}}-\boldsymbol{\mu}| = \tilde{\boldsymbol{\omega}} + \tilde{\boldsymbol{v}} = [\boldsymbol{I}_{n\times n}\ \boldsymbol{I}_{n\times n}]\,\tilde{\boldsymbol{\zeta}}$. Define

$$\boldsymbol{y}_{opt}(\boldsymbol{z}) = \boldsymbol{\Theta}_{opt}\begin{bmatrix}(\boldsymbol{\mu}-\boldsymbol{z})^+ \\ (\boldsymbol{z}-\boldsymbol{\mu})^+\end{bmatrix} + \boldsymbol{\phi}_{opt} = \boldsymbol{\Theta}_{opt}\boldsymbol{\zeta} + \boldsymbol{\phi}_{opt}.$$

For all $i\in[2n+1]$,

$$\begin{aligned} \boldsymbol{y}_{opt}(\boldsymbol{z}^i) &= \boldsymbol{y}_{opt}^{2n+1} + \boldsymbol{\Theta}_{opt}(\boldsymbol{\zeta}^i - \boldsymbol{\zeta}^{2n+1}) \\ &= \boldsymbol{y}_{opt}^{2n+1} + \begin{bmatrix}\boldsymbol{y}_{opt}^1 - \boldsymbol{y}_{opt}^{2n+1} & \cdots & \boldsymbol{y}_{opt}^{2n} - \boldsymbol{y}_{opt}^{2n+1}\end{bmatrix}\boldsymbol{D}^{-1}(\boldsymbol{\zeta}^i - \boldsymbol{\zeta}^{2n+1}) \\ &= \boldsymbol{y}_{opt}^{2n+1} + \begin{bmatrix}\boldsymbol{y}_{opt}^1 - \boldsymbol{y}_{opt}^{2n+1} & \cdots & \boldsymbol{y}_{opt}^{2n} - \boldsymbol{y}_{opt}^{2n+1}\end{bmatrix}\boldsymbol{e}_i \\ &= \boldsymbol{y}_{opt}^i, \end{aligned} \tag{44}$$

where the third last equality holds because $\boldsymbol{\zeta}^i - \boldsymbol{\zeta}^{2n+1} = \boldsymbol{D}\boldsymbol{e}_i$ for all $i \in [2n]$. We notice that $\boldsymbol{b}^\top \boldsymbol{y}_{opt}(\boldsymbol{z})$, as a linear combination of $(\boldsymbol{\mu} - \boldsymbol{z})^+$ and $(\boldsymbol{z} - \boldsymbol{\mu})^+$, is supermodular in $\boldsymbol{z}$ because it is separable. Now, utilizing the worst-case distribution given by Algorithm 1, we get

$$\begin{aligned}
\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{b}^\top \boldsymbol{y}_{opt}(\tilde{\boldsymbol{z}})\right] &= \sum_{i\in[2n+1]} p_i \boldsymbol{b}^\top \boldsymbol{y}_{opt}(\boldsymbol{z}^i) \\
&= \sum_{i\in[2n+1]} p_i \boldsymbol{b}^\top \boldsymbol{y}^i_{opt} \\
&= \sum_{i\in[2n+1]} p_i \min\left\{\boldsymbol{b}^\top \boldsymbol{y} \;\middle|\; \boldsymbol{W}\boldsymbol{x}_{opt} + \boldsymbol{U}\boldsymbol{y} \geq \boldsymbol{V}\boldsymbol{z}^i + \boldsymbol{v}^0\right\} \\
&= \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[g(\boldsymbol{x}_{opt}, \tilde{\boldsymbol{z}})\right].
\end{aligned}$$

The first and last equalities follow from the supermodularity of $\boldsymbol{b}^\top \boldsymbol{y}_{opt}(\boldsymbol{z})$ and $g(\boldsymbol{x}_{opt}, \boldsymbol{z})$ defined by (1), respectively. The second equality holds since $\boldsymbol{y}_{opt}(\boldsymbol{z}^i) = \boldsymbol{y}^i_{opt}, i \in [2n+1]$ (as shown in (44)), while the third one follows from the definition of $\boldsymbol{y}^i_{opt}$. It follows that the worst-case expected cost returned by $\boldsymbol{x}_{opt}, \boldsymbol{y}_{opt}(\boldsymbol{z})$ is the same as the optimal value of Problem (9). Further, we can observe easily that the solution $\boldsymbol{x}_{opt}, \boldsymbol{y}_{opt}(\boldsymbol{z})$ is feasible for Problem (11).

We next consider the case of uncertain scenarios. Following the above proof, we define $\boldsymbol{y}_{opt}(k, \boldsymbol{z})$ as

$$\boldsymbol{y}_{opt}(k, \boldsymbol{z}) = \boldsymbol{\Theta}^k_{opt} \begin{bmatrix} (\boldsymbol{\mu}^k - \boldsymbol{z})^+ \\ (\boldsymbol{z} - \boldsymbol{\mu}^k)^+ \end{bmatrix} + \boldsymbol{\phi}^k_{opt}. \tag{45}$$

It is supermodular in $\boldsymbol{z}$ and for any realized scenario $k$, $\sup_{\mathbb{P}^k\in\mathcal{F}^k} \mathbb{E}_{\mathbb{P}^k}\left[\boldsymbol{b}^\top \boldsymbol{y}_{opt}(k, \tilde{\boldsymbol{z}})\right] = \sup_{\mathbb{P}^k\in\mathcal{F}^k} \mathbb{E}_{\mathbb{P}^k}\left[\min\left\{\boldsymbol{b}^\top \boldsymbol{y} \;\middle|\; \boldsymbol{W}\boldsymbol{x}_{opt} + \boldsymbol{U}\boldsymbol{y} \geq \boldsymbol{V}\tilde{\boldsymbol{z}} + \boldsymbol{v}^0\right\}\right]$. Hence

$$\begin{aligned}
& \boldsymbol{a}^\top \boldsymbol{x}_{opt} + \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{b}^\top \boldsymbol{y}_{opt}(\tilde{k}, \tilde{\boldsymbol{z}})\right] \\
= \;& \boldsymbol{a}^\top \boldsymbol{x}_{opt} + \max_{\boldsymbol{q}\in\mathcal{Q}} \sum_{k\in[K]} q_k \sup_{\mathbb{P}^k\in\mathcal{F}^k} \mathbb{E}_{\mathbb{P}^k}\left[\boldsymbol{b}^\top \boldsymbol{y}_{opt}(k, \tilde{\boldsymbol{z}})\right] \\
= \;& \boldsymbol{a}^\top \boldsymbol{x}_{opt} + \max_{\boldsymbol{q}\in\mathcal{Q}} \sum_{k\in[K]} q_k \sup_{\mathbb{P}^k\in\mathcal{F}^k} \mathbb{E}_{\mathbb{P}^k}\left[\min\left\{\boldsymbol{b}^\top \boldsymbol{y} \;\middle|\; \boldsymbol{W}\boldsymbol{x}_{opt} + \boldsymbol{U}\boldsymbol{y} \geq \boldsymbol{V}\tilde{\boldsymbol{z}} + \boldsymbol{v}^0\right\}\right] \\
= \;& \boldsymbol{a}^\top \boldsymbol{x}_{opt} + \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[g(\boldsymbol{x}_{opt}, \tilde{\boldsymbol{z}})\right] \\
= \;& V_{adapt}.
\end{aligned}$$

Similar to Equation (44), we can check that $\boldsymbol{y}_{opt}(k, \boldsymbol{z}^{k,i}) \in \min\left\{\boldsymbol{b}^\top \boldsymbol{y} \;\middle|\; \boldsymbol{W}\boldsymbol{x}_{opt} + \boldsymbol{U}\boldsymbol{y} \geq \boldsymbol{V}\boldsymbol{z} + \boldsymbol{v}^0\right\}$ for all $k \in [K], i \in [2n+1]$. It follows that $\boldsymbol{x}_{opt}, \boldsymbol{\Theta}^k_{opt}, \boldsymbol{\phi}^k_{opt}, k \in [K]$ is a feasible solution to Problem (11).

Therefore, we can conclude that $V_{ldr} \leq V_{adapt}$. Hence, we have $V_{adapt} = V_{ldr}$ and $\boldsymbol{x}_{opt}, \boldsymbol{\Theta}^k_{opt}, \boldsymbol{\phi}^k_{opt}, k \in [K]$ is an optimal solution to Problem (11). □

### C.11. Proof of Theorem 3

Following the proof for Proposition 6, we denote by $\boldsymbol{y}_{opt}(k, \boldsymbol{z})$ the linear decision rule defined by (45). To complete the proof for this theorem, based on Proposition 6, it suffices to show $\boldsymbol{y}_{opt}(k, \boldsymbol{z})$

is feasible, i.e., $(\boldsymbol{y}_{opt}(k,\boldsymbol{z}),\boldsymbol{z}) \in \mathcal{S}(\boldsymbol{x})$, for all $\boldsymbol{z} \in \bigcup_{k\in[K]} \mathcal{Z}^k$, where $\mathcal{Z}^k = \Pi_{i\in[n]}\{\underline{z}_i^k, \mu_i^k, \overline{z}_i^k\}$. We first prove the following claim and then show the feasibility by induction.

**Claim.** *Fix any scenario $k$. For all $\boldsymbol{z}', \boldsymbol{z}'' \in \mathcal{Z}^k$ with $\boldsymbol{z}^\wedge = \boldsymbol{z}' \wedge \boldsymbol{z}''$, $\boldsymbol{z}^\vee = \boldsymbol{z}' \vee \boldsymbol{z}''$, if $(\boldsymbol{y}_{opt}(k,\boldsymbol{z}^\wedge),\boldsymbol{z}^\wedge)$, $(\boldsymbol{y}_{opt}(k,\boldsymbol{z}^\vee),\boldsymbol{z}^\vee)$, $(\boldsymbol{y}_{opt}(k,\boldsymbol{z}'),\boldsymbol{z}') \in \mathcal{S}(\boldsymbol{x})$, then $(\boldsymbol{y}_{opt}(k,\boldsymbol{z}''),\boldsymbol{z}'')$ is also in $\mathcal{S}(\boldsymbol{x})$.*

*Proof of Claim.* Since the function $\boldsymbol{y}_{opt}(k,\boldsymbol{z})$ is separable in $\boldsymbol{z}$, hence, it is both supermodular and submodular in $\boldsymbol{z}$. Therefore, $\boldsymbol{y}_{opt}(k,\boldsymbol{z}') + \boldsymbol{y}_{opt}(k,\boldsymbol{z}'') = \boldsymbol{y}_{opt}(k,\boldsymbol{z}^\wedge) + \boldsymbol{y}_{opt}(k,\boldsymbol{z}^\vee)$, or equivalently, $\boldsymbol{y}_{opt}(k,\boldsymbol{z}'') = \boldsymbol{y}_{opt}(k,\boldsymbol{z}^\wedge) + \boldsymbol{y}_{opt}(k,\boldsymbol{z}^\vee) - \boldsymbol{y}_{opt}(k,\boldsymbol{z}')$. With the condition in the theorem satisfied, the claim follows directly.

For all $k \in [K], i \in [2n+1]$, recall that by the proof for Proposition 6, $\boldsymbol{y}_{opt}(k,\boldsymbol{z}^{k,i})$ is the optimal second-stage decision when the uncertainty is realized as $\boldsymbol{z}^{k,i}$. This implies $\boldsymbol{y}_{opt}(k,\boldsymbol{z})$ is feasible for all $\boldsymbol{z} \in \{\boldsymbol{z}^{k,i} \mid k \in [K], i \in [2n+1]\}$. Fix any scenario $k$, define $\mathcal{Z}^{k,i} = \left\{\boldsymbol{z} \in \mathcal{Z}^k \;\middle|\; \boldsymbol{z} \le \boldsymbol{z}^{k,i}\right\}$ for all $i \in [2n+1]$. Observe that $\mathcal{Z}^{k,1} = \{\boldsymbol{z}^{k,1}\}$, we know $\boldsymbol{y}_{opt}(k,\boldsymbol{z})$ is feasible on $\mathcal{Z}^{k,i}$ when $i=1$. Next we inductively show the feasibility on $\mathcal{Z}^{k,i}$ for all $i \in [2n+1]$. Specifically, we assume the statement holds for a given $i \in [2n+1]$, and consider for the set $\mathcal{Z}^{k,i+1}$. By definition, $\boldsymbol{z}^{k,i+1}$ deviates from $\boldsymbol{z}^{k,i}$ in only one dimension, i.e., $z_l^{k,i+1} > z_l^{k,i}$ for some $l \in [n]$ and $z_{l'}^{k,i+1} = z_{l'}^{k,i}$ for all $l' \ne l$. By assumption, it suffices to prove the feasibility for any $\hat{\boldsymbol{z}} \in \mathcal{Z}^{k,i+1} \setminus \mathcal{Z}^{k,i}$. In this case, we have $\hat{z}_l = z_l^{k,i+1}$ and $\hat{z}_{l'} \le z_{l'}^{k,i}$ for all $l' \ne l$, i.e., $\hat{\boldsymbol{z}} = (\hat{z}_1,\ldots,\hat{z}_n) = (\hat{z}_1,\ldots,\hat{z}_{l-1}, z_l^{k,i+1}, \hat{z}_{l+1},\ldots,\hat{z}_n)$. Choosing $\boldsymbol{z}' = \boldsymbol{z}^{k,i} = (z_1^{k,i},\ldots,z_{l-1}^{k,i}, z_l^{k,i}, z_{l+1}^{k,i},\ldots,z_n^{k,i})$ and $\boldsymbol{z}'' = \hat{\boldsymbol{z}}$, we obtain

$$\begin{aligned}
\boldsymbol{z}^\wedge &= \boldsymbol{z}' \wedge \boldsymbol{z}'' = (\hat{z}_1,\ldots,\hat{z}_{l-1}, z_l^{k,i}, \hat{z}_{l+1},\ldots,\hat{z}_n) \in \mathcal{Z}^{k,i},\\
\boldsymbol{z}^\vee &= \boldsymbol{z}' \vee \boldsymbol{z}'' = (z_1^{k,i},\ldots,z_{l-1}^{k,i}, z_l^{k,i+1}, z_{l+1}^{k,i},\ldots,z_n^{k,i}) = \boldsymbol{z}^{k,i+1}.
\end{aligned}$$

Since $\boldsymbol{y}_{opt}(k,\boldsymbol{z})$ is feasible when $\boldsymbol{z} = \boldsymbol{z}', \boldsymbol{z}^\wedge, \boldsymbol{z}^\vee$, by the Claim we conclude that $\boldsymbol{y}_{opt}(k,\boldsymbol{z}'') = \boldsymbol{y}_{opt}(k,\hat{\boldsymbol{z}})$ is also feasible. Notice that $\mathcal{Z}^{k,i} = \mathcal{Z}^k$ when $i = 2n+1$, and the same proof goes for any $k \in [K]$, we complete the proof. □

### C.12. Proof of Proposition 7

It is obvious that $-\boldsymbol{r}^\top\boldsymbol{x} + (\boldsymbol{r}-\boldsymbol{s})(\boldsymbol{x}-\boldsymbol{z})^+$ is decreasing and supermodular in $\boldsymbol{z}$. Thus, we can apply Proposition 12. Specifically, by substituting $\boldsymbol{a} = -\boldsymbol{r}$ and $g(\boldsymbol{x},\boldsymbol{z}) = (\boldsymbol{r}-\boldsymbol{s})(\boldsymbol{x}-\boldsymbol{z})^+ = \min\left\{(\boldsymbol{r}-\boldsymbol{s})^\top\boldsymbol{y} \;\middle|\; \boldsymbol{y} \ge \boldsymbol{x}-\boldsymbol{z}, \boldsymbol{y} \ge \boldsymbol{0}\right\}$ in the formulation (29), we obtain the following reformulation for Problem (13),

$$
\begin{aligned}
\min \quad & \boldsymbol{\nu}^\top \boldsymbol{l} & \\
\text{s.t.} \quad & \boldsymbol{R}_k^\top \boldsymbol{l} \geq \sum_{i\in[2n+1]} p_i^k f^{k,i}, & k\in[K] \\
& \boldsymbol{y}^{k,i} \geq \boldsymbol{x} - \boldsymbol{z}^{k,i}, & k\in[K], i\in[2n+1] \\
& \boldsymbol{y}^{k,i} \geq \boldsymbol{0}, & k\in[K], i\in[2n+1] \\
& f^{k,i} \geq c_j\left(-\boldsymbol{r}^\top\boldsymbol{x} + (\boldsymbol{r}-\boldsymbol{s})^\top \boldsymbol{y}^{k,i}\right) + d_j, & k\in[K], i\in[2n+1], j\in[J], \\
& \boldsymbol{l}\geq \boldsymbol{0},\ \boldsymbol{x}\in\mathcal{X}^{news}, &
\end{aligned}
$$

where $p_i^k, \boldsymbol{z}^{k,i}, k\in[K], i\in[2n+1]$ are the output of Algorithm 1 given the ambiguity sets $\mathcal{F}^k, k\in[K]$ defined by Equation (4). □

### C.13. Proof of Proposition 8

Notice that (14) is a special case of the formulation (30) with $u(w)=\frac{1}{\rho}w^+$, $\boldsymbol{a}=-\boldsymbol{r}$ and $g(\boldsymbol{x},\boldsymbol{z}) = \min\left\{(\boldsymbol{r}-\boldsymbol{s})^\top\boldsymbol{y} \;\middle|\; \boldsymbol{y}\geq\boldsymbol{x}-\boldsymbol{z}, \boldsymbol{y}\geq\boldsymbol{0}\right\}$. Hence, a direct application of Corollary 3 gives the following reformulation,

$$
\begin{aligned}
\min \quad & \theta + \boldsymbol{\nu}^\top \boldsymbol{l} & \\
\text{s.t.} \quad & \boldsymbol{R}_k^\top \boldsymbol{l} \geq \sum_{i\in[2n+1]} \frac{1}{\rho}\cdot p_i^k f^{k,i}, & k\in[K] \\
& \boldsymbol{y}^{k,i} \geq \boldsymbol{x} - \boldsymbol{z}^{k,i}, & k\in[K], i\in[2n+1] \\
& \boldsymbol{y}^{k,i} \geq \boldsymbol{0}, & k\in[K], i\in[2n+1] \\
& f^{k,i} \geq -\boldsymbol{r}^\top\boldsymbol{x} + (\boldsymbol{r}-\boldsymbol{s})^\top \boldsymbol{y}^{k,i} - \theta, & k\in[K], i\in[2n+1] \\
& f^{k,i} \geq 0, & k\in[K], i\in[2n+1] \\
& \boldsymbol{l}\geq \boldsymbol{0},\ \boldsymbol{x}\in\mathcal{X}^{news}, &
\end{aligned}
$$

where $p_i^k, \boldsymbol{z}^{k,i}, k\in[K], i\in[2n+1]$ are the output of Algorithm 1 given the ambiguity sets $\mathcal{F}^k, k\in[K]$ defined by Equation (4). □

### C.14. Proof of Proposition 9

We prove the supermodularity by showing that Problem (15) satisfies the conditions in Theorem 2. We first reformulate Problem (15) as the sum of $m$ sub-problems. Denote

$$
g^j(\boldsymbol{x},\boldsymbol{z}) = \min\left\{\sum_{i\in[n]} c_{ij}y_{ij} \;\middle|\; \begin{array}{l} \sum_{i\in[n]} y_{ij} = 1 \\ 0\leq y_{ij}\leq x_i z_i,\ i\in[n] \end{array}\right\}.
$$

Then it can be verified that $g(\boldsymbol{x},\boldsymbol{z}) = \sum_{j\in[m]} g^j(\boldsymbol{x},\boldsymbol{z})$. Hence, it suffices to prove the supermodularity of $g^j(\boldsymbol{x},\boldsymbol{z})$ for all $j\in[m]$. Observing that $\boldsymbol{x}\in\{0,1\}^n$, we denote $S=\{i\in[n] \mid x_i=1\}$ and $T=$

$[n]\backslash S$. It follows that the constraints in defining $g^j(\boldsymbol{x},\boldsymbol{z})$ can be reformulated as $\boldsymbol{U}(y_{1j},\ldots,y_{nj}) - \boldsymbol{V}\boldsymbol{z} \geq \boldsymbol{v}^0$, where

$$\boldsymbol{U} = \begin{bmatrix} \boldsymbol{1}^\top \\ -\boldsymbol{1}^\top \\ \boldsymbol{I}_{n\times n} \\ -\boldsymbol{I}_T \\ -\boldsymbol{I}_S \end{bmatrix} \in \Re^{(2+2n)\times n},\ \boldsymbol{V} = \begin{bmatrix} \boldsymbol{0}_{(2+n+|T|)\times n} \\ -\boldsymbol{I}_S \end{bmatrix} \in \Re^{(2+2n)\times n},\ \boldsymbol{v}^0 = \begin{bmatrix} 1 \\ -1 \\ \boldsymbol{0}_{2n\times 1} \end{bmatrix} \in \Re^{2+2n}.$$

Here $\boldsymbol{I}_T, \boldsymbol{I}_S$ are the submatrices of $\boldsymbol{I}_{n\times n}$ consisting of rows which are indexed by elements in $T, S$, separately. Note that $\operatorname{rank}(\boldsymbol{U}) = n < 2n+2$. We hence can apply Theorem 2 to prove the supermodularity of $g^j$. To this end, we consider any index set $\mathcal{I}$ such that $|\mathcal{I}| = n+1$ and $\operatorname{rank}(\boldsymbol{U}_\mathcal{I}) = n$, any $\boldsymbol{\beta} \geq \boldsymbol{0} \in \Re^n, \boldsymbol{\alpha} \in \Re^n$ such that

$$\boldsymbol{V}_\mathcal{I}\boldsymbol{\beta} = \boldsymbol{U}_\mathcal{I}\boldsymbol{\alpha}. \tag{46}$$

Consider any $j \in [n]$, we need to show $\beta_j(\boldsymbol{V}_\mathcal{I})_j \in \operatorname{span}(\boldsymbol{U}_\mathcal{I})$. If $\beta_j(\boldsymbol{V}_\mathcal{I})_j = \boldsymbol{0}$, the result is straightforward. We now consider only the case that $\beta_j > 0$ and $(\boldsymbol{V}_\mathcal{I})_j \neq \boldsymbol{0}$.

