E-companion: "On the Trade-Off Between Distributional Belief and Ambiguity: Conservatism, Finite-Sample Guarantees, and Asymptotic Properties"

EC.1. Proofs

EC.1.1. Proof of Theorem 1

Proof. First, we prove part (i). Suppose that $\mathcal{P}_N$ is star-shaped and $\widehat{\mathbb{P}}_N \in \mathcal{P}_N$ is a star center of $\mathcal{P}_N$. We show that $\{\mathcal{P}_{N,\theta} \mid \theta \in [0,1]\}$ satisfies the hierarchical property. For any $0 \le \theta_1 < \theta_2 \le 1$ and $\mathbb{Q} \in \mathcal{P}_N$, we have

$$\begin{aligned}(1-\theta_1)\,\widehat{\mathbb{P}}_N + \theta_1\,\mathbb{Q} &= (1-\theta_1)\,\widehat{\mathbb{P}}_N + \theta_2\,\widehat{\mathbb{P}}_N - \theta_2\,\widehat{\mathbb{P}}_N + \theta_1\,\mathbb{Q}\\ &= (1-\theta_2)\,\widehat{\mathbb{P}}_N + (\theta_2-\theta_1)\,\widehat{\mathbb{P}}_N + \theta_1\,\mathbb{Q}\\ &= (1-\theta_2)\widehat{\mathbb{P}}_N + \theta_2\left[\left(1-\frac{\theta_1}{\theta_2}\right)\widehat{\mathbb{P}}_N + \frac{\theta_1}{\theta_2}\mathbb{Q}\right], \end{aligned} \tag{EC.1}$$

where $\theta_1/\theta_2 \in [0,1)$. Since $\mathcal{P}_N$ is star-shaped and $\widehat{\mathbb{P}}_N \in \mathcal{P}_N$ is a star center of $\mathcal{P}_N$, the measure $(1-\theta_1/\theta_2)\widehat{\mathbb{P}}_N + (\theta_1/\theta_2)\mathbb{Q}$ in the second term of (EC.1) belongs to $\mathcal{P}_N$. It follows from the definition of $\mathcal{P}'_{N,\theta_2}$ that $(1-\theta_1)\,\widehat{\mathbb{P}}_N + \theta_1\,\mathbb{Q} \in \mathcal{P}'_{N,\theta_2}$ for any $\mathbb{Q} \in \mathcal{P}_N$, and thus, $\mathcal{P}'_{N,\theta_1} \subseteq \mathcal{P}'_{N,\theta_2}$. This shows that $\{\mathcal{P}_{N,\theta} \mid \theta \in [0,1]\}$ satisfies the hierarchical property.

Now, suppose that $\{\mathcal{P}_{N,\theta} \mid \theta \in [0,1]\}$ satisfies the hierarchical property, i.e., $\mathcal{P}'_{N,\theta_1} \subseteq \mathcal{P}'_{N,\theta_2}$ for all $0 \le \theta_1 < \theta_2 \le 1$. We show that $\mathcal{P}_N$ is star-shaped with a star center $\widehat{\mathbb{P}}_N$. For any $\mathbb{Q} \in \mathcal{P}_N$, by (EC.1), we have that

$$(1-\theta_1)\,\widehat{\mathbb{P}}_N + \theta_1\,\mathbb{Q} = (1-\theta_2)\widehat{\mathbb{P}}_N + \theta_2\left[\left(1-\frac{\theta_1}{\theta_2}\right)\widehat{\mathbb{P}}_N + \frac{\theta_1}{\theta_2}\mathbb{Q}\right] \in \mathcal{P}'_{N,\theta_2} \tag{EC.2}$$

since $(1-\theta_1)\,\widehat{\mathbb{P}}_N + \theta_1\,\mathbb{Q} \in \mathcal{P}'_{N,\theta_1} \subseteq \mathcal{P}'_{N,\theta_2}$. By definition of $\mathcal{P}'_{N,\theta_2}$, the inclusion in (EC.2) implies that $(1-\theta_1/\theta_2)\widehat{\mathbb{P}}_N + (\theta_1/\theta_2)\mathbb{Q} \in \mathcal{P}_N$. Since $\theta_1/\theta_2 \in [0,1)$ is arbitrary, we have $(1-\alpha)\widehat{\mathbb{P}}_N + \alpha\mathbb{Q} \in \mathcal{P}_N$ for all $\alpha \in [0,1]$ and $\mathbb{Q} \in \mathcal{P}_N$. (Note that when $\alpha = 1$, we have $(1-\alpha)\widehat{\mathbb{P}}_N + \alpha\mathbb{Q} = \mathbb{Q} \in \mathcal{P}_N$.) This shows that $\mathcal{P}_N$ is star-shaped and $\widehat{\mathbb{P}}_N$ is a star center.

Next, we prove part (ii). Suppose that $\{\mathcal{P}_{N,\theta} \mid \theta \in [0,1]\}$ satisfies the strict hierarchical property, i.e., $\mathcal{P}'_{N,\theta}$ is increasing in θ. We show that $\mathcal{P}_N$ is star-shaped with a star center $\widehat{\mathbb{P}}_N$ and $\mathcal{P}_N \neq \{\widehat{\mathbb{P}}_N\}$. From part (i), we have that $\mathcal{P}_N$ is star-shaped with a star center $\widehat{\mathbb{P}}_N$. Suppose, on the contrary, that $\mathcal{P}_N = \{\mathbb{P}_N\}$. Then, we have $\mathbb{P}'_{N,\theta} = \{\widehat{\mathbb{P}}_N\}$ for all $\theta \in [0,1]$, contradicting that $\mathcal{P}'_{N,\theta}$ is increasing in θ. This shows that $\mathcal{P}_N \neq \{\mathbb{P}_N\}$.

Now, suppose that $\mathcal{P}_N$ is star-shaped with a star center $\widehat{\mathbb{P}}_N \in \mathcal{P}_N$ and $\mathcal{P}_N \neq \{\widehat{\mathbb{P}}_N\}$. We show that $\{\mathcal{P}_{N,\theta} \mid \theta \in [0,1]\}$ satisfies the strict hierarchical property. From part (i), we have $\mathcal{P}'_{N,\theta_1} \subseteq \mathcal{P}'_{N,\theta_2}$, for any $0 \le \theta_1 < \theta_2 \le 1$. To show the strict inclusion $\mathcal{P}'_{N,\theta_1} \subset \mathcal{P}'_{N,\theta_2}$, we need to show that there exists a probability measure $\mathbb{M}$ such that $\mathbb{M} \in \mathcal{P}'_{N,\theta_2}$ but $\mathbb{M} \notin \mathcal{P}'_{N,\theta_1}$. We first claim that this is equivalent to

$$\mathcal{P}_N \setminus \left\{ \left(1 - \frac{\theta_1}{\theta_2}\right) \widehat{\mathbb{P}}_N + \frac{\theta_1}{\theta_2}\,\mathbb{Q} \,\middle|\, \mathbb{Q} \in \mathcal{P}_N \right\} \neq \emptyset. \tag{EC.3}$$

Indeed, (EC.3) is equivalent to the existence of $\mathbb{Q}' \in \mathcal{P}_N$ such that $\mathbb{Q}' \neq (1-\theta_1/\theta_2)\widehat{\mathbb{P}}_N + (\theta_1/\theta_2)\mathbb{Q}$ for all $\mathbb{Q} \in \mathcal{P}_N$. Note that $\mathbb{Q}' \neq (1-\theta_1/\theta_2)\widehat{\mathbb{P}}_N + (\theta_1/\theta_2)\mathbb{Q}$ for all $\mathbb{Q} \in \mathcal{P}_N$ is equivalent to $\theta_2\mathbb{Q}' + (1-\theta_2)\widehat{\mathbb{P}}_N \neq \theta_1\mathbb{Q} + (1-\theta_1)\widehat{\mathbb{P}}_N$ for all $\mathbb{Q} \in \mathcal{P}_N$. Thus, we have that $\theta_2\mathbb{Q}' + (1-\theta_2)\widehat{\mathbb{P}}_N \in \mathcal{P}'_{N,\theta_2}$ cannot be expressed as $\theta_1\mathbb{Q} + (1-\theta_1)\widehat{\mathbb{P}}_N$ for any $\mathbb{Q} \in \mathcal{P}_N$, i.e., $\theta_2\mathbb{Q}' + (1-\theta_2)\widehat{\mathbb{P}}_N \notin \mathcal{P}'_{N,\theta_1}$. This completes the proof of the claim.

Next, suppose, for the sake of contradiction, that condition (EC.3) does not hold, i.e.,

$$\mathcal{P}_N \subseteq \left\{ \left(1 - \frac{\theta_1}{\theta_2}\right) \widehat{\mathbb{P}}_N + \frac{\theta_1}{\theta_2}\,\mathbb{Q} \,\middle|\, \mathbb{Q} \in \mathcal{P}_N \right\}. \tag{EC.4}$$

We claim that (EC.4) is equivalent to $\mathcal{P}_N = \{\widehat{\mathbb{P}}_N\}$ is a singleton, which contradicts with $\mathcal{P}_N \neq \{\widehat{\mathbb{P}}_N\}$. Consider an arbitrary probability measure $\mathbb{M} \in \mathcal{P}_N$. By (EC.4), we can write $\mathbb{M} = (1 - \theta_1/\theta_2)\widehat{\mathbb{P}}_N + (\theta_1/\theta_2)\mathbb{Q}_1$ for some $\mathbb{Q}_1 \in \mathcal{P}_N$. Moreover, note that for any $i \in \mathbb{N}$ and probability measure $\mathbb{Q}_i \in \mathcal{P}_N$, by (EC.4), we can write $\mathbb{Q}_i = (1-\theta_1/\theta_2)\widehat{\mathbb{P}}_N + (\theta_1/\theta_2)\mathbb{Q}_{i+1}$ for some $\mathbb{Q}_{i+1} \in \mathcal{P}_N$. Using this recursion, we have

$$\begin{aligned}
\mathbb{M} = \left(1 - \frac{\theta_1}{\theta_2}\right)\widehat{\mathbb{P}}_N + \frac{\theta_1}{\theta_2}\,\mathbb{Q}_1 &= \left(1 - \frac{\theta_1}{\theta_2}\right)\widehat{\mathbb{P}}_N \cdot \left(1 + \frac{\theta_1}{\theta_2}\right) + \left(\frac{\theta_1}{\theta_2}\right)^2 \mathbb{Q}_2 \\
&= \cdots \\
&= \left(1 - \frac{\theta_1}{\theta_2}\right)\widehat{\mathbb{P}}_N \cdot \sum_{i=0}^{n-1} \left(\frac{\theta_1}{\theta_2}\right)^i + \left(\frac{\theta_1}{\theta_2}\right)^n \mathbb{Q}_n
\end{aligned}$$

for any $n \in \mathbb{N}$. Since $\theta_1 < \theta_2$, $\sum_{i=0}^{n-1}(\theta_1/\theta_2)^i \to 1/(1-\theta_1/\theta_2)$ and $(\theta_1/\theta_2)^n \to 0$ as $n \to \infty$. Thus, for any $\varepsilon > 0$, there exists $n' \in \mathbb{N}$ such that $|(1-\theta_1/\theta_2)\sum_{i=0}^{n-1}(\theta_1/\theta_2)^i - 1| < \varepsilon/2$ and $|(\theta_1/\theta_2)^n| < \varepsilon/2$ for all $n > n'$. Therefore, for any $B \in \mathcal{B}$, we have that

$$|\mathbb{M}(B) - \widehat{\mathbb{P}}_N(B)| = \left| \left(1 - \frac{\theta_1}{\theta_2}\right)\widehat{\mathbb{P}}_N(B) \cdot \sum_{i=0}^{n-1} \left(\frac{\theta_1}{\theta_2}\right)^i + \left(\frac{\theta_1}{\theta_2}\right)^n \mathbb{Q}_n(B) - \widehat{\mathbb{P}}_N(B) \right|$$

$$
\begin{aligned}
&\leq \left| \left(1-\frac{\theta_1}{\theta_2}\right) \sum_{i=0}^{n-1} \left(\frac{\theta_1}{\theta_2}\right)^i - 1 \right| \cdot \widehat{\mathbb{P}}_N(B) + \left| \left(\frac{\theta_1}{\theta_2}\right)^n \right| \cdot \mathbb{Q}_n(B) \\
&\leq \frac{\varepsilon}{2} \widehat{\mathbb{P}}_N(B) + \frac{\varepsilon}{2} \mathbb{Q}_n(B) \\
&\leq \varepsilon
\end{aligned}
$$

for all $n > n'$, where the first inequality follows from triangular inequality. Since $\varepsilon > 0$ is arbitrary, it follows that $\mathbb{M} = \widehat{\mathbb{P}}_N$ for any $\mathbb{M} \in \mathcal{P}_N$, implying that $\mathcal{P}_N = \{\widehat{\mathbb{P}}_N\}$. This contradicts the assumption that $\mathcal{P}_N \neq \{\widehat{\mathbb{P}}_N\}$ and completes the proof. $\square$

EC.1.2. Relationship between Star-Shapedness and Convexity

In Lemma EC.1, we show that convexity of a set implies star-shapedness.

LEMMA EC.1. *If $\mathcal{P}_N$ is convex and $\widehat{\mathbb{P}}_N \in \mathcal{P}_N$, then $\mathcal{P}_N$ is star-shaped with a star center $\widehat{\mathbb{P}}_N$*

Proof. Since $\widehat{\mathbb{P}}_N \in \mathcal{P}_N$, it follows from the convexity of $\mathcal{P}_N$ that

$$
(1-\alpha)\widehat{\mathbb{P}}_N + \alpha \mathbb{P} \in \mathcal{P}_N, \quad \forall \alpha \in [0,1],\, \mathbb{P} \in \mathcal{P}_N.
$$

This shows that $\mathcal{P}_N$ is star-shaped with a star center $\widehat{\mathbb{P}}_N$. $\square$

EC.1.3. Proof of Proposition 1

Proof. Suppose that $\mathcal{K}_i$ is star-shaped on $\mathcal{S}_i$ with a star center $\mathbb{E}_{\widehat{\mathbb{P}}_N}[\Phi_i(\boldsymbol{\xi})]$ for all $i \in \{1, \ldots, p\}$. We show that $\mathcal{P}_N$ is star-shaped with a star center $\widehat{\mathbb{P}}_N$. Indeed, for all $i \in \{1, \ldots, p\}$, by the star-shapedness of $\mathcal{K}_i$, we have $(1-\alpha)\mathbb{E}_{\widehat{\mathbb{P}}_N}[\Phi_i(\boldsymbol{\xi})] + \alpha \mathbb{E}_{\mathbb{Q}}[\Phi_i(\boldsymbol{\xi})] = \mathbb{E}_{(1-\alpha)\widehat{\mathbb{P}}_N + \alpha\mathbb{Q}}[\Phi_i(\boldsymbol{\xi})] \in \mathcal{K}_i$ for any $\alpha \in [0,1]$ and $\mathbb{Q} \in \mathcal{P}_N$. It follows from the definition of $\mathcal{P}_N$ that $(1-\alpha)\widehat{\mathbb{P}}_N + \alpha\mathbb{Q} \in \mathcal{P}_N$ for any $\alpha \in [0,1]$ and $\mathbb{Q} \in \mathcal{P}_N$, i.e., $\mathcal{P}_N$ is star-shaped with a star center $\widehat{\mathbb{P}}_N$.

Now, suppose that $\mathcal{P}_N$ is star-shaped with a star center $\widehat{\mathbb{P}}_N$. We show that $\mathcal{K}_i$ is star-shaped on $\mathcal{S}_i := \{\mathbb{E}_{\mathbb{Q}}[\Phi_i(\boldsymbol{\xi})] \mid \mathbb{Q} \in \mathcal{P}_N\} \subseteq \mathbb{R}^{d_i \times d_i}$ with a star center $\mathbb{E}_{\widehat{\mathbb{P}}_N}[\Phi_i(\boldsymbol{\xi})] \in \mathcal{S}_i$ for all $i \in \{1, \ldots, p\}$. Since $(1-\alpha)\widehat{\mathbb{P}}_N + \alpha\mathbb{Q} \in \mathcal{P}_N$ for any $\alpha \in [0,1]$ and $\mathbb{Q} \in \mathcal{P}_N$, we have $\mathbb{E}_{(1-\alpha)\widehat{\mathbb{P}}_N + \alpha\mathbb{Q}}[\Phi_i(\boldsymbol{\xi})] = (1-\alpha)\mathbb{E}_{\widehat{\mathbb{P}}_N}[\Phi_i(\boldsymbol{\xi})] + \alpha\mathbb{E}_{\mathbb{Q}}[\Phi_i(\boldsymbol{\xi})] \in \mathcal{K}_i$ for all $i \in \{1, \ldots, p\}$. This, in turn, implies that $(1-\alpha)\mathbb{E}_{\widehat{\mathbb{P}}_N}[\Phi_i(\boldsymbol{\xi})] + \alpha\Psi \in \mathcal{K}_i$ for any $\alpha \in [0,1]$ and $\Psi \in \mathcal{S}_i$. This completes the proof. $\square$

EC.1.4. Proof of Proposition 2

Proof. Suppose that d is quasi-convex about $\widehat{\mathbb{P}}_N$ in the first argument. We show that $\mathcal{P}_N(\varepsilon)$ is star-shaped with a star center $\widehat{\mathbb{P}}_N$ for all $\varepsilon \geq 0$. Indeed, for any given $\varepsilon \geq 0$, by quasi-convexity, we have $\mathsf{d}\big((1-\alpha)\widehat{\mathbb{P}}_N + \alpha\mathbb{Q}, \widehat{\mathbb{P}}_N\big) \leq \mathsf{d}(\mathbb{Q}, \widehat{\mathbb{P}}_N) \leq \varepsilon$ for any $\alpha \in [0,1]$ and $\mathbb{Q} \in \mathcal{P}_N(\varepsilon)$. It follows from the definition of $\mathcal{P}_N(\varepsilon)$ that $(1-\alpha)\widehat{\mathbb{P}}_N+\alpha\mathbb{Q}\in\mathcal{P}_N(\varepsilon)$ for any $\alpha\in[0,1]$ and $\mathbb{Q}\in\mathcal{P}_N(\varepsilon)$, i.e., $\mathcal{P}_N(\varepsilon)$ is star-shaped with a star-center $\widehat{\mathbb{P}}_N$.

Now, suppose that $\mathcal{P}_N(\varepsilon)$ is star-shaped with a star center $\widehat{\mathbb{P}}_N$ for all $\varepsilon\geq 0$. We show that d is quasi-convex about $\widehat{\mathbb{P}}_N$. Since $\mathcal{P}_N(\varepsilon)$ is star-shaped, we have $(1-\alpha)\widehat{\mathbb{P}}_N+\alpha\mathbb{Q}\in\mathcal{P}_N(\varepsilon)$ for any $\alpha\in[0,1]$ and $\mathbb{Q}\in\mathcal{P}_N(\varepsilon)$. It follows from the definition of $\mathcal{P}_N(\varepsilon)$ that $\mathsf{d}\big((1-\alpha)\widehat{\mathbb{P}}_N+\alpha\mathbb{Q},\widehat{\mathbb{P}}_N\big)\leq\varepsilon$ for any $\alpha\in[0,1]$ and $\mathbb{Q}\in\mathcal{P}_N(\varepsilon)$. Suppose, for the sake of contradiction, that d is not quasi-convex about $\widehat{\mathbb{P}}_N$ in the first argument. That is, there exist $\alpha\in(0,1)$ and $\mathbb{Q}\in\mathcal{P}(\Xi)$ such that $\mathsf{d}\big((1-\alpha)\widehat{\mathbb{P}}_N+\alpha\mathbb{Q},\widehat{\mathbb{P}}_N\big)>\mathsf{d}(\mathbb{Q},\widehat{\mathbb{P}}_N)$. Let $\bar{\varepsilon}:=\mathsf{d}(\mathbb{Q},\widehat{\mathbb{P}}_N)\in[0,\infty)$. Since $\mathcal{P}_N(\bar{\varepsilon})$ is star-shaped with a star center $\widehat{\mathbb{P}}_N$ and $\mathbb{Q}\in\mathcal{P}_N(\bar{\varepsilon})$, we have $(1-\alpha)\widehat{\mathbb{P}}_N+\alpha\mathbb{Q}\in\mathcal{P}_N(\bar{\varepsilon})$. Thus, $\mathsf{d}\big((1-\alpha)\widehat{\mathbb{P}}_N+\alpha\mathbb{Q},\widehat{\mathbb{P}}_N\big)\leq\bar{\varepsilon}=\mathsf{d}(\mathbb{Q},\widehat{\mathbb{P}}_N)$, contradicting the assumption that d is not quasi-convex. This completes the proof. □

EC.1.5. Proof of Theorem 2

Proof. First, we show that for any $\{\theta_1,\theta_2\}\subset[0,1]$, we have

$$\mathbb{H}(\mathcal{P}'_{N,\theta_1},\mathcal{P}'_{N,\theta_2})\leq 2C_N|\theta_1-\theta_2|. \tag{EC.5}$$

