

## 7. Supplementary Material

### 7.1. Proofs

**Proof of Proposition 1.** Consider 1. Let  $(F, G) \in \mathcal{F} \times \mathcal{G}$  satisfy (1)-(3). Since  $\mathcal{R}_k$  is the worst-case risk measure, one has for each  $\omega \in \Omega$

$$\psi_k(\xi(\omega), F(\xi(\omega)), \mathbb{H}(G(\xi(\omega)))) \leq 0, \quad k = 1, \dots, q.$$

Since  $(\bar{F}(\xi(\omega)), \mathbb{H}(\bar{G}(\xi(\omega))))$  is a minimizer of  $(\text{AP})(\xi(\omega))$  and  $(F(\xi(\omega)), \mathbb{H}(G(\xi(\omega))))$  is feasible in the same problem, one has  $(\bar{F}(\xi(\omega)), \mathbb{H}(\bar{G}(\xi(\omega)))) \in C$  and

$$\begin{aligned} & \psi_0(\xi(\omega), \bar{F}(\xi(\omega)), \mathbb{H}(\bar{G}(\xi(\omega)))) \\ & \leq \psi_0(\xi(\omega), F(\xi(\omega)), \mathbb{H}(G(\xi(\omega)))) \\ & \psi_k(\xi(\omega), \bar{F}(\xi(\omega)), \mathbb{H}(\bar{G}(\xi(\omega)))) \leq 0, \quad k = 1, \dots, q, \end{aligned}$$

with these conditions holding for each  $\omega \in \Omega$ . By monotonicity of  $\mathcal{R}_0, \mathcal{R}_1, \dots, \mathcal{R}_q$ , this implies that

$$\begin{aligned} & \mathcal{R}_0(\psi_0(\xi, \bar{F}(\xi), \mathbb{H}(\bar{G}(\xi)))) \\ & \leq \mathcal{R}_0(\psi_0(\xi, F(\xi), \mathbb{H}(G(\xi)))) \\ & \mathcal{R}_k(\psi_k(\xi, \bar{F}(\xi), \mathbb{H}(\bar{G}(\xi)))) \leq 0, \quad k = 1, \dots, q. \end{aligned}$$

Thus,  $(\bar{F}, \bar{G})$  is feasible in (TP) and, in fact, a minimizer because  $(F, G)$  is chosen arbitrarily.

Next, consider 2. In view of the fact that  $\mathcal{R}_1, \dots, \mathcal{R}_q$  are worst-case risk measures, the decision  $(F^*(\xi(\omega)), \mathbb{H}(G^*(\xi(\omega))))$  is feasible for  $(\text{AP})(\xi(\omega))$  regardless of  $\omega \in \Omega$ . Since  $(\bar{F}(\xi(\omega)), \mathbb{H}(\bar{G}(\xi(\omega))))$  is a minimizer of  $(\text{AP})(\xi(\omega))$ , this implies that

$$\begin{aligned} & \psi_0(\xi(\omega), \bar{F}(\xi(\omega)), \mathbb{H}(\bar{G}(\xi(\omega)))) \\ & \leq \psi_0(\xi(\omega), F^*(\xi(\omega)), \mathbb{H}(G^*(\xi(\omega)))) \quad \forall \omega \in \Omega. \end{aligned} \tag{16}$$

We observe that  $(\bar{F}, \bar{G})$  is feasible in (TP) as argued above. This fact together with the optimality of  $(F^*, G^*)$  in (TP), with  $\mathcal{R}_0$  being the expectation, imply that

$$\begin{aligned} & \sum_{\omega \in \Omega} P(\omega) \psi_0(\xi(\omega), F^*(\xi(\omega)), \mathbb{H}(G^*(\xi(\omega)))) \\ & \leq \sum_{\omega \in \Omega} P(\omega) \psi_0(\xi(\omega), \bar{F}(\xi(\omega)), \mathbb{H}(\bar{G}(\xi(\omega)))) \end{aligned}$$

Since  $P(\omega) > 0$  for all  $\omega \in \Omega$ , the inequalities in (16) must hold with equality. We conclude that  $(F^*(\xi(\omega)), \mathbb{H}(G^*(\xi(\omega))))$  is actually a minimizer of  $(\text{AP})(\xi(\omega))$  for each  $\omega \in \Omega$ .  $\square$

**Proof of Theorem 1.** The argument leverages epi-convergence. We define  $\phi^\nu, \phi : \mathcal{F} \times \mathcal{G} \rightarrow [-\infty, \infty]$  by setting

$$\begin{aligned}
 \phi^\nu(F, G) &= \mathcal{R}_0\left(\psi_0^\nu(\xi^\nu, F(\xi^\nu), \mathbb{H}(G(\xi^\nu)))\right) \\
 &\quad + h^\nu(F, G) \\
 &\quad + \sum_{k=1}^q \iota_{(-\infty, 0]}\left(\mathcal{R}_k\left(\psi_k^\nu(\xi^\nu, F(\xi^\nu), \mathbb{H}(G(\xi^\nu)))\right)\right) \\
 &\quad + \sum_{\omega \in \Omega} \iota_C\left(F(\xi^\nu(\omega)), \mathbb{H}(G(\xi^\nu(\omega)))\right) \\
 &\quad + \sum_{\omega \in \Omega} \iota_{D_\epsilon(\delta^\nu)}\left(G(\xi^\nu(\omega))\right) \\
 \phi(F, G) &= \mathcal{R}_0\left(\psi_0(\bar{\xi}, F(\bar{\xi}), \mathbb{H}(G(\bar{\xi})))\right) + h(F, G) \\
 &\quad + \sum_{k=1}^q \iota_{(-\infty, 0]}\left(\mathcal{R}_k\left(\psi_k(\bar{\xi}, F(\bar{\xi}), \mathbb{H}(G(\bar{\xi})))\right)\right) \\
 &\quad + \sum_{\omega \in \Omega} \iota_C\left(F(\bar{\xi}(\omega)), \mathbb{H}(G(\bar{\xi}(\omega)))\right) \\
 &\quad + \sum_{\omega \in \Omega} \iota_{D_\epsilon(\delta)}\left(G(\bar{\xi}(\omega))\right),
 \end{aligned} \tag{17}$$

where for any set  $D$ ,  $\iota_D(x) = 0$  when  $x \in D$  and  $\iota_D(x) = \infty$  otherwise. We recall that  $\phi^\nu$  epi-converges to  $\phi$  provided that (see, e.g., Roysset (2018))

$$\begin{aligned}
 \forall (F^\nu, G^\nu) \rightarrow (F, G) \\
 \liminf \phi^\nu(F^\nu, G^\nu) \geq \phi(F, G)
 \end{aligned} \tag{18}$$