By $(\boldsymbol{V}_\mathcal{I})_j \neq \boldsymbol{0}$, we have $\boldsymbol{V}_j \neq \boldsymbol{0}$. Based on the structure of $\boldsymbol{V}$, $\boldsymbol{V}_j$ has only one nonzero element and indeed, there exists a unique $i$ such that $\boldsymbol{V}_j = -\boldsymbol{e}_i \in \Re^{2n+2}$. Moreover, $i \in \mathcal{I}$, $i > 2+n+|T|$, $V_{ij} = -1$ is the only nonzero element in the $i$th row, i.e., $\boldsymbol{v}_i = -\boldsymbol{e}_j \in \Re^n$. Therefore, by (46), we have $\boldsymbol{u}_i^\top\boldsymbol{\alpha} = \boldsymbol{v}_i^\top\boldsymbol{\beta} = -\beta_j$, implying $\alpha_j = \beta_j$ since $U_{ij} = -1$ is also the only nonzero element in $\boldsymbol{u}_i$. We now show that for $\boldsymbol{U}_j$, only zero element from blocks $\boldsymbol{I}_{n\times n}$ and $-\boldsymbol{I}_T$ are included in $(\boldsymbol{U}_\mathcal{I})_j$. Assume to the contrary, i.e., there is a $k \in \{3,\ldots,2+n+|T|\} \cap \mathcal{I}$ with $U_{kj} \neq 0$. Note that $U_{kj}$ is the only nonzero element in $\boldsymbol{u}_k$. Hence, $\boldsymbol{u}_k^\top\boldsymbol{\alpha} = U_{kj}\alpha_j \neq 0$, $\boldsymbol{v}_k^\top\boldsymbol{\beta} = \boldsymbol{0}^\top\boldsymbol{\beta} = 0$, contradicts with (46) and $k \in \mathcal{I}$. Therefore, $U_{kj} = 0$ for all $k \in \{3,\ldots,2+n+|T|\} \cap \mathcal{I}$. Now we consider two scenarios.

In the first scenario, $\{1,2\} \cap \mathcal{I} = \emptyset$, then $(\boldsymbol{U}_\mathcal{I})_j$ has the only one nonzero element which is from $-\boldsymbol{I}_S$, $(\boldsymbol{U}_\mathcal{I})_j = (\boldsymbol{V}_\mathcal{I})_j$, and hence $\beta_j(\boldsymbol{V}_\mathcal{I})_j \in \operatorname{span}(\boldsymbol{U}_\mathcal{I})$.

In the second scenario, $\{1,2\} \cap \mathcal{I} \neq \emptyset$. WLOG, let $1 \in \mathcal{I}$. We then have $\sum_{k\in[n]} \alpha_k = \boldsymbol{1}^\top\boldsymbol{\alpha} = \boldsymbol{u}_1^\top\boldsymbol{\alpha} = \boldsymbol{v}_1^\top\boldsymbol{\beta} = \boldsymbol{0}^\top\boldsymbol{\beta} = 0$, where the third equality is due to (46). From the above analysis, we already have $\alpha_j = \beta_j > 0$, which implies that there exists $k \neq j$ such that $\alpha_k < 0$. We now prove that for $\boldsymbol{U}_k$, only zero elements from blocks $\boldsymbol{I}_{n\times n}$ and $-\boldsymbol{I}_T$ are included in $(\boldsymbol{U}_\mathcal{I})_k$. This can be done with the same logic as that in above when we show only zero elements from blocks $\boldsymbol{I}_{n\times n}$ and $-\boldsymbol{I}_T$ are included in $(\boldsymbol{U}_\mathcal{I})_j$. We next show that for $\boldsymbol{U}_k$, only zero elements from blocks $-\boldsymbol{I}_S$ is included in $(\boldsymbol{U}_\mathcal{I})_k$. Assume to the contrary, i.e., there is an $l \in \{2+n+|T|+1,\ldots,2+2n\} \cap \mathcal{I}$ such that $u_{lk} \neq 0$. Notice that $u_{lk} = -1$ and $v_{lk} = -1$ are the only nonzero elements in $\boldsymbol{u}_l$ and $\boldsymbol{v}_l$, respectively. We have $\boldsymbol{u}_l^\top\boldsymbol{\alpha} = -\alpha_k$, $\boldsymbol{v}_l^\top\boldsymbol{\alpha} = -\beta_k$, and hence $\boldsymbol{u}_l^\top\boldsymbol{\alpha} \neq \boldsymbol{v}_l^\top\boldsymbol{\alpha}$ since $\alpha_k < 0$ and $\boldsymbol{\beta} \geq \boldsymbol{0}$. It contradicts with (46), and we have that only zero elements from blocks $-\boldsymbol{I}_S$ are included in $(\boldsymbol{U}_\mathcal{I})_k$. Therefore, from the two observations above we can conclude that $(\boldsymbol{U}_\mathcal{I})_k$ has all elements as zero from the blocks $\boldsymbol{I}_{n\times n}$, $-\boldsymbol{I}_T$ and $-\boldsymbol{I}_S$. In other words, $(\boldsymbol{U}_\mathcal{I})_k$ and $(\boldsymbol{U}_\mathcal{I})_j$ only differs at $u_{ik} = 0$, $u_{ij} = -1$. We can then easily have $\beta_j(\boldsymbol{V}_\mathcal{I})_j = \beta_j(\boldsymbol{U}_\mathcal{I})_j - \beta_j(\boldsymbol{U}_\mathcal{I})_k \in \operatorname{span}(\boldsymbol{U}_\mathcal{I})$. □

### C.15. Proof of Proposition 10

Denote $\boldsymbol{y} = (y_{11}, \dots, y_{n1}, \ \dots, \ y_{1n}, \dots, y_{nn}) \in \Re^{n^2}$. Then the second-stage problem can be expressed as

$$g(\boldsymbol{x}, \boldsymbol{z}) = \min \sum_{s,j \in [n]} b_{sj} y_{sj}$$
$$\text{s.t.} \begin{bmatrix} \boldsymbol{U}^1 & \cdots & \boldsymbol{U}^n \\ & \boldsymbol{I}_{n^2 \times n^2} & \end{bmatrix} \boldsymbol{y} - \begin{bmatrix} \boldsymbol{I}_{n \times n} \\ \boldsymbol{0}_{n^2 \times n} \end{bmatrix} \boldsymbol{z} \geq \begin{bmatrix} -\boldsymbol{x} \\ \boldsymbol{0}_{n^2 \times 1} \end{bmatrix}.$$

For any $s \in [n]$, the matrix $\boldsymbol{U}^s \in \Re^{n \times n}$ has $\boldsymbol{e}_s - \boldsymbol{e}_j$ as its $j$-th column, $j \in [n]$. Denote $\boldsymbol{U}^0 = [\boldsymbol{U}^1 \ \cdots \ \boldsymbol{U}^n], \boldsymbol{V}^0 = \boldsymbol{I}_{n \times n}$ and $\boldsymbol{U} = \begin{bmatrix} \boldsymbol{U}^0 \\ \boldsymbol{I}_{n^2 \times n^2} \end{bmatrix} \in \Re^{(n+n^2) \times n^2}, \boldsymbol{V} = \begin{bmatrix} \boldsymbol{V}^0 \\ \boldsymbol{0}_{n^2 \times n} \end{bmatrix}$. Obviously, $\operatorname{rank}(\boldsymbol{U}) = n^2$ which is less than the number of rows in $\boldsymbol{U}$. Therefore, to complete the proof, we now show that $\boldsymbol{U}, \boldsymbol{V}$ meet the second condition in Theorem 2.

Consider any index set $\mathcal{I}$ such that $|\mathcal{I}| = \operatorname{rank}(\boldsymbol{U}) + 1 = n^2 + 1, \operatorname{rank}(\boldsymbol{U}_{\mathcal{I}}) = n^2$. Denote $\mathcal{I}_0 = \mathcal{I} \cap [n], \mathcal{I}_1 = \mathcal{I} \backslash \mathcal{I}_0$, then the rows of $\boldsymbol{U}_{\mathcal{I}_0}$ (or $\boldsymbol{V}_{\mathcal{I}_0}$) are extracted from $\boldsymbol{U}^0$ (or $\boldsymbol{V}^0$); the rows of $\boldsymbol{U}_{\mathcal{I}_1}$ (or $\boldsymbol{V}_{\mathcal{I}_1}$) are extracted from $\boldsymbol{I}_{n^2 \times n^2}$ (or $\boldsymbol{0}_{n^2 \times n}$).

We first let the column index set $\mathcal{J}_0$ be such that the submatrix $\boldsymbol{U}_{\mathcal{I}_0, \mathcal{J}_0} = \boldsymbol{0}$, and let $\mathcal{J}_1 = [n^2] \setminus \mathcal{J}_0$. Hence, $\boldsymbol{U}_{\mathcal{I}}$ can be decomposed into four submatrices $\boldsymbol{U}_{\mathcal{I}_0, \mathcal{J}_0}, \boldsymbol{U}_{\mathcal{I}_0, \mathcal{J}_1}, \boldsymbol{U}_{\mathcal{I}_1, \mathcal{J}_0}, \boldsymbol{U}_{\mathcal{I}_1, \mathcal{J}_1}$. Recalling that $\boldsymbol{U}_{\mathcal{I}_1}$ is a submatrix of $\boldsymbol{I}$, there is exactly one entry being one in its each row, and at most one entry being 1 in its each column. Hence, in $\boldsymbol{U}_{\mathcal{I}_1}$, the number of columns being $\boldsymbol{0}$ is $n^2 - |\mathcal{I}_1| = n^2 - (|\mathcal{I}| - |\mathcal{I}_0|) = |\mathcal{I}_0| - 1$. Noticing $\boldsymbol{U}_{\mathcal{I}}$ is full column rank and $\boldsymbol{U}_{\mathcal{I}_0, \mathcal{J}_0} = \boldsymbol{0}$, all of the $|\mathcal{I}_0| - 1$ zero columns in $\boldsymbol{U}_{\mathcal{I}_1}$ must be in $\boldsymbol{U}_{\mathcal{I}_1, \mathcal{J}_1}$. Denote the index set $\mathcal{K}_1$ as the set of column index for those zero columns in $\boldsymbol{U}_{\mathcal{I}_1}$, and $\mathcal{K}_2 = \mathcal{J}_1 \setminus \mathcal{K}_1$. Then $\boldsymbol{U}_{\mathcal{I}_1, \mathcal{J}_1}$ can be further decomposed into two submatrices $\boldsymbol{U}_{\mathcal{I}_1, \mathcal{K}_1}, \boldsymbol{U}_{\mathcal{I}_1, \mathcal{K}_2}$ where $\boldsymbol{U}_{\mathcal{I}_1, \mathcal{K}_1} = \boldsymbol{0}_{|\mathcal{I}_1| \times (|\mathcal{I}_0| - 1)}$.

Since $\boldsymbol{U}_{\mathcal{I}_0, \mathcal{K}_1} \in \Re^{|\mathcal{I}_0| \times (|\mathcal{I}_0| - 1)}$ and it is of full column rank (otherwise it contradicts with $\boldsymbol{U}_{\mathcal{I}}$ being full column rank and $\boldsymbol{U}_{\mathcal{I}_1, \mathcal{K}_1} = \boldsymbol{0}$), we have that $\operatorname{null}(\boldsymbol{U}_{\mathcal{I}_0, \mathcal{K}_1}^\top)$ is of dimension 1. Recalling that $\boldsymbol{U}_{\mathcal{I}_0, \mathcal{K}_1}$ is a submatrix of $\boldsymbol{U}^0$, each column can only be either $\pm \boldsymbol{e}_s$ or $\boldsymbol{e}_{s_1} - \boldsymbol{e}_{s_2}$ for some $s, s_1, s_2 \in [|\mathcal{I}_0|]$. Let $\mathcal{N}_j \subseteq [|\mathcal{I}_0|]$ be the index set $\left\{ s \mid \text{the } s\text{-th entry of } (\boldsymbol{U}_{\mathcal{I}_0})_j \text{ is non-zero} \right\}$ for any $j \in \mathcal{K}_1$. We observe that

$$\operatorname{null}\left(\boldsymbol{U}_{\mathcal{I}_0, \mathcal{K}_1}^\top\right) = \left\{ \boldsymbol{\gamma} \;\middle|\; \begin{array}{ll} \forall j \in \mathcal{K}_1 \text{ with } |\mathcal{N}_j| = 1 : \gamma_s = 0 & \text{for } s \in \mathcal{N}_j \\ \forall j \in \mathcal{K}_1 \text{ with } |\mathcal{N}_j| = 2 : \gamma_{s_1} = \gamma_{s_2} & \text{for } s_1, s_2 \in \mathcal{N}_j \end{array} \right\}. \tag{47}$$

Consider any nonzero $\boldsymbol{\gamma} \in \operatorname{null}\left(\boldsymbol{U}_{\mathcal{I}_0, \mathcal{K}_1}^\top\right)$, we now prove that there is no $s_1, s_2$ such that $\gamma_{s_1}, \gamma_{s_2}$ are both nonzero and $\gamma_{s_1} \neq \gamma_{s_2}$. Assume to the contrary, i.e., we can find $s_1, s_2$ such that $\gamma_{s_1} \gamma_{s_2} \neq 0$ and $\gamma_{s_1} \neq \gamma_{s_2}$. We construct a vector $\hat{\boldsymbol{\gamma}}$ such that $\hat{\gamma}_i = 0$ for all $i$ such that $\gamma_i = 0$, and $\hat{\gamma}_i = 1$ for all $i$ such that $\gamma_i \neq 0$. As $\boldsymbol{\gamma}$ satisfies the condition in (47), so does $\hat{\boldsymbol{\gamma}}$, and hence $\hat{\boldsymbol{\gamma}} \in \operatorname{null}\left(\boldsymbol{U}_{\mathcal{I}_0, \mathcal{K}_1}^\top\right)$. Nevertheless, $\boldsymbol{\gamma}$ and $\hat{\boldsymbol{\gamma}}$ are obviously linearly independent, and hence we have contradiction to that $\operatorname{null}\left(\boldsymbol{U}_{\mathcal{I}_0, \mathcal{K}_1}^\top\right)$ is of dimension 1. Therefore, we can conclude that all nonzero elements in $\boldsymbol{\gamma}$ have the same value.

Consider any $\boldsymbol{\eta}^0 \in \Re^{|\mathcal{I}_0|}, \boldsymbol{\eta}^1 \in \Re^{|\mathcal{I}_1|}, \boldsymbol{\eta} = (\boldsymbol{\eta}^0, \boldsymbol{\eta}^1)$ such that $\boldsymbol{U}_{\mathcal{I}}^\top \boldsymbol{\eta} = \boldsymbol{0}$. It implies $\boldsymbol{0} = \boldsymbol{U}_{\mathcal{I},\mathcal{K}_1}^\top \boldsymbol{\eta} = \boldsymbol{U}_{\mathcal{I}_0,\mathcal{K}_1}^\top \boldsymbol{\eta}^0 + \boldsymbol{U}_{\mathcal{I}_1,\mathcal{K}_1}^\top \boldsymbol{\eta}^1 = \boldsymbol{U}_{\mathcal{I}_0,\mathcal{K}_1}^\top \boldsymbol{\eta}^0$, where the last equality is due to $\boldsymbol{U}_{\mathcal{I}_1,\mathcal{K}_1} = \boldsymbol{0}$. Hence, $\boldsymbol{\eta}^0 \in \operatorname{null}\left(\boldsymbol{U}_{\mathcal{I}_0,\mathcal{K}_1}^\top\right)$, whose dimension has been shown as 1. Therefore, $\boldsymbol{\eta}^0 = k\boldsymbol{\gamma}$ for some $k \in \Re$. As we have shown above, all nonzero elements in $\boldsymbol{\gamma}$ are equal, WLOG, we can have $\boldsymbol{\eta}^0 \geq \boldsymbol{0}$. We are now ready to verify the second condition in Theorem 2.

Given any $\boldsymbol{\beta} \in \Re_+^n$ with $\boldsymbol{V}_{\mathcal{I}}\boldsymbol{\beta} \in \operatorname{span}(\boldsymbol{U}_{\mathcal{I}})$, as $\boldsymbol{\eta} \in \operatorname{null}(\boldsymbol{U}_{\mathcal{I}}^\top)$, we have $0 = \boldsymbol{\eta}^\top \boldsymbol{V}_{\mathcal{I}}\boldsymbol{\beta} = (\boldsymbol{\eta}^0)^\top \boldsymbol{V}_{\mathcal{I}_0}\boldsymbol{\beta} + (\boldsymbol{\eta}^1)^\top \boldsymbol{V}_{\mathcal{I}_1}\boldsymbol{\beta} = (\boldsymbol{\eta}^0)^\top \boldsymbol{V}_{\mathcal{I}_0}\boldsymbol{\beta} = \sum_{i\in[n]} \beta_i (\boldsymbol{\eta}^0)^\top (\boldsymbol{V}_{\mathcal{I}_0})_i$, where the third equality is due to $\boldsymbol{V}_{\mathcal{I}_1} = \boldsymbol{0}$. Since $\boldsymbol{\eta}^0 \geq \boldsymbol{0}, \boldsymbol{V}_{\mathcal{I}_0} \geq \boldsymbol{0}, \boldsymbol{\beta} \geq \boldsymbol{0}$, we have that $\boldsymbol{\eta}^\top \beta_i (\boldsymbol{V}_{\mathcal{I}})_i = \beta_i (\boldsymbol{\eta}^0)^\top (\boldsymbol{V}_{\mathcal{I}_0})_i = 0$ for all $i \in [n]$. Recall that $\operatorname{null}(\boldsymbol{U}_{\mathcal{I}}^\top)$ is of dimension 1, we then have $\beta_i (\boldsymbol{V}_{\mathcal{I}})_i \in \operatorname{span}(\boldsymbol{U}_{\mathcal{I}})$, and the second condition in Theorem 2 is satisfied. Thus $g(\boldsymbol{x},\boldsymbol{z})$ is supermodular in $\boldsymbol{z}$ for all $\boldsymbol{x}$, and we obtain the following reformulation as a simple corollary of Theorem 1,

$$\begin{aligned}
\min\quad & \boldsymbol{a}^\top \boldsymbol{x} + \boldsymbol{\nu}^\top \boldsymbol{l} \\
\text{s.t.}\quad & \boldsymbol{R}_k^\top \boldsymbol{l} \geq \sum_{i\in[2n+1]} p_i^k \sum_{s,j\in[n]} b_{sj} y_{sj}^{k,i}, && k \in [K] \\
& \sum_{j\in[n]} y_{js}^{k,i} - \sum_{j\in[n]} y_{sj}^{k,i} \geq z_s^{k,i} - x_s, && s \in [n], k \in [K], i \in [2n+1] \\
& y_{sj}^{k,i} \geq 0, && s \in [n], j \in [n], k \in [K], i \in [2n+1] \\
& \boldsymbol{l} \geq \boldsymbol{0},\ \boldsymbol{x} \in \mathcal{X}^{lot},
\end{aligned}$$

where $p_i^k, \boldsymbol{z}^{k,i}, k \in [K], i \in [2n+1]$ are the output of Algorithm 1 given the ambiguity sets $\mathcal{F}^k, k \in [K]$ defined by Equation (4). □

### C.16. Proof of Proposition 11

Let $\hat{\boldsymbol{z}} \in \Re^n$ be such that $\hat{z}_i = \xi_i z_i$ for all $i \in [n]$, and we define $\hat{g}(\boldsymbol{x},\hat{\boldsymbol{z}}) = \min\left\{ \boldsymbol{1}^\top \boldsymbol{y} \;\middle|\; \begin{array}{ll} y_t \geq \sum_{s=j}^t (\hat{z}_s - x_s), & j \in [t], t \in [n] \\ y_t \geq 0, & t \in [n] \end{array} \right\}$. Notice that $\hat{g}(\boldsymbol{x},\hat{\boldsymbol{z}}) = \hat{g}(\boldsymbol{x}, (\xi_1 z_1, \ldots, \xi_n z_n)) = g(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{z})$ defined by Equation

$$g(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{z}) = \min\left\{ \boldsymbol{1}^\top \boldsymbol{y} \;\middle|\; \begin{array}{ll} y_t \geq \sum_{s=j}^t (\xi_s z_s - x_s), & j \in [t], t \in [n] \\ y_t \geq 0, & t \in [n] \end{array} \right\}. \tag{48}$$

To prove the supermodularity of $g$, we first show $\hat{g}(\boldsymbol{x},\hat{\boldsymbol{z}})$ is supermodular in $\hat{\boldsymbol{z}}$, and then prove that $g(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{z}) = \hat{g}(\boldsymbol{x}, (\xi_1 z_1, \ldots, \xi_n z_n))$ is supermodular in $(\xi_1, \ldots, \xi_n, z_1, \ldots, z_n)$.