For any $\mathbb{P}_1\in\mathcal{P}'_{N,\theta_1}$ and $\mathbb{P}_2\in\mathcal{P}'_{N,\theta_2}$, we can write $\mathbb{P}_i=(1-\theta_i)\widehat{\mathbb{P}}_N+\theta_i\mathbb{Q}_i$ for some $\mathbb{Q}_i\in\mathcal{P}_N$ and $i\in\{1,2\}$. Then,

$$\begin{aligned}\Big|\mathbb{E}_{\mathbb{P}_1}[f(\boldsymbol{x},\boldsymbol{\xi})]-\mathbb{E}_{\mathbb{P}_2}[f(\boldsymbol{x},\boldsymbol{\xi})]\Big| &= \left|(\theta_2-\theta_1)\cdot\frac{1}{N}\sum_{i=1}^N f(\boldsymbol{x},\widehat{\boldsymbol{\xi}}_i)+\theta_1\cdot\mathbb{E}_{\mathbb{Q}_1}[f(\boldsymbol{x},\boldsymbol{\xi})]-\theta_2\cdot\mathbb{E}_{\mathbb{Q}_2}[f(\boldsymbol{x},\boldsymbol{\xi})]\right|\\ &\leq|\theta_1-\theta_2|\left|\frac{1}{N}\sum_{i=1}^N f(\boldsymbol{x},\widehat{\boldsymbol{\xi}}_i)\right|+\Big|\theta_1\mathbb{E}_{\mathbb{Q}_1}[f(\boldsymbol{x},\boldsymbol{\xi})]-\theta_2\mathbb{E}_{\mathbb{Q}_2}[f(\boldsymbol{x},\boldsymbol{\xi})]\Big|\\ &\leq C_N|\theta_1-\theta_2|+\Big|\theta_1\mathbb{E}_{\mathbb{Q}_1}[f(\boldsymbol{x},\boldsymbol{\xi})]-\theta_2\mathbb{E}_{\mathbb{Q}_2}[f(\boldsymbol{x},\boldsymbol{\xi})]\Big|,\end{aligned} \tag{EC.6}$$

where the last inequality follows from Assumption 1 and the definition of C_N in (13). Note that

$$\begin{aligned}\sup_{\mathbb{P}_1\in\mathcal{P}'_{N,\theta_1}}\inf_{\mathbb{P}_2\in\mathcal{P}'_{N,\theta_2}}\mathbb{d}(\mathbb{P}_1,\mathbb{P}_2) &= \sup_{\mathbb{P}_1\in\mathcal{P}'_{N,\theta_1}}\inf_{\mathbb{P}_2\in\mathcal{P}'_{N,\theta_2}}\sup_{\boldsymbol{x}\in\mathcal{X}}\Big|\mathbb{E}_{\mathbb{P}_1}[f(\boldsymbol{x},\boldsymbol{\xi})]-\mathbb{E}_{\mathbb{P}_2}[f(\boldsymbol{x},\boldsymbol{\xi})]\Big|\\ &\leq C_N|\theta_1-\theta_2|+\sup_{\mathbb{Q}_1\in\mathcal{P}_N}\inf_{\mathbb{Q}_2\in\mathcal{P}_N}\sup_{\boldsymbol{x}\in\mathcal{X}}\Big|\theta_1\mathbb{E}_{\mathbb{Q}_1}[f(\boldsymbol{x},\boldsymbol{\xi})]-\theta_2\mathbb{E}_{\mathbb{Q}_2}[f(\boldsymbol{x},\boldsymbol{\xi})]\Big| &&\text{(EC.7a)}\\ &\leq C_N|\theta_1-\theta_2|+\sup_{\mathbb{Q}_1\in\mathcal{P}_N}\sup_{\boldsymbol{x}\in\mathcal{X}}\Big|\theta_1\mathbb{E}_{\mathbb{Q}_1}[f(\boldsymbol{x},\boldsymbol{\xi})]-\theta_2\mathbb{E}_{\mathbb{Q}_1}[f(\boldsymbol{x},\boldsymbol{\xi})]\Big| &&\text{(EC.7b)}\\ &\leq C_N|\theta_1-\theta_2|+\sup_{\mathbb{Q}_1\in\mathcal{P}_N}\sup_{\boldsymbol{x}\in\mathcal{X}}|\theta_1-\theta_2|\Big|\mathbb{E}_{\mathbb{Q}_1}[f(\boldsymbol{x},\boldsymbol{\xi})]\Big|\end{aligned}$$

$$\leq 2C_N|\theta_1 - \theta_2|. \tag{EC.7c}$$

Inequality (EC.7a) follows from (EC.6), where we note that the arguments in supremum and infimum are changed to $\mathbb{Q}_1 \in \mathcal{P}_N$ and $\mathbb{Q}_2 \in \mathcal{P}_N$, respectively; inequality (EC.7b) follows from the fact that choosing $\mathbb{Q}_2 = \mathbb{Q}_1$ for the infimum problem yields an upper bound; inequality (EC.7c) follows from Assumption 1 and the definition of C_N in (13). Interchanging the role of $\mathbb{P}_1$ and $\mathbb{P}_2$, as well as $\mathcal{P}'_{N,\theta_1}$ and $\mathcal{P}'_{N,\theta_2}$, we can obtain a similar inequality:

$$\sup_{\mathbb{P}_2 \in \mathcal{P}'_{N,\theta_2}} \inf_{\mathbb{P}_1 \in \mathcal{P}'_{N,\theta_1}} \mathbb{d}(\mathbb{P}_1, \mathbb{P}_2) \leq 2C_N|\theta_1 - \theta_2|.$$

Therefore, from the definition of $\mathbb{H}$ in (12), we have

$$\mathbb{H}(\mathcal{P}'_{N,\theta_1}, \mathcal{P}'_{N,\theta_2}) = \max\left\{ \sup_{\mathbb{P}_1 \in \mathcal{P}'_{N,\theta_1}} \inf_{\mathbb{P}_2 \in \mathcal{P}'_{N,\theta_2}} \mathbb{d}(\mathbb{P}_1, \mathbb{P}_2), \sup_{\mathbb{P}_2 \in \mathcal{P}'_{N,\theta_2}} \inf_{\mathbb{P}_1 \in \mathcal{P}'_{N,\theta_1}} \mathbb{d}(\mathbb{P}_1, \mathbb{P}_2) \right\} \leq 2C_N|\theta_1 - \theta_2|.$$

Now, we prove assertion (i):

$$|\widehat{v}_N(\theta_1) - \widehat{v}_N(\theta_2)| \leq \mathbb{H}(\mathcal{P}'_{N,\theta_1}, \mathcal{P}'_{N,\theta_2}) \leq 2C_N|\theta_1 - \theta_2|,$$

where the first inequality follows from the quantitative stability analysis; see Proposition EC.4 in EC.6.

For (ii), under the second-order growth condition, we have

$$D\Big(\widehat{\mathcal{X}}_N(\theta_2), \widehat{\mathcal{X}}_N(\theta_1)\Big) \leq \sqrt{\frac{3}{\tau}\mathbb{H}(\mathcal{P}'_{N,\theta_1}, \mathcal{P}'_{N,\theta_2})} \leq \sqrt{\frac{6C_N}{\tau}|\theta_1 - \theta_2|},$$

where the first inequality follows again from the quantitative stability analysis; see Proposition EC.4 in EC.6. This completes the proof. □

EC.1.6. Proof of Theorem 3

Proof. First, we prove part (i). To show the concavity of $\widehat{v}_N(\theta)$, we define the function $g(\theta; \boldsymbol{x}) := \mathbb{E}_{\widehat{\mathbb{P}}_N}[f(\boldsymbol{x}, \boldsymbol{\xi})] + \theta\big\{\sup_{\mathbb{P} \in \mathcal{P}_N} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \boldsymbol{\xi})] - \mathbb{E}_{\widehat{\mathbb{P}}_N}[f(\boldsymbol{x}, \boldsymbol{\xi})]\big\}$, which is linear in θ for any $\boldsymbol{x} \in \mathcal{X}$. Since $\widehat{v}_N(\theta)$ is the pointwise minimum of the linear functions $\{g(\theta; \boldsymbol{x}) \mid \boldsymbol{x} \in \mathcal{X}\}$, i.e., $\widehat{v}_N(\theta) = \min_{\boldsymbol{x} \in \mathcal{X}} g(\theta; \boldsymbol{x})$, the function $\widehat{v}_N(\theta)$ is concave. Next, to show (15), note that $\theta = (1-\theta) \cdot 0 + \theta \cdot 1$. Thus, concavity of $\widehat{v}_N$ implies

$$\widehat{r}_N(\theta) = \widehat{v}_N(\theta) - \Big[(1-\theta) \cdot \widehat{v}_N(0) + \theta \cdot \widehat{v}_N(1)\Big] \geq 0.$$

Also, by Theorem 2, we have

$$\begin{aligned}\widehat{r}_N(\theta) &= \widehat{\upsilon}_N(\theta) - \Big[(1-\theta)\cdot\widehat{\upsilon}_N(0) + \theta\cdot\widehat{\upsilon}_N(1)\Big]\\ &\le (1-\theta)\cdot\Big|\widehat{\upsilon}_N(\theta) - \widehat{\upsilon}_N(0)\Big| + \theta\cdot\Big|\widehat{\upsilon}_N(\theta) - \widehat{\upsilon}_N(1)\Big|\\ &\le (1-\theta)\cdot 2C_N\theta + \theta\cdot 2C_N(1-\theta)\\ &\le 4C_N\theta(1-\theta).\end{aligned}$$

This completes the proof of part (i).

Finally, for part (ii), we have

$$\begin{aligned}&D\Big(\widehat{\mathcal{X}}_N(\theta), (1-\theta)\widehat{\mathcal{X}}_N(0) + \theta\widehat{\mathcal{X}}_N(1)\Big)\\ &= \sup_{\boldsymbol{x}\in\widehat{\mathcal{X}}_N(\theta)} \inf_{\overline{\boldsymbol{x}}\in(1-\theta)\widehat{\mathcal{X}}_N(0)+\theta\widehat{\mathcal{X}}_N(1)} \|\boldsymbol{x}-\overline{\boldsymbol{x}}\|\\ &= \sup_{\boldsymbol{x}\in\widehat{\mathcal{X}}_N(\theta)} \inf_{\boldsymbol{x}'\in\widehat{\mathcal{X}}_N(0),\boldsymbol{x}''\in\widehat{\mathcal{X}}_N(1)} \Big\|\boldsymbol{x} - \big[(1-\theta)\boldsymbol{x}' + \theta\boldsymbol{x}''\big]\Big\|\\ &\le \sup_{\boldsymbol{x}\in\widehat{\mathcal{X}}_N(\theta)} \inf_{\boldsymbol{x}'\in\widehat{\mathcal{X}}_N(0),\boldsymbol{x}''\in\widehat{\mathcal{X}}_N(1)} \Big\{(1-\theta)\|\boldsymbol{x}-\boldsymbol{x}'\| + \theta\|\boldsymbol{x}-\boldsymbol{x}''\|\Big\} \qquad \text{(EC.8a)}\\ &\le (1-\theta)\sup_{\boldsymbol{x}\in\widehat{\mathcal{X}}_N(\theta)} \inf_{\boldsymbol{x}'\in\widehat{\mathcal{X}}_N(0)} \|\boldsymbol{x}-\boldsymbol{x}'\| + \theta \sup_{\boldsymbol{x}\in\widehat{\mathcal{X}}_N(\theta)} \inf_{\boldsymbol{x}''\in\widehat{\mathcal{X}}_N(1)} \|\boldsymbol{x}-\boldsymbol{x}''\| \qquad \text{(EC.8b)}\\ &= (1-\theta)D\Big(\widehat{\mathcal{X}}_N(\theta), \widehat{\mathcal{X}}_N(0)\Big) + \theta D\Big(\widehat{\mathcal{X}}_N(\theta), \widehat{\mathcal{X}}_N(1)\Big)\\ &\le (1-\theta)\sqrt{\frac{6C_N\theta}{\tau}} + \theta\sqrt{\frac{6C_N(1-\theta)}{\tau}} \qquad \text{(EC.8c)}\\ &= \sqrt{\frac{6C_N\theta(1-\theta)}{\tau}}\Big(\sqrt{\theta} + \sqrt{1-\theta}\Big).\end{aligned}$$

Inequality (EC.8a) follows from triangle inequality; inequality (EC.8b) follows from separating the supremum operator to each summand; inequality (EC.8c) follows from Theorem 2. □

EC.1.7. Proof of Proposition 3

Proof. Since $\mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(0)] \le \upsilon^\star$, we have

$$\begin{aligned}\mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(\theta)] - \upsilon^\star &\le \mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(\theta)] - \mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(0)]\\ &= \mathbb{E}_{\mathbb{P}^N}[(1-\theta)\cdot\widehat{\upsilon}_N(0) + \theta\cdot\widehat{\upsilon}_N(1) + \widehat{r}_N(\theta)] - \mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(0)] \quad \text{(by (15) in Theorem 3)}\\ &= \theta\Big\{\mathbb{E}_{\mathbb{P}^N}\big[\widehat{\upsilon}_N(1)\big] - \mathbb{E}_{\mathbb{P}^N}\big[\widehat{\upsilon}_N(0)\big]\Big\} + \mathbb{E}_{\mathbb{P}^N}[\widehat{r}_N(\theta)].\end{aligned}$$

From Theorem 3, we have $\widehat{r}_N(\theta) \in [0, 4C_N\theta(1-\theta)]$. Thus, $R_N(\theta) := \mathbb{E}_{\mathbb{P}^N}[\widehat{r}_N(\theta)] \in [0, 4\overline{C}_N\theta(1-\theta)]$. Finally, if $\widehat{\mathbb{P}}_N \in \mathcal{P}_N$, then $\widehat{\upsilon}_N(0) \le \widehat{\upsilon}_N(1)$. It follows that the upper bound on the bias in (17) is non-negative. □

EC.1.8. Proof of Theorem 4

Proof. From Theorem 2, we have that $\widehat{\upsilon}_N(\theta)$ is Lipschitz continuous with Lipschitz constant $C_N < \infty$ (almost surely). Thus, $\mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(\theta)]$ is also Lipschitz continuous with Lipschitz constant $\overline{C}_N = \mathbb{E}_{\mathbb{P}^N}(C_N) < \infty$. Since $\mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(0)] \leq \upsilon^\star \leq \mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(1)]$, it follows by the intermediate value theorem (see Theorem 4.23 of Rudin 1976) that there exists $\theta_N^{\mathsf{u}} \in [0,1]$ for which $\mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(\theta_N^{\mathsf{u}})] = \upsilon^\star$. This completes the proof. $\square$

EC.1.9. Proof of Corollary 1

Proof. By Theorem 3, $\widehat{\upsilon}_N(\theta)$ is concave on $[0,1]$, and so is $\mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(\theta)]$. Since $\mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(0)] \leq \mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(1)]$, we have that $\mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(\theta)]$ is either (a) non-decreasing on $[0,1]$ or (b) first non-decreasing and then non-increasing (see Lemma 1.1.4 of Niculescu and Persson 2018). By Theorem 4, there exists θ_N^{u} such that $\mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(\theta)] = \upsilon^\star$. When it is not unique, take $\theta_N^{\mathsf{u}} = \inf\{\theta \in [0,1] \mid \mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(\theta)] = \upsilon^\star\}$. It follows that $\mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(\theta)]$ is non-decreasing on $[0,\theta_N^{\mathsf{u}}]$ and thus, $|\mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(\theta)] - \upsilon^\star| \leq |\mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(0)] - \upsilon^\star|$ for all $\theta \in [0,\theta_N^{\mathsf{u}}]$. $\square$

EC.1.10. Proof of Theorem 5

Proof. First, note that θ_N^{u} satisfies $\mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(\theta_N^{\mathsf{u}})] = \upsilon^\star$, where Theorem 4 ensures the existence of θ_N^{u}. Moreover, Theorem 3 implies that

$$\mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(\theta_N^{\mathsf{u}})] = (1-\theta_N^{\mathsf{u}})\mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(0)] + \theta_N^{\mathsf{u}}\mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(1)] + R_N(\theta_N^{\mathsf{u}}),$$

where $R_N(\theta_N^{\mathsf{u}}) \in [0, 4\overline{C}_N\theta_N^{\mathsf{u}}(1-\theta_N^{\mathsf{u}})]$. Combining these two equations, we have

$$\upsilon^\star - \mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(0)] = \theta_N^{\mathsf{u}}\Big\{\mathbb{E}_{\mathbb{P}^N}\left[\widehat{\upsilon}_N(1)\right] - \mathbb{E}_{\mathbb{P}^N}\left[\widehat{\upsilon}_N(0)\right]\Big\} + R_N(\theta_N^{\mathsf{u}}). \tag{EC.9}$$

Note that left-hand-side of (EC.9) corresponds to the bias of the SAA estimator. Under assumptions (a)–(c), Theorem 6 of Banholzer et al. (2022) gives $\upsilon^\star - \mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(0)] = o(\sqrt{\log\log N}/\sqrt{N})$. For the ease of notation, let $\Delta_N = \mathbb{E}_{\mathbb{P}^N}\left[\widehat{\upsilon}_N(1)\right] - \mathbb{E}_{\mathbb{P}^N}\left[\widehat{\upsilon}_N(0)\right]$ and $b_N = \sqrt{\log\log N}/\sqrt{N}$. By the assumption that $\inf_{N\in\mathbb{N}} \Delta_N > 0$, we know that $\Delta_N \geq \Delta$ for some $\Delta > 0$. Then,

$$0 \leq \theta_N^{\mathsf{u}}\Delta \leq \theta_N^{\mathsf{u}}\Delta_N + R_N(\theta_N^{\mathsf{u}}) = o(b_N), \tag{EC.10}$$

where the second inequality follows from $R_N(\theta_N^{\mathsf{u}}) \geq 0$. In other words, (EC.10) implies that

$$0 \leq \left(b_N^{-1}\theta_N^{\mathsf{u}}\right)\Delta \leq o(1),$$

showing that $b_N^{-1}\theta_N^{\mathsf{u}} \to 0$ as $N \to \infty$. This completes the proof. $\square$

EC.1.11. Proof of Theorem 6

Proof. Theorem 5.7 of Shapiro et al. (2014) gives the following asymptotic normality of $\widehat{\upsilon}_N(0)$:

$$X_N = \sqrt{N}\Big(\widehat{\upsilon}_N(0) - \upsilon^\star\Big) \Rightarrow \inf_{\boldsymbol{x}\in\mathcal{X}^\star} \mathbb{G}(\boldsymbol{x}), \tag{EC.11}$$

where "$\Rightarrow$" denotes the convergence in distribution, $\mathcal{X}^\star$ is the set of optimal solutions to (1), and $\mathbb{G}$ is a Gaussian process indexed by $\mathcal{X}$ with mean zero and covariance function $\mathrm{Cov}(\mathbb{G}(\boldsymbol{x}_1), \mathbb{G}(\boldsymbol{x}_2)) = \mathrm{Cov}_{\mathbb{P}^\star}(f(\boldsymbol{x}_1, \boldsymbol{\xi}), f(\boldsymbol{x}_2, \boldsymbol{\xi}))$. Together with the asymptotic uniform integrability of $\{X_N\}_{N\in\mathbb{N}}$, (EC.11) implies that

$$\mathbb{E}_{\mathbb{P}^N}\Big[\sqrt{N}\Big(\widehat{\upsilon}_N(0) - \upsilon^\star\Big)\Big] = \mathbb{E}\Big[\inf_{\boldsymbol{x}\in\mathcal{X}^\star} \mathbb{G}(\boldsymbol{x})\Big] + o(1)$$

(see, e.g., Theorem 2.20 of van der Vaart 2000). Hence, the bias of the SAA estimator is given by

$$\mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(0)] - \upsilon^\star = \frac{1}{\sqrt{N}}\mathbb{E}\Big[\inf_{\boldsymbol{x}\in\mathcal{X}^\star} \mathbb{G}(\boldsymbol{x})\Big] + o(1/\sqrt{N}). \tag{EC.12}$$

If $\mathcal{X}^\star$ is a singleton, then $\inf_{\boldsymbol{x}\in\mathcal{X}^\star} \mathbb{G}(\boldsymbol{x})$ reduces to a normal distribution with mean zero. Therefore, (EC.12) implies that the bias of the SAA estimator is of order $o(1/\sqrt{N})$. A similar argument in the proof of Theorem 5 shows that $\theta_N^{\mathsf{u}} = o(1/\sqrt{N})$.

Next, we consider that $\mathcal{X}^\star$ is not a singleton. In the trivial case when $\mathbb{E}_{\mathbb{P}^N}[\inf_{\boldsymbol{x}\in\mathcal{X}^\star} \mathbb{G}(\boldsymbol{x})] = 0$, we immediately have $\theta_N^{\mathsf{u}} = o(1/\sqrt{N})$, which directly implies $\theta_N^{\mathsf{u}} = O(1/\sqrt{N})$. Now, consider the case when $\mathbb{E}_{\mathbb{P}^N}[\inf_{\boldsymbol{x}\in\mathcal{X}^\star} \mathbb{G}(\boldsymbol{x})] < 0$. From (EC.9) and (EC.12), we have

$$\theta_N^{\mathsf{u}}\Big\{\mathbb{E}_{\mathbb{P}^N}\big[\widehat{\upsilon}_N(1)\big] - \mathbb{E}_{\mathbb{P}^N}\big[\widehat{\upsilon}_N(0)\big]\Big\} + R_N(\theta_N^{\mathsf{u}}) = -\frac{1}{\sqrt{N}}\mathbb{E}\Big[\inf_{\boldsymbol{x}\in\mathcal{X}^\star} \mathbb{G}(\boldsymbol{x})\Big] + o(1/\sqrt{N}), \tag{EC.13}$$

where $R_N(\theta_N^{\mathsf{u}}) \in [0, 4C\theta_N^{\mathsf{u}}(1-\theta_N^{\mathsf{u}})]$. Again, for the ease of notation, let $\Delta_N = \mathbb{E}_{\mathbb{P}^N}\big[\widehat{\upsilon}_N(1)\big] - \mathbb{E}_{\mathbb{P}^N}\big[\widehat{\upsilon}_N(0)\big] > 0$. By the assumption that $\inf_{N\in\mathbb{N}} \Delta_N > 0$, we know that $\Delta_N \geq \Delta$ for some $\Delta > 0$. Since $R_N(\theta_N^{\mathsf{u}}) \geq 0$, (EC.13) implies

$$0 \leq \big(\sqrt{N}\theta_N^{\mathsf{u}}\big)\Delta \leq -\mathbb{E}\Big[\inf_{\boldsymbol{x}\in\mathcal{X}^\star} \mathbb{G}(\boldsymbol{x})\Big] + o(1). \tag{EC.14}$$

Thus, (EC.14) shows that $\limsup_{N\to\infty} \sqrt{N}\theta_N^{\mathsf{u}}$ is upper bounded, concluding that $\theta_N^{\mathsf{u}} = O(1/\sqrt{N})$.

□

EC.1.12. Proof of Theorem 7

Proof. First, we show that C_N defined in (13) satisfies $C_N \le L\cdot \mathrm{diam}(\Xi)+M$ almost surely. Note that

$$f(\boldsymbol{x},\boldsymbol{\xi}) = \big|f(\boldsymbol{x},\boldsymbol{\xi}) - f(\boldsymbol{x},\boldsymbol{\xi}_0)\big| + \big|f(\boldsymbol{x},\boldsymbol{\xi}_0)\big| \le L\cdot\|\boldsymbol{\xi}-\boldsymbol{\xi}_0\| + M \le L\cdot \mathrm{diam}(\Xi)+M, \tag{EC.15}$$

where the first inequality follows from assumptions (a) and (c), and the second inequality follows from assumption (b). Therefore, we have $\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\boldsymbol{\xi})] \le L\cdot \mathrm{diam}(\Xi)+M$ for any $\mathbb{P}\in\mathcal{P}(\Xi)$. It follows from the definition of C_N that $C_N \le L\cdot \mathrm{diam}(\Xi)+M$ almost surely.