$\forall (F, G) \in \mathcal{F} \times \mathcal{G}$ ,  $\exists (F^\nu, G^\nu) \rightarrow (F, G)$  such that

$$\limsup \phi^\nu(F^\nu, G^\nu) \leq \phi(F, G). \tag{19}$$

We first establish (18). Suppose that  $(F^\nu, G^\nu) \rightarrow (F, G)$ . Thus,  $\max_{\xi \in \Xi} \|F^\nu(\xi) - F(\xi)\|_2 \rightarrow 0$ . As a consequence of  $\xi^\nu \rightarrow \bar{\xi}$ , one has  $\xi^\nu(\omega) \in \Xi \rightarrow \bar{\xi}(\omega)$  for each  $\omega \in \Omega$ , with the right-hand side being equal to  $\bar{\xi} \in \Xi$ . This implies that  $F^\nu(\xi^\nu(\omega)) \rightarrow F(\bar{\xi})$  for each  $\omega \in \Omega$ . A similar argument establishes that  $G^\nu(\xi^\nu(\omega)) \rightarrow G(\bar{\xi})$  for all  $\omega \in \Omega$ .

If  $(F^\nu(\xi^\nu(\omega)), \mathbb{H}(G^\nu(\xi^\nu(\omega)))) \notin C$  for some  $\omega \in \Omega$ , then  $\phi^\nu(F^\nu, G^\nu) = \infty$ . Similarly, if  $G^\nu(\xi^\nu(\omega)) \notin D_\epsilon(\delta^\nu)$  for some  $\omega \in \Omega$ , then  $\phi^\nu(F^\nu, G^\nu) = \infty$ . We can therefore assume without loss of generality that  $(F^\nu(\xi^\nu(\omega)), \mathbb{H}(G^\nu(\xi^\nu(\omega)))) \in C$  and  $G^\nu(\xi^\nu(\omega)) \in D_\epsilon(\delta^\nu)$  for every  $\omega \in \Omega$  and  $\nu \in \mathbb{N}$  in verification of (18). This in turn implies that  $G(\bar{\xi}(\omega)) \in D_\epsilon(\delta)$  for every  $\omega \in \Omega$  because  $\delta^\nu \rightarrow \delta$ . Hence, there is  $\bar{\nu}$  such that  $\mathbb{H}(G^\nu(\xi^\nu(\omega))) = \mathbb{H}(G(\bar{\xi}(\omega)))$  for all  $\nu \geq \bar{\nu}$  and  $\omega \in \Omega$ . Since  $F^\nu(\xi^\nu(\omega)) \rightarrow F(\bar{\xi}(\omega))$  and  $C$  is closed, we also have  $(F(\bar{\xi}(\omega)), \mathbb{H}(G(\bar{\xi}(\omega)))) \in C$  for all  $\omega \in \Omega$ . Moreover, by assumption 2,

$$\begin{aligned} & \psi_k^\nu(\xi^\nu(\omega), F^\nu(\xi^\nu(\omega)), \mathbb{H}(G(\xi^\nu(\omega)))) \rightarrow \\ & \psi_k(\bar{\xi}(\omega), F(\bar{\xi}(\omega)), \mathbb{H}(G(\bar{\xi}(\omega)))) \end{aligned}, \quad k = 0, 1, \dots, q,$$

and this holds for each  $\omega \in \Omega$ . Thus, the random variables  $\psi_k^\nu(\xi^\nu, F^\nu(\xi^\nu), \mathbb{H}(G(\xi^\nu)))$  converge pointwise to  $\psi_k(\bar{\xi}, F(\bar{\xi}), \mathbb{H}(G(\bar{\xi})))$ ,  $k = 0, 1, \dots, q$ . The continuity of  $\mathcal{R}_0, \mathcal{R}_1, \dots, \mathcal{R}_q$  implies that

$$\begin{aligned} & \mathcal{R}_0(\psi_0^\nu(\xi^\nu, F^\nu(\xi^\nu), \mathbb{H}(G(\xi^\nu)))) \rightarrow \\ & \mathcal{R}_0(\psi_0(\bar{\xi}, F(\bar{\xi}), \mathbb{H}(G(\bar{\xi})))) \end{aligned}$$

and, for  $k = 1, \dots, q$ ,

$$\begin{aligned} & \liminf_{\nu \rightarrow \infty} \mathcal{R}_k(\psi_k^\nu(\xi^\nu, F^\nu(\xi^\nu), \mathbb{H}(G(\xi^\nu)))) \\ & \geq \mathcal{R}_k(\psi_k(\bar{\xi}, F(\bar{\xi}), \mathbb{H}(G(\bar{\xi})))) \end{aligned}$$

Since none of the terms in the sum defining  $\phi(F, G)$  equals  $-\infty$  and, by assumption 4,  $\liminf h^\nu(F^\nu, G^\nu) \geq h(F, G)$ , we obtain that

$$\begin{aligned}
& \liminf \phi^\nu(F^\nu, G^\nu) \\
& \geq \liminf \mathcal{R}_0\left(\psi_0^\nu(\xi^\nu, F^\nu(\xi^\nu)), \mathbb{H}(G^\nu(\xi^\nu))\right) \\
& \quad + \liminf h^\nu(F^\nu, G^\nu) \\
& \quad + \sum_{k=1}^q \liminf_{\iota_{(-\infty, 0]}} \left( \mathcal{R}_k\left(\psi_k^\nu(\xi^\nu, F^\nu(\xi^\nu)), \right. \right. \\
& \qquad \qquad \qquad \left. \left. \mathbb{H}(G^\nu(\xi^\nu))\right) \right) \\
& \geq \mathcal{R}_0\left(\psi_0(\bar{\xi}, F(\bar{\xi})), \mathbb{H}(G(\bar{\xi}))\right) + h(F, G) \\
& \quad + \sum_{k=1}^q \liminf_{\iota_{(-\infty, 0]}} \left( \mathcal{R}_k\left(\psi_k(\bar{\xi}, F(\bar{\xi})), \mathbb{H}(G(\bar{\xi}))\right) \right) \\
& = \phi(F, G)
\end{aligned}$$

and thus (18) holds.