To show the supermodularity of $\hat{g}$, we first rewrite the problem defining $\hat{g}$ in its matrix form, i.e., $\hat{g}(\boldsymbol{x},\hat{\boldsymbol{z}}) = \min\left\{ \boldsymbol{1}^\top \boldsymbol{y} \;\middle|\; \boldsymbol{U}\boldsymbol{y} - \boldsymbol{V}\hat{\boldsymbol{z}} \geq -\boldsymbol{W}\boldsymbol{x} \right\}$, where $\boldsymbol{U} = \begin{bmatrix} \bar{\boldsymbol{U}}^1 \\ \vdots \\ \bar{\boldsymbol{U}}^n \\ \bar{\boldsymbol{U}}^{n+1} \end{bmatrix} \in \Re^{\frac{n^2+3n}{2} \times n}$, $\boldsymbol{V} = \boldsymbol{W} = \begin{bmatrix} \bar{\boldsymbol{V}}^1 \\ \vdots \\ \bar{\boldsymbol{V}}^n \\ \bar{\boldsymbol{V}}^{n+1} \end{bmatrix} \in \Re^{\frac{n^2+3n}{2} \times n}$ are such that

$$\bar{\boldsymbol{U}}^t \in \Re^{t\times n} \text{ are with elements of } \bar{u}_{js}^t = \begin{cases} 1 & \text{if } s = t \\ 0 & \text{otherwise} \end{cases} \text{ for } j \in [t],\ s,t \in [n],$$

$$\bar{\boldsymbol{V}}^t \in \Re^{t\times n} \text{ are with elements of } \bar{v}^t_{js} = \begin{cases} 1 & \text{if } j \le s \le t \\ 0 & \text{otherwise} \end{cases} \text{ for } j \in [t],\ s,t \in [n],$$

$$\bar{\boldsymbol{U}}^{n+1} = \boldsymbol{I}_{n\times n},\ \bar{\boldsymbol{V}}^{n+1} = \boldsymbol{0}_{n\times n}.$$

We prove $\hat{g}(\boldsymbol{x},\hat{\boldsymbol{z}})$ is supermodular in $\hat{\boldsymbol{z}}$ by verify that $\boldsymbol{U},\boldsymbol{V}$ satisfy the condition in Theorem 2. To this end, consider any $\mathcal{I} \subseteq [(n^2+3n)/2], \boldsymbol{\beta} \in \Re^n_+$ with $|\mathcal{I}| = n+1$, $\text{rank}(\boldsymbol{U}_{\mathcal{I}}) = n$, and $\boldsymbol{V}_{\mathcal{I}}\boldsymbol{\beta} \in \text{span}(\boldsymbol{U}_{\mathcal{I}})$. Note that $\text{rank}(\boldsymbol{U}) = n$, and each row of $\boldsymbol{U}_{\mathcal{I}} \in \Re^{(n+1)\times n}$ has only one nonzero element which takes the value of 1. Therefore, there exists $\omega \in [n]$ such that $\boldsymbol{U}$ has two row vectors being $\boldsymbol{e}_\omega$, and exactly one row vector being $\boldsymbol{e}_i$ for each $i \in [n]\setminus\{\omega\}$. WLOG, we let $RI_1,\ldots,RI_{n+1}$ be the distinct row indices such that $\mathcal{I} = \{RI_1,\ldots,RI_{n+1}\}$, $\boldsymbol{u}_{RI_i} = \boldsymbol{e}_i$ for all $i \in [n]$, $\boldsymbol{u}_{RI_{n+1}} = \boldsymbol{e}_\omega$, and $RI_\omega < RI_{n+1}$. Moreover, for notational brevity, we arrange the rows in $\boldsymbol{U}_I$, $\boldsymbol{V}_I$ with the order of $RI_1,\ldots,RI_{n+1}$, which would not change the satisfaction/violation of the condition in Theorem 2. Therefore, $\boldsymbol{U}_{\mathcal{I}} = \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{e}_\omega^\top \end{bmatrix}$. In this case, for any $\boldsymbol{\alpha} \in \Re^n$, $\boldsymbol{U}_{\mathcal{I}}\boldsymbol{\alpha} = \begin{bmatrix} \boldsymbol{\alpha} \\ \alpha_\omega \end{bmatrix}$. This implies that, given any $\boldsymbol{\gamma} \in \Re^{n+1}$, we have $\boldsymbol{\gamma} \in \text{span}(\boldsymbol{U})$ if and only if $\gamma_\omega = \gamma_{n+1}$. Therefore, consider any $\boldsymbol{\beta}$ with $\boldsymbol{V}_{\mathcal{I}}\boldsymbol{\beta} \in \text{span}(\boldsymbol{U}_{\mathcal{I}})$, we know $\boldsymbol{v}^\top_{RI_\omega}\boldsymbol{\beta} = \boldsymbol{v}^\top_{RI_{n+1}}\boldsymbol{\beta}$. Our objective is to show $\beta_i(\boldsymbol{V}_{\mathcal{I}})_i \in \text{span}(\boldsymbol{U}_{\mathcal{I}})$, i.e, $\beta_i v_{RI_\omega,i} = \beta_i v_{RI_{n+1},i}$, for all $i \in [n]$. To see this, we consider two cases.

- Case 1: both $\boldsymbol{u}_{RI_\omega}$ and $\boldsymbol{u}_{RI_{n+1}}$ are extracted from $\bar{\boldsymbol{U}}^\omega$, i.e., $RI_\omega, RI_{n+1} \in \left\{\frac{\omega(\omega-1)}{2}+1,\ldots,\frac{\omega(\omega-1)}{2}+\omega\right\}$. We denote $j_\omega = RI_\omega - \frac{\omega(\omega-1)}{2}$ and $j_{n+1} = RI_{n+1} - \frac{\omega(\omega-1)}{2}$, i.e., $\boldsymbol{u}^\top_{RI_\omega}$ and $\boldsymbol{u}^\top_{RI_{n+1}}$ are the $j_\omega$-th and $j_{n+1}$-th rows in $\bar{\boldsymbol{U}}^\omega$, respectively. By the structure of $\bar{\boldsymbol{V}}^\omega$, we know for all $s \in [n]$,

$$v_{RI_\omega,s} = \bar{v}^\omega_{j_\omega,s} = \begin{cases} 1 & s = j_\omega,\ldots,\omega \\ 0 & s = 1,\ldots,j_\omega - 1 \text{ or } s = \omega+1,\ldots,n, \end{cases}$$
$$v_{RI_{n+1},s} = \bar{v}^\omega_{j_{n+1},s} = \begin{cases} 1 & s = j_{n+1},\ldots,\omega \\ 0 & s = 1,\ldots,j_{n+1} - 1 \text{ or } s = \omega+1,\ldots,n. \end{cases}$$

  In this case, $\boldsymbol{v}^\top_{RI_\omega}\boldsymbol{\beta} = \boldsymbol{v}^\top_{RI_{n+1}}\boldsymbol{\beta}$ implies $\sum\limits_{j=j_\omega}^{\omega}\beta_j = \sum\limits_{j=j_{n+1}}^{\omega}\beta_j$; and hence $\beta_j = 0$ for all $j \in \{j_\omega,\ldots,j_{n+1}-1\}$ since $\boldsymbol{\beta} \ge \boldsymbol{0}$. Now for any arbitrary $i \in [n]$, the equation $\beta_i v_{RI_\omega,i} = \beta_i v_{RI_{n+1},i}$ always holds since 1) $v_{RI_\omega,i} = v_{RI_{n+1},i} = 0$ when $i = 1,\ldots j_\omega - 1$ or $i = \omega+1,\ldots,n$; 2) $\beta_i = 0$ when $i = j_\omega,\ldots,j_{n+1}-1$; 3) $v_{RI_\omega,i} = v_{RI_{n+1},i} = 1$ when $i = j_{n+1},\ldots,\omega$.
- Case 2: $\boldsymbol{u}_{RI_\omega}$ is extracted from $\bar{\boldsymbol{U}}^\omega$ while $\boldsymbol{u}_{RI_{n+1}}$ is extracted from $\bar{\boldsymbol{U}}^{n+1}$. The submatrix $\bar{\boldsymbol{V}}^{n+1} = \boldsymbol{0}_{n\times n}$ implies in this case $\boldsymbol{v}_{RI_{n+1}} = \boldsymbol{0}$. Hence, $\boldsymbol{v}^\top_{RI_\omega}\boldsymbol{\beta} = \boldsymbol{v}^\top_{RI_{n+1}}\boldsymbol{\beta}$ implies $0 = \boldsymbol{v}^\top_{RI_\omega}\boldsymbol{\beta} = \sum\limits_{i\in[n]}\beta_i v_{RI_\omega,i}$. Since $\boldsymbol{v}_{RI_\omega} \ge \boldsymbol{0}$ and $\boldsymbol{\beta} \ge \boldsymbol{0}$, we then have $\beta_i v_{RI_\omega,i} = 0 = \beta_i v_{RI_{n+1},i}$ for all $i \in [n]$.

Therefore, $\hat{g}(\boldsymbol{x},\hat{\boldsymbol{z}})$ is supermodular in $\hat{\boldsymbol{z}}$ for all $\boldsymbol{x}$. We next prove that $g(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{z})$ is supermodular in every two distinct components of $(\boldsymbol{\xi},\boldsymbol{z})$, and hence is jointly supermodular in $(\boldsymbol{\xi},\boldsymbol{z})$.

We first consider argument as the pair $(\xi_i, z_i)$ for some $i \in [n]$ and fix all $\xi_s, z_s$ with $s \in [n]\setminus\{i\}$. As all the remaining elements are fixed, we define $g^i : \Re^2 \to \Re$ and $h : \Re \to \Re$ to be such that

$g^i(\xi_i, z_i) = h(\hat{z}_i) = h(\xi_i z_i) = \hat{g}\left(\boldsymbol{x}, (\xi_1 z_1, \ldots, \xi_n z_n)\right) = g(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{z})$. Hence, it is equivalent to show that $g^i$, as a function of $\xi_i, z_i$, is supermodular in its arguments. To this end, we first observe that $\xi_i z_i$ is increasing and supermodular in $(\xi_i, z_i)$ (recall that $\xi_i, z_i \geq 0$). Further, as a second-stage cost function, $\hat{g}(\boldsymbol{x}, \hat{\boldsymbol{z}})$ has been shown as convex in $\hat{\boldsymbol{z}}$ by literature (e.g., Birge and Louveaux 2011, Theorem 2), and it implies that $h(\hat{z}_i)$ is convex in $\hat{z}_i$. In addition, $h(\hat{z}_i)$ is also increasing in $\hat{z}_i$ by the definition in (48). Therefore, the supermodularity of $g^i$ follows as a corollary of Lemma 3, and $g(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{z})$ is supermodular in $(\xi_i, z_i)$ for all $i \in [n]$.

Next, if the argument is the pair $(\xi_i, z_j)$ for some distinct $i, j \in [n]$, we prove the supermodularity of $g^{ij}(\xi_i, z_j) = \hat{g}\left(\boldsymbol{x}, (\xi_1 z_1, \ldots, \xi_n z_n)\right)$. Consider $\boldsymbol{\xi}', \boldsymbol{\xi}'', \boldsymbol{z}', \boldsymbol{z}'' \in \Re^n$ with $\xi_i' < \xi_i'', z_j' > z_j''$, $\xi_s' = \xi_s'', z_s' = z_s''$ and we denote their common values as $\xi_s, z_s$, respectively, for all $s \in [n] \setminus \{i, j\}$. Since $\boldsymbol{\xi} \in \{0,1\}^n$, by $\xi_i' < \xi_i''$ we know $\xi_i' = 0, \xi_i'' = 1$. Define $\hat{\boldsymbol{z}}'$, $\hat{\boldsymbol{z}}''$, $\hat{\boldsymbol{z}}^{\min}$, $\hat{\boldsymbol{z}}^{\min} \in \Re^n$ such that $\hat{z}_k' = \xi_k' z_k'$, $\hat{z}_k'' = \xi_k'' z_k''$, $\hat{z}_k^{\min} = (\xi_k' \wedge \xi_k'')(z_k' \wedge z_k'')$, $\hat{z}_k^{\max} = (\xi_k' \vee \xi_k'')(z_k' \vee z_k'')$ for all $k \in [n]$. Then these four vectors differ only in their $i$th, $j$th elements. In particular, $(\hat{z}_i', \hat{z}_j') = (0, \xi_j z_j')$, $(\hat{z}_i'', \hat{z}_j'') = (z_i, \xi_j z_j'')$, $(\hat{z}_i^{\min}, \hat{z}_j^{\min}) = (0, \xi_j z_j'')$, $(\hat{z}_i^{\max}, \hat{z}_j^{\max}) = (z_i, \xi_j z_j')$. Hence, denoting $\hat{\boldsymbol{z}}^\circ \in \Re^n$ such that $\hat{z}_i^\circ = \hat{z}_j^\circ = 0$ and $\hat{z}_s^\circ = \xi_s z_s$ for all $s \in [n] \setminus \{i, j\}$, we have

$$
\begin{aligned}
& g^{ij}(\xi_i' \wedge \xi_i'', z_j' \wedge z_j'') + g^{ij}(\xi_i' \vee \xi_i'', z_j' \vee z_j'') - g^{ij}(\xi_i', z_j') - g^{ij}(\xi_i'', z_j'') \\
= \; & \hat{g}\left(\boldsymbol{x}, \hat{\boldsymbol{z}}^{\min}\right) + \hat{g}\left(\boldsymbol{x}, \hat{\boldsymbol{z}}^{\max}\right) - \hat{g}\left(\boldsymbol{x}, \hat{\boldsymbol{z}}'\right) - \hat{g}\left(\boldsymbol{x}, \hat{\boldsymbol{z}}''\right) \\
= \; & \hat{g}\left(\boldsymbol{x}, \hat{\boldsymbol{z}}^\circ + \xi_j z_j'' \boldsymbol{e}_j\right) + \hat{g}\left(\boldsymbol{x}, \hat{\boldsymbol{z}}^\circ + z_i \boldsymbol{e}_i + \xi_j z_j' \boldsymbol{e}_j\right) - \hat{g}\left(\boldsymbol{x}, \hat{\boldsymbol{z}}^\circ + \xi_j z_j' \boldsymbol{e}_j\right) - \hat{g}\left(\boldsymbol{x}, \hat{\boldsymbol{z}}^\circ + z_i \boldsymbol{e}_i + \xi_j z_j'' \boldsymbol{e}_j\right) \\
\geq \; & 0,
\end{aligned}
$$

where the inequality holds because $g\left(\boldsymbol{x}, \hat{\boldsymbol{z}}\right)$ is supermodular in $\hat{\boldsymbol{z}}$. Hence, $g^{ij}$ is supermodular and therefore $g(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{z})$ is supermodular in $(\xi_i, z_j)$.

For $(z_i, z_j)$ or $(\xi_i, \xi_j)$ with $1 \leq i < j \leq n$, the proof is similar to the second case. We now conclude that $g(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{z})$ is supermodular in $(\boldsymbol{\xi}, \boldsymbol{z})$.

Noticing that $\mathcal{F}_{\boldsymbol{\xi}}^k$ (or $\mathcal{F}_{\boldsymbol{z}}^k$) determine a set of 0-1 (or three-point) worst-case marginals for $\tilde{\boldsymbol{\xi}}$ (or $\tilde{\boldsymbol{z}}$), we claim that applying Algorithm 1 yields a $(3n+1)$-point joint distribution of $(\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{z}})$ for each realized scenario. The number of points follows from one plus the number of steps it takes when moving from $(\boldsymbol{0}, \underline{\boldsymbol{z}}^k)$ to $(\boldsymbol{1}, \overline{\boldsymbol{z}}^k)$ only in the positive directions. The number of steps is $3n$, since there are exactly 3 steps on the $i$-th dimension—from $\xi_i = 0 \to 1$, and from $z_i = \underline{z}_i^k \to \mu_i \to \overline{z}_i^k$. We then utilize the results in Theorem 1 and obtain the reformulation as follows.

$$\begin{aligned}
\min \quad & \boldsymbol{\nu}^\top \boldsymbol{l} \\
\text{s.t.} \quad & \boldsymbol{R}_k^\top \boldsymbol{l} \geq \sum_{i\in[3n+1]} p_i^k \mathbf{1}^\top \boldsymbol{y}^{k,i}, && k\in[K] \\
& y_t^{k,i} \geq \sum_{s=j}^{t} (\xi^{k,i} z_s^{k,i} - x_s), && j\in[t], t\in[n], k\in[K], i\in[3n+1] \\
& \boldsymbol{y}^{k,i} \geq \mathbf{0}, && k\in[K], i\in[3n+1] \\
& \boldsymbol{l}\geq \mathbf{0},\ \boldsymbol{x}\in\mathcal{X}^{app},
\end{aligned}$$

where $p_i^k, \boldsymbol{\xi}^{k,i}, \boldsymbol{z}^{k,i}, k\in[K], i\in[3n+1]$ are the output of Algorithm 1, given the ambiguity sets $\mathcal{G}^k, k\in[K]$ defined as $\mathcal{G}^k = \left\{\mathbb{P}^k \;\middle|\; \boldsymbol{\Pi}_{\boldsymbol{\xi}}\mathbb{P}^k \in \mathcal{F}_{\boldsymbol{\xi}}^k, \boldsymbol{\Pi}_{\boldsymbol{z}}\mathbb{P}^k \in \mathcal{F}_{\boldsymbol{z}}^k\right\}$, where $\boldsymbol{\Pi}_{\boldsymbol{\xi}}\mathbb{P}^k$, $\boldsymbol{\Pi}_{\boldsymbol{z}}\mathbb{P}^k$ denotes the marginal distribution of $\tilde{\boldsymbol{\xi}}$ and $\tilde{\boldsymbol{z}}$, respectively under $\mathbb{P}^k$. $\mathcal{F}_{\boldsymbol{\xi}}^k$ is the conditional ambiguity set of $\mathcal{F}_{\boldsymbol{\xi}}$ when $\tilde{k}$ is realized as $k$, and $\mathcal{F}_{\boldsymbol{z}}^k$ is defined by (4). □

### C.17. Proof of Theorem 4

The constraint of the second-stage problem

$$\begin{aligned}
g(\boldsymbol{x},\boldsymbol{z}) = \min \quad & \boldsymbol{h}^\top(\boldsymbol{x}-\boldsymbol{A}\boldsymbol{y}) + \boldsymbol{p}^\top(\boldsymbol{z}-\boldsymbol{y}) - \boldsymbol{r}^\top\boldsymbol{y} \\
\text{s.t.} \quad & \boldsymbol{A}\boldsymbol{y}\leq\boldsymbol{x},\ \boldsymbol{y}\leq\boldsymbol{z},\ \boldsymbol{y}\geq\mathbf{0}.
\end{aligned}$$

can be represented as $\boldsymbol{U}\boldsymbol{y}-\boldsymbol{V}\boldsymbol{z}\geq-\boldsymbol{W}\boldsymbol{x}+\boldsymbol{v}^0$, where $\boldsymbol{U}=\begin{bmatrix}-\boldsymbol{I}\\ \boldsymbol{I}\\ -\boldsymbol{A}\end{bmatrix}, \boldsymbol{V}=\begin{bmatrix}-\boldsymbol{I}\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix}$ and $\boldsymbol{W}=\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\\ \boldsymbol{I}\end{bmatrix}$.

We first prove the "if" direction. Suppose the condition for $\boldsymbol{A}$ in this theorem is satisfied. By Theorem 2, whether $\boldsymbol{U},\boldsymbol{V}$ lead to supermodularity of $g$ is equivalent to whether $-\boldsymbol{U},-\boldsymbol{V}$ do so. Therefore, here we verify the supermodularity based on $-\boldsymbol{U},-\boldsymbol{V}$. Observing that $-\boldsymbol{U}$ and $-\boldsymbol{V}$ have the structure as in Proposition 5 with $\boldsymbol{U}^\circ=\begin{bmatrix}-\boldsymbol{I}\\ \boldsymbol{A}\end{bmatrix}$, we now show that every $2\times3$ submatrix of $\boldsymbol{U}^\circ$ contains at least one pair of columns which are linearly dependent. If both rows of the $2\times3$ submatrix are extracted from $\boldsymbol{A}$, then this submatrix must have two linearly dependent columns by the assumption on $\boldsymbol{A}$. If at least one of the rows are from $-\boldsymbol{I}$, since the rows from $-\boldsymbol{I}$ have at least two zero elements, then this submatrix must have two linearly dependent columns.

We now prove the "only if" direction by contradiction. We first consider the case that $\boldsymbol{A}\in\Re_+^{2\times3}$. Assume the contrary, i.e., every two columns in $\boldsymbol{A}$ are in different directions. Given that $\boldsymbol{A}\geq\mathbf{0}$, there must be one column in $\boldsymbol{A}$ being a conical combination of the other two columns. WLOG, let $\boldsymbol{A}_3$ be a conical combination of $\boldsymbol{A}_1,\boldsymbol{A}_2$. We remark that multiplying any strictly positive constant by a row/column in $\boldsymbol{A}$, or switching rows, or switching columns does not affect whether the corresponding function $g$ is supermodular. Therefore, we can make the following simplification on $\boldsymbol{A}$. Since $\boldsymbol{A}_1,\boldsymbol{A}_2$ are linearly independent, WLOG, we can let $\boldsymbol{A}=\begin{bmatrix}1 & a & c\\ b & 1 & d\end{bmatrix}$ with

$ab < 1$. Since $\boldsymbol{A}_3$ is a conical combination of $\boldsymbol{A}_1, \boldsymbol{A}_2$, we have $cd > 0$; WLOG, we can let $d = 1$, i.e., $\boldsymbol{A} = \begin{bmatrix} 1 & a & c \\ b & 1 & 1 \end{bmatrix}$. Multiplying the first row by $1/c$, and then multiplying the first column by $c$, we have $\boldsymbol{A} = \begin{bmatrix} 1 & a/c & 1 \\ bc & 1 & 1 \end{bmatrix}$. Let $a/c$, $bc$ be the new $a, b$, we have $\boldsymbol{A} = \begin{bmatrix} 1 & a & 1 \\ b & 1 & 1 \end{bmatrix}$ with $ab < 1$. Again, since $\boldsymbol{A}_3$ is a conical combination of $\boldsymbol{A}_1, \boldsymbol{A}_2$, we have either $a, b < 1$ or $a, b > 1$. Together with $ab < 1$, we know $a, b < 1$. In summary, WLOG, we let $\boldsymbol{A} = \begin{bmatrix} 1 & a & 1 \\ b & 1 & 1 \end{bmatrix}$ with $a, b \in [0, 1)$.