Next, we show the desired inequality. Note that almost surely, we have

$$\widehat{v}_N(0) \le \widehat{v}_N(\theta) \le \widehat{v}_N(0) + \Big\{\theta\big[\widehat{v}_N(1)-\widehat{v}_N(0)\big] + 4\theta C_N(1-\theta)\Big\}, \tag{EC.16a}$$
$$\le \widehat{v}_N(0) + \Big[2\theta C_N + 4\theta C_N(1-\theta)\Big], \tag{EC.16b}$$
$$\le \widehat{v}_N(0) + 2\theta(3-2\theta)\big(L\cdot \mathrm{diam}(\Xi)+M\big). \tag{EC.16c}$$

The first inequality in (EC.16a) follows from assumption (d), the second inequality in (EC.16a) follows from part (i) of Theorem 3, (EC.16b) follows from Theorem 2, and (EC.16c) follows from (EC.15). Multiplying both sides of inequality (EC.16c) by -1 and taking the expectation with respect to $\mathbb{P}^N$, we obtain

$$-\mathbb{E}_{\mathbb{P}^N}[\widehat{v}_N(0)] - 2\theta(3-2\theta)\big(L\cdot \mathrm{diam}(\Xi)+M\big) \le -\mathbb{E}_{\mathbb{P}^N}[\widehat{v}_N(\theta)] \le -\mathbb{E}_{\mathbb{P}^N}[\widehat{v}_N(0)]. \tag{EC.17}$$

Together with (EC.16a)–(EC.16c) and (EC.17), we have

$$\begin{aligned}
&\widehat{v}_N(0) - \mathbb{E}_{\mathbb{P}^N}[\widehat{v}_N(0)] - 2\theta(3-2\theta)\big(L\cdot \mathrm{diam}(\Xi)+M\big)\\
\le\ &\widehat{v}_N(\theta) - \mathbb{E}_{\mathbb{P}^N}[\widehat{v}_N(\theta)]\\
\le\ &\widehat{v}_N(0) - \mathbb{E}_{\mathbb{P}^N}[\widehat{v}_N(0)] + 2\theta(3-2\theta)\big(L\cdot \mathrm{diam}(\Xi)+M\big),
\end{aligned}$$

which implies that

$$\bigg|\Big\{\widehat{v}_N(\theta) - \mathbb{E}_{\mathbb{P}^N}[\widehat{v}_N(\theta)]\Big\} - \Big\{\widehat{v}_N(0) - \mathbb{E}_{\mathbb{P}^N}[\widehat{v}_N(0)]\Big\}\bigg| \le 2\theta(3-2\theta)\big(L\cdot \mathrm{diam}(\Xi)+M\big) \tag{EC.18}$$

almost surely. Therefore, we have

$$\begin{aligned}
&\sqrt{\mathrm{Var}_{\mathbb{P}^N}\left(\widehat{v}_N(\theta)\right)}\\
=&\sqrt{\mathbb{E}_{\mathbb{P}^N}\Big[\big(\widehat{v}_N(\theta) - \mathbb{E}_{\mathbb{P}^N}[\widehat{v}_N(\theta)]\big)^2\Big]}
\end{aligned}$$

$$
\begin{aligned}
&= \sqrt{\mathbb{E}_{\mathbb{P}^N}\left[\left(\left(\widehat{\upsilon}_N(0) - \mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(0)]\right) + \left[\left(\widehat{\upsilon}_N(\theta) - \mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(\theta)]\right) - \left(\widehat{\upsilon}_N(0) - \mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(0)]\right)\right]\right)^2\right]} \\
&\le \sqrt{\mathbb{E}_{\mathbb{P}^N}\left[\left(\widehat{\upsilon}_N(0) - \mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(0)]\right)^2\right]} + \sqrt{\mathbb{E}_{\mathbb{P}^N}\left[\left(\left\{\widehat{\upsilon}_N(\theta) - \mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(\theta)]\right\} - \left\{\widehat{\upsilon}_N(0) - \mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(0)]\right\}\right)^2\right]} \\
&\le \sqrt{\mathrm{Var}_{\mathbb{P}^N}\left(\widehat{\upsilon}_N(0)\right)} + 2\theta(3-2\theta)\big(L\cdot \mathrm{diam}(\Xi) + M\big),
\end{aligned}
$$

where the first inequality follows from Minkowski's inequality, and the second inequality follows from (EC.18). This completes the proof. □

EC.1.13. Proof of Theorem 8

Proof. First, we prove the following lemma.

LEMMA EC.2. *Let A_1 and A_2 be random variables, $\mathbb{P}$ be any probability distribution, and $\theta \in (0,1)$. For any $\delta > 0$, the following inequality holds.*

$$
\mathbb{P}\big((1-\theta)A_1 + \theta A_2 > \delta\big) \le \inf_{\gamma\in[0,\delta]} \left\{ \mathbb{P}\left(A_1 > \frac{\gamma}{1-\theta}\right) + \mathbb{P}\left(A_2 > \frac{\delta-\gamma}{\theta}\right)\right\} \tag{EC.19}
$$

Proof of Lemma EC.2. Consider the event $(1-\theta)A_1 + \theta A_2 > \delta$. We have either $A_1 > \gamma/(1-\theta)$ or $A_2 > (\delta-\gamma)/\theta$ for any $\gamma \in [0,\delta]$. It follows from the union bound (Boole's inequality) that

$$
\mathbb{P}\big((1-\theta)A_1 + \theta A_2 > \delta\big) \le \mathbb{P}\left(A_1 > \frac{\gamma}{1-\theta}\right) + \mathbb{P}\left(A_2 > \frac{\delta-\gamma}{\theta}\right) \tag{EC.20}
$$

for any $\gamma \in [0,\delta]$. Inequality (EC.19) follows directly from (EC.20), as the right-hand side of (EC.20) holds for any $\gamma \in [0,\delta]$. This completes the proof of Lemma EC.2. □

With Lemma EC.2, we are ready to derive the desired generalization bound as follows:

$$
\begin{aligned}
&\mathbb{P}^N\left(\sup_{\boldsymbol{x}\in\mathcal{X}}\left\{\mathbb{E}_{\mathbb{P}^\star}[f(\boldsymbol{x},\boldsymbol{\xi})] - \sup_{\mathbb{P}\in\mathcal{P}'_{N,\theta}} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\boldsymbol{\xi})]\right\} > \delta\right) \\
&= \mathbb{P}^N\left(\sup_{\boldsymbol{x}\in\mathcal{X}}\left\{(1-\theta)\cdot\Big(\mathbb{E}_{\mathbb{P}^\star}[f(\boldsymbol{x},\boldsymbol{\xi})] - \mathbb{E}_{\widehat{\mathbb{P}}_N}[f(\boldsymbol{x},\boldsymbol{\xi})]\Big) + \theta\cdot\Big(\mathbb{E}_{\mathbb{P}^\star}[f(\boldsymbol{x},\boldsymbol{\xi})] - \sup_{\mathbb{P}\in\mathcal{P}_N}\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\boldsymbol{\xi})]\Big)\right\} > \delta\right) \\
&\le \mathbb{P}^N\left((1-\theta)\sup_{\boldsymbol{x}\in\mathcal{X}}\left\{\mathbb{E}_{\mathbb{P}^\star}[f(\boldsymbol{x},\boldsymbol{\xi})] - \mathbb{E}_{\widehat{\mathbb{P}}_N}[f(\boldsymbol{x},\boldsymbol{\xi})]\right\} + \theta\sup_{\boldsymbol{x}\in\mathcal{X}}\left\{\mathbb{E}_{\mathbb{P}^\star}[f(\boldsymbol{x},\boldsymbol{\xi})] - \sup_{\mathbb{P}\in\mathcal{P}_N}\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\boldsymbol{\xi})]\right\} > \delta\right) \\
&\le \inf_{\gamma\in[0,\delta]}\left\{\mathbb{P}^N\left(\sup_{\boldsymbol{x}\in\mathcal{X}}\left\{\mathbb{E}_{\mathbb{P}^\star}[f(\boldsymbol{x},\boldsymbol{\xi})] - \mathbb{E}_{\widehat{\mathbb{P}}_N}[f(\boldsymbol{x},\boldsymbol{\xi})]\right\} > \frac{\gamma}{1-\theta}\right)\right. \\
&\qquad\qquad \left. + \mathbb{P}^N\left(\sup_{\boldsymbol{x}\in\mathcal{X}}\left\{\mathbb{E}_{\mathbb{P}^\star}[f(\boldsymbol{x},\boldsymbol{\xi})] - \sup_{\mathbb{P}\in\mathcal{P}_N}\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\boldsymbol{\xi})]\right\} > \frac{\delta-\gamma}{\theta}\right)\right\} \qquad \text{(EC.21a)} \\
&\le \inf_{\gamma\in[0,\delta]}\left\{\alpha_{N,1}\left(\frac{\gamma}{1-\theta}\right) + \alpha_{N,2}\left(\frac{\delta-\gamma}{\theta}\right)\right\}. \qquad \text{(EC.21b)}
\end{aligned}
$$

Inequality (EC.21a) follows from (EC.19) and inequality (EC.21b) follows from (19) and (20).

□

EC.1.14. Proof of Lemma 1

Proof. For brevity, all convergence, equalities and inequalities hold almost surely in this proof. If f is uniformly bounded, i.e., $\sup_{\boldsymbol{x}\in\mathcal{X}}\sup_{\boldsymbol{\xi}\in\Xi} f(\boldsymbol{x},\boldsymbol{\xi}) < \infty$, then the desired inequality follows immediately. Now, suppose that Assumption 3(a) holds. We can directly obtain an upper bound from (22):

$$\sup_{\mathbb{P}\in\mathcal{P}_N} \mathbb{E}_{\mathbb{P}}|f(\boldsymbol{x},\boldsymbol{\xi})| \le \sup_{\mathbb{P}\in\widehat{\mathcal{P}}} \mathbb{E}_{\mathbb{P}}[\kappa(\boldsymbol{\xi})]\cdot \operatorname{diam}(\mathcal{X}) + \sup_{\mathbb{P}\in\widehat{\mathcal{P}}} \mathbb{E}_{\mathbb{P}}|f(\boldsymbol{x}_0,\boldsymbol{\xi})| := M < \infty \tag{EC.22}$$

for all sufficiently large N. Thus, we obtain $\limsup_{N\to\infty}\sup_{\boldsymbol{x}\in\mathcal{X}}\sup_{\mathbb{P}\in\mathcal{P}_N}\mathbb{E}_{\mathbb{P}}|f(\boldsymbol{x},\boldsymbol{\xi})| \le M$ by taking supremum over $\boldsymbol{x}\in\mathcal{X}$ on both sides of (EC.22).

Now, suppose that Assumption 3(b) holds. Then, for any $\boldsymbol{x}\in\mathcal{X}$, we have

$$\begin{aligned}
\sup_{\mathbb{P}\in\mathcal{P}_N} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\boldsymbol{\xi})] &\le \left|\sup_{\mathbb{P}\in\mathcal{P}_N} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\boldsymbol{\xi})] - \sup_{\mathbb{P}\in\widehat{\mathcal{P}}} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\boldsymbol{\xi})]\right| + \sup_{\mathbb{P}\in\widehat{\mathcal{P}}} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\boldsymbol{\xi})] \\
&\le \mathbb{H}(\mathcal{P}_N,\widehat{\mathcal{P}}) + \left\{ \sup_{\mathbb{P}\in\widehat{\mathcal{P}}} \mathbb{E}_{\mathbb{P}}[\kappa(\boldsymbol{\xi})]\cdot\operatorname{diam}(\mathcal{X}) + \sup_{\mathbb{P}\in\widehat{\mathcal{P}}} \mathbb{E}_{\mathbb{P}}|f(\boldsymbol{x}_0,\boldsymbol{\xi})| \right\},
\end{aligned} \tag{EC.23}$$

where the first term in (EC.23) follows from the definition of $\mathbb{H}(\mathcal{P}_N,\widehat{\mathcal{P}})$ and the second term in (EC.23) follows from the same argument in (EC.22). Since $\mathbb{H}(\mathcal{P}_N,\widehat{\mathcal{P}})\to 0$, (EC.23) implies that

$$\limsup_{N\to\infty}\sup_{\boldsymbol{x}\in\mathcal{X}}\sup_{\mathbb{P}\in\mathcal{P}_N} \mathbb{E}_{\mathbb{P}}|f(\boldsymbol{x},\boldsymbol{\xi})| \le \sup_{\mathbb{P}\in\widehat{\mathcal{P}}} \mathbb{E}_{\mathbb{P}}[\kappa(\boldsymbol{\xi})]\cdot\operatorname{diam}(\mathcal{X}) + \sup_{\mathbb{P}\in\widehat{\mathcal{P}}} \mathbb{E}_{\mathbb{P}}|f(\boldsymbol{x}_0,\boldsymbol{\xi})| < \infty.$$

Finally, suppose that Assumption 3(c) holds. For any $\mathbb{P}\in\mathcal{P}_N$ identified as $\boldsymbol{p}\in\mathbb{R}^N_+$,

$$\begin{aligned}
\left|\mathbb{E}_{\mathbb{P}}[\kappa(\boldsymbol{\xi})] - \mathbb{E}_{\widehat{\mathbb{P}}_N}[\kappa(\boldsymbol{\xi})]\right| = \left|\sum_{i=1}^N \left(p_i - \frac{1}{N}\right)\kappa(\widehat{\boldsymbol{\xi}}_i)\right| &\le \left\|\boldsymbol{p} - \frac{1}{N}\mathbf{1}\right\|_\infty \cdot \sum_{i=1}^N \kappa(\widehat{\boldsymbol{\xi}}_i) \\
&= \frac{1}{N}\|N\boldsymbol{p}-\mathbf{1}\|_\infty \cdot \sum_{i=1}^N \kappa(\widehat{\boldsymbol{\xi}}_i) \\
&= \|N\boldsymbol{p}-\mathbf{1}\|_\infty \cdot \mathbb{E}_{\widehat{\mathbb{P}}_N}[\kappa(\boldsymbol{\xi})].
\end{aligned} \tag{EC.24}$$

Hence, we have

$$\begin{aligned}
\limsup_{N\to\infty}\sup_{\mathbb{P}\in\mathcal{P}_N} \mathbb{E}_{\mathbb{P}}[\kappa(\boldsymbol{\xi})] &\le \limsup_{N\to\infty}\sup_{\mathbb{P}\in\mathcal{P}_N} \left\{\mathbb{E}_{\widehat{\mathbb{P}}_N}[\kappa(\boldsymbol{\xi})] + \left|\mathbb{E}_{\mathbb{P}}[\kappa(\boldsymbol{\xi})] - \mathbb{E}_{\widehat{\mathbb{P}}_N}[\kappa(\boldsymbol{\xi})]\right|\right\} \\
&\le \limsup_{N\to\infty} \mathbb{E}_{\widehat{\mathbb{P}}_N}[\kappa(\boldsymbol{\xi})] + \limsup_{N\to\infty}\left\{\sup_{\mathbb{P}\in\mathcal{P}_N}\|N\boldsymbol{p}-\mathbf{1}\|_\infty\cdot\mathbb{E}_{\widehat{\mathbb{P}}_N}[\kappa(\boldsymbol{\xi})]\right\} \qquad \text{(EC.25)} \\
&\le \limsup_{N\to\infty} \mathbb{E}_{\widehat{\mathbb{P}}_N}[\kappa(\boldsymbol{\xi})] + \left(\limsup_{N\to\infty}\sup_{\mathbb{P}\in\mathcal{P}_N}\|N\boldsymbol{p}-\mathbf{1}\|_\infty\right)\left(\limsup_{N\to\infty}\mathbb{E}_{\widehat{\mathbb{P}}_N}[\kappa(\boldsymbol{\xi})]\right)
\end{aligned}$$

$$= \left(1 + \limsup_{N\to\infty} \sup_{\mathbb{P}\in\mathcal{P}_N} \|N\boldsymbol{p} - \mathbf{1}\|_\infty\right)\left(\limsup_{N\to\infty} \mathbb{E}_{\widehat{\mathbb{P}}_N}[\kappa(\boldsymbol{\xi})]\right),$$

where the second term in (EC.25) follows from (EC.24). Since $\{\widehat{\boldsymbol{\xi}}_i\}_{i=1}^N$ are i.i.d. following $\mathbb{P}^\star$, the strong law of large numbers gives $\limsup_{N\to\infty} \mathbb{E}_{\widehat{\mathbb{P}}_N}[\kappa(\boldsymbol{\xi})] = \mathbb{E}_{\mathbb{P}^\star}[\kappa(\boldsymbol{\xi})] < \infty$. Moreover, since $\limsup_{N\to\infty} \sup_{\mathbb{P}\in\mathcal{P}_N} \|N\boldsymbol{p} - \mathbf{1}\|_\infty < \infty$, we have $\limsup_{N\to\infty} \sup_{\mathbb{P}\in\mathcal{P}_N} \mathbb{E}_{\mathbb{P}}[\kappa(\boldsymbol{\xi})] < \infty$. Using a similar argument, we can obtain $\limsup_{N\to\infty} \sup_{\mathbb{P}\in\mathcal{P}_N} \mathbb{E}_{\mathbb{P}}|f(\boldsymbol{x}_0, \boldsymbol{\xi})| < \infty$. Therefore, we have

$$\begin{aligned}\limsup_{N\to\infty} \sup_{\boldsymbol{x}\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{P}_N} \mathbb{E}_{\mathbb{P}}|f(\boldsymbol{x},\boldsymbol{\xi})| &\le \limsup_{N\to\infty} \left\{ \sup_{\mathbb{P}\in\mathcal{P}_N} \mathbb{E}_{\mathbb{P}}[\kappa(\boldsymbol{\xi})] \cdot \operatorname{diam}(\mathcal{X}) + \sup_{\mathbb{P}\in\mathcal{P}_N} \mathbb{E}_{\mathbb{P}}|f(\boldsymbol{x}_0,\boldsymbol{\xi})| \right\} \\ &\le \limsup_{N\to\infty} \sup_{\mathbb{P}\in\mathcal{P}_N} \mathbb{E}_{\mathbb{P}}[\kappa(\boldsymbol{\xi})] \cdot \operatorname{diam}(\mathcal{X}) + \limsup_{N\to\infty} \sup_{\mathbb{P}\in\mathcal{P}_N} \mathbb{E}_{\mathbb{P}}|f(\boldsymbol{x}_0,\boldsymbol{\xi})| < \infty.\end{aligned}$$

This completes the proof. □

EC.1.15. Proof of Theorem 9

Proof. For brevity, all convergence, equalities and inequalities hold almost surely in this proof. First, we claim that the following uniform convergence holds:

$$\sup_{\boldsymbol{x}\in\mathcal{X}} \left| \left\{ (1-\theta_N) \cdot \frac{1}{N} \sum_{i=1}^N f(\boldsymbol{x}, \widehat{\boldsymbol{\xi}}_i) + \theta_N \cdot \sup_{\mathbb{P}\in\mathcal{P}_N} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\boldsymbol{\xi})] \right\} - \mathbb{E}_{\mathbb{P}^\star}[f(\boldsymbol{x},\boldsymbol{\xi})] \right| \to 0 \tag{EC.26}$$

as $N \to \infty$. To prove (EC.26), note that

$$\begin{aligned}&\sup_{\boldsymbol{x}\in\mathcal{X}} \left| \left\{ (1-\theta_N) \cdot \frac{1}{N} \sum_{i=1}^N f(\boldsymbol{x}, \widehat{\boldsymbol{\xi}}_i) + \theta_N \cdot \sup_{\mathbb{P}\in\mathcal{P}_N} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\boldsymbol{\xi})] \right\} - \mathbb{E}_{\mathbb{P}^\star}[f(\boldsymbol{x},\boldsymbol{\xi})] \right| \\ &\le (1-\theta_N) \cdot \sup_{\boldsymbol{x}\in\mathcal{X}} \left| \frac{1}{N} \sum_{i=1}^N f(\boldsymbol{x}, \widehat{\boldsymbol{\xi}}_i) - \mathbb{E}_{\mathbb{P}^\star}[f(\boldsymbol{x},\boldsymbol{\xi})] \right| + \theta_N \cdot \sup_{\boldsymbol{x}\in\mathcal{X}} \left| \sup_{\mathbb{P}\in\mathcal{P}_N} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\boldsymbol{\xi})] - \mathbb{E}_{\mathbb{P}^\star}[f(\boldsymbol{x},\boldsymbol{\xi})] \right|.\end{aligned} \tag{EC.27}$$

Assumption 2 implies that the first term $\sup_{\boldsymbol{x}\in\mathcal{X}} \left| N^{-1} \sum_{i=1}^N f(\boldsymbol{x}, \widehat{\boldsymbol{\xi}}_i) - \mathbb{E}_{\mathbb{P}^\star}[f(\boldsymbol{x},\boldsymbol{\xi})] \right|$ converges to zero. Next, for the second term, note that

$$\theta_N \cdot \sup_{\boldsymbol{x}\in\mathcal{X}} \left| \sup_{\mathbb{P}\in\mathcal{P}_N} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\boldsymbol{\xi})] - \mathbb{E}_{\mathbb{P}^\star}[f(\boldsymbol{x},\boldsymbol{\xi})] \right| \le \theta_N \left(\sup_{\boldsymbol{x}\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{P}_N} \mathbb{E}_{\mathbb{P}}|f(\boldsymbol{x},\boldsymbol{\xi})| + \sup_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}_{\mathbb{P}^\star}|f(\boldsymbol{x},\boldsymbol{\xi})| \right). \tag{EC.28}$$

Since $\limsup_{N\to\infty} \sup_{\boldsymbol{x}\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{P}_N} \mathbb{E}_{\mathbb{P}}|f(\boldsymbol{x},\boldsymbol{\xi})| \le M$ by Lemma 1 and $\sup_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}_{\mathbb{P}^\star}|f(\boldsymbol{x},\boldsymbol{\xi})| < \infty$, the upper bound in (EC.28) converges to zero by $\theta_N = o(1)$. Therefore, the upper bound in (EC.27) converges to zero, which shows (EC.26).