We next establish (19). Let  $(F, G) \in \mathcal{F} \times \mathcal{G}$ . Without loss of generality we assume that  $G(\bar{\xi}(\omega)) \in D_\epsilon(\delta)$  and  $(F(\bar{\xi}(\omega)), \mathbb{H}(G(\bar{\xi}(\omega)))) \in C$  for all  $\omega \in \Omega$  and  $\mathcal{R}_k(\psi_k(\bar{\xi}, F(\bar{\xi})), \mathbb{H}(G(\bar{\xi}))) \leq 0$  for  $k = 1, \dots, q$  because otherwise  $\phi(F, G) = \infty$  and (19) holds trivially. For each  $\nu \in \mathbb{N}$ , we construct  $G^\nu : \Xi \rightarrow \mathbb{R}^m$  by setting

$$G^\nu(\xi) = G(\xi) + b^\nu, \quad \xi \in \Xi,$$

where  $b^\nu \in \mathbb{R}^m$  is to be determined. By assumption 3,  $G^\nu \in \mathcal{G}$ . We define  $b^\nu = (b_1^\nu, \dots, b_m^\nu)$  componentwise and let

$$b_i^\nu = \begin{cases} \max_{\omega \in \Omega} \max \{0, -\delta^\nu - g_i(\xi^\nu(\omega))\} & \text{if } g_i(\bar{\xi}) = -\delta \\ \min_{\omega \in \Omega} \min \{0, -\epsilon - g_i(\xi^\nu(\omega))\} & \text{if } g_i(\bar{\xi}) = -\epsilon \\ \max_{\omega \in \Omega} \max \{0, \epsilon - g_i(\xi^\nu(\omega))\} & \text{if } g_i(\bar{\xi}) = \epsilon \\ \min_{\omega \in \Omega} \min \{0, \delta^\nu - g_i(\xi^\nu(\omega))\} & \text{if } g_i(\bar{\xi}) = \delta \\ 0 & \text{otherwise,} \end{cases}$$

where  $g_i$  is the  $i$ th component function of  $G$ . Since  $\delta^\nu \rightarrow \delta$  and  $g_i(\boldsymbol{\xi}^\nu(\omega)) \rightarrow g_i(\bar{\xi})$  for every  $\omega \in \Omega$ , one has  $b_i^\nu \rightarrow 0$ . Thus,  $\max_{\xi \in \Xi} \|G^\nu(\xi) - G(\xi)\|_2 \rightarrow 0$ . Moreover, for any  $\omega \in \Omega$ ,

$$g_i^\nu(\boldsymbol{\xi}^\nu(\omega)) = g_i(\boldsymbol{\xi}^\nu(\omega)) + b_i^\nu \in \begin{cases} [-\delta^\nu, \infty) \\ \text{if } g_i(\bar{\xi}) = -\delta \\ (-\infty, -\epsilon] \\ \text{if } g_i(\bar{\xi}) = -\epsilon \\ [\epsilon, \infty) \\ \text{if } g_i(\bar{\xi}) = \epsilon \\ (-\infty, \delta^\nu] \\ \text{if } g_i(\bar{\xi}) = \delta, \end{cases}$$

where  $g_i^\nu$  is the  $i$ th component function of  $G^\nu$ . This means that  $G^\nu(\boldsymbol{\xi}^\nu(\omega)) \in D_\epsilon(\delta^\nu)$  for sufficiently large  $\nu$  regardless of  $\omega \in \Omega$ . In fact, there exists  $\bar{\nu}$  such that  $\mathbb{H}(G^\nu(\boldsymbol{\xi}^\nu(\omega))) = \mathbb{H}(G(\bar{\xi}(\omega)))$  for all  $\nu \geq \bar{\nu}$  and  $\omega \in \Omega$ .

Next, we turn to the construction of  $F^\nu$ . Set  $\bar{y} = \mathbb{H}(G(\bar{\xi}))$ . By assumption 5, there exist sequences  $\{x^\mu \in \mathbb{R}^n, \mu \in \mathbb{N}\}$  and  $\{\gamma^\mu \in (0, \infty), \mu \in \mathbb{N}\}$  with the properties that  $x^\mu \rightarrow F(\bar{\xi})$ ,  $\gamma^\mu \rightarrow 0$ , and

$$\begin{aligned} \psi_k(\bar{\xi}, x^\mu, \bar{y}) &< 0, \quad k = 1, \dots, q, \quad \mu \in \mathbb{N} \\ (z^\mu, \bar{y}) &\in C \quad \forall z^\mu \in \mathbb{B}(x^\mu, \gamma^\mu), \quad \mu \in \mathbb{N}. \end{aligned}$$

We define  $\{\bar{F}^\mu, \mu \in \mathbb{N}\}$  by setting

$$\bar{F}^\mu(\xi) = F(\xi) - F(\bar{\xi}) + x^\mu \quad \text{for } \xi \in \Xi.$$