We define $\bar{g}(\boldsymbol{x}, \boldsymbol{z}) = g(\boldsymbol{x}, \boldsymbol{z}) - \boldsymbol{p}^\top \boldsymbol{z}$, then it is equivalent to prove that $\bar{g}(\boldsymbol{x}, \boldsymbol{z})$ is not supermodular in $\boldsymbol{z}$. We now construct such a counterexample. Let $\boldsymbol{h} = \boldsymbol{0}$, $\boldsymbol{r} = \boldsymbol{0}$, $\boldsymbol{p} = (1, 1, \epsilon)$ with any $\epsilon \in (0, 1)$. We choose $\boldsymbol{x} = (1 - ab)\boldsymbol{1}$, $\boldsymbol{z}' = (1 - a, 0, 1 - ab)$, $\boldsymbol{z}'' = (0, 1 - b, 1 - ab)$. Denote $\boldsymbol{z}^\wedge = \boldsymbol{z}' \wedge \boldsymbol{z}''$, $\boldsymbol{z}^\vee = \boldsymbol{z}' \vee \boldsymbol{z}''$, we have $\boldsymbol{z}^\wedge = (0, 0, 1 - ab)$, $\boldsymbol{z}^\vee = (1 - a, 1 - b, 1 - ab)$. We notice that

$$\begin{aligned} \bar{g}(\boldsymbol{x}, \boldsymbol{z}) &= \min \left\{ -\boldsymbol{p}^\top \boldsymbol{y} \;\middle|\; \boldsymbol{A}\boldsymbol{y} \le \boldsymbol{x}, \boldsymbol{0} \le \boldsymbol{y} \le \boldsymbol{z} \right\} \\ &= \min \left\{ -y_1 - y_2 - \epsilon y_3 \;\middle|\; \begin{array}{l} y_1 + a y_2 + y_3 \le 1 - ab \\ b y_1 + y_2 + y_3 \le 1 - ab \\ (0,0,0) \le (y_1, y_2, y_3) \le (z_1, z_2, z_3). \end{array} \right\} \end{aligned} \tag{49}$$

Hence,

$$\begin{aligned} \bar{g}(\boldsymbol{x}, \boldsymbol{z}') &= \min \left\{ -y_1 - \epsilon y_3 \;\middle|\; \begin{array}{l} y_1 + y_3 \le 1 - ab, \\ 0 \le y_1 \le 1 - a,\; y_2 = 0,\; y_3 \ge 0 \end{array} \right\}, \\ \bar{g}(\boldsymbol{x}, \boldsymbol{z}'') &= \min \left\{ -y_2 - \epsilon y_3 \;\middle|\; \begin{array}{l} y_2 + y_3 \le 1 - ab, \\ y_1 = 0,\; 0 \le y_2 \le 1 - b,\; y_3 \ge 0 \end{array} \right\}, \\ \bar{g}(\boldsymbol{x}, \boldsymbol{z}^\wedge) &= \min \left\{ -\epsilon y_3 \;\middle|\; \begin{array}{l} y_3 \le 1 - ab, \\ y_1 = y_2 = 0,\; y_3 \ge 0 \end{array} \right\}, \\ \bar{g}(\boldsymbol{x}, \boldsymbol{z}^\vee) &= \min \left\{ -y_1 - y_2 - \epsilon y_3 \;\middle|\; \begin{array}{l} y_1 + a y_2 + y_3 \le 1 - ab, \\ b y_1 + y_2 + y_3 \le 1 - ab, \\ 0 \le y_1 \le 1 - a,\; 0 \le y_2 \le 1 - b,\; y_3 \ge 0 \end{array} \right\}. \end{aligned}$$

Since $0 < \epsilon < 1$, in the optimization problem for $\bar{g}(\boldsymbol{x}, \boldsymbol{z}')$, the optimal solution should be that $y_1$ goes to the upper bound, i.e, $y_1 = 1 - a$, $y_2 = 0$ and $y_3 = (1 - ab) - (1 - a) = a(1 - b)$. Similarly, in the optimization problem for $\bar{g}(\boldsymbol{x}, \boldsymbol{z}'')$, the optimal $\boldsymbol{y} = (0, 1 - b, b(1 - a))$; in that for $\bar{g}(\boldsymbol{x}, \boldsymbol{z}^\wedge)$, the optimal $\boldsymbol{y} = (0, 0, 1 - ab)$; in that for $\bar{g}(\boldsymbol{x}, \boldsymbol{z}^\vee)$, the optimal $\boldsymbol{y} = (1 - a, 1 - b, 0)$. We then have

$$\begin{aligned} & \bar{g}(\boldsymbol{x}, \boldsymbol{z}') + \bar{g}(\boldsymbol{x}, \boldsymbol{z}'') - \bar{g}(\boldsymbol{x}, \boldsymbol{z}^\wedge) - \bar{g}(\boldsymbol{x}, \boldsymbol{z}^\vee) \\ = & -\left( (1 - a + \epsilon a(1 - b)) + (1 - b + \epsilon b(1 - a)) - \epsilon(1 - ab) - (1 - a + 1 - b) \right) \\ = & \; \epsilon(1 - a)(1 - b) > 0, \end{aligned}$$

where the last equality holds since $0 < a, b < 1$. Therefore, $\bar{g}(\boldsymbol{x}, \boldsymbol{z}^\wedge) + \bar{g}(\boldsymbol{x}, \boldsymbol{z}^\vee) < \bar{g}(\boldsymbol{x}, \boldsymbol{z}') + \bar{g}(\boldsymbol{x}, \boldsymbol{z}'')$, this function $\bar{g}$ is not supermodular.

For the general case of $\boldsymbol{A} \in \Re_+^{l \times n}$, we can prove the result by the same contradiction. WLOG, we assume the $2 \times 3$ submatrix of $\boldsymbol{A}$, which is obtained by deleting all rows except the first two and all columns except the first three, is such that each pair of columns in it are linearly independent. We can then let $z_i' = z_i'' = 0$ for $i \in \{4,5,\dots,n\}$ and $x_i$ be sufficiently large for $i \in \{3,4,\dots,l\}$ such that it would not affect the feasible region of $\boldsymbol{y}$. We then have $\bar{g}(\boldsymbol{x},\boldsymbol{z})$ with exactly the same expression in Equation (49). Therefore, we still have $\bar{g}(\boldsymbol{x},\boldsymbol{z}^{\wedge}) + \bar{g}(\boldsymbol{x},\boldsymbol{z}^{\vee}) < \bar{g}(\boldsymbol{x},\boldsymbol{z}') + \bar{g}(\boldsymbol{x},\boldsymbol{z}'')$. □

### C.18. Proof of Corollary 2

We first prove the "if" direction using Theorem 4. Consider any $2\times 3$ submatrix of $\boldsymbol{A}$, which, WLOG, is $\boldsymbol{C} = \boldsymbol{A}_{\{1,2\},\{1,2,3\}}$. Let $\hat{S}_i = S_i \cap \{1,2,3\},\ i = 1,2$. If $\hat{S}_1 \cap \hat{S}_2 = \emptyset$, then at least one of the rows in $\boldsymbol{C}$ has two zero elements, and hence $\boldsymbol{C}$ has at least one pair of columns which are linearly dependent. If $\hat{S}_1 \cap \hat{S}_2 \neq \emptyset$, by the definition of Tree Family, we have either $\hat{S}_1 \subseteq \hat{S}_2$ or $\hat{S}_2 \subseteq \hat{S}_1$. WLOG, we let $\hat{S}_1 \subseteq \hat{S}_2$. If $|\hat{S}_1| = 1$, then the first row of $\boldsymbol{C}$ has two zero elements and hence $\boldsymbol{C}$ has at least one pair of columns which are linearly dependent. If $|\hat{S}_1| \geq 2$, WLOG, $\{1,2\} \subseteq \hat{S}_1$, by the definition of Proportional Tree Family, we have $a_{11}/a_{21} = a_{12}/a_{22}$, hence $\boldsymbol{C}$ has at least one pair of columns which are linearly dependent. In summary, $\boldsymbol{C}$ always have at least one pair of columns which are linearly dependent. □

### C.19. Proof of Theorem 5

The case for $\operatorname{rank}(\boldsymbol{U}) = r$ is straightforward, so we only consider the case where $\operatorname{rank}(\boldsymbol{U}) < r$. In that case, we only need to verify whether $\boldsymbol{U},\boldsymbol{V}$ satisfy the second part of the condition in Theorem 2, which depends solely on the relationship between $\boldsymbol{V}$ and $\operatorname{span}(\boldsymbol{U})$. Thus, removing the dependent columns in $\boldsymbol{U}$ does not change the satisfaction or violation of the conditions. Therefore, the procedure in line 4 of the algorithm does not change the result and WLOG, we can assume $\boldsymbol{U}$ has $r_0$ columns, i.e., $m = r_0$.

First we look into the case where Algorithm 2 returns $s = 0$. This implies that there exists an index set $\mathcal{I} \subseteq [r]$ and indices $i \in [r] \setminus \mathcal{I}, a, b \in [r_0]$ with $|\mathcal{I}| = r_0$, $\boldsymbol{U}_{\mathcal{I}}$ invertible and $d_{ia}d_{ib} < 0$. WLOG, we let $d_{ia} > 0, d_{ib} < 0$.

Denote $\boldsymbol{\beta} = \dfrac{\boldsymbol{e}_a}{d_{ia}} - \dfrac{\boldsymbol{e}_b}{d_{ib}} \geq \boldsymbol{0}$, $\boldsymbol{\alpha} = \boldsymbol{U}_{\mathcal{I}}^{-1}\left(\dfrac{(\boldsymbol{V}_{\mathcal{I}})_a}{d_{ia}} - \dfrac{(\boldsymbol{V}_{\mathcal{I}})_b}{d_{ib}}\right)$, then

$$\begin{bmatrix} \boldsymbol{V}_{\mathcal{I}} \\ \boldsymbol{v}_i^{\top} \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} \dfrac{(\boldsymbol{V}_{\mathcal{I}})_a}{d_{ia}} - \dfrac{(\boldsymbol{V}_{\mathcal{I}})_b}{d_{ib}} \\ \dfrac{v_{ia}}{d_{ia}} - \dfrac{v_{ib}}{d_{ib}} \end{bmatrix} = \begin{bmatrix} \dfrac{(\boldsymbol{V}_{\mathcal{I}})_a}{d_{ia}} - \dfrac{(\boldsymbol{V}_{\mathcal{I}})_b}{d_{ib}} \\ \dfrac{d_{ia} + \boldsymbol{u}_i^{\top}\boldsymbol{U}_{\mathcal{I}}^{-1}(\boldsymbol{V}_{\mathcal{I}})_a}{d_{ia}} - \dfrac{d_{ib} + \boldsymbol{u}_i^{\top}\boldsymbol{U}_{\mathcal{I}}^{-1}(\boldsymbol{V}_{\mathcal{I}})_b}{d_{ib}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{U}_{\mathcal{I}} \\ \boldsymbol{u}_i^{\top} \end{bmatrix} \boldsymbol{\alpha},$$

We let $\hat{\mathcal{I}} = \mathcal{I} \cup \{i\}$. The above equality implies $\boldsymbol{V}_{\hat{\mathcal{I}}}\boldsymbol{\beta} = \boldsymbol{U}_{\hat{\mathcal{I}}}\boldsymbol{\alpha} \in \operatorname{span}(\boldsymbol{U}_{\hat{\mathcal{I}}})$. On the other hand, for $\beta_a(\boldsymbol{V}_{\hat{\mathcal{I}}})_a$ we have

$$\beta_a \begin{bmatrix} (\boldsymbol{V}_{\mathcal{I}})_a \\ v_{ia} \end{bmatrix} = \begin{bmatrix} \dfrac{(\boldsymbol{V}_{\mathcal{I}})_a}{d_{ia}} \\ \dfrac{d_{ia} + \boldsymbol{u}_i^{\top}\boldsymbol{U}_{\mathcal{I}}^{-1}(\boldsymbol{V}_{\mathcal{I}})_a}{d_{ia}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{U}_{\mathcal{I}} \\ \boldsymbol{u}_i^{\top} \end{bmatrix} \frac{\boldsymbol{U}_{\mathcal{I}}^{-1}(\boldsymbol{V}_{\mathcal{I}})_a}{d_{ia}} + \begin{bmatrix} \boldsymbol{0} \\ 1 \end{bmatrix} = \boldsymbol{U}_{\hat{\mathcal{I}}} \frac{\boldsymbol{U}_{\mathcal{I}}^{-1}(\boldsymbol{V}_{\mathcal{I}})_a}{d_{ia}} + \begin{bmatrix} \boldsymbol{0} \\ 1 \end{bmatrix}.$$

Since $\boldsymbol{U}_{\mathcal{I}}$ is invertible, there is no $\boldsymbol{\gamma} \in \Re^{r_0}$ such that $\boldsymbol{U}_{\hat{\mathcal{I}}}\boldsymbol{\gamma} = \begin{bmatrix} \boldsymbol{U}_{\mathcal{I}} \\ \boldsymbol{u}_i^\top \end{bmatrix} \boldsymbol{\gamma} = \begin{bmatrix} \boldsymbol{0} \\ 1 \end{bmatrix}$. Hence $\beta_a(\boldsymbol{V}_{\hat{\mathcal{I}}})_a \notin \text{span}(\boldsymbol{U}_{\hat{\mathcal{I}}})$ and the second part of the condition in Theorem 2 is violated.

We now investigate the case where the second part of the condition in Theorem 2 is violated. That means, there exist $\hat{\mathcal{I}} \subseteq [r]$, $\boldsymbol{\beta} \geq \boldsymbol{0}$ and $a \in [r_0]$ such that $|\hat{\mathcal{I}}| = r_0 + 1, \text{rank}(\boldsymbol{U}_{\hat{\mathcal{I}}}) = r_0$, $\boldsymbol{V}_{\hat{\mathcal{I}}}\boldsymbol{\beta} \in \text{span}(\boldsymbol{U}_{\hat{\mathcal{I}}})$ but $\beta_a(\boldsymbol{V}_{\hat{\mathcal{I}}})_a \notin \text{span}(\boldsymbol{U}_{\hat{\mathcal{I}}})$. We choose $\mathcal{I} \subseteq \hat{\mathcal{I}}$ such that $|\mathcal{I}| = r_0$ and $\boldsymbol{U}_{\mathcal{I}}$ is invertible, and denote $i$ as the unique index in $\hat{\mathcal{I}} \backslash \mathcal{I}$. It follows that

$$\boldsymbol{V}_{\hat{\mathcal{I}}}\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{V}_{\mathcal{I}} \\ \boldsymbol{v}_i^\top \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{v}_i^\top - \boldsymbol{u}_i^\top \boldsymbol{U}_{\mathcal{I}}^{-1} \boldsymbol{V}_{\mathcal{I}} \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{U}_{\mathcal{I}} \\ \boldsymbol{u}_i^\top \end{bmatrix} \boldsymbol{U}_{\mathcal{I}}^{-1} \boldsymbol{V}_{\mathcal{I}} \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{d}_i^\top \end{bmatrix} \boldsymbol{\beta} + \boldsymbol{U}_{\hat{\mathcal{I}}} \boldsymbol{U}_{\mathcal{I}}^{-1} \boldsymbol{V}_{\mathcal{I}} \boldsymbol{\beta},$$

$$\beta_a(\boldsymbol{V}_{\hat{\mathcal{I}}})_a = \beta_a \begin{bmatrix} (\boldsymbol{V}_{\mathcal{I}})_a \\ v_{ia} \end{bmatrix} = \beta_a \left( \begin{bmatrix} \boldsymbol{0} \\ v_{ia} - \boldsymbol{u}_i^\top \boldsymbol{U}_{\mathcal{I}}^{-1} (\boldsymbol{V}_{\mathcal{I}})_a \end{bmatrix} + \begin{bmatrix} \boldsymbol{U}_{\mathcal{I}} \\ \boldsymbol{u}_i^\top \end{bmatrix} \boldsymbol{U}_{\mathcal{I}}^{-1} (\boldsymbol{V}_{\mathcal{I}})_a \right) = \begin{bmatrix} \boldsymbol{0} \\ \beta_a d_{ia} \end{bmatrix} + \beta_a \boldsymbol{U}_{\hat{\mathcal{I}}} \boldsymbol{U}_{\mathcal{I}}^{-1} (\boldsymbol{V}_{\mathcal{I}})_a.$$

Since $\boldsymbol{V}_{\hat{\mathcal{I}}}\boldsymbol{\beta}, \boldsymbol{U}_{\hat{\mathcal{I}}}\boldsymbol{U}_{\mathcal{I}}^{-1}\boldsymbol{V}_{\mathcal{I}}\boldsymbol{\beta}, \beta_a \boldsymbol{U}_{\hat{\mathcal{I}}}\boldsymbol{U}_{\mathcal{I}}^{-1}(\boldsymbol{V}_{\mathcal{I}})_a \in \text{span}(\boldsymbol{U}_{\hat{\mathcal{I}}})$ and $\beta_a(\boldsymbol{V}_{\hat{\mathcal{I}}})_a \notin \text{span}(\boldsymbol{U}_{\hat{\mathcal{I}}})$, the above equations imply $\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{d}_i^\top \end{bmatrix} \boldsymbol{\beta} \in \text{span}(\boldsymbol{U}_{\hat{\mathcal{I}}})$ and $\begin{bmatrix} \boldsymbol{0} \\ \beta_i d_{ia} \end{bmatrix} \notin \text{span}(\boldsymbol{U}_{\hat{\mathcal{I}}})$. According to $\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{d}_i^\top \end{bmatrix} \boldsymbol{\beta} \in \text{span}(\boldsymbol{U}_{\hat{\mathcal{I}}})$, there exists $\boldsymbol{\alpha} \in \Re^{r_0}$ with $\boldsymbol{U}_{\mathcal{I}}\boldsymbol{\alpha} = \boldsymbol{0}, \boldsymbol{u}_i^\top \boldsymbol{\alpha} = \boldsymbol{d}_i^\top \boldsymbol{\beta}$. Since $\boldsymbol{U}_{\mathcal{I}}$ is invertible, $\boldsymbol{\alpha} = \boldsymbol{0}$ and hence $\boldsymbol{d}_i^\top \boldsymbol{\beta} = \boldsymbol{u}_i^\top \boldsymbol{\alpha} = 0$. According to $\begin{bmatrix} \boldsymbol{0} \\ \beta_i d_{ia} \end{bmatrix} \notin \text{span}(\boldsymbol{U}_{\hat{\mathcal{I}}})$, we obtain $\beta_a d_{ia} \neq 0$. As $\boldsymbol{\beta} \geq \boldsymbol{0}$, $\beta_a d_{ia} \neq 0$ and $\boldsymbol{d}_i^\top \boldsymbol{\beta} = 0$, we must have an index $b \in [r_0]$ such that $d_{ib}, d_{ia}$ are of different signs. Hence the algorithm returns $s = 0$.

□

### C.20. Proof of Theorem 6

We first reformulate the second-stage problem as

$$\begin{aligned} g^W(\boldsymbol{x}, \boldsymbol{z}) = \min \ & \boldsymbol{b}^\top \boldsymbol{y} \\ \text{s.t.} \ & \boldsymbol{U}\boldsymbol{y} - \left(\boldsymbol{V} - \left[\boldsymbol{W}^1\boldsymbol{x} \cdots \boldsymbol{W}^n\boldsymbol{x}\right]\right) \boldsymbol{z} \geq -\boldsymbol{W}^0\boldsymbol{x} + \boldsymbol{v}^0, \end{aligned}$$

where $[\boldsymbol{W}^1\boldsymbol{x} \cdots \boldsymbol{W}^n\boldsymbol{x}]$ stands for an $r \times n$ matrix with its $i$-th column being $\boldsymbol{W}^i\boldsymbol{x}$. We denote $\bar{\boldsymbol{V}}^{\boldsymbol{x}} = \boldsymbol{V} - [\boldsymbol{W}^1\boldsymbol{x} \cdots \boldsymbol{W}^n\boldsymbol{x}]$ for convenience. Following Theorem 2, it suffices to show that the proposed conditions hold if and only if $\boldsymbol{U}, \bar{\boldsymbol{V}}^{\boldsymbol{x}}$ satisfy the conditions in Theorem 2 for any $\boldsymbol{x}$. The case of $\text{rank}(\boldsymbol{U}) = r$ is straightforward. Hence, in the rest of the prove, we only focus on the case of $\text{rank}(\boldsymbol{U}) < r$, i.e., the second condition in this theorem and that in Theorem 2, which are called Condition 2) and Condition $\tilde{2}$) throughout this proof. In particular, Condition $\tilde{2}$) can be stated as

*$\tilde{2}$) for all* $\mathcal{I} \subseteq [r]$, $\boldsymbol{\beta} \in \Re^n_+$, $\boldsymbol{x} \in \Re^l$ *with* $|\mathcal{I}| = \text{rank}(\boldsymbol{U}) + 1$, $\text{rank}(\boldsymbol{U}_{\mathcal{I}}) = \text{rank}(\boldsymbol{U})$ *and* $\bar{\boldsymbol{V}}^{\boldsymbol{x}}_{\mathcal{I}}\boldsymbol{\beta} \in \text{span}(\boldsymbol{U}_{\mathcal{I}})$*, we must have* $\beta_i(\bar{\boldsymbol{V}}^{\boldsymbol{x}}_{\mathcal{I}})_i \in \text{span}(\boldsymbol{U}_{\mathcal{I}})$ *holds for every* $i \in [n]$.

We now prove that Condition 2) is equivalent to Condition $\tilde{2}$).