Now, with the use of (EC.26), we can prove the desired asymptotic convergence results in a way similar to Theorem 5.3 of Shapiro et al. (2014). For completeness, we also provide the details here. To show assertion (i), note that

$$\begin{aligned}|\widehat{\upsilon}_N(\theta_N)-\upsilon^\star| &= \left|\min_{\boldsymbol{x}\in\mathcal{X}}\left\{(1-\theta_N)\cdot\frac{1}{N}\sum_{i=1}^N f(\boldsymbol{x},\widehat{\boldsymbol{\xi}}_i)+\theta_N\cdot\sup_{\mathbb{P}\in\mathcal{P}_N}\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\boldsymbol{\xi})]\right\}-\min_{\boldsymbol{x}\in\mathcal{X}}\mathbb{E}_{\mathbb{P}^\star}[f(\boldsymbol{x},\boldsymbol{\xi})]\right|\\ &\le \sup_{\boldsymbol{x}\in\mathcal{X}}\left|\left\{(1-\theta_N)\cdot\frac{1}{N}\sum_{i=1}^N f(\boldsymbol{x},\widehat{\boldsymbol{\xi}}_i)+\theta_N\cdot\sup_{\mathbb{P}\in\mathcal{P}_N}\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\boldsymbol{\xi})]\right\}-\mathbb{E}_{\mathbb{P}^\star}[f(\boldsymbol{x},\boldsymbol{\xi})]\right|,\end{aligned}$$

which converges to zero by (EC.26). Next, to show assertion (ii), suppose, on the contrary, that $D\big(\widehat{\mathcal{X}}_N(\theta_N),\mathcal{X}^\star\big)\to 0$ does not hold almost surely, i.e., $\mathbb{P}^\infty\big(D\big(\widehat{\mathcal{X}}_N(\theta_N),\mathcal{X}^\star\big)\not\to 0\big)>0$. Consider a data sequence such that the event $D\big(\widehat{\mathcal{X}}_N(\theta_N),\mathcal{X}^\star\big)\not\to 0$ holds. Then, for some $\epsilon>0$, there exists a sequence $\{\boldsymbol{x}_{N_j}\}_{j\in\mathbb{N}}$ such that $\boldsymbol{x}_{N_j}\in\widehat{\mathcal{X}}_{N_j}(\theta_{N_j})$ and $d\big(\boldsymbol{x}_{N_j},\mathcal{X}^\star\big)>\epsilon$. Since $\mathcal{X}$ is compact by Assumption 1, without loss of generality, we can assume that $\boldsymbol{x}_{N_j}\to\overline{\boldsymbol{x}}$ for some $\overline{\boldsymbol{x}}\in\mathcal{X}$. By continuity of the distance function d, we have $d(\overline{\boldsymbol{x}},\mathcal{X}^\star)=\lim_{j\to\infty}d(\boldsymbol{x}_{N_j},\mathcal{X}^\star)>\epsilon$. Thus, $\overline{\boldsymbol{x}}\notin\mathcal{X}^\star$, which implies that $\mathbb{E}_{\mathbb{P}^\star}[f(\overline{\boldsymbol{x}},\boldsymbol{\xi})]>\upsilon^\star$. However, note that

$$\left|\mathbb{E}_{\mathbb{P}^\star}[f(\overline{\boldsymbol{x}},\boldsymbol{\xi})]-\widehat{\upsilon}_{N_j}(\theta_{N_j})\right|\le\left|\mathbb{E}_{\mathbb{P}^\star}[f(\overline{\boldsymbol{x}},\boldsymbol{\xi})]-\mathbb{E}_{\mathbb{P}^\star}[f(\boldsymbol{x}_{N_j},\boldsymbol{\xi})]\right|+\left|\mathbb{E}_{\mathbb{P}^\star}[f(\boldsymbol{x}_{N_j},\boldsymbol{\xi})]-\widehat{\upsilon}_{N_j}(\theta_{N_j})\right|,$$

where the first term converges to zero by the continuity of $\mathbb{E}_{\mathbb{P}^\star}[f(\cdot,\boldsymbol{\xi})]$, and the second term converges to zero by (EC.26). Therefore, we arrive at $\lim_{j\to\infty}\widehat{\upsilon}_{N_j}(\theta_{N_j})=\mathbb{E}_{\mathbb{P}^\star}[f(\overline{\boldsymbol{x}},\boldsymbol{\xi})]>\upsilon^\star$, which contradicts assertion (i) that $\lim_{j\to\infty}\widehat{\upsilon}_{N_j}(\theta_{N_j})=\upsilon^\star$. □

EC.1.16. Proof of Lemma 2

Proof. Recall that, following the convention in empirical process theory, we view $\mathbb{P}\in\mathcal{P}(\Xi)$ as an element in $\ell^\infty(\mathcal{H})$ defined as $\mathbb{P}(h)=\mathbb{E}_{\mathbb{P}}(h)$ for $h\in\mathcal{H}$. We divide the proof of the desired weak convergence $\mathbb{S}_N\Rightarrow\mathbb{G}'$ into two steps.

Step 1. We first show that $\big(\mathbb{S}_N(h_1),\ldots,\mathbb{S}_N(h_k)\big)\Rightarrow\big(\mathbb{G}'(h_1),\ldots,\mathbb{G}'(h_k)\big)$ for any finite subset $\{h_1,\ldots,h_k\}\subset\mathcal{H}$. To prove this, we write $\mathbb{S}_N=\sqrt{N}(1-\theta_N)(\widehat{\mathbb{P}}_N-\mathbb{P}^\star)+\sqrt{N}\theta_N(\mathbb{P}_N-\mathbb{P}^\star)$. Note that $\sqrt{N}(\widehat{\mathbb{P}}_N-\mathbb{P}^\star)(h_1,\ldots,h_k)\Rightarrow\big(\mathbb{G}'(h_1),\ldots,\mathbb{G}'(h_k)\big)$ by Assumption 4. Also, from the assumption that $\theta_N=o(N^{-1/2})$, we have $(1-\theta_N)\to 1$. Hence, for the first term in $\mathbb{S}_N$, we have $(1-\theta_N)\cdot\sqrt{N}(\widehat{\mathbb{P}}_N-\mathbb{P}^\star)(h_1,\ldots,h_k)\Rightarrow\big(\mathbb{G}'(h_1),\ldots,\mathbb{G}'(h_k)\big)$ by Slutsky's Theorem (see, e.g., Example 1.4.7 of van der Vaart and Wellner 1996). Next, for the second term in $\mathbb{S}_N$, Lemma 1 implies that $\limsup_{N\to\infty}\sup_{h\in\mathcal{H}}\sup_{\mathbb{P}\in\mathcal{P}_N}\mathbb{E}_{\mathbb{P}}|h|\le M$ almost surely for some constant M. Therefore, there exists constant M' such that $\limsup_{N\to\infty}|\mathbb{E}_{\mathbb{P}_N}(h)-\mathbb{E}_{\mathbb{P}^\star}(h)|\le$

$\limsup_{N\to\infty}\sup_{h\in\mathcal{H}}\sup_{\mathbb{P}\in\mathcal{P}_N}|\mathbb{E}_{\mathbb{P}}(h)-\mathbb{E}_{\mathbb{P}^\star}(h)|\leq M'$ for any $\mathbb{P}_N\in\mathcal{P}_N$ and $h\in\mathcal{H}$ (since $\mathbb{E}_{\mathbb{P}^\star}|h|<\infty$). By assumption (b), $\sqrt{N}\theta_N$ converges to zero as $N\to\infty$ and hence, $\sqrt{N}\theta_N(\mathbb{P}_N-\mathbb{P}^\star)(h)$ also converges to zero almost surely. This implies that $\sqrt{N}\theta_N(\mathbb{P}_N-\mathbb{P}^\star)(h_1,\ldots,h_k)\to(0,\ldots,0)\in\mathbb{R}^k$ almost surely. Therefore, provoking Slutsky's Theorem again, we obtain the desired convergence, i.e., $\big(\mathbb{S}_N(h_1),\ldots,\mathbb{S}_N(h_k)\big)\Rightarrow\big(\mathbb{G}'(h_1),\ldots,\mathbb{G}'(h_k)\big)$. This completes step 1.

Step 2. We show that $\mathbb{S}_N$ is asymptotically tight. Note that for any $\varepsilon>0$,

$$\limsup_{\delta\to 0}\limsup_{N\to\infty}\mathbb{P}^N\left(\sup_{\|h-h'\|_{L^2(\mathbb{P}^\star)}<\delta}\big|\mathbb{S}_N(h-h')\big|\geq\varepsilon\right) \tag{EC.29a}$$

$$\leq\limsup_{\delta\to 0}\limsup_{N\to\infty}\mathbb{P}^N\left(\sup_{\|h-h'\|_{L^2(\mathbb{P}^\star)}<\delta}\big|(1-\theta_N)\sqrt{N}(\widehat{\mathbb{P}}_N-\mathbb{P}^\star)(h-h')\big|\geq\frac{\varepsilon}{2}\right) \tag{EC.29b}$$

$$+\limsup_{\delta\to 0}\limsup_{N\to\infty}\mathbb{P}^N\left(\sup_{\|h-h'\|_{L^2(\mathbb{P}^\star)}<\delta}\big|\theta_N\sqrt{N}(\mathbb{P}_N-\mathbb{P}^\star)(h-h')\big|\geq\frac{\varepsilon}{2}\right). \tag{EC.29c}$$

Since $\sqrt{N}(1-\theta_N)(\widehat{\mathbb{P}}_N-\mathbb{P}^\star)\Rightarrow\mathbb{G}'$, the sequence $\{\sqrt{N}(1-\theta_N)(\widehat{\mathbb{P}}_N-\mathbb{P}^\star)\}_{N\in\mathbb{N}}$ is asymptotically tight. Therefore, (EC.29b) equals zero by Theorem 1.5.7 of van der Vaart and Wellner (1996). Now, we show that (EC.29c) also vanishes. It suffices to show that the sequence $\{\sqrt{N}\theta_N(\mathbb{P}_N-\mathbb{P}^\star)\}_{N\in\mathbb{N}}$ is asymptotically tight. Note that

$$\mathbb{Q}_N(h):=\sqrt{N}\theta_N(\mathbb{P}_N-\mathbb{P}^\star)(h)=\sqrt{N}\theta_N[\mathbb{E}_{\mathbb{P}_N}(h)-\mathbb{E}_{\mathbb{P}^\star}(h)].$$

By assumption (a), the metric space $(\mathcal{H},\|\cdot\|_{L^2(\mathbb{P}^\star)})$ is totally bounded. Also, as shown earlier (in step 1), $\mathbb{Q}_N(h)$ converges to zero almost surely. It follows by Theorem 1.5.4 of van der Vaart and Wellner (1996) that $\mathbb{Q}_N(h)$ is asymptotically tight for any $h\in\mathcal{H}$. Next, we claim that $\mathbb{Q}_N$ is asymptotically uniform equicontinuous in probability, i.e., for any $\varepsilon>0$, and $\eta>0$, there exists $\delta>0$ such that

$$\limsup_{N\to\infty}\mathbb{P}^N\left(\sup_{\|h-h'\|_{L^2(\mathbb{P}^\star)}<\delta}|\mathbb{Q}_N(h-h')|>\varepsilon\right)<\eta.$$

For notational simplicity, we write $\|h-h'\|=\|h-h'\|_{L^2(\mathbb{P}^\star)}$. Note that

$$\begin{aligned}\sup_{\|h-h'\|<\delta}\big|\mathbb{Q}_N(h-h')\big|&=\sup_{\|h-h'\|<\delta}\big|\sqrt{N}\theta_N(\mathbb{P}_N-\mathbb{P}^\star)(h-h')\big|\\&\leq\sqrt{N}\theta_N\sup_{\|h-h'\|<\delta}\Big\{\big|\mathbb{P}_N(h-h')\big|+\big|\mathbb{P}^\star(h-h')\big|\Big\}\\&\leq\sqrt{N}\theta_N\sup_{\|h-h'\|<\delta}\big|\mathbb{P}_N(h-h')\big|+\sqrt{N}\theta_N\sup_{\|h-h'\|<\delta}\big|\mathbb{P}^\star(h-h')\big|.\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&\mathbb{P}^N\left(\sup_{\|h-h'\|_{L^2(\mathbb{P}^\star)}<\delta}|\mathbb{Q}_N(h-h')|>\varepsilon\right)\\
&\le \mathbb{P}^N\left(\sqrt{N}\theta_N\sup_{\|h-h'\|<\delta}|\mathbb{P}_N(h-h')|>\frac{\varepsilon}{2}\right)+\mathbb{P}^N\left(\sqrt{N}\theta_N\sup_{\|h-h'\|<\delta}|\mathbb{P}^\star(h-h')|>\frac{\varepsilon}{2}\right)\\
&=: A_N+B_N.
\end{aligned}\tag{EC.30}$$

Consider the term B_N in (EC.30). For any h and h' such that $\|h-h'\|<\delta$, we have $|\mathbb{P}^\star(h-h')|=|\mathbb{E}_{\mathbb{P}^\star}(h-h')|\le\mathbb{E}_{\mathbb{P}}|h-h'|\le\sqrt{\mathbb{E}_{\mathbb{P}^\star}[(h-h')^2]}=\|h-h'\|<\delta$, where the second inequality follows from Cauchy-Schwartz inequality. Also, by assumption (b), we have $\sqrt{N}\theta_N=o(1)$. Therefore,

$$\limsup_{N\to\infty}B_N\le\mathbb{P}^\infty\left(\limsup_{N\to\infty}\left\{\sqrt{N}\theta_N\delta\right\}>\frac{\varepsilon}{2}\right)=0.\tag{EC.31}$$

Consider the term A_N in (EC.30). Note that for any $\mathbb{P}_N\in\mathcal{P}_N$,

$$\sup_{\|h-h'\|<\delta}\mathbb{E}_{\mathbb{P}_N}|h-h'|\le 2\sup_{h\in\mathcal{H}}\mathbb{E}_{\mathbb{P}_N}|h|\le 2\sup_{h\in\mathcal{H}}\sup_{\mathbb{P}\in\mathcal{P}_N}\mathbb{E}_{\mathbb{P}}|h|.\tag{EC.32}$$

By Lemma 1, (EC.32) implies that $\limsup_{N\to\infty}\sup_{\|h-h'\|<\delta}\mathbb{E}_{\mathbb{P}_N}|h-h'|\le M''$ almost surely for some constant M''. Since $\sqrt{N}\theta_N=o(1)$, we have $\limsup_{N\to\infty}\left\{\sqrt{N}\theta_N\sup_{\|h-h'\|<\delta}\mathbb{E}_{\mathbb{P}_N}|h-h'|\right\}=0$. Therefore, we have

$$\limsup_{N\to\infty}A_N\le\mathbb{P}^\infty\left(\limsup_{N\to\infty}\left\{\sqrt{N}\theta_N\sup_{\|h-h'\|<\delta}\mathbb{E}_{\mathbb{P}_N}|h-h'|\right\}>\frac{\varepsilon}{2}\right)=0.\tag{EC.33}$$

Combining (EC.31) and (EC.33) with (EC.30), we have

$$0\le\limsup_{N\to\infty}\mathbb{P}^N\left(\sup_{\|h-h'\|_{L^2(\mathbb{P}^\star)}<\delta}|\mathbb{Q}_N(h-h')|>\varepsilon\right)\le\limsup_{N\to\infty}A_N+\limsup_{N\to\infty}B_N\le 0.\tag{EC.34}$$

Since (EC.34) holds for any $\delta>0$, this shows that $\mathbb{Q}_N$ is asymptotically uniform continuous in probability. Since (a) the metric space $(\mathcal{H},\|\cdot\|_{L^2(\mathbb{P}^\star)})$ is totally bounded, (b) $\mathbb{Q}_N(h)$ is asymptotically tight for any $h\in\mathcal{H}$, and (c) $\mathbb{Q}_N$ is asymptotically uniform continuous in probability, $\mathbb{Q}_N$ is asymptotically tight by Theorem 1.5.7 of van der Vaart and Wellner (1996), implying that (EC.29c) equals zero. Since both (EC.29b) and (EC.29c) equal zero, (EC.29a) also equals zero, showing that the sequence $\{\mathbb{S}_N\}_{N\in\mathbb{N}}$ is asymptotically tight. This completes step 2.

Combining the two steps, we have (a) $\{\mathbb{S}_N\}$ is asymptotically tight and (b) the marginals $\big(\mathbb{S}_N(h_1),\ldots,\mathbb{S}_N(h_k)\big)$ converge weakly to $\big(\mathbb{G}'(h_1),\ldots,\mathbb{G}'(h_k)\big)$. It follows from Theorem 1.5.4 of van der Vaart and Wellner (1996) that $\mathbb{S}_N\Rightarrow\mathbb{G}'$. $\square$

EC.1.17. Proof of Theorem 10

Proof. By our assumption, there exists $\mathbb{P}^\star_N \in \arg\max_{\mathbb{P}\in\mathcal{P}_N} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\boldsymbol{\xi})]$ such that $\mathbb{P}^\star_N \in \mathcal{P}_N$ for any $\boldsymbol{x}\in\mathcal{X}$. Thus, by Lemma 2, we have

$$\sqrt{N}\left[(1-\theta_N)\widehat{\mathbb{P}}_N(\cdot)+\theta_N \sup_{\mathbb{P}\in\mathcal{P}_N}\mathbb{P}(\cdot)-\mathbb{P}^\star(\cdot)\right]\Rightarrow \mathbb{G}'(\cdot) \text{ in } \ell^\infty(\mathcal{H}), \tag{EC.35}$$

where $\sup_{\mathbb{P}\in\mathcal{P}_N}\mathbb{P}(\cdot)\in\mathcal{P}_N$ denotes the worst-case distribution of the input function $h\in\mathcal{H}$. Recall that $\mathcal{H}=\{f(\boldsymbol{x},\cdot)\mid \boldsymbol{x}\in\mathcal{X}\}$. Note that the map from $\ell^\infty(\mathcal{H})$ to $\ell^\infty(\mathcal{X})$ given by $g(\cdot)\mapsto g(h(\cdot,\cdot))$ is continuous, where $\ell^\infty(\mathcal{X})$ is the Banach space of bounded functions $\psi:\mathcal{X}\to\mathbb{R}$ equipped with the supremum norm $\|\psi\|=\sup_{\boldsymbol{x}\in\mathcal{X}}|\psi(\boldsymbol{x})|$. By continuous mapping theorem (see Theorem 1.3.6 of van der Vaart and Wellner 1996), (EC.35) implies that

$$\sqrt{N}\left[(1-\theta_N)\mathbb{E}_{\widehat{\mathbb{P}}_N}[f(\cdot,\boldsymbol{\xi})]+\theta_N\sup_{\mathbb{P}\in\mathcal{P}_N}\mathbb{E}_{\mathbb{P}}[f(\cdot,\boldsymbol{\xi})]-\mathbb{E}_{\mathbb{P}^\star}[f(\cdot,\boldsymbol{\xi})]\right]\Rightarrow \mathbb{G}(\cdot) \text{ in } \ell^\infty(\mathcal{X}). \tag{EC.36}$$

Consider the functional $V:\ell^\infty(\mathcal{X})\to\mathbb{R}$ by $V(\psi)=\inf_{\boldsymbol{x}\in\mathcal{X}}\psi(\boldsymbol{x})$. Since $\mathcal{X}$ is compact by Assumption 1, the Hadamard directional derivative of V at ψ is given by $V'_\psi(\phi)=\inf_{\boldsymbol{x}\in B(\psi)}\phi(\boldsymbol{x})$, where $B(\psi)=\arg\min_{\boldsymbol{x}\in\mathcal{X}}\psi(\boldsymbol{x})$ (see, e.g., Corollary 2.2 of Cárcamo et al. 2020). Thus, together with (EC.36), applying the Delta's method (see, e.g., Theorem 2.2 of Cárcamo et al. 2020), we obtain the desired assertions: (i) $\sqrt{N}(\widehat{v}_N-v^\star)\Rightarrow\inf_{\boldsymbol{x}\in\mathcal{X}^\star}\mathbb{G}(\boldsymbol{x})$ and (ii)

$$\widehat{v}_N-v^\star=\inf_{\boldsymbol{x}\in\mathcal{X}^\star}\left\{(1-\theta_N)\mathbb{E}_{\widehat{\mathbb{P}}_N}[f(\boldsymbol{x},\boldsymbol{\xi})]+\theta_N\sup_{\mathbb{P}\in\mathcal{P}_N}\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\boldsymbol{\xi})]-\mathbb{E}_{\mathbb{P}^\star}[f(\boldsymbol{x},\boldsymbol{\xi})]\right\}+o_{\mathbb{P}^\star}(N^{-1/2}). \tag{EC.37}$$

Finally, since $-\infty<v^\star=\mathbb{E}_{\mathbb{P}^\star}[f(\boldsymbol{x},\boldsymbol{\xi})]$ for any $\boldsymbol{x}\in\mathcal{X}^\star$, (EC.37) directly implies $\widehat{v}_N=\inf_{\boldsymbol{x}\in\mathcal{X}^\star}\left\{(1-\theta_N)\mathbb{E}_{\mathbb{P}^\star}[f(\boldsymbol{x},\boldsymbol{\xi})]+\theta_N\sup_{\mathbb{P}\in\mathcal{P}_N}\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\boldsymbol{\xi})]\right\}+o_{\mathbb{P}^\star}(N^{-1/2})$. This completes the proof. □

EC.1.18. Proof of Theorem 11

Proof. Using (23), we can rewrite the objective function of the TRO model with shape parameter $\mathcal{P}_{N,r_N}$ as follows:

$$\begin{aligned}\sup_{\mathbb{P}\in\mathcal{P}'_{N,\theta}}\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\boldsymbol{\xi})]&=(1-\theta_N)\mathbb{E}_{\widehat{\mathbb{P}}_N}[f(\boldsymbol{x},\boldsymbol{\xi})]+\theta_N\sup_{\mathbb{P}\in\mathcal{P}_{N,r_N}}\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\boldsymbol{\xi})]\\&=\mathbb{E}_{\widehat{\mathbb{P}}_N}[f(\boldsymbol{x},\boldsymbol{\xi})]+\left(\theta_N r_N^\gamma\right)g_N(\boldsymbol{x})+\left(\theta_N r_N^\gamma\right)\varepsilon_N(\boldsymbol{x}),\end{aligned}$$

which resembles the expansion (23) with r_N^γ replaced by $\theta_N r_N^\gamma$. Thus, we can prove the desired assertions by following the same proof techniques of Theorem 1 in Blanchet and Shapiro (2023). □

EC.2. Additional Discussions

EC.2.1. Differences between the Huber Contamination Model and Our TRO Model

As mentioned in Section 1, the form of our TRO ambiguity set resembles those considered in robust statistics, particularly in Huber contamination models (Huber 1964). However, the underlying idea of the Huber contamination model fundamentally differs from that of our TRO model. In robust statistics, the Huber contamination model is used to address data contamination or outliers. Specifically, in this model, the data $\{\widehat{\boldsymbol{\xi}}_i\}_{i=1}^N$ is assumed to be drawn from a mixture distribution $(1-\varepsilon)\mathbb{P}_0 + \varepsilon\mathbb{Q}$, where $\mathbb{P}_0$ is the distribution of interest (e.g., the true distribution), $\mathbb{Q} \in \mathcal{P}(\Xi)$ represents some arbitrary distribution, and ε represents the contamination ratio (see, e.g., Chen et al. 2018, Copas 1988, Huber 1964, Mu and Xiong 2023). This stream of literature often focuses on developing statistical procedures to estimate the distribution of interest $\mathbb{P}_0$. In contrast, as discussed in Section 1, we consider the case where the true distribution $\mathbb{P}^\star$ of $\boldsymbol{\xi}$ is unknown. We assume that one has a (potentially small) set of historical observations $\{\widehat{\boldsymbol{\xi}}_i\}_{i=1}^N$ of $\boldsymbol{\xi}$ from the unknown true distribution $\mathbb{P}^\star$ (i.e., we do not assume that the data is contaminated). Our TRO model is an alternative approach for modeling uncertainty in problem (1) that serves as a middle ground between the optimistic approach, which adopts a distributional belief, and the pessimistic approach, which protects against distributional ambiguity. The TRO ambiguity set $\mathcal{P}'_{N,\theta} = \left\{(1-\theta)\widehat{\mathbb{P}}_N + \theta\mathbb{Q} \mid \mathbb{Q} \in \mathcal{P}_N\right\}$ is a key ingredient of our TRO model in (4). The size parameter $\theta \in [0,1]$ controls the trade-off between solving the problem under a distributional belief and solving it under the worst-case distribution that resides in the shape parameter $\mathcal{P}_N$. Hence, θ plays a different role in our model than ε in the Huber contamination model (which quantifies the contamination level in the data).