By assumption 3,  $\bar{F}^\mu \in \mathcal{F}$ . By construction,  $\max_{\xi \in \Xi} \|\bar{F}^\mu(\xi) - F(\xi)\|_\infty \rightarrow 0$  as  $\mu \rightarrow \infty$  and, for each  $\mu \in \mathbb{N}$ , one has

$$\begin{aligned} \psi_k(\bar{\xi}, \bar{F}^\mu(\bar{\xi}), \bar{y}) &< 0, \quad k = 1, \dots, q \\ (z^\mu, \bar{y}) &\in C \quad \forall z^\mu \in \mathbb{B}(\bar{F}^\mu(\bar{\xi}), \gamma^\mu). \end{aligned}$$

Let  $\bar{\nu}_0 = \bar{\nu}$ . For each fixed  $\mu \in \mathbb{N}$ , there exists  $\bar{\nu}_\mu > \bar{\nu}_{\mu-1}$  such that for all  $\nu \geq \bar{\nu}_\mu$  one has

$$\begin{aligned} \psi_k^\nu(\boldsymbol{\xi}^\nu(\omega), \bar{F}^\mu(\boldsymbol{\xi}^\nu(\omega)), \bar{y}) &\leq 0 \quad \forall \omega \in \Omega, \quad k = 1, \dots, q \\ (\bar{F}^\mu(\boldsymbol{\xi}^\nu(\omega)), \bar{y}) &\in C \quad \forall \omega \in \Omega. \end{aligned}$$

We recall that  $\bar{F}^\mu(\boldsymbol{\xi}^\nu(\omega)) \rightarrow \bar{F}^\mu(\bar{\xi})$  as  $\nu \rightarrow \infty$ . Thus, the inequality holds for sufficiently large  $\nu$  because of assumption 2. The inclusion in  $C$  must hold for sufficiently large  $\nu$  because  $\|\bar{F}^\mu(\boldsymbol{\xi}^\nu(\omega)) - \bar{F}^\mu(\bar{\xi})\|_2 \leq \gamma^\mu$

for sufficiently large  $\nu$ . We then construct  $\{F^\nu\}_{\nu=\bar{\nu}_1}^\infty$  by setting  $F^\nu = \bar{F}^\mu$  for  $\bar{\nu}_{\mu+1} > \nu \geq \bar{\nu}_\mu$ ,  $\mu \in \mathbb{N}$ . By construction,  $(F^\nu(\xi^\nu(\omega)), \bar{y}) \in C$  for all  $\omega \in \Omega$  and  $\nu \geq \bar{\nu}_1$  and  $\max_{\xi \in \Xi} \|F^\nu(\xi) - F(\xi)\|_\infty \rightarrow 0$ . Moreover, because  $\mathcal{R}_1, \dots, \mathcal{R}_q$  are monotone by assumption 1, we find that for  $\nu \geq \bar{\nu}_1$ ,

$$\mathcal{R}_k\left(\psi_k^\nu(\xi^\nu, F^\nu(\xi^\nu), \bar{y})\right) \leq 0, \quad k = 1, \dots, q.$$

We have constructed  $(F^\nu, G^\nu) \rightarrow (F, G)$  with the property that

$$\begin{aligned} & \limsup \phi^\nu(F^\nu, G^\nu) \\ & \leq \limsup \mathcal{R}_0\left(\psi_0^\nu(\xi^\nu, F^\nu(\xi^\nu), \mathbb{H}(G^\nu(\xi^\nu)))\right) \\ & + \limsup h^\nu(F^\nu, G^\nu). \end{aligned} \tag{20}$$

Since  $\psi_0^\nu(\xi^\nu(\omega), F^\nu(\xi^\nu(\omega)), \bar{y}) \rightarrow \psi_0(\bar{\xi}(\omega), F(\bar{\xi}(\omega)), \bar{y})$  for all  $\omega \in \Omega$  by assumption 2, we find that the random variables

$$\begin{aligned} & \psi_0^\nu\left(\xi^\nu, F^\nu(\xi^\nu), \mathbb{H}(G^\nu(\xi^\nu))\right) \\ & \rightarrow \psi_0\left(\bar{\xi}, F(\bar{\xi}), \mathbb{H}(G(\bar{\xi}))\right). \end{aligned}$$

The continuity of  $\mathcal{R}_0$  (cf. assumption 1) ensures that

$$\begin{aligned} & \mathcal{R}_0\left(\psi_0^\nu(\xi^\nu, F^\nu(\xi^\nu), \mathbb{H}(G^\nu(\xi^\nu)))\right) \\ & \rightarrow \mathcal{R}_0\left(\psi_0(\bar{\xi}, F(\bar{\xi}), \mathbb{H}(G(\bar{\xi})))\right). \end{aligned}$$

These facts together with assumption 4 establish that (20) implies (19). We have confirmed that  $\phi^\nu$  epi-converges to  $\phi$ .

Next, let  $N \subset \mathbb{N}$  be a subsequence and  $\{(\hat{F}^\nu, \hat{G}^\nu), \nu \in \mathbb{N}\}$  be a sequence satisfying assumption 6 with  $(\hat{F}^\nu, \hat{G}^\nu) \xrightarrow{N} (F^*, G^*)$ . It is well known that

$$\begin{aligned} & \phi^\nu(\hat{F}^\nu, \hat{G}^\nu) \xrightarrow{N} \inf_{(F,G) \in \mathcal{F} \times \mathcal{G}} \phi(F, G) \\ & (F^*, G^*) \in \operatorname{argmin}_{(F,G) \in \mathcal{F} \times \mathcal{G}} \phi(F, G) \end{aligned}$$

whenever  $\inf_{(F,G) \in \mathcal{F} \times \mathcal{G}} \phi(F, G) < \infty$ ; see, e.g., (Royset 2018, Proposition 2.1) or (Royset and Wets 2021, Theorem 5.5) for a proof that extends to the present setting nearly verbatim. Directly from the definition of risk measures, we find that for any  $(F, G) \in \mathcal{F} \times \mathcal{G}$ ,

$$\begin{aligned} \phi(F, G) &= \psi_0\left(\bar{\xi}, F(\bar{\xi}), \mathbb{H}(G(\bar{\xi}))\right) + h(F, G) \\ &\quad + \sum_{k=1}^q \iota_{(-\infty, 0]} \left( \psi_k\left(\bar{\xi}, F(\bar{\xi}), \mathbb{H}(G(\bar{\xi}))\right) \right) \\ &\quad + \iota_C\left(F(\bar{\xi}), \mathbb{H}(G(\bar{\xi}))\right) + \iota_{D_\epsilon(\delta)}(G(\bar{\xi})). \end{aligned}$$