First, we make an equivalent interpretation for Condition 2) and Condition $\tilde{2}$). Notice that both conditions are for the same set of index sets. We consider any such index set $\mathcal{I}$. Since $|\mathcal{I}| = \text{rank}(\boldsymbol{U}_{\mathcal{I}}) + 1$, $\text{span}(\boldsymbol{U}_{\mathcal{I}})$ is a hyperplane in $\Re^{|\mathcal{I}|}$. Therefore, there exists a unit vector $\boldsymbol{\eta} \in \Re^{|\mathcal{I}|}$ such that it is orthogonal to all vectors in span($\boldsymbol{U}_\mathcal{I}$), and all elements in $\Re^{|\mathcal{I}|}$ can be represented as linear combinations of $\boldsymbol{\eta}$ and a vector in span($\boldsymbol{U}_\mathcal{I}$). Therefore, for any $i,j \in [n]$,

$$\begin{aligned}(\boldsymbol{V}_\mathcal{I})_i &= \boldsymbol{\xi}_i + \lambda_i \boldsymbol{\eta}, \quad \boldsymbol{W}_\mathcal{I}^i \boldsymbol{x} = \boldsymbol{\zeta}_i^{\boldsymbol{x}} + \mu_i^{\boldsymbol{x}} \boldsymbol{\eta},\\ (\boldsymbol{V}_\mathcal{I})_j &= \boldsymbol{\xi}_j + \lambda_j \boldsymbol{\eta}, \quad \boldsymbol{W}_\mathcal{I}^j \boldsymbol{x} = \boldsymbol{\zeta}_j^{\boldsymbol{x}} + \mu_j^{\boldsymbol{x}} \boldsymbol{\eta}.\end{aligned} \tag{50}$$

for some $\lambda_i, \lambda_j, \mu_i^{\boldsymbol{x}}, \mu_j^{\boldsymbol{x}} \in \Re$ and $\boldsymbol{\xi}_i, \boldsymbol{\xi}_j, \boldsymbol{\zeta}_i^{\boldsymbol{x}}, \boldsymbol{\zeta}_j^{\boldsymbol{x}} \in \text{span}(\boldsymbol{U}_\mathcal{I})$. Since $\boldsymbol{\eta}$ is a unit vector, we have

$$0 = \boldsymbol{\eta}^\top \left( (\boldsymbol{V}_\mathcal{I})_i - \boldsymbol{\eta}\boldsymbol{\eta}^\top (\boldsymbol{V}_\mathcal{I})_i \right) = \boldsymbol{\eta}^\top (\boldsymbol{V}_\mathcal{I})_i - \boldsymbol{\eta}^\top \boldsymbol{\eta}\boldsymbol{\eta}^\top (\boldsymbol{\xi}_i + \lambda_i \boldsymbol{\eta}) = \boldsymbol{\eta}^\top (\boldsymbol{V}_\mathcal{I})_i - \lambda_i,$$

and hence $\lambda_i = \boldsymbol{\eta}^\top (\boldsymbol{V}_\mathcal{I})_i$. The same logic applies to $(\boldsymbol{V}_\mathcal{I})_j$ and $\boldsymbol{W}_\mathcal{I}^i \boldsymbol{x}, \boldsymbol{W}_\mathcal{I}^j \boldsymbol{x}$.

In Condition 2), we notice that $(\boldsymbol{\eta}^\top (\boldsymbol{V}_\mathcal{I})_i) \cdot (\boldsymbol{\eta}^\top (\boldsymbol{V}_\mathcal{I})_j) \geq 0$ is equivalent to $\lambda_i \lambda_j \geq 0$; moreover, $(\boldsymbol{W}_\mathcal{I}^i)^\top \boldsymbol{\eta}\boldsymbol{\eta}^\top \boldsymbol{W}_\mathcal{I}^j$ is positive semidefinite if and only if $(\boldsymbol{\eta}^\top \boldsymbol{W}_\mathcal{I}^i \boldsymbol{x}) \cdot \left(\boldsymbol{\eta}^\top \boldsymbol{W}_\mathcal{I}^j \boldsymbol{x}\right) \geq 0$ for all $\boldsymbol{x}$. Hence, we conclude that Condition 2a) is equivalent to "$\lambda_i \lambda_j \geq 0, \mu_i^{\boldsymbol{x}} \mu_j^{\boldsymbol{x}} \geq 0$ for all $\boldsymbol{x}$ and $i,j \in [n]$". For Condition 2b), since the equality holds if and only if $(\boldsymbol{\eta}^\top (\boldsymbol{V}_\mathcal{I})_i) \cdot \left(\boldsymbol{\eta}^\top \boldsymbol{W}_\mathcal{I}^j \boldsymbol{x}\right) = (\boldsymbol{\eta}^\top (\boldsymbol{V}_\mathcal{I})_j) \cdot (\boldsymbol{\eta}^\top \boldsymbol{W}_\mathcal{I}^i \boldsymbol{x})$ for all $\boldsymbol{x}$, we conclude that it is equivalent to the condition "$\lambda_i \mu_j^{\boldsymbol{x}} = \lambda_j \mu_i^{\boldsymbol{x}}$ for all $\boldsymbol{x}$ and $i,j \in [n]$".

For Condition $\tilde{2}$), by the definition of $\bar{\boldsymbol{V}}_\mathcal{I}^{\boldsymbol{x}}$,

$$\begin{aligned}(\bar{\boldsymbol{V}}_\mathcal{I}^{\boldsymbol{x}})_i &= (\boldsymbol{V}_\mathcal{I})_i - \boldsymbol{W}_\mathcal{I}^i \boldsymbol{x} = (\boldsymbol{\xi}_i - \boldsymbol{\zeta}_i^{\boldsymbol{x}}) + (\lambda_i - \mu_i^{\boldsymbol{x}}) \boldsymbol{\eta},\\ (\bar{\boldsymbol{V}}_\mathcal{I}^{\boldsymbol{x}})_j &= (\boldsymbol{V}_\mathcal{I})_j - \boldsymbol{W}_\mathcal{I}^j \boldsymbol{x} = (\boldsymbol{\xi}_j - \boldsymbol{\zeta}_j^{\boldsymbol{x}}) + (\lambda_j - \mu_j^{\boldsymbol{x}}) \boldsymbol{\eta}.\end{aligned}$$

Observing that Condition $\tilde{2}$) is violated if and only if there exist $i,j \in [n]$ with $(\lambda_i - \mu_i^{\boldsymbol{x}})(\lambda_j - \mu_j^{\boldsymbol{x}}) < 0$, we obtain its equivalent condition as

$$(\lambda_i - \mu_i^{\boldsymbol{x}})(\lambda_j - \mu_j^{\boldsymbol{x}}) \geq 0 \qquad \forall \boldsymbol{x} \in \Re^l, i,j \in [n]. \tag{51}$$

We now prove the direction "Condition $\tilde{2}$) $\implies$ Condition 2)". Consider any two distinct indexes $i,j \in [n]$. We first choose $\boldsymbol{x} = \boldsymbol{0}$, hence $\bar{\boldsymbol{V}}^{\boldsymbol{x}} = \boldsymbol{V}$. We assume the contrary to the first argument of Condition 2a), i.e., $\lambda_i \lambda_j < 0$, WLOG, $\lambda_i > 0$, $\lambda_j < 0$. We then choose $\boldsymbol{\beta} = -\lambda_j \boldsymbol{e}_i + \lambda_i \boldsymbol{e}_j \in \Re_+^n$, and have $\boldsymbol{V}_\mathcal{I} \boldsymbol{\beta} = -\lambda_j \boldsymbol{\xi}_i + \lambda_i \boldsymbol{\xi}_j \in \text{span}(\boldsymbol{U}_\mathcal{I})$. However, $\beta_j (\boldsymbol{V}_\mathcal{I})_j = \lambda_i \boldsymbol{\xi}_j + \lambda_i \lambda_j \boldsymbol{\eta} \notin \text{span}(\boldsymbol{U}_\mathcal{I})$. We hence have contradiction with Condition $\tilde{2}$), and conclude $\lambda_i \lambda_j \geq 0$, the first argument of Condition 2a) is true.

Next we show the second argument of Condition 2a) by contradiction. Notice for any constant $\theta \in \Re$, $(\bar{\boldsymbol{V}}_\mathcal{I}^{\theta\boldsymbol{x}})_i = (\boldsymbol{\xi}_i - \theta \boldsymbol{\zeta}_i^{\boldsymbol{x}}) + (\lambda_i - \theta \mu_i^{\boldsymbol{x}}) \boldsymbol{\eta}, (\bar{\boldsymbol{V}}_\mathcal{I}^{\theta\boldsymbol{x}})_j = (\boldsymbol{\xi}_j - \theta \boldsymbol{\zeta}_j^{\boldsymbol{x}}) + (\lambda_j - \theta \mu_j^{\boldsymbol{x}}) \boldsymbol{\eta}$. If $\mu_i^{\boldsymbol{x}} \mu_j^{\boldsymbol{x}} < 0$, we can always find $\theta$ such that $(\lambda_i - \theta \mu_i^{\boldsymbol{x}})(\lambda_j - \theta \mu_j^{\boldsymbol{x}}) = \mu_i^{\boldsymbol{x}} \mu_j^{\boldsymbol{x}} \theta^2 - (\lambda_i \mu_j^{\boldsymbol{x}} + \mu_i^{\boldsymbol{x}} \lambda_j)\theta + \lambda_i \lambda_j < 0$. Therefore, the equivalent condition for Condition $\tilde{2}$), i.e., (51), is violated for $\theta \boldsymbol{x}$. Hence, we have contradiction, and conclude that $\mu_i^{\boldsymbol{x}} \mu_j^{\boldsymbol{x}} \geq 0$, the second argument of Condition 2a) is true.

We now prove Condition 2b). By Condition 2a), we already have $\lambda_i \lambda_j \geq 0, \mu_i^{\boldsymbol{x}} \mu_j^{\boldsymbol{x}} \geq 0$. WLOG, we assume $\lambda_i, \lambda_j, \mu_i^{\boldsymbol{x}}, \mu_j^{\boldsymbol{x}} \geq 0$. Assume the opposite to Condition 2b), i.e., $\lambda_i \mu_j^{\boldsymbol{x}} \neq \lambda_j \mu_i^{\boldsymbol{x}}$. WLOG, we let $0 \leq \lambda_i \mu_j^{\boldsymbol{x}} < \lambda_j \mu_i^{\boldsymbol{x}}$, which implies $\lambda_j, \mu_i^{\boldsymbol{x}} > 0$. By Condition $\tilde{2}$), i.e., (51), we have $(\lambda_i - \mu_i^{\boldsymbol{x}})(\lambda_j - \mu_j^{\boldsymbol{x}}) \geq 0$. Combining with $\lambda_j, \mu_i^{\boldsymbol{x}} > 0$, we know that at least one of $\lambda_i, \mu_j^{\boldsymbol{x}}$ is nonzero. Consider the case that $\mu_j^{\boldsymbol{x}} \neq 0$. Define $\theta_i = \lambda_i/\mu_i^{\boldsymbol{x}}, \theta_j = \lambda_j/\mu_j^{\boldsymbol{x}}$, then following the assumptions of $\lambda_i\mu_j^{\boldsymbol{x}} < \lambda_j\mu_i^{\boldsymbol{x}}$ we have $\theta_i < \theta_j$. Choosing any $\theta \in (\theta_i, \theta_j)$, we have $\lambda_i < \theta\mu_i^{\boldsymbol{x}}, \lambda_j > \theta\mu_j^{\boldsymbol{x}}$. Hence, the equivalent condition for Condition $\tilde{2}$), i.e., (51), is violated for $\theta\boldsymbol{x}$. The case of $\lambda_j \neq 0$ can be proved similarly. Hence, we always have contradiction, and conclude that Condition 2b) is true.

Now it remains to prove the direction "Condition 2) $\implies$ Condition $\tilde{2}$)". Given any $\boldsymbol{x} \in \Re^l$, we let $\lambda_i, \lambda_j, \mu_i^{\boldsymbol{x}}, \mu_j^{\boldsymbol{x}} \in \Re$ and $\boldsymbol{\xi}_i, \boldsymbol{\xi}_j, \boldsymbol{\zeta}_i^{\boldsymbol{x}}, \boldsymbol{\zeta}_j^{\boldsymbol{x}} \in \operatorname{span}(\boldsymbol{U}_{\mathcal{I}})$ be constants as defined in (50). By Condition 2), we know $\lambda_i\lambda_j \geq 0, \mu_i^{\boldsymbol{x}}\mu_j^{\boldsymbol{x}} \geq 0$ and $\lambda_i\mu_j^{\boldsymbol{x}} = \lambda_j\mu_i^{\boldsymbol{x}}$. WLOG, we let $\lambda_i, \lambda_j \geq 0$. Possible realizations of the parameters are as follows.

- $\lambda_i = \lambda_j = 0$. Then either $\mu_i^{\boldsymbol{x}}, \mu_j^{\boldsymbol{x}} \geq 0$ or $\mu_i^{\boldsymbol{x}}, \mu_j^{\boldsymbol{x}} \leq 0$, it always implies $(\lambda_i - \mu_i^{\boldsymbol{x}})(\lambda_j - \mu_j^{\boldsymbol{x}}) \geq 0$.
- $\lambda_i = 0, \lambda_j > 0$ (or $\lambda_i > 0, \lambda_j = 0$). Then $\lambda_i\mu_j^{\boldsymbol{x}} = \lambda_j\mu_i^{\boldsymbol{x}} = 0$, implying $\mu_i^{\boldsymbol{x}} = 0$ (or $\mu_j^{\boldsymbol{x}} = 0$). In either case we have $(\lambda_i - \mu_i^{\boldsymbol{x}})(\lambda_j - \mu_j^{\boldsymbol{x}}) \geq 0$.
- $\lambda_i > 0, \lambda_j > 0$. Denote $\theta = \lambda_i/\mu_i^{\boldsymbol{x}} = \lambda_j/\mu_j^{\boldsymbol{x}}$, then $(\lambda_i - \mu_i^{\boldsymbol{x}})(\lambda_j - \mu_j^{\boldsymbol{x}}) = \mu_i^{\boldsymbol{x}}\mu_j^{\boldsymbol{x}}(\theta - 1)^2 \geq 0$.

So we always have $(\lambda_i - \mu_i^{\boldsymbol{x}})(\lambda_j - \mu_j^{\boldsymbol{x}}) \geq 0$, Condition $\tilde{2}$) holds. □

### C.21. Proof of Lemma 3

Consider any $\boldsymbol{z}', \boldsymbol{z}'' \in \Re^n$, we denote $a = h(\boldsymbol{z}' \wedge \boldsymbol{z}''), b = h(\boldsymbol{z}'), c = h(\boldsymbol{z}''), d = h(\boldsymbol{z}' \vee \boldsymbol{z}'')$ and $d_0 = b + c - a$. From the supermodularity of $f$ we know $b + c \leq a + d$; hence, $d_0 \leq d$. We then have

$$\phi(\boldsymbol{z}') + \phi(\boldsymbol{z}'') = u(b) + u(c) \leq u(a) + u(d_0) \leq u(a) + u(d) = \phi(\boldsymbol{z}' \wedge \boldsymbol{z}'') + \phi(\boldsymbol{z}' \vee \boldsymbol{z}''), \tag{52}$$

where the second inequality holds since $u$ is non-decreasing, and the first equality can be proved as follows. We notice that either $a \leq \min\{b, c\} \leq \max\{b, c\} \leq d_0$ (if $h$ is increasing) or $a \geq \max\{b, c\} \geq \min\{b, c\} \geq d_0$ (if $h$ is decreasing) holds; since $a + d_0 = b + c$ and $u$ is convex, we then have the first inequality in Equation (52). That proves the supermodularity of $\phi$. □

### C.22. Proof of Proposition 12

Applying Lemma 3, we have that $u\left(\boldsymbol{a}^\top\boldsymbol{x} + g(\boldsymbol{x}, \boldsymbol{z})\right)$ is supermodular in $\boldsymbol{z}$ for all $\boldsymbol{x} \in \mathcal{X}$. Hence, following Theorem 1 by treating $u\left(\boldsymbol{a}^\top\boldsymbol{x} + g(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right)$ in Problem (27) as the $g(\boldsymbol{x}, \tilde{\boldsymbol{z}})$ in Problem (2), Problem (27) can be solved equivalently by

$$\begin{aligned}
\min \quad & \boldsymbol{\nu}^\top \boldsymbol{l} \\
\text{s.t.} \quad & \boldsymbol{R}_k^\top \boldsymbol{l} \geq \sum_{i \in [2n+1]} p_i^k u\left(\boldsymbol{a}^\top\boldsymbol{x} + \boldsymbol{b}^\top\boldsymbol{y}^{k,i}\right), && k \in [K] \\
& \boldsymbol{W}\boldsymbol{x} + \boldsymbol{U}\boldsymbol{y}^{k,i} \geq \boldsymbol{V}\boldsymbol{z}^{k,i} + \boldsymbol{v}^0, && k \in [K], i \in [2n+1] \\
& \boldsymbol{l} \geq \boldsymbol{0}, \ \boldsymbol{x} \in \mathcal{X},
\end{aligned}$$

Introducing auxiliary variables $f^{k,i}$ with $f^{k,i} \geq u\left(\boldsymbol{a}^\top\boldsymbol{x} + \boldsymbol{b}^\top\boldsymbol{y}^{k,i}\right) = \max_{j \in [j]}\{c_j\left(\boldsymbol{a}^\top\boldsymbol{x} + \boldsymbol{b}^\top\boldsymbol{y}^{k,i}\right) + d_j\}$, we then obtain the equivalent reformulation as in (29). □

### C.23. Proof of Corollary 3

By the minimax Theorem in Sion (1958), in Problem (30), we can interchange the maximization over $\mathbb{P}\in\mathcal{F}$ and the minimization over $\theta\in\Re$. Hence, Problem (30) is equivalent to

$$\begin{array}{ll}\min & \theta+\sup\limits_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[u\left(\boldsymbol{a}^\top\boldsymbol{x}+g(\boldsymbol{x},\tilde{\boldsymbol{z}})-\theta\right)\right]\\ \text{s.t.} & \boldsymbol{x}\in\mathcal{X}.\end{array}$$

Its equivalent reformulation (31) can be obtained as a direct application of Proposition 12. □

### C.24. Proof of Theorem 7

We first prove the direction of "1" → "2", by contradiction. Assume "2" is false, i.e., there exists $i\in[n],\ j\in[J]$ such that $h_i^j$ has at least three pieces on $[\underline{z}_i,\overline{z}_i]$, then it suffices to show there are $f^1,f^2$ that are associated with worst-case distributions with distinct marginals.

WLOG, let $h_1^1$ be the function which at least three pieces on $[\underline{z}_1,\overline{z}_1]$. We choose functions $f^1,f^2:\Re^n\to\Re$ such that $f^1(\boldsymbol{z})=g^1(z_1), f^2(\boldsymbol{z})=g^2(z_1), \boldsymbol{z}\in\Re^n$ for some $g^1,g^2:\Re\to\Re$. Moreover, for all $j\in\{2,\ldots,J\}$, we choose $\delta_1^j$ to be sufficiently large such that $\mathbb{E}_{\mathbb{P}}\left[h_1^j(\tilde{z}_1)\right]\le\delta_1^j$ holds for all $\mathbb{P}\in\{\mathbb{P}\mid\mathbb{P}(\underline{z}_1\le\tilde{z}_1\le\overline{z}_1)=1,\mathbb{E}_{\mathbb{P}}[\tilde{z}_1]=\mu_1\}$. We then have for $i=1,2$,

$$\sup_{\mathbb{P}\in\mathcal{F}^G}\mathbb{E}_{\mathbb{P}}\left[f^i(\tilde{\boldsymbol{z}})\right]=\sup_{\mathbb{P}\in\mathcal{G}}\mathbb{E}_{\mathbb{P}}\left[g^i(\tilde{z}_1)\right]$$

where

$$\mathcal{G}=\left\{\mathbb{P}\ \left|\ \begin{array}{l}\mathbb{P}\big(\underline{z}_1\le\tilde{z}_1\le\overline{z}_1\big)=1\\ \mathbb{E}_{\mathbb{P}}[\tilde{z}_1]=\mu_1\\ \mathbb{E}_{\mathbb{P}}\big[h_1^1(\tilde{z}_1)\big]\le\delta_1^1.\end{array}\right.\right\},$$

For notational simplicity, we omit the superscript and subscript of $h$ and $\delta$, as well as the subscript of $z$ and $\mu$. That is, we consider $\sup_{\mathbb{P}\in\mathcal{G}}\mathbb{E}_{\mathbb{P}}\left[g^i(\tilde{z})\right]$ with $\mathcal{G}=\left\{\mathbb{P}\ \left|\ \begin{array}{l}\mathbb{P}\big(\underline{z}\le\tilde{z}\le\overline{z}\big)=1\\ \mathbb{E}_{\mathbb{P}}[\tilde{z}]=\mu\\ \mathbb{E}_{\mathbb{P}}\big[h(\tilde{z})\big]\le\delta\end{array}\right.\right\}$, where $h$ has at least three pieces on $[\underline{z},\overline{z}]$. Now it suffices to find $g^1,g^2:\Re\to\Re$ such that there does not exist a common worst-case distribution.

Let $J+1$ be the number of pieces of $h$ on $[\underline{z},\overline{z}]$ for some $J\ge 2$, and denote the corresponding breakpoints by $\underline{z}=z^0<\cdots<z^{J+1}=\overline{z}$. We define two functions $l^1,l^2:\Re_+\to\Re$ such that

$$\begin{aligned}&l^1(p_1)\in\left\{p_0h(z^0)+p_1h(z^1)+p_{J+1}h(z^{J+1})\mid p_0+p_1+p_{J+1}=1,p_0z^0+p_1z^1+p_{J+1}z^{J+1}=\mu\right\}\\&l^2(p_2)\in\left\{p_0h(z^0)+p_2h(z^2)+p_{J+1}h(z^{J+1})\mid p_0+p_2+p_{J+1}=1,p_0z^0+p_2z^2+p_{J+1}z^{J+1}=\mu\right\}\end{aligned}\qquad(53)$$

Notice that the sets in Equation (53) are singleton since for any given $p_1$ or $p_2$, we have unique $p_0$ and $p_{J+1}$. Therefore, the functions $l^1,l^2$ are indeed uniquely determined by Equation (53). We have two observations on $l^1,l^2$. First, $l^1(0)=l^2(0)$, and when $p_1=p_2=0$, their corresponding $p_0$ and $p_{J+1}$ (in the set defined in Equation (53)) are strictly positive. Second, both $l^1,l^2$ are continuous function, and they are also increasing function due to the convexity of $h$. By the two observations, we can find $\epsilon_1, \epsilon_2 > 0$ which are sufficiently small and such that $l^1(\epsilon_1) = l^2(\epsilon_2)$, and when $p_1 = \epsilon_1$ and $p_2 = \epsilon_2$, their corresponding $p_0$ and $p_{J+1}$ are strictly positive. Define

$$\boldsymbol{H}^1 = \begin{bmatrix} 1 & 1 & 1 \\ z^0 & z^1 & z^{J+1} \\ h\big(z^0\big) & h\big(z^1\big) & h\big(z^{J+1}\big) \end{bmatrix}, \qquad \boldsymbol{H}^2 = \begin{bmatrix} 1 & 1 & 1 \\ z^0 & z^2 & z^{J+1} \\ h\big(z^0\big) & h\big(z^2\big) & h\big(z^{J+1}\big) \end{bmatrix}.$$

We hence can find $\boldsymbol{p}^1, \boldsymbol{p}^2 \in \Re^3_{++}$ and choose $\delta \in \Re$ such that $\boldsymbol{H}^1\boldsymbol{p}^1 = \boldsymbol{H}^2\boldsymbol{p}^2 = (1, \mu, \delta)$. Let the discrete probability $\mathbb{P}^1, \mathbb{P}^2$ be with

$$\mathbb{P}^1\left(\tilde{z} = w\right) = \begin{cases} p^1_1 \text{ if } w = z^0 \\ p^1_2 \text{ if } w = z^1 \\ p^1_3 \text{ if } w = z^{J+1} \\ 0 \;\; \text{otherwise} \end{cases}, \qquad \mathbb{P}^2\left(\tilde{z} = w\right) = \begin{cases} p^2_1 \text{ if } w = z^0 \\ p^2_2 \text{ if } w = z^2 \\ p^2_3 \text{ if } w = z^{J+1} \\ 0 \;\; \text{otherwise} \end{cases}.$$

Then $\mathbb{P}^1, \mathbb{P}^2$ have the support $\mathcal{Z}^1 = \{z^0, z^1, z^{J+1}\}, \mathcal{Z}^2 = \{z^0, z^2, z^{J+1}\}$, respectively.