EC.2.2. Additional Clarification Related to Remark 1

Suppose that the (data-driven) ambiguity set $\mathcal{P}_{N,\alpha}$ satisfies $\mathbb{P}^N(\mathbb{P}^\star \in \mathcal{P}_{N,\alpha}) \geq 1-\alpha$ for some small $\alpha \in (0,1)$. In this case, the set $\mathcal{P}_{N,\alpha}$ potentially consists of a wide range of distributions such that $\mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(\theta) \mid \mathbb{P}^\star \in \mathcal{P}_{N,\alpha}] \geq \upsilon^\star + \Delta(\alpha)$ for some $\Delta(\alpha) > 0$, i.e., the expected value of $\widehat{\upsilon}_N(\theta)$ (conditional on $\mathbb{P}^\star \in \mathcal{P}_{N,\alpha}$) is strictly greater than the true optimal value $\upsilon^\star$. If the function f is bounded from below, say, $f(\boldsymbol{x},\boldsymbol{\xi}) \geq M$ for all $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{\xi} \in \Xi$, then

$$\begin{aligned}\mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(\theta)] &= \mathbb{P}^N(\mathbb{P}^\star \in \mathcal{P}_{N,\alpha})\mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(\theta) \mid \mathbb{P}^\star \in \mathcal{P}_{N,\alpha}] + \mathbb{P}^N(\mathbb{P}^\star \notin \mathcal{P}_{N,\alpha})\mathbb{E}_{\mathbb{P}^N}[\widehat{\upsilon}_N(\theta) \mid \mathbb{P}^\star \notin \mathcal{P}_{N,\alpha}] \\ &\geq (1-\alpha)[\upsilon^\star + \Delta(\alpha)] + M\mathbb{P}^N(\mathbb{P}^\star \notin \mathcal{P}_{N,\alpha}). \end{aligned} \tag{EC.38}$$

If we pick a sufficiently small α, the probability $\mathbb{P}^N(\mathbb{P}^\star \notin \mathcal{P}_{N,\alpha})$ can be arbitrarily small since it satisfies $\mathbb{P}^N(\mathbb{P}^\star \notin \mathcal{P}_{N,\alpha}) \leq \alpha$. As α decreases, the value of $\Delta(\alpha)$ is non-decreasing since $\mathcal{P}_{N,\alpha}$

consists of a larger number of distributions. Thus, the first term in (EC.38) increases when α decreases. Therefore, we can choose α sufficiently small such that $\mathbb{E}_{\mathbb{P}^N}[\widehat{v}_N(\theta)] \geq v^\star$.

EC.3. An Example of a Sequence of TRO Ambiguity Sets

Let $\widehat{\mathbb{P}}_N = \delta_0$, i.e., the Dirac measure on 0, and $\mathcal{P}_N = \{(1-t)\delta_1 + t\delta_e \mid t \in [0,1],\, e \in \{0,2\}\}$. That is, $\mathcal{P}_N$ contains the one-point distributions $\{\delta_0, \delta_1, \delta_2\}$, as well as all two-point distributions with support on either $\{0,1\}$ or $\{1,2\}$. Note that $\mathcal{P}_N$ is star-shaped with a star center δ_1. Indeed, for any $\alpha \in [0,1]$ and $\mathbb{Q} = (1-t)\delta_1 + t\delta_e \in \mathcal{P}_N$, we have

$$(1-\alpha)\delta_1 + \alpha\mathbb{Q} = (1-\alpha)\delta_1 + \alpha\big[(1-t)\delta_1 + t\delta_e\big] = (1-\alpha t)\delta_1 + \alpha t\delta_e \in \mathcal{P}_N$$

since $\alpha t \in [0,1]$. However, $\widehat{\mathbb{P}}_N = \delta_0$ is not a star center. To see this, note that

$$\frac{1}{2}\delta_0 + \frac{1}{2}\left(\frac{1}{2}\delta_1 + \frac{1}{2}\delta_2\right) = \frac{1}{2}\delta_0 + \frac{1}{4}\delta_1 + \frac{1}{4}\delta_2 \notin \mathcal{P}_N$$

since it is a three-point distribution.

Now, we show that $\mathcal{P}'_{N,\theta}$ is *not* non-decreasing. In particular, we show that for any $0 < \theta_1 < \theta_2 \leq 1$, there exists $\mathbb{M} \in \mathcal{P}'_{N,\theta_1}$ but $\mathbb{M} \notin \mathcal{P}'_{N,\theta_2}$. Indeed, since $\frac{1}{2}\delta_1 + \frac{1}{2}\delta_2 \in \mathcal{P}_N$, we construct the measure $\mathbb{M} \in \mathcal{P}'_{N,\theta_1}$ as follows:

$$\mathbb{M} = (1-\theta_1)\delta_0 + \theta_1\left(\frac{1}{2}\delta_1 + \frac{1}{2}\delta_2\right). \tag{EC.39}$$

We show that $\mathbb{M}$ defined in (EC.39) does not belong to $\mathcal{P}'_{N,\theta_2}$. Note that

$$\begin{aligned}\mathbb{M} = (1-\theta_1)\delta_0 + \theta_1\left(\frac{1}{2}\delta_1 + \frac{1}{2}\delta_2\right) &= (1-\theta_2)\delta_0 + (\theta_2-\theta_1)\delta_0 + \theta_1\left(\frac{1}{2}\delta_1 + \frac{1}{2}\delta_2\right)\\ &= (1-\theta_2)\delta_0 + \theta_2\left\{\left(1-\frac{\theta_1}{\theta_2}\right)\delta_0 + \frac{\theta_1}{\theta_2}\left(\frac{1}{2}\delta_1 + \frac{1}{2}\delta_2\right)\right\}.\end{aligned}$$

Since $(1-\frac{\theta_1}{\theta_2})\delta_0 + \frac{\theta_1}{\theta_2}(\frac{1}{2}\delta_1 + \frac{1}{2}\delta_2)$ is a three-point distribution that does not belong to $\mathcal{P}_N$, we have $\mathbb{M} \notin \mathcal{P}'_{N,\theta_2}$.

EC.4. Robust Optimization Ambiguity Set

PROPOSITION EC.1. *For a fixed $\boldsymbol{x} \in \mathcal{X}$, if there exists $\boldsymbol{\xi}_0 \in \mathcal{U}$ such that $f(\boldsymbol{x},\boldsymbol{\xi}_0) \geq f(\boldsymbol{x},\widehat{\boldsymbol{\xi}}_i)$ for all $i \in \{1,\dots,N\}$, then $\sup_{\boldsymbol{\xi}\in\mathcal{U}} f(\boldsymbol{x},\boldsymbol{\xi}) = \sup_{\mathbb{P}\in\mathcal{P}_N} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\boldsymbol{\xi})]$, where $\mathcal{P}_N = \operatorname{conv}\left(\widehat{\mathbb{P}}_N \cup \left\{\delta_{\boldsymbol{\xi}} \mid \boldsymbol{\xi} \in \mathcal{U}\right\}\right)$.*

Proof. First, if $\sup_{\boldsymbol{\xi}\in\mathcal{U}} f(\boldsymbol{x},\boldsymbol{\xi}) = \infty$, then there exists a sequence $\{\boldsymbol{\xi}_j\}_{j\in\mathbb{N}} \subset \mathcal{U}$ such that $f(\boldsymbol{x},\boldsymbol{\xi}_j) \to \infty$ as $j \to \infty$. Since $\boldsymbol{\xi}_j \in \mathcal{U}$, we have $\delta_{\boldsymbol{\xi}_j} \in \mathcal{P}_N$ and $\mathbb{E}_{\delta_{\boldsymbol{\xi}_j}}[f(\boldsymbol{x},\boldsymbol{\xi})] = f(\boldsymbol{x},\boldsymbol{\xi}_j) \to \infty$ as $j \to \infty$. This shows that $\sup_{\mathbb{P}\in\mathcal{P}_N} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\boldsymbol{\xi})] = \infty$.

Now, assume that $\sup_{\boldsymbol{\xi}\in\mathcal{U}} f(\boldsymbol{x},\boldsymbol{\xi}) < \infty$, and thus $f(\boldsymbol{x},\boldsymbol{\xi}) < \infty$ for all $\boldsymbol{\xi} \in \mathcal{U}$. Note that for any $\mathbb{P} \in \mathcal{P}_N$, by definition of $\mathcal{P}_N$, there exists $K \in \mathbb{N}$, $\lambda \in [0,1]$, and $\{\alpha_k\}_{k=1}^K \subseteq [0,1]$ with $\alpha_k \geq 0$ and $\sum_{k=1}^K \alpha_k = 1$ such that $\mathbb{P} = (1-\lambda)\widehat{\mathbb{P}}_N + \lambda \sum_{k=1}^K \alpha_i \delta_{\bar{\boldsymbol{\xi}}_k}$ for some $\bar{\boldsymbol{\xi}}_k \in \mathcal{U}$, $k \in \{1,\dots,K\}$. Then,

$$\begin{aligned}
&\sup_{\mathbb{P}\in\mathcal{P}_N} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\boldsymbol{\xi})] \\
&= \sup \left\{ (1-\lambda)\cdot \frac{1}{N}\sum_{i=1}^N f(\boldsymbol{x},\widehat{\boldsymbol{\xi}}_i) + \lambda \sum_{k=1}^K \alpha_k f(\boldsymbol{x},\bar{\boldsymbol{\xi}}_k) \;\middle|\; \begin{array}{l} K\in\mathbb{N},\, \lambda\in[0,1],\, \alpha_k\in[0,1], \\ \sum_{k=1}^K \alpha_k = 1,\, \bar{\boldsymbol{\xi}}_k \in \mathcal{U},\, k\in\{1,\dots,K\} \end{array} \right\} \quad \text{(EC.40a)} \\
&= \sup_{\lambda\in[0,1]} \left\{ (1-\lambda)\cdot \frac{1}{N}\sum_{i=1}^N f(\boldsymbol{x},\widehat{\boldsymbol{\xi}}_i) + \lambda\cdot \sup\left\{ \sum_{k=1}^K \alpha_k f(\boldsymbol{x},\bar{\boldsymbol{\xi}}_k) \;\middle|\; \begin{array}{l} K\in\mathbb{N},\, \alpha_k\in[0,1], \sum_{k=1}^K \alpha_k = 1, \\ \bar{\boldsymbol{\xi}}_k \in \mathcal{U},\, k\in\{1,\dots,K\} \end{array} \right\} \right\} \\
&= \sup_{\lambda\in[0,1]} \left\{ (1-\lambda)\cdot \frac{1}{N}\sum_{i=1}^N f(\boldsymbol{x},\widehat{\boldsymbol{\xi}}_i) + \lambda\cdot \sup_{\boldsymbol{\xi}\in\mathcal{U}} f(\boldsymbol{x},\boldsymbol{\xi}) \right\} \quad \text{(EC.40b)} \\
&= \sup_{\boldsymbol{\xi}\in\mathcal{U}} f(\boldsymbol{x},\boldsymbol{\xi}). \quad \text{(EC.40c)}
\end{aligned}$$

Equality (EC.40a) follows from $\mathbb{P} \in \mathcal{P}_N$ and the definition of $\mathcal{P}_N$. Equality (EC.40b) follows from the fact that

$$\sum_{k=1}^K \alpha_k f(\boldsymbol{x},\bar{\boldsymbol{\xi}}_k) \leq \sup_{\boldsymbol{\xi}\in\mathcal{U}} f(\boldsymbol{x},\boldsymbol{\xi}) \leq \sup\left\{ \sum_{k=1}^K \alpha_k f(\boldsymbol{x},\bar{\boldsymbol{\xi}}_k) \;\middle|\; \begin{array}{l} K\in\mathbb{N},\, \alpha_k\in[0,1], \sum_{k=1}^K \alpha_k = 1, \\ \bar{\boldsymbol{\xi}}_k \in \mathcal{U},\, k\in\{1,\dots,K\} \end{array} \right\} \quad \text{(EC.41)}$$

for any $k \in \mathbb{N}$, $\alpha_k \in [0,1]$ and $\bar{\boldsymbol{\xi}}_k \in \mathcal{U}$ for $k \in \{1,\dots,K\}$ with $\sum_{k=1}^K \alpha_k = 1$. Here, the first inequality in (EC.41) follows from $f(\boldsymbol{x},\bar{\boldsymbol{\xi}}) \leq \sup_{\boldsymbol{\xi}\in\mathcal{U}} f(\boldsymbol{x},\boldsymbol{\xi})$ for all $\bar{\boldsymbol{\xi}} \in \mathcal{U}$ while the second inequality in (EC.41) follows from letting $K = 1$. Taking supremum over K, $\{\alpha_k\}_{k=1}^K$ and $\{\bar{\boldsymbol{\xi}}_k\}_{k=1}^K$ in (EC.41), we obtain the desired equality in (EC.40b). Finally, equality (EC.40c) follows from the fact that $\lambda = 1$ is optimal to (EC.40b) since

$$f(\boldsymbol{x},\widehat{\boldsymbol{\xi}}_i) \leq f(\boldsymbol{x},\boldsymbol{\xi}_0) \leq \sup_{\boldsymbol{\xi}\in\mathcal{U}} f(\boldsymbol{x},\boldsymbol{\xi})$$

for all $i \in \{1,\dots,N\}$, where the first inequality follows from our assumption that $f(\boldsymbol{x},\boldsymbol{\xi}_0) \geq f(\boldsymbol{x},\widehat{\boldsymbol{\xi}}_i)$ for all $i \in \{1,\dots,N\}$. This shows that $(1/N)\sum_{i=1}^N f(\boldsymbol{x},\widehat{\boldsymbol{\xi}}_i) \leq \sup_{\boldsymbol{\xi}\in\mathcal{U}} f(\boldsymbol{x},\boldsymbol{\xi})$. $\square$

EC.5. Convergence of Distance-Based Ambiguity Sets

PROPOSITION EC.2. *Consider the distance-based ambiguity set* $\mathcal{P}_N = \{\mathbb{P} \in \mathcal{P}(\Xi) \mid \mathsf{d}(\mathbb{P}, \widehat{\mathbb{P}}_N) \leq r\}$ *for some radius* $r > 0$*, where* d *is any statistical distance satisfying*

(i) (normalization) $\mathsf{d}(\mathbb{P}, \mathbb{P}) = 0$ *for any* $\mathbb{P} \in \mathcal{P}(\Xi)$*,*

(ii) (symmetry) $\mathsf{d}(\mathbb{P}_1, \mathbb{P}_2) = \mathsf{d}(\mathbb{P}_2, \mathbb{P}_1)$ *for any* $\{\mathbb{P}_1, \mathbb{P}_2\} \subseteq \mathcal{P}(\Xi)$*,*

(iii) (triangle inequality) $\mathsf{d}(\mathbb{P}_1, \mathbb{P}_2) \leq \mathsf{d}(\mathbb{P}_1, \mathbb{P}_3) + \mathsf{d}(\mathbb{P}_3, \mathbb{P}_2)$ *for any* $\{\mathbb{P}_1, \mathbb{P}_2, \mathbb{P}_3\} \subseteq \mathcal{P}(\Xi)$*, and*

(iv) (convexity) d *is convex in the first argument.*

Let $\widehat{\mathcal{P}} = \{\mathbb{P} \in \mathcal{P}(\Xi) \mid \mathsf{d}(\mathbb{P}, \mathbb{P}^\star) \leq r\}$*. If* $\Delta := \sup_{\mathbb{P}_1 \in \mathcal{P}(\Xi), \mathbb{P}_2 \in \mathcal{P}(\Xi)} \mathbb{d}(\mathbb{P}_1, \mathbb{P}_2) < \infty$ *and* $\mathsf{d}(\widehat{\mathbb{P}}_N, \mathbb{P}^\star) \to 0$ *almost surely, then* $\mathbb{H}(\mathcal{P}_N, \widehat{\mathcal{P}}) \to 0$ *almost surely as* $N \to \infty$*.*

Proof. The idea of the proof follows from the proof of Hoffman's Lemma for moment problems (see Theorem 2 in Liu et al. 2019). First, we claim that for any $\mathbb{P}_1 \in \mathcal{P}_N$, $(1-\rho)\mathbb{P}_1 + \rho\mathbb{P}^\star \in \widehat{\mathcal{P}}$, where $\rho = \mathsf{d}(\widehat{\mathbb{P}}_N, \mathbb{P}^\star)/\big(r + \mathsf{d}(\widehat{\mathbb{P}}_N, \mathbb{P}^\star)\big)$. To see this, note that

$$\begin{aligned} \mathsf{d}\big((1-\rho)\mathbb{P}_1 + \rho\mathbb{P}^\star, \mathbb{P}^\star\big) &\leq (1-\rho)\mathsf{d}(\mathbb{P}_1, \mathbb{P}^\star) \\ &\leq \frac{r}{r + \mathsf{d}(\widehat{\mathbb{P}}_N, \mathbb{P}^\star)} \left[\mathsf{d}(\mathbb{P}_1, \widehat{\mathbb{P}}_N) + \mathsf{d}(\widehat{\mathbb{P}}_N, \mathbb{P}^\star)\right] \\ &\leq \frac{r}{r + \mathsf{d}(\widehat{\mathbb{P}}_N, \mathbb{P}^\star)} \left[r + \mathsf{d}(\widehat{\mathbb{P}}_N, \mathbb{P}^\star)\right] = r, \end{aligned} \tag{EC.42}$$

where the first inequality follows from properties (iv) and (i), the second inequality follows from the definition of ρ and property (iii). Then, we have

$$\begin{aligned} \mathbb{D}(\mathbb{P}_1, \widehat{\mathcal{P}}) = \inf_{\mathbb{P} \in \widehat{\mathcal{P}}} \mathbb{d}(\mathbb{P}_1, \mathbb{P}) \leq \mathbb{d}\big(\mathbb{P}_1, (1-\rho)\mathbb{P}_1 + \rho\mathbb{P}^\star\big) & \\ = \sup_{\boldsymbol{x} \in \mathcal{X}} \left| \mathbb{E}_{\mathbb{P}_1}[f(\boldsymbol{x}, \boldsymbol{\xi})] - \Big\{(1-\rho)\mathbb{E}_{\mathbb{P}_1}[f(\boldsymbol{x}, \boldsymbol{\xi})] + \rho\mathbb{E}_{\mathbb{P}^\star}[f(\boldsymbol{x}, \boldsymbol{\xi})]\Big\} \right| & \\ = \rho \sup_{\boldsymbol{x} \in \mathcal{X}} \left| \mathbb{E}_{\mathbb{P}_1}[f(\boldsymbol{x}, \boldsymbol{\xi})] - \mathbb{E}_{\mathbb{P}^\star}[f(\boldsymbol{x}, \boldsymbol{\xi})] \right| & \\ = \frac{\mathsf{d}(\widehat{\mathbb{P}}_N, \mathbb{P}^\star)}{r + \mathsf{d}(\widehat{\mathbb{P}}_N, \mathbb{P}^\star)} \cdot \mathbb{d}(\mathbb{P}_1, \mathbb{P}^\star) \leq \frac{\Delta}{r} \cdot \mathsf{d}(\widehat{\mathbb{P}}_N, \mathbb{P}^\star), & \end{aligned} \tag{EC.43}$$

where the last inequality follows from the fact that $\mathsf{d}(\widehat{\mathbb{P}}_N, \mathbb{P}^\star) \geq 0$ and $\mathbb{d}(\mathbb{P}_1, \mathbb{P}^\star) \leq \Delta$. This implies that $\sup_{\mathbb{P}_1 \in \mathcal{P}_N} \mathbb{D}(\mathbb{P}_1, \widehat{\mathcal{P}}) \leq (\Delta/r)\mathsf{d}(\widehat{\mathbb{P}}_N, \mathbb{P}^\star)$. Similarly, following a similar argument in (EC.42), we can show that for any $\mathbb{P}_2 \in \widehat{\mathcal{P}}$, $(1-\tau)\mathbb{P}_2 + \tau\widehat{\mathbb{P}}_N \in \mathcal{P}_N$, where $\tau = \mathsf{d}(\mathbb{P}^\star, \widehat{\mathbb{P}}_N)/\big(r + \mathsf{d}(\mathbb{P}^\star, \widehat{\mathbb{P}}_N)\big)$. A similar argument as in (EC.43) shows that

$$\mathbb{D}(\mathbb{P}_2, \widehat{\mathcal{P}}_N) \leq \frac{\Delta}{r} \cdot \mathsf{d}(\mathbb{P}^\star, \widehat{\mathbb{P}}_N) = \frac{\Delta}{r} \cdot \mathsf{d}(\widehat{\mathbb{P}}_N, \mathbb{P}^\star),$$

where the last inequality follows from property (ii). This implies that $\sup_{\mathbb{P}_2 \in \widehat{\mathcal{P}}} \mathbb{D}(\mathbb{P}_2, \widehat{\mathcal{P}}_N) \leq (\Delta/r)\mathsf{d}(\widehat{\mathbb{P}}_N, \mathbb{P}^\star)$. Therefore, we have

$$\mathbb{H}(\mathcal{P}_N, \widehat{\mathcal{P}}) = \max\left\{\sup_{\mathbb{P}_1 \in \mathcal{P}_N} \mathbb{D}(\mathbb{P}_1, \widehat{\mathcal{P}}),\ \sup_{\mathbb{P}_2 \in \widehat{\mathcal{P}}} \mathbb{D}(\mathbb{P}_2, \mathcal{P}_N)\right\} \leq \frac{\Delta}{r} \cdot \mathsf{d}(\widehat{\mathbb{P}}_N, \mathbb{P}^\star) \to 0$$

almost surely as $N \to \infty$. □

EC.6. Some Quantitative Stability Analysis Results

Consider two DRO models with two different ambiguity sets:

$$\upsilon_i = \inf_{\boldsymbol{x} \in \mathcal{X}} \sup_{\mathbb{P} \in \mathcal{P}_i} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \boldsymbol{\xi})]$$

for $i \in \{1, 2\}$, where we assume that υ_i is finite and the set of optimal solutions $\mathcal{X}_i^\star$ is non-empty. In quantitative stability analysis, we analyze how the change in the ambiguity set would affect the optimal value and the set of optimal solutions to the DRO model. In this appendix, we summarize some relevant results in the existing literature, in particular, the upper bounds on the differences between the optimal values and the set of optimal solutions from the two DRO models (see, e.g., Liu and Xu 2013, Pichler and Xu 2022, Sun and Xu 2016). For the sake of completeness, we also provide the proof of these results.