The minimum value of  $\phi$  cannot be  $\infty$  by assumption 5 and the fact that  $h$  is real-valued. Thus,  $(F^*, G^*)$  is a minimizer of  $\phi$ . Moreover,  $(F^*(\bar{\xi}), \mathbb{H}(G^*(\bar{\xi}))) \in C$ ,  $G^*(\bar{\xi}) \in D_\epsilon(\delta)$ , and

$$\psi_k\left(\bar{\xi}, F^*(\bar{\xi}), \mathbb{H}(G^*(\bar{\xi}))\right) \leq 0, \quad k = 1, \dots, q,$$

which implies that  $(F^*(\bar{\xi}), \mathbb{H}(G^*(\bar{\xi})))$  is feasible in (AP)( $\bar{\xi}$ ). Let  $(x, y) \in C$  satisfy  $\psi_k(\bar{\xi}, x, y) \leq 0$ ,  $k = 1, \dots, q$ , which then is an arbitrary feasible point in (AP)( $\bar{\xi}$ ). We construct  $F$  and  $G$  by setting

$$\begin{aligned} F(\xi) &= F^*(\xi) - F^*(\bar{\xi}) + x \\ g_i(\xi) &= g_i^*(\xi) - g_i^*(\bar{\xi}) + \epsilon(2y_i - 1), \quad i = 1, \dots, m, \end{aligned}$$

where  $y_i$  is the  $i$ th component of  $y$  and  $g_i$  is the  $i$ th component function of  $G$ . Note that  $\mathbb{H}(G(\bar{\xi})) = y$ . By assumption 3,  $F \in \mathcal{F}$  and  $G \in \mathcal{G}$ . Consequently,  $(F(\bar{\xi}), \mathbb{H}(G(\bar{\xi}))) \in C$ ,  $G(\bar{\xi}) \in D_\epsilon(\delta)$ , and

$$\psi_k\left(\bar{\xi}, F(\bar{\xi}), \mathbb{H}(G(\bar{\xi}))\right) \leq 0, \quad k = 1, \dots, q,$$

by construction. Since  $(F^*, G^*)$  is a minimizer of  $\phi$ , this implies that

$$\begin{aligned} &\psi_0\left(\bar{\xi}, F^*(\bar{\xi}), \mathbb{H}(G^*(\bar{\xi}))\right) + h(F^*, G^*) \\ &= \phi(F^*, G^*) \\ &\leq \phi(F, G) \\ &= \psi_0\left(\bar{\xi}, F(\bar{\xi}), \mathbb{H}(G(\bar{\xi}))\right) + h(F, G) \\ &= \psi_0(\bar{\xi}, x, y) + h(F, G). \end{aligned}$$

Since  $h(F^*, G^*) = h(F, G)$  by assumption 4, this implies that

$$\psi_0\left(\bar{\xi}, F^*(\bar{\xi}), \mathbb{H}(G^*(\bar{\xi}))\right) \leq \psi_0(\bar{\xi}, x, y).$$

We conclude that  $(F^*(\bar{\xi}), \mathbb{H}(G^*(\bar{\xi})))$  is optimal in  $(\text{AP})(\bar{\xi})$  because  $(x, y)$  is an arbitrary feasible point for that problem. We have already confirmed that  $G^*(\bar{\xi}) \in D_\epsilon(\delta)$ . The fact that  $\mathbb{H}(\hat{G}^\nu(\xi^\nu(\omega))) = \mathbb{H}(G^*(\bar{\xi}))$  for all  $\omega \in \Omega$  when  $\nu \in N$  is sufficiently large follows immediately from  $\hat{G}^\nu(\xi^\nu(\omega)) \rightarrow G^*(\bar{\xi}) \in D_\epsilon(\delta)$  regardless of  $\omega \in \Omega$ . Since  $\phi^\nu(\hat{F}^\nu, \hat{G}^\nu)$  converges along  $\nu \in N$  to a value that is not  $\infty$ , we also conclude that

$$\begin{aligned} & \lim_{\nu \in N} \phi^\nu(\hat{F}^\nu, \hat{G}^\nu) \\ &= \lim_{\nu \in N} \mathcal{R}_0\left(\psi_0^\nu(\xi^\nu, \hat{F}^\nu(\xi^\nu), \mathbb{H}(\hat{G}^\nu(\xi^\nu)))\right) \\ & \quad + h^\nu(\hat{F}^\nu, \hat{G}^\nu) \\ &= \phi(F^*, G^*) \\ &= \psi_0\left(\bar{\xi}, F^*(\bar{\xi}), \mathbb{H}(G^*(\bar{\xi}))\right) + h(F^*, G^*). \end{aligned}$$

Under the additional assumption that both  $\mathcal{F}$  and  $\mathcal{G}$  contain the zero mappings and  $h(F, G) = 0$  only when  $F$  and  $G$  are constant mappings, we realize that  $F^*$  and  $G^*$  must be constant mappings because otherwise one can construct  $F^{**}$  and  $G^{**}$  by setting

$$F^{**}(\xi) = F^*(\bar{\xi}) \quad \text{and} \quad G^{**}(\xi) = G^*(\bar{\xi}),$$

and these mappings are in  $\mathcal{F}$  and  $\mathcal{G}$ , respectively. This achieves  $\phi(F^{**}, G^{**}) < \phi(F^*, G^*)$  because  $0 = h(F^{**}, G^{**}) < h(F^*, G^*)$ .  $\square$