Consider any $i \in \{1, 2\}$. Since $h$ is piecewise linear convex, we can choose a convex function $g^i$ such that $g^i(z) = h(z)$ for $z \in \mathcal{Z}^i$ and $g^i(z) < h(z)$ for all $z \in [\underline{z}, \overline{z}] \setminus \mathcal{Z}^i$. Therefore, we have

$$\mathbb{E}_{\mathbb{P}^i}\left[g^i(z)\right] = \sum_{z \in \mathcal{Z}^i} \mathbb{P}(\tilde{z} = z) g^i(z) = \sum_{z \in \mathcal{Z}^i} \mathbb{P}(\tilde{z} = z) h(z) = \mathbb{E}_{\mathbb{P}^i}\left[h(z)\right] = \delta,$$

where the first and and third equalities are by the definition of $\mathbb{P}^i$, the second equality holds since $g^i(z) = h(z)$ when $z \in \mathcal{Z}^i$, and the last equality is due to the way we choose $\boldsymbol{p}^i$. Since $g^i(z) \le h(z)$, we have $\mathbb{E}_{\mathbb{P}}\left[g^i(z)\right] \le \mathbb{E}_{\mathbb{P}}\left[h(z)\right] \le \delta$ for any $\mathbb{P} \in \mathcal{G}$. Hence $\mathbb{P}^i$ is a worst-case distribution to $\sup_{\mathbb{P} \in \mathcal{G}} \mathbb{E}_{\mathbb{P}}\left[g^i(\tilde{z})\right]$. In what follows, we show that $\mathbb{P}^i$ is the only worst-case distribution.

We first consider any $\mathbb{P} \in \mathcal{G}$ with support $\mathcal{Z}$ such that $\mathcal{Z} \setminus \mathcal{Z}^i \neq \emptyset$, then there exists $[z', z''] \subseteq [\underline{z}, \overline{z}] \setminus \mathcal{Z}^i$ such that $\mathbb{P}(\tilde{z} \in [z', z'']) > 0$. Therefore,

$$\mathbb{E}_{\mathbb{P}}\left[g^i(z)\right] = \int_{[\underline{z}, \overline{z}]} g^i(z) d\mathbb{P} < \int_{[\underline{z}, \overline{z}]} h(z) d\mathbb{P} = \mathbb{E}_{\mathbb{P}}\left[h(z)\right] \le \delta,$$

where the first inequality follows from that $g^i(z) < h(z)$ for all $z \in [\underline{z}, \overline{z}] \setminus \mathcal{Z}^i$, the last inequality is due to $\mathbb{P} \in \mathcal{G}$. Hence, $\mathbb{P} \notin \arg\sup_{\mathbb{P} \in \mathcal{G}} \mathbb{E}_{\mathbb{P}}\left[g^i(\tilde{z})\right]$. It implies that for any $\mathbb{P}^* \in \arg\sup_{\mathbb{P} \in \mathcal{G}} \mathbb{E}_{\mathbb{P}}\left[g^i(\tilde{z})\right]$, the support of $\mathbb{P}^*$ must be a subset of $\mathcal{Z}^i$, and $\mathbb{P}^*$ can be fully characterized by a vector $\boldsymbol{p}^* \in \Re^3_+$ such that $\boldsymbol{H}^i\boldsymbol{p}^* = (1, \mu, \delta)$. Observing that $\boldsymbol{H}^i$ is invertible (due to that $h$ is not linear), $\boldsymbol{p}^\star$ is unique and is exactly the aforementioned $\boldsymbol{p}^i$. Therefore, $\mathbb{P}^i$ is the unique worst-case distribution to $\sup_{\mathbb{P} \in \mathcal{G}} \mathbb{E}_{\mathbb{P}}\left[g^i(\tilde{z})\right]$. Hence, there does not exist a common worst-case distribution to $\sup_{\mathbb{P} \in \mathcal{G}} \mathbb{E}_{\mathbb{P}}\left[g^1(\tilde{z})\right]$ and $\sup_{\mathbb{P} \in \mathcal{G}} \mathbb{E}_{\mathbb{P}}\left[g^2(\tilde{z})\right]$. "1" is false.

We next prove the direction of "2"→"1".

By strong duality,

$$\sup_{\mathbb{P} \in \mathcal{F}^G} \mathbb{E}_{\mathbb{P}}\left[f(\tilde{\boldsymbol{z}})\right] = \inf\left\{ s + \boldsymbol{\mu}^\top \boldsymbol{t} + \sum_{i=1}^{n}\sum_{j=1}^{J_i} \delta_i^j r_i^j \;\middle|\; \begin{array}{l} s + \boldsymbol{z}^\top \boldsymbol{t} + \sum_{i=1}^{n}\sum_{j=1}^{J_i} h_i^j(z_i) r_i^j \ge f(\boldsymbol{z}),\ \forall \boldsymbol{z} \in [\underline{\boldsymbol{z}}, \overline{\boldsymbol{z}}] \\ r_i^j \ge 0,\ i \in [n], j \in [J_i] \end{array} \right\}.$$

Let $\mathcal{Z} = \{\boldsymbol{z} \mid z_i \in \{z_i^1, \ldots, z_i^{S_i}\}, i \in [n]\}$ which contains all $\boldsymbol{z}$ such that each of its dimension is on the breakpoints. Then we observe that $[\underline{\boldsymbol{z}}, \overline{\boldsymbol{z}}]$ can be decomposed as $[\underline{\boldsymbol{z}}, \overline{\boldsymbol{z}}] = \cup_{i=1}^S \mathcal{Z}^i$ for some $S$ and disjoint $\mathcal{Z}^1, \ldots, \mathcal{Z}^S$ such that all $\mathcal{Z}^i$ are boxes with extreme points in $\mathcal{Z}$ and $\sum_{i=1}^n \sum_{j=1}^{J_i} h_i^j(z_i)$ are linear within each $\mathcal{Z}^i$. Together with the convexity of $f$, the dual problem is equivalent to

$$\inf \left\{ s + \boldsymbol{\mu}^\top \boldsymbol{t} + \sum_{i=1}^n \sum_{j=1}^{J_i} \delta_i^j r_i^j \;\middle|\; \begin{array}{l} s + \boldsymbol{z}^\top \boldsymbol{t} + \sum_{i=1}^n \sum_{j=1}^{J_i} h_i^j(z_i) r_i^j \geq f(\boldsymbol{z}), \ \forall \boldsymbol{z} \in \mathcal{Z} \\ r_i^j \geq 0, \ i \in [n], j \in [J_i] \end{array} \right\}.$$

Writing its dual form again, we conclude that there exists a worst-case distribution with its support as $\mathcal{Z}$. Hence, for $\sup_{\mathbb{P} \in \mathcal{F}^G} \mathbb{E}_{\mathbb{P}}[f(\tilde{\boldsymbol{z}})]$, it suffices to consider the probability distributions with support $\mathcal{Z}$.

Assuming "2" is true, i.e., $h_i^j, i \in [n], j \in [J_i]$ are piecewise linear convex functions with exactly two pieces on $[\underline{z}_i, \overline{z}_i]$, we will show "1" is true. In other words, we will show the existence of a $\mathbb{P}^* \in \arg\sup_{\mathbb{P} \in \mathcal{F}^G} \mathbb{E}_{\mathbb{P}}[f(\tilde{\boldsymbol{z}})]$ such that for any dimension $i$, $\mathbb{P}^*(\tilde{z}_i = w)$ has the structure as in "1". WLOG, we let such $i$ be $n$. Further, for notational simplicity, we drop the subscript $n$ for $\tilde{z}_n, z_n, \underline{z}_n, \overline{z}_n, \mu_n, h_n^j, \delta_n^j, J_n$. Hence, we have $\tilde{\boldsymbol{z}} = (\tilde{z}_1, \ldots, \tilde{z}_{n-1}, \tilde{z})$, $\boldsymbol{z} = (z_1, \ldots, z_{n-1}, z)$ and so on, and we will prove that $\mathbb{P}^*(\tilde{z} = w)$ has the structure as in "1".

The proof will be done by induction. Starting from the case of $J = 1$, with an approach almost the same as that in the proof for Proposition 1, we can show that $\mathbb{P}^*$ has the structure in "1". More specifically, denoting the breakpoint of $h^1$ by $\hat{z} \in (\underline{z}, \overline{z})$, then we move the probability mass on $\boldsymbol{z}$ with $z = \hat{z}$ to $\boldsymbol{z} - (\hat{z} - \underline{z})\boldsymbol{e}_n$ and $\boldsymbol{z} + (\overline{z} - \hat{z})\boldsymbol{e}_n$ until we cannot move any further. Such move will terminate at a probability distribution which has marginals in the form given by "1", and the expected value of $\mathbb{E}_{\mathbb{P}}[f(\tilde{\boldsymbol{z}})]$ is no less.

Suppose when $J = \hat{J} - 1$ for some $\hat{J} \geq 2$, we have "1" being true. We now consider the case of $J = \hat{J}$. We separately analyze the following two scenarios.

- **Scenario I**: There are distinct $i, j \in [J]$ such that $h^i, h^j$ have the same breakpoint in $(\underline{z}, \overline{z})$.

WLOG, we let $h^1, h^2$ be both with breakpoint $\hat{z} \in (\underline{z}, \overline{z})$. Define $\hat{\mathcal{G}} = \{\mathbb{P} \mid \mathbb{P}(\underline{z} \leq \tilde{z} \leq \overline{z}) = 1, \mathbb{E}_{\mathbb{P}}[\tilde{z}] = \mu\}$, $\mathcal{G}^i = \{\mathbb{P} \mid \mathbb{E}_{\mathbb{P}}[h^i(\tilde{z})] \leq \delta^i\}, \ i \in \{1, 2\}$.

If $\hat{\mathcal{G}} \cap \mathcal{G}^1 \cap \mathcal{G}^2 = \hat{\mathcal{G}} \cap \mathcal{G}^2$, then denote $\mathcal{G}'$ to be the ambiguity set obtained from $\mathcal{F}^G$ by removing the constraint on $\mathbb{E}_{\mathbb{P}}[h^1(\tilde{z})]$. We have $\mathcal{G}' = \mathcal{F}^G$, and hence $\sup_{\mathbb{P} \in \mathcal{F}^G} \mathbb{E}_{\mathbb{P}}[f(\tilde{\boldsymbol{z}})] = \sup_{\mathbb{P} \in \mathcal{G}'} \mathbb{E}_{\mathbb{P}}[f(\tilde{\boldsymbol{z}})]$. Therefore, we have a problem with $J = \hat{J} - 1$, in which case we know "1" is true by the induction assumption.

If $\hat{\mathcal{G}} \cap \mathcal{G}^1 \cap \mathcal{G}^2 \neq \hat{\mathcal{G}} \cap \mathcal{G}^2$, we next show $\hat{\mathcal{G}} \cap \mathcal{G}^1 \cap \mathcal{G}^2 = \hat{\mathcal{G}} \cap \mathcal{G}^1$.

Consider any $\mathbb{P} \in \hat{\mathcal{G}}$, we define a vector $(s_1^{\mathbb{P}}, s_2^{\mathbb{P}}, \underline{s}^{\mathbb{P}}, \hat{s}^{\mathbb{P}}, \overline{s}^{\mathbb{P}})$ which is uniquely determined by the following system of equations,

$$\begin{cases} \displaystyle\int_{z \le \hat{z}} z d\mathbb{P}(\boldsymbol{z}) = \underline{z}\underline{s}^{\mathbb{P}} + \hat{z} s_1^{\mathbb{P}} \\ \mathbb{P}(\tilde{z} \le \hat{z}) = \underline{s}^{\mathbb{P}} + s_1^{\mathbb{P}} \\ \displaystyle\int_{z > \hat{z}} z d\mathbb{P}(\boldsymbol{z}) = \hat{z} s_2^{\mathbb{P}} + \overline{z}\overline{s}^{\mathbb{P}}. \\ \mathbb{P}(\tilde{z} > \hat{z}) = s_2^{\mathbb{P}} + \overline{s}^{\mathbb{P}} \\ \hat{s}^{\mathbb{P}} = s_1^{\mathbb{P}} + s_2^{\mathbb{P}} \end{cases} \tag{54}$$

In this case, for any piecewise linear convex function with two pieces and with breakpoint at $\hat{z}$, which can be denoted by $h(z) = \begin{cases} \underline{a}z + \underline{b} \text{ if } z \le \hat{z} \\ \overline{a}z + \overline{b} \text{ if } z \ge \hat{z} \end{cases}$ where $\underline{a} < \overline{a}$, we have

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}[h(\tilde{z})] &= \int_{z \le \hat{z}} (\underline{a}z + \underline{b})\, d\mathbb{P}(\boldsymbol{z}) + \int_{z > \hat{z}} (\overline{a}z + \overline{b})\, d\mathbb{P}(\boldsymbol{z}) \\ &= \underline{a}\int_{z \le \hat{z}} z d\mathbb{P}(\boldsymbol{z}) + \underline{b}\mathbb{P}(\tilde{z} \le \hat{z}) + \overline{a}\int_{z > \hat{z}} z d\mathbb{P}(\boldsymbol{z}) + \overline{b}\mathbb{P}(\tilde{z} > \hat{z}) \\ &= \underline{a}\left(\underline{z}\underline{s}^{\mathbb{P}} + \hat{z}s_1^{\mathbb{P}}\right) + \underline{b}\left(\underline{s}^{\mathbb{P}} + s_1^{\mathbb{P}}\right) + \overline{a}\left(\hat{z}s_2^{\mathbb{P}} + \overline{z}\overline{s}^{\mathbb{P}}\right) + \overline{b}\left(s_2^{\mathbb{P}} + \overline{s}^{\mathbb{P}}\right) \\ &= \underline{s}^{\mathbb{P}} h(\underline{z}) + \hat{s}^{\mathbb{P}} h(\hat{z}) + \overline{s}^{\mathbb{P}} h(\overline{z}), \end{aligned}$$

where the third and fourth inequalities are due to (54). Moreover, by (54) we can easily have $\underline{s}^{\mathbb{P}} + \hat{s}^{\mathbb{P}} + \overline{s}^{\mathbb{P}} = 1$ and $\underline{s}^{\mathbb{P}}\underline{z} + \hat{s}^{\mathbb{P}}\hat{z} + \overline{s}^{\mathbb{P}}\overline{z} = \mu$, which imply

$$\underline{s}^{\mathbb{P}} = \frac{\overline{z} - \mu - (\overline{z} - \hat{z})\hat{s}^{\mathbb{P}}}{\overline{z} - \underline{z}}, \quad \overline{s}^{\mathbb{P}} = \frac{\mu - \underline{z} - (\hat{z} - \underline{z})\hat{s}^{\mathbb{P}}}{\overline{z} - \underline{z}}.$$

Therefore,

$$\mathbb{E}_{\mathbb{P}}[h(\tilde{z})] = \frac{\overline{z} - \mu - (\overline{z} - \hat{z})\hat{s}^{\mathbb{P}}}{\overline{z} - \underline{z}} h(\underline{z}) + \hat{s}^{\mathbb{P}} h(\hat{z}) + \frac{\mu - \underline{z} - (\hat{z} - \underline{z})\hat{s}^{\mathbb{P}}}{\overline{z} - \underline{z}} h(\overline{z}) = c^h + \Delta^h \hat{s}^{\mathbb{P}}, \tag{55}$$

where $c^h, \Delta^h$ are constants depending on $h$ but independent from $\mathbb{P}$; moreover,

$$\Delta^h = h(\hat{z}) - \left(\frac{\overline{z} - \hat{z}}{\overline{z} - \underline{z}} h(\underline{z}) + \frac{\hat{z} - \underline{z}}{\overline{z} - \underline{z}} h(\overline{z})\right) < h(\hat{z}) - h\left(\frac{\overline{z} - \hat{z}}{\overline{z} - \underline{z}}\underline{z} + \frac{\hat{z} - \underline{z}}{\overline{z} - \underline{z}}\overline{z}\right) = h(\hat{z}) - h(\hat{z}) = 0, \tag{56}$$

where the inequality follows from the convexity of $h$.

Recall that $\hat{\mathcal{G}} \cap \mathcal{G}^1 \cap \mathcal{G}^2 \neq \hat{\mathcal{G}} \cap \mathcal{G}^2$, then there exists $\mathbb{P}^o \in \left(\hat{\mathcal{G}} \cap \mathcal{G}^2\right) \setminus \mathcal{G}^1$. Therefore, consider any $\hat{\mathbb{P}} \in \hat{\mathcal{G}} \cap \mathcal{G}^1$,

$$c^{h^1} + \Delta^{h^1} \hat{s}^{\mathbb{P}^o} = \mathbb{E}_{\mathbb{P}^o}\left[h^1(\tilde{z})\right] > \delta^1 \ge \mathbb{E}_{\hat{\mathbb{P}}}\left[h^1(\tilde{z})\right] = c^{h^1} + \Delta^{h^1} \hat{s}^{\hat{\mathbb{P}}},$$

where the two equalities follow from (55), the two inequalities are due to $\mathbb{P}^o \notin \mathcal{G}^1$ and $\hat{\mathbb{P}} \in \mathcal{G}^1$. Hence, we have $\hat{s}^{\mathbb{P}^o} < \hat{s}^{\hat{\mathbb{P}}}$ since (56) results in $\Delta^{h^1} < 0$. It then implies

$$\mathbb{E}_{\hat{\mathbb{P}}}\left[h^2(\tilde{z})\right] = c^{h^2} + \Delta^{h^2} \hat{s}^{\hat{\mathbb{P}}} < c^{h^2} + \Delta^{h^2} \hat{s}^{\mathbb{P}^o} = \mathbb{E}_{\mathbb{P}^o}\left[h^2(\tilde{z})\right] \le \delta^2,$$

where the last inequality holds since $\mathbb{P}^o \in \mathcal{G}^2$. Therefore, $\hat{\mathbb{P}} \in \mathcal{G}^2$, and we have $\hat{\mathcal{G}} \cap \mathcal{G}^1 \cap \mathcal{G}^2 = \hat{\mathcal{G}} \cap \mathcal{G}^1$. Similar to the case of $\hat{\mathcal{G}} \cap \mathcal{G}^1 \cap \mathcal{G}^2 = \hat{\mathcal{G}} \cap \mathcal{G}^2$, we now can reduce the problem $\sup_{\mathbb{P} \in \mathcal{F}^G} \mathbb{E}_{\mathbb{P}}[f(\tilde{\boldsymbol{z}})]$ to one with $J = \hat{J} - 1$, and hence "1" is true by induction.

• **Scenario II**: All $h^j$, $j \in [\hat{J}]$, have distinct breakpoints in $(\underline{z}, \overline{z})$. In this case, denote the breakpoint of $h^j$ by $z^j$, $j \in [\hat{J}]$. WLOG, assume $\underline{z} = z^0 < z^1 < \cdots < z^{\hat{J}} < z^{\hat{J}+1} = \overline{z}$. Consider any $\mathbb{P} \in \mathcal{F}^G$. Denote by $p_j = \mathbb{P}(\tilde{z} = z^j)$ the marginal probability mass at $z = z^j, j = 0, \ldots, \hat{J}+1$. Recalling that we just focus on the distribution with support at the breakpoints, then the constraint $\mathbb{P} \in \mathcal{F}^G$ is equivalent to the following system,

$$\sum_{j=0}^{\hat{J}+1} p_j = 1, \tag{57a}$$

$$\sum_{j=0}^{\hat{J}+1} z^j p_j = \mu, \tag{57b}$$

$$\sum_{j=0}^{\hat{J}+1} h^i(z^j) p_j \le \delta^i, \qquad i \in [\hat{J}], \tag{57c}$$

$$p_j \ge 0, \qquad j \in \{0, \ldots, \hat{J}+1\}. \tag{57d}$$

By (57a) and (57b) we have

$$p_0 = \frac{\overline{z} - \mu - \sum_{j=1}^{\hat{J}} (\overline{z} - z^j) p_j}{\overline{z} - \underline{z}}, \ p_{\hat{J}+1} = \frac{\mu - \underline{z} - \sum_{j=1}^{\hat{J}} (z^j - \underline{z}) p_j}{\overline{z} - \underline{z}}, \tag{58}$$

which implies that $p_0, p_{\hat{J}+1}$ can be determined by $\boldsymbol{p} = (p_1, \ldots, p_{\hat{J}})$. In what follows, we simplify the constraints (57a)-(57d).