PROPOSITION EC.3. *The (pointwise) absolute difference between the two objective functions is upper bounded by the Hausdorff distance between the two ambiguity sets, i.e., for any* $\boldsymbol{x} \in \mathcal{X}$,

$$\left|\sup_{\mathbb{Q} \in \mathcal{P}_1} \mathbb{E}_{\mathbb{Q}}[f(\boldsymbol{x}, \boldsymbol{\xi})] - \sup_{\mathbb{Q}' \in \mathcal{P}_2} \mathbb{E}_{\mathbb{Q}'}[f(\boldsymbol{x}, \boldsymbol{\xi})]\right| \leq \mathbb{H}(\mathcal{P}_1, \mathcal{P}_2). \tag{EC.44}$$

Proof. By definition of the pseudometric $\mathbb{d}$ in (11), for any $\mathbb{Q} \in \mathcal{P}_1$ and $\mathbb{Q}' \in \mathcal{P}_2$, we have $\left|\mathbb{E}_{\mathbb{Q}}[f(\boldsymbol{x}, \boldsymbol{\xi})] - \mathbb{E}_{\mathbb{Q}'}[f(\boldsymbol{x}, \boldsymbol{\xi})]\right| \leq \mathbb{d}(\mathbb{Q}, \mathbb{Q}')$ for all $\boldsymbol{x} \in \mathcal{X}$. Since

$$\begin{aligned}
\sup_{\mathbb{Q} \in \mathcal{P}_1} \mathbb{E}_{\mathbb{Q}}[f(\boldsymbol{x}, \boldsymbol{\xi})] - \sup_{\mathbb{Q}' \in \mathcal{P}_2} \mathbb{E}_{\mathbb{Q}'}[f(\boldsymbol{x}, \boldsymbol{\xi})] &= \sup_{\mathbb{Q} \in \mathcal{P}_1} \inf_{\mathbb{Q}' \in \mathcal{P}_2} \left\{\mathbb{E}_{\mathbb{Q}}[f(\boldsymbol{x}, \boldsymbol{\xi})] - \mathbb{E}_{\mathbb{Q}'}[f(\boldsymbol{x}, \boldsymbol{\xi})]\right\} \\
&\leq \sup_{\mathbb{Q} \in \mathcal{P}_1} \inf_{\mathbb{Q}' \in \mathcal{P}_2} \mathbb{d}(\mathbb{Q}, \mathbb{Q}')
\end{aligned}$$

and

$$\begin{aligned}
\sup_{\mathbb{Q}' \in \mathcal{P}_2} \mathbb{E}_{\mathbb{Q}'}[f(\boldsymbol{x}, \boldsymbol{\xi})] - \sup_{\mathbb{Q} \in \mathcal{P}_1} \mathbb{E}_{\mathbb{Q}}[f(\boldsymbol{x}, \boldsymbol{\xi})] &= \sup_{\mathbb{Q}' \in \mathcal{P}_2} \inf_{\mathbb{Q} \in \mathcal{P}_1} \left\{\mathbb{E}_{\mathbb{Q}'}[f(\boldsymbol{x}, \boldsymbol{\xi})] - \mathbb{E}_{\mathbb{Q}}[f(\boldsymbol{x}, \boldsymbol{\xi})]\right\} \\
&\leq \sup_{\mathbb{Q}' \in \mathcal{P}_2} \inf_{\mathbb{Q} \in \mathcal{P}_1} \mathbb{d}(\mathbb{Q}, \mathbb{Q}'),
\end{aligned}$$

we obtain

$$\left|\sup_{\mathbb{Q}\in\mathcal{P}_1}\mathbb{E}_{\mathbb{Q}}[f(\boldsymbol{x},\boldsymbol{\xi})]-\sup_{\mathbb{Q}'\in\mathcal{P}_2}\mathbb{E}_{\mathbb{Q}'}[f(\boldsymbol{x},\boldsymbol{\xi})]\right|\le\max\left\{\sup_{\mathbb{Q}\in\mathcal{P}_1}\inf_{\mathbb{Q}'\in\mathcal{P}_2}\mathbb{d}(\mathbb{Q},\mathbb{Q}'),\ \sup_{\mathbb{Q}'\in\mathcal{P}_2}\inf_{\mathbb{Q}\in\mathcal{P}_1}\mathbb{d}(\mathbb{Q},\mathbb{Q}')\right\}$$
$$=\mathbb{H}(\mathcal{P}_1,\mathcal{P}_2).$$

This completes the proof. □

PROPOSITION EC.4. *The following assertions hold.*

(i) $|\upsilon_1-\upsilon_2|\le\mathbb{H}(\mathcal{P}_1,\mathcal{P}_2)$.

(ii) If, in addition, $\sup_{\mathbb{P}\in\mathcal{P}_1}\mathbb{E}[f(\boldsymbol{x},\boldsymbol{\xi})]$ *satisfies the second order growth condition at* $\mathcal{X}_1^\star$*, i.e., there exists* $\tau>0$ *such that*

$$\sup_{\mathbb{P}\in\mathcal{P}_1}\mathbb{E}[f(\boldsymbol{x},\boldsymbol{\xi})]\ge\upsilon_1+\tau[d(\boldsymbol{x},\mathcal{X}_1^\star)]^2$$

for all $\boldsymbol{x}\in\mathcal{X}$*, then*

$$D(\mathcal{X}_2^\star,\mathcal{X}_1^\star)\le\sqrt{\frac{3}{\tau}\mathbb{H}(\mathcal{P}_1,\mathcal{P}_2)}.$$

Proof. Part (i) follows directly from

$$\begin{aligned}|\upsilon_1-\upsilon_2|&=\left|\inf_{\boldsymbol{x}\in\mathcal{X}}\sup_{\mathbb{P}\in\mathcal{P}_1}\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\boldsymbol{\xi})]-\inf_{\boldsymbol{x}\in\mathcal{X}}\sup_{\mathbb{P}\in\mathcal{P}_2}\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\boldsymbol{\xi})]\right|\\&\le\sup_{\boldsymbol{x}\in\mathcal{X}}\left|\sup_{\mathbb{Q}\in\mathcal{P}_1}\mathbb{E}_{\mathbb{Q}}[f(\boldsymbol{x},\boldsymbol{\xi})]-\sup_{\mathbb{Q}'\in\mathcal{P}_2}\mathbb{E}_{\mathbb{Q}'}[f(\boldsymbol{x},\boldsymbol{\xi})]\right|\\&\le\mathbb{H}(\mathcal{P}_1,\mathcal{P}_2),\end{aligned}$$

where the last inequality follows from Proposition EC.3. For part (ii), suppose, on the contrary, that

$$D(\mathcal{X}_2^\star,\mathcal{X}_1^\star)=\sup_{\boldsymbol{x}\in\mathcal{X}_2^\star}d(\boldsymbol{x},\mathcal{X}_1^\star)>\sqrt{\frac{3}{\tau}\mathbb{H}(\mathcal{P}_1,\mathcal{P}_2)}.$$

Then, there exists $\overline{\boldsymbol{x}}\in\mathcal{X}_2^\star$ such that $d(\overline{\boldsymbol{x}},\mathcal{X}_1^\star)>\sqrt{3\tau^{-1}\mathbb{H}(\mathcal{P}_1,\mathcal{P}_2)}$. The second order growth condition implies that

$$\sup_{\mathbb{P}\in\mathcal{P}_1}\mathbb{E}[f(\overline{\boldsymbol{x}},\boldsymbol{\xi})]-\upsilon_1\ge\tau\big[d(\overline{\boldsymbol{x}},\mathcal{X}_1^\star)\big]^2>3\,\mathbb{H}(\mathcal{P}_1,\mathcal{P}_2). \tag{EC.45}$$

However, by Proposition EC.3 and part (i), we have

$$\begin{aligned}\sup_{\mathbb{P}\in\mathcal{P}_1}\mathbb{E}[f(\overline{\boldsymbol{x}},\boldsymbol{\xi})]-\upsilon_1&\le\left\{\sup_{\mathbb{P}\in\mathcal{P}_2}\mathbb{E}[f(\overline{\boldsymbol{x}},\boldsymbol{\xi})]+\mathbb{H}(\mathcal{P}_1,\mathcal{P}_2)\right\}-\big[\upsilon_2-\mathbb{H}(\mathcal{P}_1,\mathcal{P}_2)\big]\\&=\sup_{\mathbb{P}\in\mathcal{P}_2}\mathbb{E}[f(\overline{\boldsymbol{x}},\boldsymbol{\xi})]-\upsilon_2+2\,\mathbb{H}(\mathcal{P}_1,\mathcal{P}_2)\\&=2\,\mathbb{H}(\mathcal{P}_1,\mathcal{P}_2),\end{aligned}$$

where the first equality follows from $\overline{\boldsymbol{x}}\in\mathcal{X}_2^\star$. Thus, this contradicts with (EC.45). □

EC.7. Details of Numerical Experiments

EC.7.1. Inventory Control – Reformulations

Recall the TRO model (24) under shape parameters (a)–(d) in Table 1. This model cannot be solved directly because of the inner supremum problem $\sup_{\mathbb{P}\in\mathcal{P}_N}\mathbb{E}_{\mathbb{P}}\left[(\xi-x)_+-\xi\right]$. In this section, we provide tractable reformulation to this inner supremum problem, and thus our TRO model.

First, consider ambiguity set (a) in Table 1. Using the worst-case distribution derived in Gallego and Moon (1993), we have

$$\sup_{\mathbb{P}\in\mathcal{P}_N}\mathbb{E}_{\mathbb{P}}\left[(\xi-x)_+-\xi\right]=\begin{cases}\dfrac{1}{2}\left[-(x-\widehat{\mu}_N)+\sqrt{\widehat{\sigma}_N^2+(x-\widehat{\mu}_N)^2}\right]-\widehat{\mu}_N & \text{if } x\geq\dfrac{\widehat{\sigma}_N^2+\widehat{\mu}_N^2}{2\widehat{\mu}_N},\\ -\dfrac{\widehat{\mu}_N^2}{\widehat{\sigma}_N^2+\widehat{\mu}_N^2}x & \text{if } 0\leq x\leq\dfrac{\widehat{\sigma}_N^2+\widehat{\mu}_N^2}{2\widehat{\mu}_N}.\end{cases}$$

Therefore, the TRO model under set (a) is equivalent to

$$\begin{aligned}\underset{x\geq 0}{\text{minimize}}\ (c-h)x+(p-h)\Bigg\{&(1-\theta)\frac{1}{N}\sum_{i=1}^N\left[(\widehat{\xi}_i-x)_+-\widehat{\xi}_i\right]\\&+\theta\left[\left\{\frac{1}{2}\left[\sqrt{\widehat{\sigma}_N^2+(x-\widehat{\mu}_N)^2}-(x-\widehat{\mu}_N)\right]-\widehat{\mu}_N\right\}\cdot\mathbf{1}_{x\geq\frac{\widehat{\mu}_N^2+\widehat{\sigma}_N^2}{2\widehat{\mu}_N}}-\frac{\widehat{\mu}_N^2}{\widehat{\mu}_N^2+\widehat{\sigma}_N^2}x\cdot\mathbf{1}_{0\leq x\leq\frac{\widehat{\mu}_N^2+\widehat{\sigma}_N^2}{2\widehat{\mu}_N}}\right]\Bigg\}.\end{aligned}$$

Next, consider ambiguity set (b) in Table 1. Applying the reformulation in Lee et al. (2021), we have

$$\sup_{\mathbb{P}\in\mathcal{P}_N}\mathbb{E}_{\mathbb{P}}\left[(\xi-x)_+-\xi\right]=r+\frac{1}{N}\sum_{i=1}^N\left[(\widehat{\xi}_i-x)_+-\widehat{\xi}_i\right].$$

Therefore, the TRO model under set (b) is equivalent to

$$\underset{x\geq 0}{\text{minimize}}\quad (c-h)x+(p-h)\left\{(1-\theta)\frac{1}{N}\sum_{i=1}^N\left[(\widehat{\xi}_i-x)_+-\widehat{\xi}_i\right]+\theta\left\{r+\frac{1}{N}\sum_{i=1}^N\left[(\widehat{\xi}_i-x)_+-\widehat{\xi}_i\right]\right\}\right\}.$$

Now, consider ambiguity set (c) in Table 1. By the strong duality result for ϕ-divergence DRO problems in Bayraksan and Love (2015), we have

$$\sup_{\mathbb{P}\in\mathcal{P}_N}\mathbb{E}_{\mathbb{P}}\left[(\xi-x)_+-\xi\right]=\inf_{\lambda\geq 0,\,\tau}\left\{\tau+\lambda r-\frac{\lambda}{N}\sum_{i=1}^N\log\left(1-\frac{(\widehat{\xi}_i-x)_+-\widehat{\xi}_i-\tau}{\lambda}\right)\right\},$$

where $-0\log(1-s/0)=0$ for $s\leq 0$ and $-0\log(1-s/0)=\infty$ for $s>0$. Therefore, the TRO model under set (c) is equivalent to

$$\begin{aligned}\underset{x\geq 0,\,\lambda\geq 0,\,\tau}{\text{minimize}}\quad (c-h)x+(p-h)\Bigg\{&(1-\theta)\frac{1}{N}\sum_{i=1}^N\left[(\widehat{\xi}_i-x)_+-\widehat{\xi}_i\right]\\&+\theta\left\{\tau+\lambda r-\frac{\lambda}{N}\sum_{i=1}^N\log\left(1-\frac{(\widehat{\xi}_i-x)_+-\widehat{\xi}_i-\tau}{\lambda}\right)\right\}\Bigg\}.\end{aligned}$$

Finally, consider the ambiguity set (d) in Table 1. For notational simplicity, let $\gamma := t_{N-1,\alpha/2}\widehat{\sigma}_N/\sqrt{N}$. Then, we have

$$\sup_{\mathbb{P}\in\mathcal{P}_N}\mathbb{E}_{\mathbb{P}}\big[(\xi-x)_+-\xi\big]=\sup_{\xi\geq 0:|\xi-\widehat{\mu}_N|\leq\gamma}\big\{(\xi-x)_+-\xi\big\}=\max\big\{-x,-(\widehat{\mu}_N-\gamma)_+\big\},$$

where the last equality follows from the fact that $(\xi-x)_+-\xi=\max\{-x,-\xi\}$ and the objective function is non-increasing in ξ. Thus, the supremum over $\xi\in[(\widehat{\mu}_N-\gamma)_+,\widehat{\mu}_N+\gamma]$ is attained at the lower bound. Therefore, the TRO model under set (d) is equivalent to

$$\underset{x\geq 0}{\text{minimize}}\quad (c-h)x+(p-h)\left\{(1-\theta)\frac{1}{N}\sum_{i=1}^{N}\big[(\widehat{\xi}_i-x)_+-\widehat{\xi}_i\big]+\theta\max\big\{-x,-(\widehat{\mu}_N-\gamma)_+\big\}\right\}.$$

EC.7.2. Inventory Control – Ambiguity Set (d)

We show that ambiguity set (d) in Table 1, i.e., $\mathcal{P}_N=\{\delta_\xi \mid |\xi-\widehat{\mu}_N|\leq t_{N-1,\alpha/2}\widehat{\sigma}_N/\sqrt{N}\}$, satisfies Assumption 3(b). Recall that $f(x,\xi)=(\xi-x)_+-\xi$, which is Lipschitz in ξ with modulus 1. We claim that $\mathbb{H}(\mathcal{P}_N,\widehat{\mathcal{P}})\to 0$ as $N\to\infty$ almost surely, where $\widehat{\mathcal{P}}=\{\delta_{\mu^\star}\}$ with $\mu^\star=\mathbb{E}_{\mathbb{P}^\star}(\xi)$. Indeed,

$$\begin{aligned}
\mathbb{H}(\mathcal{P}_N,\widehat{\mathcal{P}})=\sup_{\mathbb{P}\in\mathcal{P}_N}\mathbb{D}(\mathbb{P},\widehat{\mathcal{P}})&=\sup_{\xi\geq 0:|\xi-\widehat{\mu}_N|\leq t_{N-1,\alpha/2}\widehat{\sigma}_N/\sqrt{N}}\ \sup_{x\geq 0}\big|f(x,\xi)-f(x,\mu^\star)\big| && \text{(EC.46a)}\\
&\leq \sup_{\xi\geq 0:|\xi-\widehat{\mu}_N|\leq t_{N-1,\alpha/2}\widehat{\sigma}_N/\sqrt{N}}|\xi-\mu^\star| && \text{(EC.46b)}\\
&=\max\left\{\left|\left(\widehat{\mu}_N-t_{N-1,\alpha/2}\frac{\widehat{\sigma}_N}{\sqrt{N}}\right)_+-\mu^\star\right|,\left|\widehat{\mu}_N+t_{N-1,\alpha/2}\frac{\widehat{\sigma}_N}{\sqrt{N}}-\mu^\star\right|\right\} && \text{(EC.46c)}\\
&\to 0
\end{aligned}$$

almost surely. Here, (EC.46a) follows from the definitions in (12) and (11), and that $\sup_{\mathbb{P}\in\widehat{\mathcal{P}}}\mathbb{D}(\mathbb{P},\widehat{\mathcal{P}})\leq\sup_{\mathbb{P}\in\mathcal{P}_N}\mathbb{D}(\mathbb{P},\widehat{\mathcal{P}})$; (EC.46b) follows from the Lipschitz continuity of $f(x,\xi)$ in ξ; (EC.46c) follows from the fact that the supremum over a compact interval of the absolute value function $|\xi-\mu^\star|$ is attained either at the lower or the upper bound of the interval. Finally, by the strong law of large numbers, $|\widehat{\mu}_N-\mu^\star|\to 0$ and $|\widehat{\sigma}_N-\sigma^\star|\to 0$ almost surely, where $\sigma^\star=\mathrm{Var}_{\mathbb{P}^\star}(\xi)<\infty$. Also, $t_{N-1,\alpha/2}\to z_{\alpha/2}<\infty$, where $z_{\alpha/2}$ is the upper $(1-\alpha)/2$-th quantile of a standard normal distribution. Thus, we have (EC.46c) converges to zero almost surely. Thus, ambiguity set (d) satisfies Assumption 3(b).

EC.7.3. Portfolio Optimization – Reformulations

Recall the TRO model (28) under shape paraemters (a)–(c) in Table 4. Again, this model cannot be solved directly because of the inner supremum problem $\sup_{\mathbb{P}\in\mathcal{P}_N}\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},t,\boldsymbol{\xi})]$. In this section, we provide tractable reformulation to this inner supremum problem, and thus our TRO model.

First, consider ambiguity set (a) in Table 4. By Lemma 2.2 and Lemma 2.4 of Chen et al. (2011), we have

$$\sup_{\mathbb{P}\in\mathcal{P}_N}\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},t,\boldsymbol{\xi})]=(1-\beta)t-\beta\boldsymbol{x}^\top\widehat{\mu}_N+\frac{1-\beta}{1-\alpha}\cdot\frac{1}{2}\left[-\boldsymbol{x}^\top\widehat{\boldsymbol{\mu}}_N-t+\sqrt{\boldsymbol{x}^\top\widehat{\boldsymbol{\Sigma}}_N\boldsymbol{x}+(-\boldsymbol{x}^\top\widehat{\boldsymbol{\mu}}_N-t)^2}\right].$$

Therefore, the TRO model under set (a) is equivalent to

$$\begin{aligned}\underset{\boldsymbol{x}\in\mathcal{X},\,t\in\mathbb{R}}{\text{minimize}}\quad &(1-\theta)\cdot\frac{1}{N}\sum_{i=1}^N\left\{(1-\beta)t+\beta(-\boldsymbol{x}^\top\widehat{\boldsymbol{\xi}}_i)+\frac{1-\beta}{1-\alpha}(-\boldsymbol{x}^\top\widehat{\boldsymbol{\xi}}_i-t)^+\right\}\\ &+\theta\left\{(1-\beta)t-\beta\boldsymbol{x}^\top\widehat{\mu}_N+\frac{1-\beta}{1-\alpha}\cdot\frac{1}{2}\left[-\boldsymbol{x}^\top\widehat{\boldsymbol{\mu}}_N-t+\sqrt{\boldsymbol{x}^\top\widehat{\boldsymbol{\Sigma}}_N\boldsymbol{x}+(-\boldsymbol{x}^\top\widehat{\boldsymbol{\mu}}_N-t)^2}\right]\right\}.\end{aligned}$$

Now, consider ambiguity set (b) in Table 4. By Theorem 6.3 of Mohajerin Esfahani and Kuhn (2018), we have

$$\sup_{\mathbb{P}\in\mathcal{P}_N}\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},t,\boldsymbol{\xi})]=(1-\beta)t+r\left(\beta+\frac{1-\beta}{1-\alpha}\right)\|\boldsymbol{x}\|_\infty+\frac{1}{N}\sum_{i=1}^N\left[\beta(-\boldsymbol{x}^\top\widehat{\boldsymbol{\xi}}_i)+\frac{1-\beta}{1-\alpha}(-\boldsymbol{x}^\top\widehat{\boldsymbol{\xi}}_i-t)_+\right].$$

Therefore, the TRO model under set (b) is equivalent to

$$\begin{aligned}\underset{\boldsymbol{x}\in\mathcal{X},\,t\in\mathbb{R}}{\text{minimize}}\quad &(1-\theta)\cdot\frac{1}{N}\sum_{i=1}^N\left\{(1-\beta)t+\beta(-\boldsymbol{x}^\top\widehat{\boldsymbol{\xi}}_i)+\frac{1-\beta}{1-\alpha}(-\boldsymbol{x}^\top\widehat{\boldsymbol{\xi}}_i-t)^+\right\}\\ &+\theta\left\{(1-\beta)t+r\left(\beta+\frac{1-\beta}{1-\alpha}\right)\|\boldsymbol{x}\|_\infty+\frac{1}{N}\sum_{i=1}^N\left[\beta(-\boldsymbol{x}^\top\widehat{\boldsymbol{\xi}}_i)+\frac{1-\beta}{1-\alpha}(-\boldsymbol{x}^\top\widehat{\boldsymbol{\xi}}_i-t)_+\right]\right\}.\end{aligned}$$

Finally, consider ambiguity set (c). Note that ambiguity set (c) is characterized by the ℓ_1 norm constraint. Thus, as in Lemma 3.1 of Huang et al. (2021), using linear programming duality, $\sup_{\mathbb{P}\in\mathcal{P}_N}\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},t,\boldsymbol{\xi})]$ is equivalent to

$$\underset{\tau\in\mathbb{R},\,\lambda\in\mathbb{R}}{\text{minimize}}\quad (1-\beta)t+\tau+r\lambda+\frac{1}{N}\sum_{i=1}^N(u_i^+-u_i^-) \tag{EC.47a}$$

$$\text{subject to}\quad \tau\geq\beta(-\boldsymbol{x}^\top\widehat{\boldsymbol{\xi}}_i)+\frac{1-\beta}{1-\alpha}\big(-\boldsymbol{x}^\top\widehat{\boldsymbol{\xi}}_i-t\big)_+-u_i^++u_i^-,\quad\forall i\in\{1,\dots,n\}, \tag{EC.47b}$$

$$\lambda\geq u_i^++u_i^-,\quad\forall i\in\{1,\dots,n\}, \tag{EC.47c}$$

$$\lambda\geq 0,\ u_i^+\geq 0,\ u_i^-\geq 0,\quad\forall i\in\{1,\dots,n\}. \tag{EC.47d}$$