**Proof of Corollary 2.** The arguments leading to the verification of (18) in the proof of Theorem 1 remain valid under the present assumptions. Consequently, using the notation from that proof, we obtain

$$\begin{aligned} \infty &> \liminf_{\nu \in N} \gamma^\nu \geq \liminf_{\nu \in N} \phi^\nu(\hat{F}^\nu, \hat{G}^\nu) \\ &\geq \phi(F^*, G^*) \\ &= \psi_0\left(\bar{\xi}, F^*(\bar{\xi}), \mathbb{H}(G^*(\bar{\xi}))\right) + h(F^*, G^*), \end{aligned}$$

where we use the fact that  $\phi(F^*, G^*) < \infty$ . Thus,  $(F^*(\bar{\xi}), \mathbb{H}(G^*(\bar{\xi})))$  is feasible for  $(\text{AP})(\bar{\xi})$  and  $G^*(\bar{\xi}) \in D_\epsilon(\delta)$ . The fact that  $\mathbb{H}(\hat{G}^\nu(\xi^\nu(\omega))) = \mathbb{H}(G^*(\bar{\xi}))$  for all  $\omega \in \Omega$  when  $\nu \in N$  is sufficiently large follows because  $\hat{G}^\nu(\xi^\nu(\omega)) \rightarrow G^*(\bar{\xi}) \in D_\epsilon(\delta)$  regardless of  $\omega \in \Omega$ .  $\square$

## 7.2. Implementation Details

In view of the previous study Lejeune et al. (2024), we solve (SP2)( $\xi$ ) and the various training problems using a linearization approach that leverages the fact that minimizing  $\exp(-\alpha z)$  over  $z \in \{0, 1, 2, \dots, T\} \cap \mathcal{Z}$ , where  $\mathcal{Z}$  represents constraints, is equivalent to the problem

$$\begin{aligned} & \text{minimize} \sum_{j=0}^T w_j e^{-j\alpha} \\ & \text{subject to} \sum_{j=1}^T j w_j = z \\ & \sum_{j=0}^T w_j = 1 \\ & w_j \in \{0, 1\}, j = 0, 1, 2, \dots, T. \end{aligned}$$

Implementing this idea, we obtain the following reformulation of (SP2)( $\xi$ ):

$$\begin{aligned} & \text{minimize}_{w,y} \sum_{i=1}^I q_i(\xi) \sum_{j=0}^T w_{i,j,1} \exp(-j\alpha(\xi)) \\ & \text{subject to} \sum_{i=1}^I q_i(\xi) \sum_{j=0}^T w_{i,j,2} \exp(-j\alpha(\xi)) \leq \tau \\ & \sum_{c=1}^C y_{c,t} = 1 \\ & \quad \forall t = 1, \dots, T \\ & \sum_{c' \in N(c)} y_{c',t-1} \geq y_{c,t} \\ & \quad \forall c = 1, \dots, C, t = 2, \dots, T \\ & \sum_{j=1}^T j w_{i,j,k} = \sum_{c=1}^C \sum_{t=1}^T \zeta_{c,t,i}^k y_{c,t} \\ & \quad \forall i = 1, \dots, I, k = 1, 2 \\ & \sum_{j=0}^T w_{i,j,k} = 1 \\ & \quad \forall i = 1, \dots, I, k = 1, 2 \\ & w_{i,j,k}, y_{c,t} \in \{0, 1\} \\ & \quad \forall c = 1, \dots, C, t = 1, \dots, T \\ & i = 1, \dots, I, j = 0, \dots, T, k = 1, 2. \end{aligned}$$

The variables  $w_{i,j,k}$  can be relaxed to continuous on  $[0, 1]$  without jeopardizing the equivalence with (SP2)( $\xi$ ) because the exponential function is strictly convex. This relaxation, however, tends to cause numerical instabilities and we use the stated formulation consistently. We linearize (SP1)( $\xi$ ) and the training problems (EW-SP2) $^\nu$ , (WW-SP2) $^\nu$ , and ( $\beta$ -SP1) $^\nu$  similarly.

Under the expectation risk measure for  $\mathcal{R}_0$ , the training problem (TP) $^\nu$  supporting (SP2)( $\xi$ ) takes the following specialized form after leveraging a similar expansion of variables as in (TP-super):

$$\begin{aligned}
& \text{(EW-SP2)}^\nu \\
& \text{minimize}_{y, B, b} \frac{1}{|\Omega|} \sum_{\omega \in \Omega} \sum_{i=1}^I q_i(\xi^\nu(\omega)) \cdot \\
& \exp \left( -\alpha(\xi^\nu(\omega)) \sum_{c=1}^C \sum_{t=1}^T \zeta_{c,t,i}^1 y_{c,t}(\omega) \right) \\
& + \theta \sum_{c=1}^C \sum_{t=1}^T \|B_{c,t}\|_1 \\
& \text{subject to } \sum_{i=1}^I q_i(\xi^\nu(\omega)) \cdot \\
& \exp \left( -\alpha(\xi^\nu(\omega)) \sum_{c=1}^C \sum_{t=1}^T \zeta_{c,t,i}^2 y_{c,t}(\omega) \right) \leq \tau \\
& \quad \forall \omega \in \Omega
\end{aligned} \tag{21}$$

$$\sum_{c=1}^C y_{c,t}(\omega) = 1 \quad \forall t = 1, \dots, T, \omega \in \Omega \tag{22}$$

$$\begin{aligned}
& \sum_{c' \in N(c)} y_{c',t-1}(\omega) \geq y_{c,t}(\omega) \\
& \quad \forall c = 1, \dots, C, t = 2, \dots, T, \omega \in \Omega
\end{aligned} \tag{23}$$

$$\begin{aligned}
& -\delta + (\delta + \epsilon) y_{c,t}(\omega) \leq \langle B_{c,t}, \xi^\nu(\omega) \rangle + b_{c,t} \\
& \quad \forall c = 1, \dots, C, t = 1, \dots, T, \omega \in \Omega
\end{aligned} \tag{24}$$

$$\begin{aligned}
& -\epsilon + (\delta + \epsilon) y_{c,t}(\omega) \geq \langle B_{c,t}, \xi^\nu(\omega) \rangle + b_{c,t} \\
& \quad \forall c = 1, \dots, C, t = 1, \dots, T, \omega \in \Omega
\end{aligned} \tag{25}$$

$$B_{c,t} \in \mathbb{R}^r, b_{c,t} \in \mathbb{R} \quad \forall c = 1, \dots, C, t = 1, \dots, T \tag{26}$$

$$y_{c,t}(\omega) \in \{0, 1\} \quad \forall c = 1, \dots, C, t = 1, \dots, T, \omega \in \Omega. \tag{27}$$