We first investigate the constraint (57c) for any given $i \in [\hat{J}]$. Since $h^i$ is convex and has breakpoints $\{\underline{z}, z^i, \overline{z}\}$, we can denote $h^i(z) = \begin{cases} h^i(z^i) - \gamma_i(z^i - z) & \text{if } z \in [\underline{z}, z^i] \\ h^i(z^i) + \xi_i(z - z^i) & \text{if } z \in [z^i, \overline{z}] \end{cases}$ for some $\gamma_i < \xi_i$. It follows that

$$\begin{aligned}
&\sum_{j=0}^{\hat{J}+1} h^i(z^j) p_j \\
&= h^i(z^i) - \gamma_i(z^i - \underline{z}) p_0 - \gamma_i \sum_{j=1}^{i} (z^i - z^j) p_j + \xi_i \sum_{j=i}^{\hat{J}} (z^j - z^i) p_j + \xi_i(\overline{z} - z^i) p_{\hat{J}+1} \\
&= h^i(z^i) - \frac{\gamma_i}{\overline{z} - \underline{z}} \left( (z^i - \underline{z}) \left( \overline{z} - \mu - \sum_{j=1}^{\hat{J}} (\overline{z} - z^j) p_j \right) + (\overline{z} - \underline{z}) \sum_{j=1}^{i} (z^i - z^j) p_j \right) \\
&\quad + \frac{\xi_i}{\overline{z} - \underline{z}} \left( (\overline{z} - z^i) \left( \mu - \underline{z} - \sum_{j=1}^{\hat{J}} (z^j - \underline{z}) p_j \right) + (\overline{z} - \underline{z}) \sum_{j=i}^{\hat{J}} (z^j - z^i) p_j \right)
\end{aligned}$$

$$
\begin{aligned}
&= h^i(z^i) + \frac{\xi_i(\overline{z}-z^i)(\mu-\underline{z})}{\overline{z}-\underline{z}} - \frac{\gamma_i(\overline{z}-\mu)(z^i-\underline{z})}{\overline{z}-\underline{z}} \\
&\quad - \frac{\gamma_i}{\overline{z}-\underline{z}} \left( \sum_{j=1}^{i} \big( (\overline{z}-\underline{z})(z^i-z^j) - (\overline{z}-z^j)(z^i-\underline{z}) \big) p_j - \sum_{j=i+1}^{\hat{J}} (\overline{z}-z^j)(z^i-\underline{z}) p_j \right) \\
&\quad + \frac{\xi_i}{\overline{z}-\underline{z}} \left( -\sum_{j=1}^{i-1} (\overline{z}-z^i)(z^j-\underline{z}) p_j + \sum_{j=i}^{\hat{J}} \big( (z^j-z^i)(\overline{z}-\underline{z}) - (\overline{z}-z^i)(z^j-\underline{z}) \big) p_j \right) \\
&= h^i(z^i) + \frac{\xi_i(\overline{z}-z^i)(\mu-\underline{z})}{\overline{z}-\underline{z}} - \frac{\gamma_i(\overline{z}-\mu)(z^i-\underline{z})}{\overline{z}-\underline{z}} \\
&\quad + \frac{\gamma_i}{\overline{z}-\underline{z}} \left( -\sum_{j=1}^{i-1} (z^i-\overline{z})(z^j-\underline{z}) p_j + (\overline{z}-z^i)(z^i-\underline{z}) p_i + \sum_{j=i+1}^{\hat{J}} (\overline{z}-z^j)(z^i-\underline{z}) p_j \right) \\
&\quad - \frac{\xi_i}{\overline{z}-\underline{z}} \left( \sum_{j=1}^{i-1} (\overline{z}-z^i)(z^j-\underline{z}) p_j + (\overline{z}-z^i)(z^i-\underline{z}) p_i - \sum_{j=i+1}^{\hat{J}} (z^j-\overline{z})(z^i-\underline{z}) p_j \right) \\
&= h^i(z^i) + \frac{\xi_i(\overline{z}-z^i)(\mu-\underline{z})}{\overline{z}-\underline{z}} - \frac{\gamma_i(\overline{z}-\mu)(z^i-\underline{z})}{\overline{z}-\underline{z}} - \frac{\xi_i-\gamma_i}{\overline{z}-\underline{z}} \sum_{j=1}^{\hat{J}} \left(\overline{z}-z^{\max\{i,j\}}\right)\left(z^{\min\{i,j\}}-\underline{z}\right) p_j.
\end{aligned}
$$

Hence the $i$-th constraint of (57c) is equivalent to

$$
\sum_{j=1}^{\hat{J}} \left(\overline{z}-z^{\max\{i,j\}}\right)\left(z^{\min\{i,j\}}-\underline{z}\right) p_j \ge d_i,
$$

where $d_i = \frac{\overline{z}-\underline{z}}{\xi_i-\gamma_i}\left(h^i(z^i) + \frac{\xi_i(\overline{z}-z^i)(\mu-\underline{z})}{\overline{z}-\underline{z}} - \frac{\gamma_i(\overline{z}-\mu)(z^i-\underline{z})}{\overline{z}-\underline{z}} - \delta^i\right)$. Denote $\lambda_j = \overline{z}-z^j, \pi_j = z^j-\underline{z}$ for all $j \in [\hat{J}]$, and let

$$
\boldsymbol{A} = \left(\lambda_{\max\{i,j\}}\pi_{\min\{i,j\}}\right)_{i,j\in[\hat{J}]} = \begin{bmatrix} \lambda_1\pi_1 & \lambda_2\pi_1 & \lambda_3\pi_1 & \cdots & \lambda_{\hat{J}}\pi_1 \\ \lambda_2\pi_1 & \lambda_2\pi_2 & \lambda_3\pi_2 & \cdots & \lambda_{\hat{J}}\pi_2 \\ \lambda_3\pi_1 & \lambda_3\pi_2 & \lambda_3\pi_3 & \cdots & \lambda_{\hat{J}}\pi_3 \\ \vdots & \vdots & \vdots & & \vdots \\ \lambda_{\hat{J}}\pi_1 & \lambda_{\hat{J}}\pi_2 & \lambda_{\hat{J}}\pi_3 & \cdots & \lambda_{\hat{J}}\pi_{\hat{J}} \end{bmatrix}.
$$

Then (57c) is equivalent to $\boldsymbol{A}\boldsymbol{p} \ge \boldsymbol{d}$, where both $\boldsymbol{A}$ and $\boldsymbol{d}$ are constants determined by $\mathcal{F}^G$.

For (57d), by (58) we have that $p_0 \ge 0$ is equivalent to $\sum_{j=1}^{\hat{J}}(\overline{z}-z^j)p_j \le \overline{z}-\mu$, and $p_{\hat{J}+1} \ge 0$ is equivalent to $\sum_{j=1}^{\hat{J}}(z^j-\underline{z})p_j \le \mu-\underline{z}$. Recalling the definition of $\boldsymbol{A}$, the constraints $p_0 \ge 0, p_{\hat{J}+1} \ge 0$ can be further reformulated as

$$
\boldsymbol{a}_1^\top \boldsymbol{p} \le b_l, \boldsymbol{a}_{\hat{J}}^\top \boldsymbol{p} \le b_u,
$$

respectively, where $b_l = (\overline{z}-\mu)\pi_1,\ b_u = (\mu-\underline{z})\lambda_{\hat{J}}$.

Therefore, $p_0, \dots, p_{\hat{J}+1}$ satisfy (57a)-(57d) if and only if $(p_0, \boldsymbol{p}, p_{\hat{J}+1}) \in \mathcal{P}$ where

$$
\mathcal{P} = \left\{ (p_0, \boldsymbol{p}, p_{\hat{J}+1}) \in \Re^{\hat{J}+2} \;\middle|\; \begin{array}{l} \boldsymbol{A}\boldsymbol{p} \ge \boldsymbol{d},\ \boldsymbol{a}_1^\top \boldsymbol{p} \le b_l,\ \boldsymbol{a}_{\hat{J}}^\top \boldsymbol{p} \le b_u,\ \boldsymbol{p} \ge \boldsymbol{0} \\ p_0 = \frac{\overline{z}-\mu}{\overline{z}-\underline{z}} - \frac{\boldsymbol{a}_1^\top \boldsymbol{p}}{(\overline{z}-\underline{z})\pi_1},\quad p_{\hat{J}+1} = \frac{\mu-\underline{z}}{\overline{z}-\underline{z}} - \frac{\boldsymbol{a}_{\hat{J}}^\top \boldsymbol{p}}{(\overline{z}-\underline{z})\lambda_{\hat{J}}} \end{array} \right\}, \tag{59}
$$

where the equalities on $p_0$ and $p_{\hat{J}+1}$ are from the equalities in (58). Note that $\mathcal{P} \neq \emptyset$ as we assume $\mathcal{F}^G \neq \emptyset$. Denote by $(\mathrm{C}_i), i \in [\hat{J}]$ the $i$-th constraint of $\boldsymbol{A}\boldsymbol{p} \geq \boldsymbol{d}$, i.e., $\boldsymbol{a}_i^\top \boldsymbol{p} \geq d_i$. We say a constraint $(\mathrm{C}_i)$ is *redundant* if the strict inequality $\boldsymbol{a}_i^\top \boldsymbol{p} > d_i$ holds for any $\boldsymbol{p} \in \hat{\mathcal{P}} = \{\boldsymbol{p} \in \mathcal{R}_+^{\hat{J}} \mid \boldsymbol{A}\boldsymbol{p} \geq \boldsymbol{d}\}$.

Consider the case that there exists $i \in [\hat{J}]$ such that $(\mathrm{C}_i)$ is redundant. WLOG, we let the redundant constraint be $(\mathrm{C}_{\hat{J}})$. In this case, we define $\mathcal{P}^o = \{\boldsymbol{p} \in \mathcal{R}_+^I \mid \boldsymbol{a}_i^\top \boldsymbol{p} \geq d_i,\ i \in [\hat{J}-1]\}$ and will show that $\hat{\mathcal{P}} = \mathcal{P}^o$. Obviously, $\hat{\mathcal{P}} \subseteq \mathcal{P}^o$ since all constraints in defining $\mathcal{P}^o$ are also used in defining $\hat{\mathcal{P}}$. We now show $\mathcal{P}^o \subseteq \hat{\mathcal{P}}$ by contradiction. Assume that there exists $\boldsymbol{p}^o \in \mathcal{P}^o \setminus \hat{\mathcal{P}}$, we have $\boldsymbol{a}_{\hat{J}}^\top \boldsymbol{p}^o < d_{\hat{J}}$. Choosing any $\boldsymbol{p} \in \hat{\mathcal{P}}$, as $(\mathrm{C}_{\hat{J}})$ is redundant, $\boldsymbol{a}_{\hat{J}}^\top \boldsymbol{p} > d_{\hat{J}}$. Therefore, we can find $\lambda \in (0,1)$ such that $\boldsymbol{p}^\lambda = \lambda \boldsymbol{p} + (1-\lambda)\boldsymbol{p}^o$ such that $\boldsymbol{a}_{\hat{J}}^\top \boldsymbol{p}^\lambda = d_{\hat{J}}$. Moreover, by $\boldsymbol{p}^o \in \mathcal{P}^o$ and $\boldsymbol{p} \in \hat{\mathcal{P}}$, we have $\boldsymbol{p}^\lambda \geq \boldsymbol{0}$ and $\boldsymbol{a}_i^\top \boldsymbol{p}^\lambda \geq d_i,\ i \in [\hat{J}-1]$. Therefore, we conclude $\boldsymbol{p}^\lambda \in \hat{\mathcal{P}}$, which is a contradiction since we assume $(\mathrm{C}_{\hat{J}})$ is redundant. Hence, $\mathcal{P}^o \subseteq \hat{\mathcal{P}}$, and it implies $\mathcal{P}^o = \hat{\mathcal{P}}$. Consequently, removing the constraint $\boldsymbol{a}_{\hat{J}}^\top \boldsymbol{p} \geq d_{\hat{J}}$ from the constraints in (59) does not change the set $\mathcal{P}$. Investigating its reformulation back to the form as constraints (57a)-(57d), we can see that now the problem of $\sup_{\mathbb{P} \in \mathcal{F}^G} \mathbb{E}_{\mathbb{P}}[f(\tilde{\boldsymbol{z}})]$ is equivalent to $\sup_{\mathbb{P} \in \mathcal{G}'} \mathbb{E}_{\mathbb{P}}[f(\tilde{\boldsymbol{z}})]$ where $\mathcal{G}'$ is the ambiguity set obtained from $\mathcal{F}^G$ by removing the constraint on $h^{\hat{J}}$. Therefore, we have a problem with $J = \hat{J} - 1$, in which case we already have "1" being true by induction.

Now it suffices to consider the case that there is no redundant constraint among $(\mathrm{C}_1), \ldots, (\mathrm{C}_{\hat{J}})$. We will prove that there exists a unique $(p_0^\star, \boldsymbol{p}^\star, p_{\hat{J}+1}^\star) \in \mathcal{P}$ with $\boldsymbol{A}\boldsymbol{p}^\star = \boldsymbol{d}$. Recall that the system $\boldsymbol{A}\boldsymbol{p} = \boldsymbol{d}$ is

$$\begin{cases} \lambda_1\pi_1 p_1 + \lambda_2\pi_1 p_2 + \lambda_3\pi_1 p_3 + \cdots + \lambda_{\hat{J}}\pi_1 p_{\hat{J}} = d_1 & (\mathrm{B}_1) \\ \lambda_2\pi_1 p_1 + \lambda_2\pi_2 p_2 + \lambda_3\pi_2 p_3 + \cdots + \lambda_{\hat{J}}\pi_2 p_{\hat{J}} = d_2 & (\mathrm{B}_2) \\ \lambda_3\pi_1 p_1 + \lambda_3\pi_2 p_2 + \lambda_3\pi_3 p_3 + \cdots + \lambda_{\hat{J}}\pi_3 p_{\hat{J}} = d_3 & (\mathrm{B}_3) \\ \vdots \\ \lambda_{\hat{J}}\pi_1 p_1 + \lambda_{\hat{J}}\pi_2 p_2 + \lambda_{\hat{J}}\pi_3 p_3 + \cdots + \lambda_{\hat{J}}\pi_{\hat{J}} p_{\hat{J}} = d_{\hat{J}} & (\mathrm{B}_{\hat{J}}) \end{cases}$$

Combining $(\mathrm{B}_1)$ and $(\mathrm{B}_2)$ we have $\pi_1 p_1 = \frac{d_1\pi_2 - d_2\pi_1}{\lambda_1\pi_2 - \lambda_2\pi_1}$. Combining $(\mathrm{B}_2)$ and $(\mathrm{B}_3)$ we obtain $\pi_2 p_2 = \frac{d_2\pi_3 - d_3\pi_2}{\lambda_2\pi_3 - \lambda_3\pi_2} - \pi_1 p_1$. Continuing the same procedure, we have

$$\begin{aligned} p_1^\star &= \frac{1}{\pi_1}\frac{d_1\pi_2 - d_2\pi_1}{\lambda_1\pi_2 - \lambda_2\pi_1} \\ p_2^\star &= \frac{1}{\pi_2}\left(\frac{d_2\pi_3 - d_3\pi_2}{\lambda_2\pi_3 - \lambda_3\pi_2} - \pi_1 p_1^\star\right) \\ p_3^\star &= \frac{1}{\pi_3}\left(\frac{d_3\pi_4 - d_4\pi_3}{\lambda_3\pi_4 - \lambda_4\pi_3} - \pi_1 p_1^\star - \pi_2 p_2^\star\right) \\ &\vdots \\ p_{\hat{J}-1}^\star &= \frac{1}{\pi_{\hat{J}-1}}\left(\frac{d_{\hat{J}-1}\pi_{\hat{J}} - d_{\hat{J}}\pi_{\hat{J}-1}}{\lambda_{\hat{J}-1}\pi_{\hat{J}} - \lambda_{\hat{J}}\pi_{\hat{J}-1}} - \pi_1 p_1^\star - \cdots - \pi_{\hat{J}-2} p_{\hat{J}-2}^\star\right) \\ p_{\hat{J}}^\star &= \frac{1}{\lambda_{\hat{J}}}\frac{\lambda_{\hat{J}-1} d_{\hat{J}} - \lambda_{\hat{J}} d_{\hat{J}-1}}{\lambda_{\hat{J}-1}\pi_{\hat{J}} - \lambda_{\hat{J}}\pi_{\hat{J}-1}} \end{aligned}$$

is the unique solution to $\boldsymbol{A}\boldsymbol{p}=\boldsymbol{d}$. $p_0^\star$ and $p_{\hat{J}+1}^\star$ can be uniquely determined by the equalities in (59). Moreover, since $\mathcal{P}\neq\emptyset$, we must have $d_1\le b_l$, $d_{\hat{J}}\le b_u$, which implies $\boldsymbol{a}_1^\top\boldsymbol{p}^\star=d_1\le b_l$, $\boldsymbol{a}_{\hat{J}}^\top\boldsymbol{p}^\star=d_{\hat{J}}\le b_u$. To see $(p_0^\star,\boldsymbol{p}^\star,p_{\hat{J}+1}^\star)\in\mathcal{P}$, it remains to prove for any $j\in[\hat{J}]$, $p_j^\star\ge 0$. We show that this must be the case, otherwise the constraint ($\mathrm{C}_j$) is redundant. Recall that by the definition of $\lambda_j$ and $\pi_j$, $j\in[\hat{J}]$, we have $\lambda_1>\cdots>\lambda_{\hat{J}}>0$ and $0<\pi_1<\cdots<\pi_{\hat{J}}$.

We first show that $p_1^\star\ge 0$, i.e., $d_1\pi_2-d_2\pi_1\ge 0$. Assume to the contrary that $d_1\pi_2<d_2\pi_1$, then

$$
\begin{aligned}
\boldsymbol{a}_1^\top\boldsymbol{p}-d_1 &= \lambda_1\pi_1p_1+\pi_1(\lambda_2p_2+\cdots+\lambda_{\hat{J}}p_{\hat{J}})-d_1\\
&\ge \lambda_1\pi_1p_1+\frac{\pi_1}{\pi_2}(d_2-\lambda_2\pi_1p_1)-d_1\\
&= \frac{\pi_1}{\pi_2}d_2+\left(\lambda_1-\frac{\pi_1}{\pi_2}\lambda_2\right)\pi_1p_1-d_1\\
&\ge \frac{\pi_1}{\pi_2}d_2-d_1>0
\end{aligned}
$$

for all $\boldsymbol{p}\in\mathcal{P}$. Here the first inequality follows from $\boldsymbol{a}_2^\top\boldsymbol{p}\ge d_2$; the second inequality holds because $\lambda_1>\lambda_2,\pi_1<\pi_2$, and the last inequality follows from the assumption $d_1\pi_2<d_2\pi_1$. Hence ($\mathrm{C}_1$) is redundant.

Next, for $p_{\hat{J}}^\star$, we show $\lambda_{\hat{J}-1}\pi_{\hat{J}}-\lambda_{\hat{J}}\pi_{\hat{J}-1}\ge 0$ by contradiction. Assume $\lambda_{\hat{J}-1}\pi_{\hat{J}}<\lambda_{\hat{J}}\pi_{\hat{J}-1}$, then similar as above we have

$$
\begin{aligned}
\boldsymbol{a}_{\hat{J}}^\top\boldsymbol{p}-d_{\hat{J}} &= \lambda_{\hat{J}}(\pi_1p_1+\cdots+\pi_{\hat{J}-1}p_{\hat{J}-1})+\lambda_{\hat{J}}\pi_{\hat{J}}p_{\hat{J}}-d_{\hat{J}}\\
&\ge \frac{\lambda_{\hat{J}}}{\lambda_{\hat{J}-1}}(d_{\hat{J}-1}-\lambda_{\hat{J}}\pi_{\hat{J}-1}p_{\hat{J}})+\lambda_{\hat{J}}\pi_{\hat{J}}p_{\hat{J}}-d_{\hat{J}}\\
&= \frac{\lambda_{\hat{J}}}{\lambda_{\hat{J}-1}}d_{\hat{J}-1}+\left(\pi_{\hat{J}}-\frac{\lambda_{\hat{J}}}{\lambda_{\hat{J}-1}}\pi_{\hat{J}-1}\right)\lambda_{\hat{J}}p_{\hat{J}}-d_{\hat{J}}\\
&\ge \frac{\lambda_{\hat{J}}}{\lambda_{\hat{J}-1}}d_{\hat{J}-1}-d_{\hat{J}}>0
\end{aligned}
$$

for all $\boldsymbol{p}\in\mathcal{P}$. Here the first inequality follows from $\boldsymbol{a}_{\hat{J}-1}^\top\boldsymbol{p}\ge d_{\hat{J}-1}$; the second inequality holds because $\lambda_{\hat{J}}<\lambda_{\hat{J}-1},\pi_{\hat{J}}>\pi_{\hat{J}-1}$, and the last inequality follows from the assumption $\lambda_{\hat{J}-1}\pi_{\hat{J}}<\lambda_{\hat{J}}\pi_{\hat{J}-1}$. Hence ($\mathrm{C}_{\hat{J}}$) is redundant.