Therefore, the TRO model under set (c) is equivalent to

$$\begin{aligned}
\underset{\boldsymbol{x}\in\mathcal{X},t,\tau,\lambda,u^+,u^-}{\text{minimize}} \quad & (1-\theta)\cdot\frac{1}{N}\sum_{i=1}^{N}\left\{(1-\beta)t+\beta(-\boldsymbol{x}^\top\widehat{\boldsymbol{\xi}}_i)+\frac{1-\beta}{1-\alpha}(-\boldsymbol{x}^\top\widehat{\boldsymbol{\xi}}_i-t)^+\right\} \\
& \theta\left\{(1-\beta)t+\tau+r\lambda+\frac{1}{N}\sum_{i=1}^{N}(u_i^+-u_i^-)\right\} \qquad \text{(EC.48a)} \\
\text{subject to} \quad & \tau\geq\beta(-\boldsymbol{x}^\top\widehat{\boldsymbol{\xi}}_i)+\frac{1-\beta}{1-\alpha}\big(-\boldsymbol{x}^\top\widehat{\boldsymbol{\xi}}_i-t\big)_+-u_i^++u_i^-, \quad \forall i\in\{1,\dots,n\}, \qquad \text{(EC.48b)} \\
& \lambda\geq u_i^++u_i^-, \quad \forall i\in\{1,\dots,n\}, \qquad \text{(EC.48c)} \\
& \lambda\geq 0,\ u_i^+\geq 0,\ u_i^-\geq 0, \quad \forall i\in\{1,\dots,n\}. \qquad \text{(EC.48d)}
\end{aligned}$$

EC.7.4. Portfolio Optimization – Lipschitz Continuity

In this section, we show that the function $f(\boldsymbol{x},t,\boldsymbol{\xi})=(1-\beta)t+\beta(-\boldsymbol{x}^\top\boldsymbol{\xi})+[(1-\beta)/(1-\alpha)](-\boldsymbol{x}^\top\boldsymbol{\xi}-t)^+$ is Lipschitz continuous in $(\boldsymbol{x},t)$. Indeed, for any $\{(\boldsymbol{x}_1,t_1),(\boldsymbol{x}_2,t_2)\}\subseteq\mathcal{X}\times\mathbb{R}$, we have

$$\begin{aligned}
& |f(\boldsymbol{x}_1,t_1,\boldsymbol{\xi})-f(\boldsymbol{x}_2,t_2,\boldsymbol{\xi})| \\
=& \left|\left[(1-\beta)t_1+\beta(-\boldsymbol{x}_1^\top\boldsymbol{\xi})+\frac{1-\beta}{1-\alpha}(-\boldsymbol{x}_1^\top\boldsymbol{\xi}-t_1)^+\right]-\left[(1-\beta)t_2+\beta(-\boldsymbol{x}_2^\top\boldsymbol{\xi})+\frac{1-\beta}{1-\alpha}(-\boldsymbol{x}_2^\top\boldsymbol{\xi}-t_2)^+\right]\right| \\
\leq& (1-\beta)\big|t_1-t_2\big|+\beta\big|\boldsymbol{x}_1^\top\boldsymbol{\xi}-\boldsymbol{x}_2^\top\boldsymbol{\xi}\big|+\frac{1-\beta}{1-\alpha}\Big|(-\boldsymbol{x}_1^\top\boldsymbol{\xi}-t_1)^+-(-\boldsymbol{x}_2^\top\boldsymbol{\xi}-t_2)^+\Big| \\
\leq& (1-\beta)\big|t_1-t_2\big|+\beta\big|\boldsymbol{x}_1^\top\boldsymbol{\xi}-\boldsymbol{x}_2^\top\boldsymbol{\xi}\big|+\frac{1-\beta}{1-\alpha}\Big(\big|t_1-t_2\big|+\big|\boldsymbol{x}_1^\top\boldsymbol{\xi}-\boldsymbol{x}_2^\top\boldsymbol{\xi}\big|\Big) \qquad \text{(EC.49a)} \\
\leq& (1-\beta)\left(1+\frac{1}{1-\alpha}\right)|t_1-t_2|+\frac{1-\alpha\beta}{1-\alpha}\|\boldsymbol{\xi}\|_1\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_\infty \qquad \text{(EC.49b)} \\
\leq& \left[(1-\beta)\left(1+\frac{1}{1-\alpha}\right)+\frac{1-\alpha\beta}{1-\alpha}\|\boldsymbol{\xi}\|_1\right]\Big(|t_1-t_2|+\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_\infty\Big) \\
\leq& \left[(1-\beta)\left(1+\frac{1}{1-\alpha}\right)+\frac{1-\alpha\beta}{1-\alpha}\|\boldsymbol{\xi}\|_1\right]\left\|\begin{pmatrix}\boldsymbol{x}_1-\boldsymbol{x}_2\\ t_1-t_2\end{pmatrix}\right\|_1 =:\kappa(\boldsymbol{\xi})\left\|\begin{pmatrix}\boldsymbol{x}_1-\boldsymbol{x}_2\\ t_1-t_2\end{pmatrix}\right\|_1. \qquad \text{(EC.49c)}
\end{aligned}$$

Here, (EC.49a) follows from the fact that $|(a_1)^+-(a_2)^+|\leq|a_1-a_2|$ for any $\{a_1,a_2\}\subset\mathbb{R}$; (EC.49b) follows from $\big|\boldsymbol{x}_1^\top\boldsymbol{\xi}-\boldsymbol{x}_2^\top\boldsymbol{\xi}\big|\leq\|\boldsymbol{\xi}\|_1\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_\infty$; the inequality in (EC.49c) follows from $\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_\infty\leq\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_1$.

EC.8. Additional Results

In this section, we provide additional experiment results. Specifically, in Section EC.8.1, we analyze the spectra of solutions obtained from the DRO model with distance-based ambiguity sets and

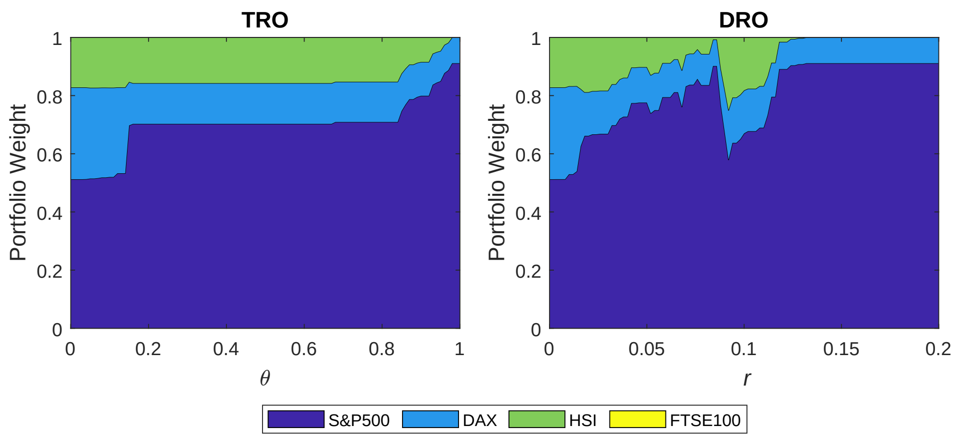


Figure EC.1 **Spectrum of optimal solutions from the TRO model with TRO ambiguity set constructed using total variation ambiguity set as its shape parameter and the total variation DRO model in the portfolio optimization problem.**

those obtained from our TRO model with the TRO ambiguity set constructed using these distance-based ambiguity sets as its shape parameter. In Section EC.8.2, we analyze the bias-variance trade-off of the TRO estimator. In Section EC.8.3, we analyze the conservatism of our TRO model. In Section EC.8.4, we analyze the out-of-sample performance of our TRO model.

EC.8.1. Spectra of Solutions Obtained from Distance-Based DRO Models

As noted by a reviewer of our paper, one could obtain a spectrum of optimal solutions to DRO models equipped with distance-based ambiguity sets by changing the radii of such sets. In this section, we compare the spectra of solutions obtained from the DRO model with distance-based ambiguity sets, namely the 1-Wasserstein ambiguity set and the total variation ambiguity set (i.e., set (b) and set (c) in Table 4), and those obtained from our TRO model with TRO ambiguity set constructed using these distance-based ambiguity sets as its shape parameter. For illustrative purposes, we focus on the portfolio optimization problem discussed in Section 5.2. We follow the same experimental settings detailed in Section 5.2. to obtain the spectra of optimal solution to the DRO and TRO models for this problem. Specifically, we solve the TRO model with different size parameters ($\theta \in \{0, 0.01, 0.02, \dots, 1\}$), the 1-Wasserstein DRO model with different radii ($r \in \{0, 0.002, 0.004, \dots, 0.1\}$), and the total variation DRO model with different radii ($r \in \{0, 0.002, 0.004, \dots, 0.2\}$).

Figure EC.1 presents the spectra of solutions obtained from the TRO model employing the total variation ambiguity set as the shape parameter in the TRO ambiguity set and the DRO model with the total variation ambiguity set. Clearly, the spectra of TRO and DRO solutions could be different.

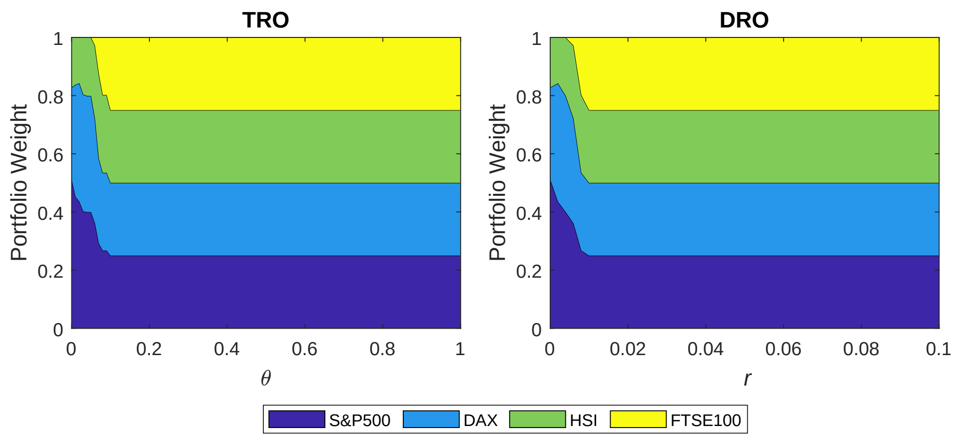
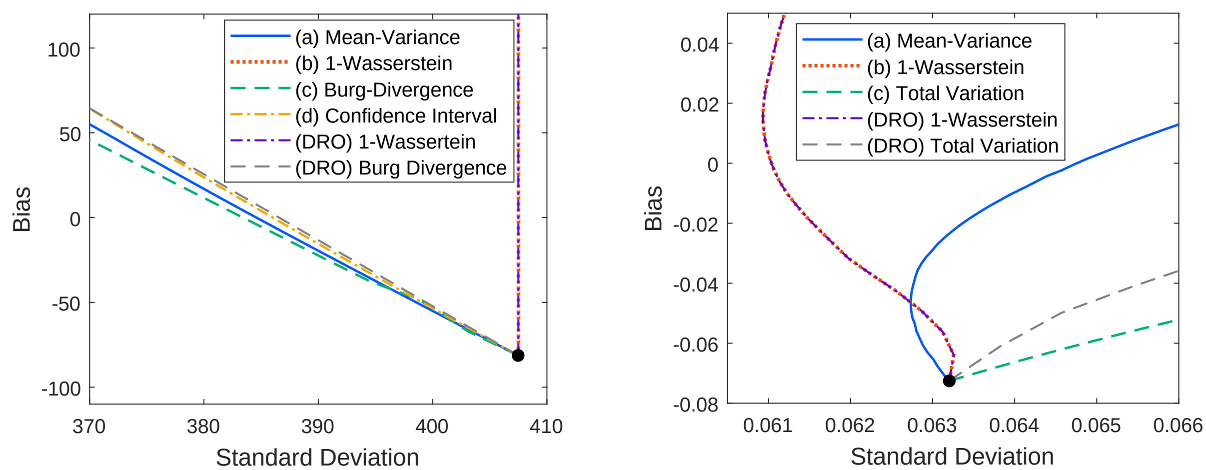


Figure EC.2 **Spectrum of optimal solutions from the TRO model with TRO ambiguity set constructed using 1-Wasserstein ambiguity set as its shape parameter and the 1-Wasserstein DRO model in the portfolio optimization problem.**

Specifically, the portfolio weights change gradually as the TRO ambiguity set's size parameter θ increases. In contrast, the shape of the portfolio weights changes when the radius r of the total variation set increases from 0.086 to 0.092.

Figure EC.2 presents the spectra of solutions obtained from the TRO model employing the 1-Wasserstein ambiguity set as the shape parameter of the TRO ambiguity set and the DRO model with 1-Wasserstein ambiguity set. In contrast to the spectra obtained with the total variation ambiguity set, the spectra of the TRO and DRO models with the 1-Wasserstein set are approximately the same. Specifically, the portfolio weights of S&P500, DAX, and HSI decrease to 0.25, and the portfolio weight of FTSE 100 increases to 0.25 when the TRO ambiguity set's size parameter θ increases from 0 to 0.1 or when the radius r in the DRO model increases from 0 to 0.01.

These results suggest that the spectra obtained from our TRO model with a TRO ambiguity set characterized by a distance-based shape parameter and the DRO model equipped with that shape parameter could be different for some choices of the statistical distance in the distance-based shape parameter. Finally, we note that it is not possible to obtain a spectrum of optimal solutions to the DRO model with other ambiguity sets, such as the mean-variance ambiguity set since such sets do not have a parameter that allows for controlling the conservatism. In contrast, our TRO model with TRO ambiguity set constructed using mean-variance ambiguity set as its shape parameter will enable decision-makers to explore a spectrum of solutions, ranging from optimistic to conservative solutions (see results in Section 5.2). As discussed in Section 6, we leave investigating the characteristics of the spectra of TRO optimal solutions of the TRO model under different shape parameters for future work.



(a) Inventory Control **(b) Portfolio Optimization**

Figure EC.3 **The bias-variance trade-off characterized by $(\texttt{std}_N(\theta), \texttt{bias}_N(\theta))$ for different values of θ with $N = 10$. The black dot corresponds to ($\texttt{std}_N(0)$, $\texttt{bias}_N(0)$), i.e., the standard deviation and bias of the SAA estimator.**

EC.8.2. Bias-Variance Trade-off

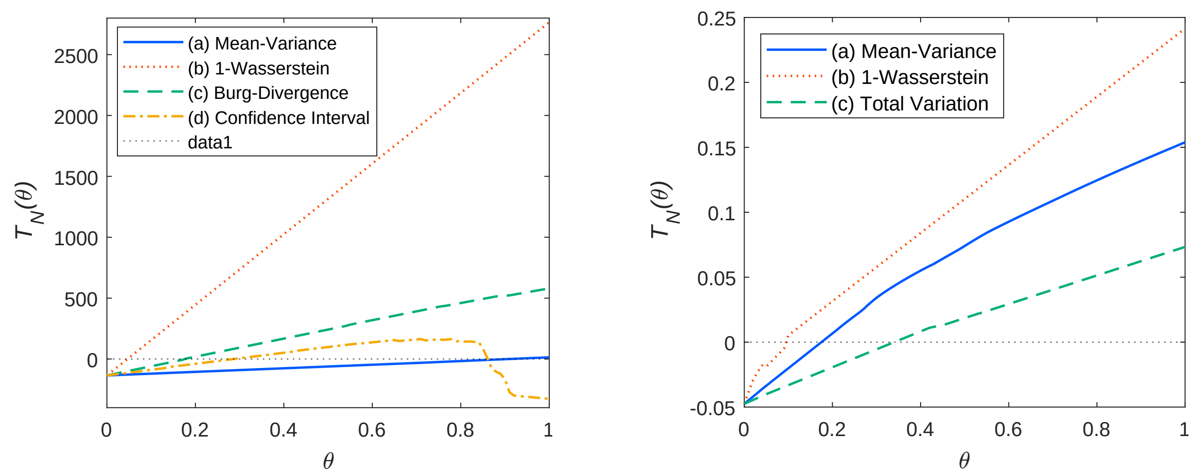
In this section, we numerically analyze the bias-variance trade-off of the TRO estimator $\widehat{v}_N(\theta)$. We follow the same experimental settings detailed in Sections 5.1 and 5.2 to estimate the bias, $\texttt{bias}_N(\theta)$, and standard deviation, $\texttt{std}_N(\theta)$, of $\widehat{v}_N(\theta)$ for $\theta \in \Theta := \{0, 0.01, 0.02, \ldots, 1\}$ with $N = 10$. Figures EC.3a and EC.3b present the bias-variance curves for the TRO model of the inventory control problem and portfolio optimization problem with TRO ambiguity set constructed using different shape parameters $\mathcal{P}_N$. The (x, y) values of each point on each curve correspond to $(\texttt{std}_N(\theta), \texttt{bias}_N(\theta))$ for some $\theta \in [0, 1]$. The black dot corresponds to ($\texttt{std}_N(0)$, $\texttt{bias}_N(0)$), i.e., the standard deviation and bias of the SAA estimator. The figure also presents the bias-variance curves of the estimators obtained from some distance-based DRO models with varying radii r. Specifically, for the inventory control problem (see Figure EC.3a), we consider the 1-Wasserstein and Burg divergence DRO models; for the portfolio optimization problem (see Figure EC.3b), we consider the 1-Wasserstein and total variation DRO models.

Not surprisingly, the bias and standard deviation vary across different shape parameters employed in the TRO model and different statistical distances employed in the distance-based DRO model. For the inventory control problem (see Figure EC.3a), TRO model employing set (b) as the shape parameter and 1-Wasserstein DRO model result in estimators with constant standard deviation but varying bias. Note that these two models yield the same bias-variance curve because they have the same optimal solution for this problem. In contrast, using sets (a), (c), or (d) as shape parameters produces TRO estimators that achieve both lower bias and standard deviation compared with the SAA estimator. This indicates that employing sets (a), (c), or (d) as shape parameters offers a superior bias-variance trade-off for this problem. Interestingly, using set (c), i.e., Burg divergence ambiguity set, as the shape parameter in the TRO model produces estimators with better bias-variance trade-off than the Burg divergence DRO model. We also observe that neither TRO models nor distance-based DRO models consistently yield estimators with the best bias-variance trade-off. For example, when all estimators have biases between -62 and -42, the TRO estimator using set (a) has the smallest standard deviation. When all estimators have biases less than -62 or greater than -42, the TRO estimator using set (c) has the smallest standard deviation.

For the portfolio optimization problem (see Figure EC.3b), we again observe that neither TRO models nor distance-based DRO models consistently produce estimators with the best bias-variance trade-off. Specifically, when all estimators have biases less than -0.0464, the TRO estimator using set (a) has the smallest standard deviation. When all estimators have biases greater than -0.0464, TRO model with set (b) as the shape parameter and 1-Wasserstein DRO model produce estimators with the smallest standard deviation. Additionally, employing some shape parameters in the TRO model and some statistical distances in the distance-based DRO model yield estimators with smaller standard deviations than the SAA estimator. For example, TRO model with sets (a) and (b) as the shape parameter, as well as 1-Wasserstein DRO model, produce estimators with smaller standard deviations for certain values of the shape parameter θ or radius r. In contrast, TRO model with set (c) as the shape parameter and total variation DRO model yield estimators with smaller biases but larger standard deviations than the SAA estimator.

EC.8.3. Conservatism of TRO Solutions

In this section, we investigate the conservatism of the spectra of optimal solutions to the TRO model. Similar to prior studies (see, e.g., Liu et al. (2022), Yin et al. (2023)), we employ the measure $T_N(\theta) = \widehat{v}_N(\theta) - \mathbb{E}_{\mathbb{P}^\star}[f(\boldsymbol{x}_N(\theta), \boldsymbol{\xi})]$ to quantify conservatism, where $\widehat{v}_N(\theta)$ and $\boldsymbol{x}_N(\theta)$ are the optimal value and optimal solution to the TRO model with size parameter θ. This measure captures the extent to which the estimated optimal value (cost), $\widehat{v}_N(\theta)$, exceeds the actual expected objective function value (cost), $\mathbb{E}_{\mathbb{P}^\star}[f(\boldsymbol{x}_N(\theta), \boldsymbol{\xi})]$, associated with implementing the optimal solution $\boldsymbol{x}_N(\theta)$. A larger positive value of $T_N(\theta)$ indicates that the estimated cost $\widehat{v}_N(\theta)$ exceeds the true expected cost, suggesting that the solution $\boldsymbol{x}_N(\theta)$ may be overly conservative. Conversely, a large negative value of $T_N(\theta)$, with a large absolute value, indicates that the actual cost is underestimated, suggesting that $\boldsymbol{x}_N(\theta)$ is overly optimistic.



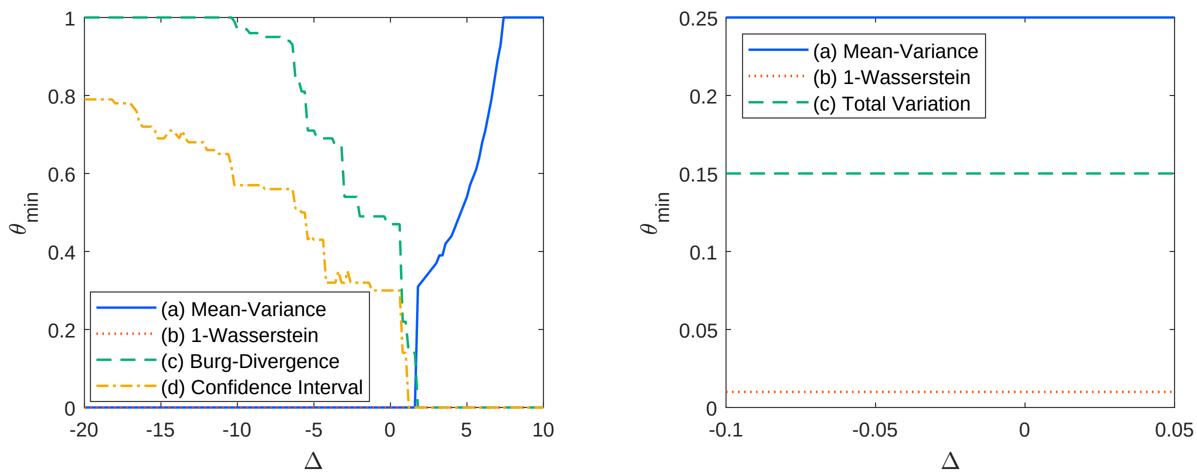
(a) Inventory Control **(b) Portfolio Optimization**

Figure EC.4 $T_N(\theta)$ **for different values of** θ

We follow the same experimental settings detailed in Sections 5.1 and 5.2 to obtain spectra of optimal solutions to our TRO models and compute the associated $T_N(\theta)$. Figures EC.4a and EC.4b present the values of $T_N(\theta)$ for the inventory control problem and portfolio optimization problem, respectively. Clearly, $T_N(0)$ is negative, suggesting that the SAA solutions $\widehat{v}_N(0)$ may be optimistic. Moreover, except for the TRO model of the inventory control problem that employs set (d) as the shape parameter, the value of $T_N(\theta)$ increases with θ. This indicates that the TRO model generates more conservative solutions as θ grows, which is consistent with the theoretical results and discussions in Section 2. For example, $T_N(1)$ associated with the DRO (TRO with $\theta = 1$) solutions to the portfolio optimization problem (Figure EC.4b) is fairly large, suggesting that the DRO solutions $\boldsymbol{x}_N(1)$ using sets (a)–(c) may be overly conservative. Solutions with an intermediate value of θ have a larger absolute value of $T_N(\theta)$ than the SAA solutions but a smaller $T_N(\theta)$ than the DRO solutions. This indicates that these solutions are less optimistic than the SAA solutions and less conservative than the DRO solutions. In the next section, we demonstrate that the out-of-sample performance of TRO solutions obtained with an intermediate value θ is better than those of the SAA and DRO solutions. Finally, we note that in the inventory control problem, since the optimal value $\widehat{v}_N(\theta)$ of TRO model with TRO ambiguity set constructed using set (d) is first increasing and then decreasing (see discussions in Section 5.1), the resulting $T_N(\theta)$ follows the same pattern.