Under the worst-case risk measure for  $\mathcal{R}_0$ , the training problem takes the form:

$$\begin{aligned}
 & (\text{WW-SP2})^\nu \\
 & \underset{y, B, b, \gamma}{\text{minimize}} \quad \gamma + \theta \sum_{c=1}^C \sum_{t=1}^T \|B_{c,t}\|_1 \\
 & \text{subject to} \quad (21)-(27), \quad \gamma \in \mathbb{R} \\
 & \sum_{i=1}^I q_i(\xi^\nu(\omega)) \cdot \\
 & \exp\left(-\alpha(\xi^\nu(\omega)) \sum_{c=1}^C \sum_{t=1}^T \zeta_{c,t,i}^1 y_{c,t}(\omega)\right) \leq \gamma \\
 & \quad \forall \omega \in \Omega.
 \end{aligned}$$

Under the  $\beta$ -superquantile risk measure for  $\mathcal{R}_0$ , the training problem (TP) $^\nu$  supporting (SP1)( $\xi$ ) takes the form:

$$\begin{aligned}
 & (\beta\text{-SP1})^\nu \\
 & \underset{y, B, b, \gamma, u}{\text{minimize}} \quad \gamma + \frac{1}{1-\beta} \frac{1}{|\Omega|} \sum_{\omega \in \Omega} u(\omega) + \theta \sum_{c=1}^C \sum_{t=1}^T \|B_{c,t}\|_1 \\
 & \text{subject to} \quad (22)-(27), \quad \gamma \in \mathbb{R}, \quad u(\omega) \geq 0 \quad \forall \omega \in \Omega. \\
 & \sum_{i=1}^I q_i(\xi^\nu(\omega)) \cdot \\
 & \exp\left(-\alpha(\xi^\nu(\omega)) \sum_{c=1}^C \sum_{t=1}^T \zeta_{c,t,i}^1 y_{c,t}(\omega)\right) \\
 & \leq \gamma + u(\omega) \quad \forall \omega \in \Omega.
 \end{aligned}$$

### 7.3. Additional Numerical Results

Figure 2 illustrates the discretized environment consisting of a 9-by-9 grid, with cell numbers, as utilized in Section 5.

Table 7 summarizes computing times for the Decomposition Algorithm across 16 instances of (EW-SP2) $^\nu$  and 16 instances of (WW-SP2) $^\nu$ .

Tables 8 and 9 supplement Tables 5 and 6 by considering the uniform test data for decision rules based on the beta training data.

We supplement the numerical results of Subsection 5.3, which utilizes superquantiles for risk measures in the training problem  $(\beta\text{-SP1})^\nu$ , by also considering expectations and worst-case risk. Let (E-SP1) $^\nu$  and

**Figure 2** Cell numbers in discretized environment with 81 cells.

1	2	3	4	5	6	7	8	9
10	11	12	13	14	15	16	17	18
19	20	21	22	23	24	25	26	27
28	29	30	31	32	33	34	35	36
37	38	39	40	41	42	43	44	45
46	47	48	49	50	51	52	53	54
55	56	57	58	59	60	61	62	63
64	65	66	67	68	69	70	71	72
73	74	75	76	77	78	79	80	81

$\nu$	$(EW-SP2)^\nu$		$(WW-SP2)^\nu$	
	$ \Omega  = 10$	$ \Omega  = 100$	$ \Omega  = 10$	$ \Omega  = 100$
1	59	831	39	902
2	330	786	26	860
3	99	780	32	902
4	26	843	19	834
5	32	841	22	1099
6	55	879	34	1133
7	39	932	37	1128
8	38	952	33	978

**Table 7** Run times (sec.) for the Decomposition Algorithm on different training problems.

rule	training data	test data	number feasible	suboptimality		
				min	avg	max
$B^\nu, b^\nu$	beta	unif	38	0.000	0.005	0.017
MDR	beta	unif	100	0.006	0.017	0.031
AMDR	beta	unif	100	0.000	0.000	0.002

**Table 8** Performance of decision rules obtain from  $(\beta-SP1)^\nu$ . Suboptimality statistics are computed over feasible decisions.

rule	training data	test data	number feasible	suboptimality		
				min	avg	max
$B^\nu, b^\nu$	beta	unif	33	0.010	0.048	0.076
MDR	beta	unif	100	0.033	0.052	0.095
AMDR	beta	unif	100	0.000	0.000	0.007

**Table 9** Performance of decision rules obtain from  $(\beta-SP1)^\nu$  against dispersed target. Suboptimality statistics are computed over feasible decisions.

$(W-SP1)^\nu$  be the training problems obtained from  $(EW-SP2)^\nu$  and  $(WW-SP2)^\nu$ , respectively, by dropping (21) pertaining to the second target.

Using the uniform training data, the Decomposition Algorithm solves  $(E-SP1)^\nu$  in 335 seconds to a relative optimality gap of 0.06%, which corresponds to an absolute gap of 0.002. This produces a decision rule  $(B^\nu, b^\nu)$ . The first six rows of Table 10 report the performance of that decision rule as well as its

derivatives MDR and AMDR. We find that the three decision rules perform equally well. This is caused by the fact that  $B^\nu$  actually is the zero matrix in this case. Thus, for the range of parameter values considered in training, no adaptation is found to be beneficial.