Finally, for all $j\in\{2,\ldots,\hat{J}-1\}$, we show that $\pi_jp_j^\star=\frac{d_j\pi_{j+1}-d_{j+1}\pi_j}{\lambda_j\pi_{j+1}-\lambda_{j+1}\pi_j}-\sum_{k=1}^{j-1}\pi_kp_k^\star\ge 0$. Suppose not, i.e., $\frac{d_j\pi_{j+1}-d_{j+1}\pi_j}{\lambda_j\pi_{j+1}-\lambda_{j+1}\pi_j}<\sum_{k=1}^{j-1}\pi_kp_k^\star=\frac{d_{j-1}\pi_j-d_j\pi_{j-1}}{\lambda_{j-1}\pi_j-\lambda_j\pi_{j-1}}$. Consider any $\boldsymbol{p}\in\mathcal{P}$. We then have

$$
\begin{aligned}
\boldsymbol{a}_j^\top\boldsymbol{p}-d_j &= \lambda_j(\pi_1p_1+\cdots+\pi_jp_j)+\pi_j(\lambda_{j+1}p_{j+1}+\cdots+\lambda_{\hat{J}}p_{\hat{J}})-d_j\\
&\ge \lambda_j(\pi_1p_1+\cdots+\pi_jp_j)+\frac{\pi_j}{\pi_{j+1}}(d_{j+1}-\lambda_{j+1}(\pi_1p_1+\cdots+\pi_jp_j)-d_j\\
&= \frac{1}{\pi_{j+1}}\left(d_{j+1}\pi_j-d_j\pi_{j+1}-(\lambda_{j+1}\pi_j-\lambda_j\pi_{j+1})(\pi_1p_1+\cdots+\pi_jp_j)\right),
\end{aligned}
$$

where the inequality follows from $\boldsymbol{a}_{j+1}^\top \boldsymbol{p} \geq d_{j+1}$. Further, we also have

$$\begin{aligned}\boldsymbol{a}_j^\top \boldsymbol{p} - d_j &= \lambda_j(\pi_1 p_1 + \cdots + \pi_{j-1}p_{j-1}) + \pi_j(\lambda_j p_j + \cdots + \lambda_{\hat{J}} p_{\hat{J}}) - d_j \\ &\geq \lambda_j(\pi_1 p_1 + \cdots + \pi_{j-1}p_{j-1}) + \frac{\pi_j}{\pi_{j-1}}(d_{j-1} - \lambda_{j-1}(\pi_1 p_1 + \cdots + \pi_{j-1}p_{j-1}) - d_j \\ &= \frac{1}{\pi_{j-1}}\left(d_{j-1}\pi_j - d_j\pi_{j-1} - (\lambda_{j-1}\pi_j - \lambda_j\pi_{j-1})(\pi_1 p_1 + \cdots + \pi_{j-1}p_{j-1})\right),\end{aligned}$$

where the inequality follows from $\boldsymbol{a}_{j-1}^\top \boldsymbol{p} \geq d_{j-1}$. Define two $\Re \to \Re$ functions $\phi'(t) = \frac{1}{\pi_{j+1}}\big(d_{j+1}\pi_j - d_j\pi_{j+1} - (\lambda_{j+1}\pi_j - \lambda_j\pi_{j+1})t\big), \phi''(t) = \frac{1}{\pi_{j-1}}\big(d_{j-1}\pi_j - d_j\pi_{j-1} - (\lambda_{j-1}\pi_j - \lambda_j\pi_{j-1})t\big)$, then $\boldsymbol{a}_j^\top \boldsymbol{p} - d_j \geq \max\big\{\phi'(\pi_1 p_1 + \cdots + \pi_j p_j), \phi''(\pi_1 p_1 + \cdots + \pi_{j-1}p_{j-1})\big\}$. By defintion, $\lambda_{j+1}\pi_j - \lambda_j\pi_{j+1} < 0$, hence $\phi'$ is increasing, which implies $\phi'(\pi_1 p_1 + \cdots + \pi_j p_j) \geq \phi'(\pi_1 p_1 + \cdots + \pi_{j-1}p_{j-1})$. Thus

$$\boldsymbol{a}_j^\top \boldsymbol{p} - d_j \geq \max\left\{\phi'(\pi_1 p_1 + \cdots + \pi_{j-1}p_{j-1}), \phi''(\pi_1 p_1 + \cdots + \pi_{j-1}p_{j-1})\right\}.$$

Notice that $\phi'(t) = 0$ if and only if $t = \frac{d_j\pi_{j+1} - d_{j+1}\pi_j}{\lambda_j\pi_{j+1} - \lambda_{j+1}\pi_j}$. Together with that $\phi'(t)$ is increasing, we have that $\phi'(t) > 0$ if $t > \frac{d_j\pi_{j+1} - d_{j+1}\pi_j}{\lambda_j\pi_{j+1} - \lambda_{j+1}\pi_j}$. Similarly, since $\lambda_{j-1}\pi_j - \lambda_j\pi_{j-1} > 0$, $\phi''$ is decreasing, and we obtain $\phi''(t) > 0$ if $t < \frac{d_{j-1}\pi_j - d_j\pi_{j-1}}{\lambda_{j-1}\pi_j - \lambda_j\pi_{j-1}}$. By assumption we have $\frac{d_j\pi_{j+1} - d_{j+1}\pi_j}{\lambda_j\pi_{j+1} - \lambda_{j+1}\pi_j} < \frac{d_{j-1}\pi_j - d_j\pi_{j-1}}{\lambda_{j-1}\pi_j - \lambda_j\pi_{j-1}}$, therefore we can find some $\tau \in \left(\frac{d_j\pi_{j+1} - d_{j+1}\pi_j}{\lambda_j\pi_{j+1} - \lambda_{j+1}\pi_j}, \frac{d_{j-1}\pi_j - d_j\pi_{j-1}}{\lambda_{j-1}\pi_j - \lambda_j\pi_{j-1}}\right)$ such that $\phi'(\tau) = \phi''(\tau) > 0$. Now, for all $t \in \Re$,

$$\begin{aligned}\max\{\phi'(t), \phi''(t)\} &\geq \phi'(t) > \phi'\left(\frac{d_j\pi_{j+1} - d_{j+1}\pi_j}{\lambda_j\pi_{j+1} - \lambda_{j+1}\pi_j}\right) = 0 \quad \text{if } t \geq \tau, \\ \max\{\phi'(t), \phi''(t)\} &\geq \phi''(t) > \phi'\left(\frac{d_{j-1}\pi_j - d_j\pi_{j-1}}{\lambda_{j-1}\pi_j - \lambda_j\pi_{j-1}}\right) = 0 \quad \text{if } t \leq \tau,\end{aligned}$$

which implies

$$\boldsymbol{a}_j^\top \boldsymbol{p} - d_j \geq \max\left\{\phi'(\pi_1 p_1 + \cdots + \pi_{j-1}p_{j-1}), \phi''(\pi_1 p_1 + \cdots + \pi_{j-1}p_{j-1})\right\} > 0.$$

Hence ($\mathrm{C}_j$) is redundant. We then conclude $\boldsymbol{p}^\star \geq \boldsymbol{0}$.

In summary, we have a unique $(p_0^\star, \boldsymbol{p}^\star, p_{\hat{J}+1}^\star) \in \mathcal{P}$ with $\boldsymbol{A}\boldsymbol{p}^\star = \boldsymbol{d}$.

Related to $(p_0^\star, \boldsymbol{p}^\star, p_{\hat{J}+1}^\star)$, we next prove the following observation.

**Observation:** Considering any $(p_0, \boldsymbol{p}, p_{\hat{J}+1}) \in \mathcal{P}$ with $(p_0, \boldsymbol{p}, p_{\hat{J}+1}) \neq (p_0^\star, \boldsymbol{p}^\star, p_{\hat{J}+1}^\star)$, there exists $i \in \{0, 1, \ldots, \hat{J}-1\}$, $i+1 \leq k \leq \hat{J}$, such that

1) $p_j = p_j^* \ \forall j \in \{0, \ldots, i-1\}$ if $i \geq 1$;
2) $p_i < p_i^*$;
3) $p_j = 0 \ \forall j \in \{i+1, \ldots, k-1\}$ if $k \geq i+2$;
4) $p_k > 0$;
5) $\boldsymbol{a}_k^\top \boldsymbol{p} > d_k$

Specifically, parts 1) and 2) mean that $i$ is the index of the first distinct component when comparing $(p_0, \boldsymbol{p}, p_{\hat{J}+1})$ and $(p_0^\star, \boldsymbol{p}^\star, p_{\hat{J}+1}^\star)$; parts 3) and 4) mean that $k$ is the index of the first nonzero component in $(p_0, \boldsymbol{p}, p_{\hat{J}+1})$ after $p_i$.

To prove parts 1) and 2), we consider any $i \in \{0, \dots, \hat{J}-1\}$, and have

$$\begin{aligned}
&\frac{\overline{z}-\mu}{\overline{z}-\underline{z}} - \frac{\boldsymbol{a}_{i+1}^\top \boldsymbol{p}}{(\overline{z}-\underline{z})\pi_{i+1}} \\
&= \frac{\overline{z}-\mu}{\overline{z}-\underline{z}} - \frac{1}{\overline{z}-\underline{z}} \left( \sum_{j=i+1}^{\hat{J}} \lambda_j p_j + \frac{1}{\pi_{i+1}} \left( \lambda_{i+1}\pi_1 p_1 + \cdots + \lambda_{i+1}\pi_i p_i \right) \right) \\
&= \frac{\overline{z}-\mu}{\overline{z}-\underline{z}} - \frac{1}{\overline{z}-\underline{z}} \left( \sum_{j=1}^{\hat{J}} \lambda_j p_j + \frac{1}{\pi_{i+1}} \left( \lambda_{i+1}\pi_1 p_1 + \cdots + \lambda_{i+1}\pi_i p_i - \lambda_1 \pi_{i+1} p_1 - \cdots - \lambda_i \pi_{i+1} p_i \right) \right) \\
&= p_0 + \frac{1}{(\overline{z}-\underline{z})\pi_{i+1}} \sum_{j=1}^{i} \alpha_{i+1,j} p_j,
\end{aligned}$$

where we define $\alpha_{i+1,j} = \lambda_j \pi_{i+1} - \lambda_{i+1}\pi_j > 0$ for all $j \le i$ since in this case $\lambda_j > \lambda_{i+1}$ and $\pi_j < \pi_{i+1}$. Hence,

$$\boldsymbol{a}_{i+1}^\top \boldsymbol{p} = (\overline{z}-\mu)\pi_{i+1} - (\overline{z}-\underline{z})\pi_{i+1} p_0 - \sum_{j=1}^{i} \alpha_{i+1,j} p_j. \tag{60}$$

Consider $i = 0$, by (60) we have

$$p_0 = \frac{\overline{z}-\mu}{\overline{z}-\underline{z}} - \frac{\boldsymbol{a}_1^\top \boldsymbol{p}}{(\overline{z}-\underline{z})\pi_1} \le \frac{\overline{z}-\mu}{\overline{z}-\underline{z}} - \frac{d_1}{(\overline{z}-\underline{z})\pi_1} = \frac{\overline{z}-\mu}{\overline{z}-\underline{z}} - \frac{\boldsymbol{a}_1^\top \boldsymbol{p}^\star}{(\overline{z}-\underline{z})\pi_1} = p_0^\star,$$

where the first inequality is due to $\boldsymbol{A}\boldsymbol{p} \ge \boldsymbol{d}$, the second equality follows from $\boldsymbol{A}\boldsymbol{p}^\star = \boldsymbol{d}$ and the last equality holds since Equation (60) also applies to $(p_0^\star, \boldsymbol{p}^\star, p_{\hat{J}+1}^\star)$. Hence, if $p_0 \ne p_0^\star$, we must have $p_0 < p_0^*$. Now, consider the case where $p_0 = p_0^\star$, we then denote $i \ge 1$ as the index of the first distinct component, i.e., $p_j = p_j^\star$ for all $j \in \{0, \dots, i-1\}$, and $p_i \ne p_i^\star$. Note that $i \le \hat{J}-1$, otherwise the only distinct components are the last two dimension, i.e., the marginal masses at $z^{\hat{J}}$ and $z^{\hat{J}+1}$, which is impossible since $(p_0, \boldsymbol{p}, p_{\hat{J}+1})$ and $(p_0^\star, \boldsymbol{p}^\star, p_{\hat{J}+1}^\star)$ correspond to the same mean. As $i \le \hat{J}-1$, by (60),

$$\begin{aligned}
\boldsymbol{a}_{i+1}^\top \boldsymbol{p} &= (\overline{z}-\mu)\pi_{i+1} - (\overline{z}-\underline{z})\pi_{i+1} p_0^\star - \textstyle\sum_{j=1}^{i} \alpha_{i+1,j} p_j^\star + \alpha_{i+1,i}(p_i^\star - p_i) \\
&= \boldsymbol{a}_{i+1}^\top \boldsymbol{p}^\star + \alpha_{i+1,i}(p_i^\star - p_i) \\
&= d_{i+1} + \alpha_{i+1,i}(p_i^\star - p_i) \\
&\le \boldsymbol{a}_{i+1}^\top \boldsymbol{p} + \alpha_{i+1,i}(p_i^\star - p_i),
\end{aligned}$$

which implies $p_i < p_i^\star$ since $p_i \ne p_i^\star$. Therefore, parts 1) and 2) in **Observation** are proved.

Parts 3) and 4) in **Observation** are straightfoward. Specifically,

$$\sum_{j=i+1}^{\hat{J}+1} p_j = 1 - \sum_{j=0}^{i} p_j = 1 - \sum_{j=0}^{i} p_j^\star + (p_i^\star - p_i) \ge p_i^\star - p_i > 0.$$

Hence, there must be a nonzero component in $p_{i+1}, \dots, p_{\hat{J}+1}$. We then just let $k$ be the index of the first nonzero component, parts 3) and 4) in **Observation** are proved.

Part 5) can be proved by the adoption of (60), which leads to

$$\boldsymbol{a}_k^\top \boldsymbol{p} = (\overline{z}-\mu)\pi_k - (\overline{z}-\underline{z})\pi_k p_0 - \sum_{j=1}^{k-1}\alpha_{k,j}p_j > (\overline{z}-\mu)\pi_k - (\overline{z}-\underline{z})\pi_k p_0^\star - \sum_{j=1}^{k-1}\alpha_{k,j}p_j^\star = d_k.$$

Here the inequality is due to parts 1) and 2), and $0 = p_j \le p_j^\star$ for all $j \in \{i+1,\dots,k-1\}$.

Now, base on **Observation**, we prove "1" is true by proposing a process to construct new distribution. Given any $\mathbb{P}\in\mathcal{F}^G$, let the associated marginals on $\tilde{z}$ at $z^0,\dots,z^{\hat{J}+1}$ be $(p_0,\boldsymbol{p},p_{\hat{J}+1})$. Consider the case where $(p_0,\boldsymbol{p},p_{\hat{J}+1}) \ne (p_0^\star,\boldsymbol{p}^\star,p_{\hat{J}+1}^\star)$. We now construct a new probability distribution $\mathbb{P}'$ with support only at the breakpoints and defined as

$$\mathbb{P}'(\tilde{\boldsymbol{z}}=\boldsymbol{z}) = \begin{cases} \mathbb{P}(\tilde{\boldsymbol{z}}=\boldsymbol{z}) & \text{if } z\notin\{z^{k-1},z^k,z^{k+1}\} \\ (1-\theta)\mathbb{P}(\tilde{\boldsymbol{z}}=\boldsymbol{z}) & \text{if } z=z^k \\ \mathbb{P}(\tilde{\boldsymbol{z}}=\boldsymbol{z}) + \frac{z^{k+1}-z^k}{z^{k+1}-z^{k-1}}\theta\mathbb{P}(\tilde{\boldsymbol{z}}=\boldsymbol{z}+(z^k-z^{k-1})\boldsymbol{e}_n) & \text{if } z=z^{k-1} \\ \mathbb{P}(\tilde{\boldsymbol{z}}=\boldsymbol{z}) + \frac{z^k-z^{k-1}}{z^{k+1}-z^{k-1}}\theta\mathbb{P}(\tilde{\boldsymbol{z}}=\boldsymbol{z}-(z^{k+1}-z^k)\boldsymbol{e}_n) & \text{if } z=z^{k+1} \end{cases} \tag{61}$$

for some $\theta\in(0,1)$. Intuitively, for all $z_1,\dots,z_{n-1}$, we move $\theta$ portion of the probability mass at $(z_1,\dots,z_{n-1},z^k)$ to $(z_1,\dots,z_{n-1},z^{k-1})$ and $(z_1,\dots,z_{n-1},z^{k+1})$, keeping the mean unchanged. Hence $\mathbb{P}'$ has the same marginal for $(\tilde{z}_1,\dots,\tilde{z}_{n-1})$ as $\mathbb{P}$. Denote the marginal of $\mathbb{P}'$ on $\tilde{z}$ by $p_0',\boldsymbol{p}',p_{\hat{J}+1}'$ such that $\mathbb{P}'(\tilde{z}=z^j)=p_j'$ for all $j=0,\dots,\hat{J}+1$. By (61),

$$\begin{cases} p_j' = p_j, & \forall j\notin\{k-1,k,k+1\}, \\ p_k' = p_k - \theta p_k, \\ p_{k-1}' = p_{k-1} + \frac{z^{k+1}-z^k}{z^{k+1}-z^{k-1}}\theta p_k, \\ p_{k+1}' = p_{k+1} + \frac{z^k-z^{k-1}}{z^{k+1}-z^{k-1}}\theta p_k. \end{cases}$$

There are three properties of $\mathbb{P}'$.

**(P1)** $\mathbb{E}_{\mathbb{P}'}[f(\tilde{\boldsymbol{z}})] \ge \mathbb{E}_{\mathbb{P}}[f(\tilde{\boldsymbol{z}})]$. This is because $\mathbb{E}_{\mathbb{P}'}[f(\tilde{\boldsymbol{z}})] - \mathbb{E}_{\mathbb{P}}[f(\tilde{\boldsymbol{z}})]$ equals

$$\sum_{z_i\in\left\{z_i^1,\dots,z_i^{S_i}\right\},i\in[n-1]} \theta\mathbb{P}\left(\tilde{\boldsymbol{z}}=(z_1,\dots,z_{n-1},z^k)\right)\left(\frac{z^{k+1}-z^k}{z^{k+1}-z^{k-1}}f(z_1,\dots,z_{n-1},z^{k-1}) + \frac{z^k-z^{k-1}}{z^{k+1}-z^{k-1}}f(z_1,\dots,z_{n-1},z^{k+1}) - f(z_1,\dots,z_{n-1},z^k)\right),$$

which is nonnegative since $f$ is convex.

**(P2)** $\boldsymbol{a}_j^\top\boldsymbol{p}' = \boldsymbol{a}_j^\top\boldsymbol{p}$ for all $j\ne k$ and $\boldsymbol{a}_k^\top\boldsymbol{p}' < \boldsymbol{a}_k^\top\boldsymbol{p}$. To see this, for any $j\in[\hat{J}]$, we observe

$$\mathbb{E}_{\mathbb{P}'}\left[h^j(\tilde{z})\right] - \mathbb{E}_{\mathbb{P}}\left[h^j(\tilde{z})\right] = \theta p_k\left(\frac{z^{k+1}-z^k}{z^{k+1}-z^{k-1}}h^j(z^{k-1}) + \frac{z^k-z^{k-1}}{z^{k+1}-z^{k-1}}h^j(z^{k+1}) - h^j(z^k)\right) \ge 0, \tag{62}$$

where the inequality is due to the convexity of $h$. Moreover, the "$\ge$" takes "$=$" if $j\ne k$ since $h^j$ is linear on $[z^{k-1},z^{k+1}]$ for such $j$; by contrast, "$\ge$" becomes "$>$" for $j=k$ since $h^k$ has a breakpoint at $z^k$. Therefore, by the definition of $\boldsymbol{A}$, this property is proved.

**(P3)** $\boldsymbol{a}_k^\top\boldsymbol{p}'$ is continuously decreasing in $\theta$, which is implied by (62) and the definition of $\boldsymbol{A}$.

Based on the **Observation** and **(P1)-(P3)**, given any $\mathbb{P} \in \mathcal{F}^G$ whose marginal on $\tilde{z}$ is different from $(p_0^\star, \boldsymbol{p}^\star, p_{\hat{J}+1}^\star)$, we can use the procedure as in (61) to construct a new probability distribution $\mathbb{P}'$. In this construction, we either choose $\theta = 1$ or the maximal value less than 1 such that $\boldsymbol{a}_k^\top \boldsymbol{p}'$ drops to the value of $d_k$ (note that when $\theta = 0$, $\boldsymbol{a}_k^\top \boldsymbol{p}' = \boldsymbol{a}_k^\top \boldsymbol{p} > d_k$, where the inequality is due to the part 5) in **Observation**). Hence, $\mathbb{P}' \in \mathcal{F}^G$. Moreover, by **(P1)**, with $\mathbb{P}'$, the expectation of $f(\tilde{\boldsymbol{z}})$ is no less. Therefore, for any $\mathbb{P} \in \mathcal{F}^G$, by this procedure we construct a new probability distribution $\mathbb{P}' \in \mathcal{F}^G$ such that the objective is improved and the marginal masses after $z^i$ is moved towards $z^i$, the smallest breakpoint where the marginal mass of $\mathbb{P}$ differs from $(p_0^\star, \boldsymbol{p}^\star, p_{\hat{J}+1}^\star)$. Repeating such process, the margin converges to $(p_0^\star, \boldsymbol{p}^\star, p_{\hat{J}+1}^\star)$. We hence conclude that there must be a worst-case distribution whose $n$-th marginal is $(p_0^\star, \boldsymbol{p}^\star, p_{\hat{J}+1}^\star)$. □

## References

Ben-Tal, Aharon, Eithan Hochman. 1972. More bounds on the expectation of a convex function of a random variable. *Journal of Applied Probability* **9**(4) 803-812.

Ben-Tal, Aharon, Marc Teboulle. 1986. Expected utility, penalty functions, and duality in stochastic nonlinear programming. *Management Science* **32**(11) 1445-1466.

Ben-Tal, Aharon, Marc Teboulle. 2007. An old-new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance* **17**(3) 449-476.

Bertsimas, Dimitris, Vineet Goyal. 2012. On the power and limitations of affine policies in two-stage adaptive optimization. *Mathematical Programming* **134**(2) 491-531.

Birge, John R, Francois Louveaux. 2011. *Introduction to Stochastic Programming*. Springer Science & Business Media.

Chen, Xin, Daniel Zhuoyu Long, Jin Qi. 2021. Preservation of supermodularity in parametric optimization: Necessary and sufficient conditions on constraint structures. *Operations Research* **69**(1) 1-12.

Shapiro, Alexander. 2001. On duality theory of conic linear problems. *Semi-infinite programming*. Springer, 135-165.

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

von Neumann, John. 1928. Zur theorie der gesellschaftsspiele. *Mathematische Annalen* **100**(1) 295-320.