For the beta training data, the Decomposition Algorithm solves (E-SP1) $^\nu$  in 936 seconds, with relative optimality gap of 4.12% and absolute gap of 0.015. The last six rows of Table 10 report the performance of the resulting decision rules. Now, direct use of  $(B^\nu, b^\nu)$  is feasible only at a third of the test points. Moreover, the importance of adaptation is clear: AMDR outperforms MDR with a wide margin.

rule	training data	test data	number feasible	suboptimality		
				min	avg	max
$B^\nu, b^\nu$	unif	unif	100	0.000	0.002	0.013
MDR	unif	unif	100	0.000	0.002	0.013
AMDR	unif	unif	100	0.000	0.002	0.013
$B^\nu, b^\nu$	unif	beta	100	0.000	0.009	0.033
MDR	unif	beta	100	0.000	0.009	0.033
AMDR	unif	beta	100	0.000	0.009	0.033
$B^\nu, b^\nu$	beta	unif	30	0.000	0.002	0.013
MDR	beta	unif	100	0.000	0.011	0.026
AMDR	beta	unif	100	0.000	0.000	0.006
$B^\nu, b^\nu$	beta	beta	37	0.000	0.009	0.033
MDR	beta	beta	100	0.001	0.017	0.049
AMDR	beta	beta	100	0.000	0.002	0.014

**Table 10** Performance of decision rules obtain from (E-SP1) $^\nu$ . Suboptimality statistics are computed over feasible decisions.

Table 11 provides parallel results to those in Table 10 after the switch from (E-SP1) $^\nu$  to (W-SP1) $^\nu$ . The solution of (W-SP1) $^\nu$  using the Decomposition Algorithm takes 361 seconds on the uniform training data, producing a relative optimality gap of 2.25% (absolute gap of 0.009). On the beta training data, the numbers are 511 seconds, 1.89%, and 0.008. (W-SP1) $^\nu$  produces mostly worst levels of suboptimality compare to those obtained from (E-SP1) $^\nu$  for the instances in Table 11, but does achieve more feasible decisions. Still, AMDR produced by (W-SP1) $^\nu$  consistently prescribes quality decisions.

We replicate the results of Tables 10 and 11 also for the more challenging dispersed target case. Now, the solution of (E-SP1) $^\nu$  using the Decomposition Algorithm takes 746 seconds on the uniform training data, producing a relative optimality gap of 2.05% (absolute gap of 0.011). On the beta training data, the numbers are 130 seconds, 4.70%, and 0.026. For (W-SP1) $^\nu$ , the Decomposition Algorithm takes 940 seconds on the uniform training data, producing a relative optimality gap of 2.17% (absolute gap of 0.022). On the beta training data, the numbers are 830 seconds, 3.97%, and 0.024. Tables 12 and 13 report on the performance of the resulting decision rules. While the level of suboptimality of the decision rules are higher compared to the original target setting, AMDR remains viable. Again, (E-SP1) $^\nu$  tends to produce better levels of suboptimality (see Table 12) as compared to (W-SP1) $^\nu$  in Table 13, but is also more likely

rule	training data	test data	number feasible	suboptimality		
				min	avg	max
$B^\nu, b^\nu$	unif	unif	92	0.000	0.009	0.023
MDR	unif	unif	100	0.003	0.011	0.023
AMDR	unif	unif	100	0.000	0.002	0.013
$B^\nu, b^\nu$	unif	beta	70	0.000	0.016	0.034
MDR	unif	beta	100	0.001	0.018	0.042
AMDR	unif	beta	100	0.000	0.008	0.033
$B^\nu, b^\nu$	beta	unif	67	0.000	0.008	0.018
MDR	beta	unif	100	0.003	0.011	0.023
AMDR	beta	unif	100	0.000	0.001	0.013
$B^\nu, b^\nu$	beta	beta	65	0.000	0.015	0.032
MDR	beta	beta	100	0.001	0.018	0.042
AMDR	beta	beta	100	0.000	0.006	0.023

**Table 11** Performance of decision rules obtain from  $(W\text{-SP1})^\nu$ . Suboptimality statistics are computed over feasible decisions.

to produce infeasible decisions.

rule	training data	test data	number feasible	suboptimality		
				min	avg	max
$B^\nu, b^\nu$	unif	unif	100	0.000	0.011	0.032
MDR	unif	unif	100	0.000	0.011	0.032
AMDR	unif	unif	100	0.000	0.011	0.032
$B^\nu, b^\nu$	unif	beta	100	0.000	0.028	0.086
MDR	unif	beta	100	0.000	0.028	0.086
AMDR	unif	beta	100	0.000	0.028	0.086
$B^\nu, b^\nu$	beta	unif	1	0.028	0.028	0.028
MDR	beta	unif	100	0.000	0.018	0.075
AMDR	beta	unif	100	0.000	0.001	0.021
$B^\nu, b^\nu$	beta	beta	5	0.000	0.023	0.094
MDR	beta	beta	100	0.000	0.018	0.075
AMDR	beta	beta	100	0.000	0.001	0.021

**Table 12** Performance of decision rules obtain from  $(E\text{-SP1})^\nu$  against dispersed target. Suboptimality statistics are computed over feasible decisions.

rule	training data	test data	number feasible	suboptimality		
				min	avg	max
$B^\nu, b^\nu$	unif	unif	100	0.006	0.023	0.049
MDR	unif	unif	100	0.009	0.024	0.049
AMDR	unif	unif	100	0.000	0.003	0.027
$B^\nu, b^\nu$	unif	beta	100	0.000	0.036	0.122
MDR	unif	beta	100	0.005	0.038	0.094
AMDR	unif	beta	100	0.000	0.016	0.068
$B^\nu, b^\nu$	beta	unif	50	0.000	0.006	0.044
MDR	beta	unif	100	0.008	0.018	0.036
AMDR	beta	unif	100	0.000	0.002	0.025
$B^\nu, b^\nu$	beta	beta	36	0.000	0.027	0.084
MDR	beta	beta	100	0.002	0.032	0.080
AMDR	beta	beta	100	0.000	0.013	0.068

**Table 13** Performance of decision rules obtain from  $(W-SP1)^\nu$  against dispersed target. Suboptimality statistics are computed over feasible decisions.

### Additional Reference

Royset JO (2018) Approximations and solution estimates in optimization. *Mathematical Programming* 170(2):479–506.