EC.8.4. Out-of-Sample Performance

In this section, we analyze the performance of the spectrum of optimal solutions to the TRO model via out-of-sample testing. First, we follow the same procedure detailed in Sections 5.1 and 5.2 to



(a) Inventory Control **(b) Portfolio Optimization**

Figure EC.5 θ_{min} **for different values of perturbation** Δ

obtain the spectrum of TRO solutions $\{\boldsymbol{x}_N(\theta) \mid \theta \in \Theta\}$, where $\Theta := \{0, 0.01, 0.02, \ldots, 1\}$. Then, we compute the out-of-sample cost $\widehat{v}_N^{\text{OS}}(\theta) = \mathbb{E}_{\mathbb{P}^{\text{OS}}}[f(\boldsymbol{x}_N(\theta), \boldsymbol{\xi})]$ of the TRO solutions for all $\theta \in \Theta$, where $\mathbb{P}^{\text{OS}}$ denotes the out-of-sample distribution. For the inventory control problem, we choose $\mathbb{P}^{\text{OS}}$ as the exponential distribution with mean $50 + \Delta$ and $\Delta \in \{-20, -19.8, -19.6, \ldots, 10\}$, where a positive (resp. negative) value of Δ corresponds to an increase (resp. decrease) in mean. For the portfolio optimization problem, we choose $\mathbb{P}^{\text{OS}}$ as the multivatiate normal distribution described in Section 5.2, and we change the mean from $\boldsymbol{\mu}$ to $\boldsymbol{\mu} + \Delta\mathbf{1}$ with $\Delta \in \{-0.1, -0.098, -0.096, \ldots, 0.05\}$. Finally, we search for $\theta_{\text{min}} := \arg\min\{\theta \in \Theta \mid \widehat{v}_N^{\text{OS}}(\theta) \leq \widehat{v}_N^{\text{OS}}(\theta'),\ \forall \theta' \in \Theta\}$, i.e., the value of θ that gives the smallest out-of-sample cost.

Figures EC.5a and EC.5b present the values of θ_{min} in the inventory control problem and portfolio optimization problem, respectively. Not surprisingly, the value of θ_{min} varies across different shape parameters employed in the TRO model. Moreover, the value of θ that yields the lowest out-of-sample costs is generally neither 0 nor 1 but lies strictly between these two extremes. This highlights that adopting solutions from the spectrum of TRO optimal solutions results in smaller out-of-sample costs than the SAA and DRO solutions. Consider the inventory control problem, for example (see Figure EC.5a). When $\Delta \in [1.8, 7.2]$, the TRO model for this problem with set (a) as the shape parameter and intermediate values of θ yields the smallest out-of-sample costs. The TRO model with sets (c) and (d) and intermediate values of θ yield the smallest out-of-sample costs when $\Delta \in [-10.2, 1.6]$ and $\Delta \in [-20, 1]$, respectively. Note that the optimal solution to our TRO model using set (b) is always the same for any $\theta \in [0, 1]$, i.e., there is only one solution for this problem. Similarly, for the portfolio optimization problem (see Figure EC.5b), our TRO model with sets (a), (b), and (c) and $\theta = 0.25$, $\theta = 0.01$, and $\theta = 0.15$ yields the smallest out-of-sample costs, respectively. Note that θ_{min} is the same for all $\Delta \in [-0.1, 0.05]$ since adding a constant Δ to each entry of the mean vector $\boldsymbol{\mu}$ shifts the entire objective function of this problem by a constant. These results demonstrate the benefits of adopting solutions on the spectrum of TRO optimal solutions over the SAA and DRO solutions.

EC.9. Comparison with Wang et al. (2023)

In this appendix, we provide a detailed comparison between our work and that of Wang et al. (2023).

- *The Proposed Model.* While our TRO model looks similar to Wang et al. (2023)'s model at the outset, our work actually generalizes the idea of Wang et al. (2023). Recall our TRO model in (4), which is equipped with the TRO ambiguity set $\mathcal{P}'_{N,\theta}$ in (5). The TRO ambiguity set (5) is characterized by two parameters: the *shape* parameter $\mathcal{P}_N$ and the *size* parameter θ. The shape parameter $\mathcal{P}_N$ represents distributional ambiguity and could be any (data-driven) ambiguity set satisfying some mild assumptions mentioned in the paper. The size parameter $\theta \in [0,1]$ controls the level of optimism, i.e., it controls the trade-off between solving the problem under a distributional belief and solving it under ambiguity. By choosing the shape parameter $\mathcal{P}_N$ as the distance-based ambiguity set $B_{\epsilon_N}(\widehat{\mathbb{P}}_N) = \{\mathbb{P} \in \mathcal{P}(\Xi) \mid \Delta(\mathbb{P}, \widehat{\mathbb{P}}_N) \leq \epsilon\}$ (where $\mathcal{P}(\Xi)$ is the set of probability measures on the support Ξ and Δ is a statistical distance), our TRO model reduces to the Bayesian distributionally robust (BDR) optimization model proposed in Wang et al. (2023) (see equation (11) in Wang et al. 2023), i.e., model (4) reduces to

$$\underset{\boldsymbol{x} \in \mathcal{X}}{\text{minimize}} \quad \beta_N \max_{\mathbb{P} \in B_{\epsilon_N}(\widehat{\mathbb{P}}_N)} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \boldsymbol{\xi})] + (1 - \beta_N) \mathbb{E}_{\widehat{\mathbb{P}}_N}[f(\boldsymbol{x}, \boldsymbol{\xi})]. \tag{BDR}$$

  Thus, Wang et al. (2023)'s BDR model is a special case of our TRO model. In particular, we emphasize that one can construct the TRO ambiguity set $\mathcal{P}'_{N,\theta}$ using any shape parameter $\mathcal{P}_N$, including general moment- and distance-based ambiguity sets. Hence, our theoretical results are valid for various types of the shape parameter. In contrast, Wang et al. (2023) analyses are limited to the case where $\mathcal{P}_N$ is a distance-based ambiguity set.
- *Hierarchical Properties*. In Section 2.1, we analyze properties of the TRO ambiguity set $\mathcal{P}'_{N,\theta}$ defined in (5) and the sequence of TRO ambiguity sets $\{\mathcal{P}'_{N,\theta} \mid \theta \in [0,1]\}$. These were not analyzed in Wang et al. (2023). We first formally introduce the notion of *hierarchical properties* of the sequence of the TRO ambiguity sets $\{\mathcal{P}'_{N,\theta} \mid \theta \in [0,1]\}$; see Definition 2. The hierarchical

properties indicate that the size of the TRO ambiguity set $\mathcal{P}'_{N,\theta}$ increases with θ, i.e., $\mathcal{P}'_{N,\theta}$ contains more distributions with a larger θ. This implies that the TRO model is more conservative when we pick a larger θ. Then, in Theorem 1, we provide necessary and sufficient conditions for the sequence of TRO ambiguity sets $\{\mathcal{P}'_{N,\theta} \mid \theta \in [0,1]\}$ to satisfy these properties. Specifically, Theorem 1 establishes that constructing the TRO ambiguity set $\mathcal{P}'_{N,\theta}$ using a star-shaped shape parameter $\mathcal{P}_N$ with a star center $\widehat{\mathbb{P}}_N$ is necessary and sufficient for the sequence of TRO ambiguity sets $\{\mathcal{P}'_{N,\theta} \mid \theta \in [0,1]\}$ to satisfy the hierarchical property. Part (i) shows that for a general star-shaped shape parameter $\mathcal{P}_N$, the TRO ambiguity set $\mathcal{P}'_{N,\theta}$ is non-decreasing in θ, i.e., $\mathcal{P}'_{N,\theta_1} \subseteq \mathcal{P}'_{N,\theta_2}$ whenever $\theta_1 \leq \theta_2$, indicating that the objective function of the trade-off model (4) is non-decreasing in θ. Part (ii) illustrates the relationship between the sets $\{\widehat{\mathbb{P}}_N\}$, $\mathcal{P}'_{N,\theta_1}$, $\mathcal{P}'_{N,\theta_2}$, and $\mathcal{P}_N$ with $0 < \theta_1 < \theta_2 < 1$. Specifically, part (ii) shows how the TRO ambiguity set $\mathcal{P}'_{N,\theta}$ enlarges with θ. This, in turn, implies that the TRO model is more conservative when we pick a larger θ. These important new results establish the connection between the specific choice of θ and the conservatism of the TRO model. However, they were not analyzed in Wang et al. (2023). In addition, in Proposition 2, we derive necessary and sufficient conditions for a general distance-based ambiguity set, i.e., the one adopted in the BDR model, to be star-shaped with a star center $\widehat{\mathbb{P}}_N$. Thus, our results provide necessary and sufficient conditions under which the sequence of ambiguity sets corresponding to the BDR model satisfies the hierarchical property. This was not studied in Wang et al. (2023). Moreover, as mentioned in the first point, our findings can be applied to TRO models with a general shape parameter. For instance, in Proposition 1, we establish the corresponding conditions for a general moment-based ambiguity set to be star-shaped, with a star center represented by $\widehat{\mathbb{P}}_N$, a novel contribution not explored in previous studies.

- *Analysis of the Conservatism.* In Section 2.2, we investigate the conservatism and properties of the optimal value $\widehat{\upsilon}_N(\theta)$ and the set of optimal solutions $\widehat{\mathcal{X}}_N(\theta)$ of the TRO model through the lens of quantitative stability analysis. First, in Theorem 2, we establish mechanisms to quantify the difference in $\widehat{\upsilon}_N(\theta)$ and $\widehat{\mathcal{X}}_N(\theta)$ (and hence conservatism) incurred by perturbation in θ. In particular, it shows that $\widehat{\upsilon}_N(\theta)$ is Lipschitz continuous in θ and $\widehat{\mathcal{X}}_N(\theta)$ is Hölder continuous with Hölder exponent $1/2$ under distance D. This shows that both the optimal value and the set of optimal solutions change gradually with $\theta \in [0,1]$. Then, in Theorem 3, we show that naively combining SAA and DRO optimal solutions, e.g., via a convex combination after solving each separately, may not yield a feasible solution to the TRO problem (4). This is particularly true in applications where $\mathcal{X}$ is not convex (e.g., problems involving integer variables such as facility location and scheduling problems). Specifically, part (i) of Theorem 3 establishes that the optimal value $\widehat{v}_N(\theta)$ to our TRO model is not less than the convex combination $(1-\theta)\widehat{v}_N(0)+\theta\widehat{v}_N(1)$ of the SAA and DRO optimal values. In addition, if $\widehat{\mathbb{P}}_N \in \mathcal{P}_N$, Theorem 3 implies that $\widehat{v}_N$ is non-decreasing in θ as illustrated in Figure 2. Part (ii) of Theorem 3 indicates that the set of optimal solutions $\widehat{\mathcal{X}}_N(\theta)$ to our TRO model can be approximated by $\overline{\mathcal{X}}_N(\theta) := (1-\theta)\widehat{\mathcal{X}}_N(0)+\theta\widehat{\mathcal{X}}_N(1)$ only when θ is close to zero or one; however, the difference could be huge for intermediate values of $\theta \in (0,1)$. These important investigations and results are new. In particular, Wang et al. (2023) only suspected that their BDR model is likely to be less conservative than the DRO model without providing any theoretical analysis. Indeed, we could apply our results to show that by changing β_N in Wang et al. (2023)'s BDR model, one can obtain a spectrum of optimal solutions, ranging from optimistic to conservative solutions.

- *Finite-Sample Properties.* As discussed in the first paragraph of Section 3.1, the optimal value of our TRO model $\widehat{v}_N(\theta)$ represents an estimator of the true optimal value of the stochastic optimization problem $v^\star = \inf_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}_{\mathbb{P}^\star}\big[f(\boldsymbol{x},\boldsymbol{\xi})\big]$. Analyzing the bias of an estimator to the true optimal value $v^\star$ is common in the related literature; see, e.g., Blanchet et al. (2019), Dentcheva and Lin (2022). Also, it is well known that the SAA estimator $\widehat{v}_N(0)$ is a downward biased estimator of $v^\star$; see (16). Thus, we and Wang et al. (2023) analyze the bias of the TRO and BDR estimators, respectively. Specifically, we and Wang et al. (2023) show that there exists $\theta_N^{\mathsf{u}} \in [0,1]$ in our TRO model and $\beta_N^{\mathsf{u}} \in [0,1]$ in the BDR model such that the optimal values of the TRO model and the BDR model are unbiased estimators of $v^\star$. However, in Section 3.1, we provide a more detailed investigation of the bias of the more general model, the TRO model, as well as new results. First, in Proposition 3, we derive an upper bound on the bias of the TRO estimator $\widehat{v}_N(\theta)$. It suggests that the bias of $\widehat{v}_N(\theta)$ may not be a downward bias as that of the SAA estimator. Second, in Corollary 1, we show that for sufficiently small θ, the TRO estimator $\widehat{v}_N(\theta)$ has a smaller bias than the SAA estimator, which was not discussed in Wang et al. (2023). Third, we show that $\mathbb{E}_{\mathbb{P}_N}[\widehat{v}_N(\theta)]$ can be decomposed as the sum of three terms: (a) the expected value of the SAA estimator $\mathbb{E}_{\mathbb{P}^N}[\widehat{v}_N(0)]$, (b) the DRO effect $\theta\big\{\mathbb{E}_{\mathbb{P}^N}\big[\widehat{v}_N(1)\big] - \mathbb{E}_{\mathbb{P}^N}\big[\widehat{v}_N(0)\big]\big\}$, and (c) the concavity effect $R_N(\theta)$ (see Figure 3). Finally, as pointed out by Wang et al. (2023), parameter β_N^{u} is typically hard to obtain. Similarly, θ_N^{u} is hard to obtain. However, in Theorems 5 and 6, we analyze the asymptotic behavior of θ_N^{u}. In particular, we prove the convergence of θ_N^{u} as $N \to \infty$ and derive its convergence rate. Wang et al. (2023) did not conduct such analyses. In Section 3.2, we derive the generalization bound for our TRO model in Theorem 8 based on that of the SAA and DRO models. In particular, we show that the probability (18) (i.e., the generalization error) is upper bounded by the sum of the probabilities $\alpha_{N,1}$ in (19) from the SAA model and $\alpha_{N,2}$ in (20) from the DRO model. This indeed corrects the generalization bound derived in Theorem 3.5 of Wang et al. (2023), where they did not take the sum of the two probability bounds from the SAA and DRO models; see Appendix C.6 of Wang et al. (2023). In addition, we show that for specific choices of the shape parameter $\mathcal{P}_N$, such as popular distance-based ambiguity sets as the one employed in Wang et al. (2023)'s BDR model, the generalization error exhibits an exponentially decaying tail. This important and attractive finite-sample property was not mentioned in Wang et al. (2023).

- *Asymptotic Convergence.* In Section 4, we show the almost sure convergence of the optimal value $\widehat{v}_N(\theta_N)$ and the set of optimal solutions $\widehat{\mathcal{X}}_N(\theta_N)$ of the TRO model to their true counterparts when $N \to \infty$, and we derive the asymptotic distribution of $\widehat{v}_N(\theta_N)$ when $N \to \infty$. Our asymptotic convergence results hold for TRO models with TRO ambiguity sets constructed using general shape parameters $\mathcal{P}_N$, such as moment- and distance-based ambiguity sets. This differs from results in the existing literature focusing on a specific ambiguity set, including Wang et al. (2023). Note that the asymptotic convergence and distribution are two basic asymptotic properties that are commonly analyzed in the existing literature for data-driven optimization models (Blanchet and Shapiro 2023, Kuhn et al. 2019, Shapiro et al. 2014). Hence, Wang et al. (2023) analyzed these asymptotic properties of their BDR model. However, the following are some differences between our analyses and those of Wang et al. (2023).
  - First, Wang et al. (2023) assumed the DRO objective $\sup_{\mathbb{P} \in B_{\epsilon_N}(\widehat{\mathbb{P}}_N)} \mathbb{E}[h(\boldsymbol{x}, \boldsymbol{\xi})]$ is $\mathbb{P}^\star$-bounded and attainable for $\boldsymbol{x} \in \mathcal{X}$, where we recall that $\mathbb{P}^\star$ is the true distribution; see assumption C1 of Theorem 3.3 in Wang et al. (2023). To justify this assumption, Wang et al. (2023) provided one example based on the Wasserstein ambiguity set; see Appendix C.2 in Wang et al. (2023). Our convergence analyses also require the DRO objective to be upper-bounded for sufficiently large N. However, we adopt a more general set of assumptions under which the desired boundedness condition holds. Specifically, we impose assumptions on the objective function or the sequence of the ambiguity sets $\{\mathcal{P}_N\}_{N \in \mathbb{N}}$; see Assumption 3. Examples 6–10 provide a wide range of settings under which Assumption 3 holds. These examples include the case where $\mathcal{P}_N$ is constructed based on the Wasserstein ambiguity sets, as well as other

distance-based and moment-based ambiguity sets. In Lemma 1, we formally prove that under Assumption 3, the DRO objective is *asymptotically* bounded, which is weaker than assumption C1 adopted by Wang et al. (2023) (which requires the DRO objective to be bounded for all $N \in \mathbb{N}$).

— Second, in Theorem 9, we prove the almost-sure convergence of our TRO model. Specifically, we show that the optimal value $\widehat{v}_N(\theta_N)$ and the set of optimal solutions $\widehat{\mathcal{X}}_N(\theta_N)$ of our TRO model converges almost surely to the true optimal value $v^\star$ and the set of optimal solutions $\mathcal{X}^\star$ to (1), respectively. In contrast, in Theorem 3.3 of Wang et al. (2023), they only provided the convergence of the optimal value and the set of optimal solutions of the BDR model in probability, which is weaker than our almost-sure convergence. We would like to highlight that the asymptotic convergence holds for TRO models with TRO ambiguity sets constructed using general shape parameters, such as moment-based ambiguity sets. This differs from the existing convergence results established for data-driven DRO models, which mainly employ distance-based ambiguity sets.

— Third, in Theorem 3.3 of Wang et al. (2023), they derived the asymptotic normality of the optimal value of the BDR model. To show this, they assumed that the optimal solution $\widehat{\boldsymbol{x}}_{b,N}$ to the BDR model converges (in probability) to an optimal solution $\boldsymbol{x}_0$ to the stochastic optimization problem under the true distribution $\mathbb{P}^\star$. This assumption is not common in relevant literature when deriving the asymptotic distribution of the optimal value of data-driven optimization models (Blanchet and Shapiro 2023, Guigues et al. 2018, Shapiro et al. 2014). In contrast, in Theorem 10, we derive the asymptotic distribution of $\widehat{v}_N(\theta_N)$ without imposing such an assumption. Specifically, we apply the Delta's method (see, e.g., Cárcamo et al. 2020) to derive the following asymptotic distribution of the optimal value $\widehat{v}_N = \widehat{v}_N(\theta_N)$ of the TRO model: $\sqrt{N}(\widehat{v}_N - v^\star) \Rightarrow \inf_{\boldsymbol{x}\in\mathcal{X}^\star} \mathbb{G}(\boldsymbol{x})$, where $\mathbb{G}$ is a tight Gaussian process indexed by $\mathcal{X}$ with mean zero and covariance function $\mathrm{Cov}(\mathbb{G}(\boldsymbol{x}_1)), \mathbb{G}(\boldsymbol{x}_2)) = \mathrm{Cov}_{\mathbb{P}^\star}(f(\boldsymbol{x}_1,\boldsymbol{\xi}), f(\boldsymbol{x}_2,\boldsymbol{\xi}))$. Thus, our results on the asymptotic distribution of the optimal value are different from those of Wang et al. (2023).

— Finally, for the special case when the shape parameter is chosen as some popular distance-based ambiguity set $\mathcal{P}_{N,r_N} = \{\mathbb{P} \in \mathcal{P}(\Xi) \mid \mathsf{d}(\mathbb{P}, \widehat{\mathbb{P}}_N) \le r_N\}$ as in Wang et al. (2023)'s BDR model, we can recover the asymptotics of the optimal value in classical distance-based DRO models. Specifically, in Theorem 11, we derive the asymptotic distribution of the optimal value $\widehat{v}_N(\theta_N)$ of our TRO model under different convergence rates of the size parameter θ_N

and the radius r_N in the shape parameter $\mathcal{P}_{N,r_N}$. This generalizes the asymptotic convergence of the optimal solution derived in Theorem 3.3 of Wang et al. (2023). In particular, the asymptotic distribution results in Theorem 3.3 of Wang et al. (2023) essentially correspond to the case (i) in Theorem 11 only. Cases (ii) and (iii) in Theorem 11 were not investigated in Wang et al. (2023). Moreover, we also investigate the connection between the size parameter θ_N in our TRO model (with TRO ambiguity set constructed using the shape parameter $\mathcal{P}_{N,r}$) and the radius r_N in classical distance-based DRO models. This was not analyzed in Wang et al. (2023).

References

Banholzer D, Fliege J, Werner R (2022) On rates of convergence for sample average approximations in the almost sure sense and in mean. *Mathematical Programming* 191(1):307–345.

Bayraksan G, Love DK (2015) Data-driven stochastic programming using phi-divergences. *The Operations Research Revolution*, 1–19 (INFORMS).

Blanchet J, Shapiro A (2023) Statistical limit theorems in distributionally robust optimization. *2023 Winter Simulation Conference (WSC)*, 31–45 (IEEE).

Blanchet JH, Glynn PW, Pei Y (2019) Unbiased multilevel Monte Carlo: Stochastic optimization, steady-state simulation, quantiles, and other applications. *arXiv preprint arXiv:1904.09929* .

Cárcamo J, Cuevas A, Rodríguez LA (2020) Directional differentiability for supremum-type functionals: Statistical applications. *Bernoulli* 26(3):2143–2175.

Chen L, He S, Zhang S (2011) Tight bounds for some risk measures, with applications to robust portfolio selection. *Operations Research* 59(4):847–865.

Chen M, Gao C, Ren Z (2018) Robust covariance and scatter matrix estimation under Huber's contamination model. *The Annals of Statistics* 46(5):1932–1960.

Copas JB (1988) Binary regression models for contaminated data. *Journal of the Royal Statistical Society: Series B (Methodological)* 50(2):225–253.

Dentcheva D, Lin Y (2022) Bias reduction in sample-based optimization. *SIAM Journal on Optimization* 32(1):130–151.

Gallego G, Moon I (1993) The distribution free newsboy problem: Review and extensions. *Journal of the Operational Research Society* 44(8):825–834.

Guigues V, Krätschmer V, Shapiro A (2018) A central limit theorem and hypotheses testing for risk-averse stochastic programs. *SIAM Journal on Optimization* 28(2):1337–1366.

Huang R, Qu S, Yang X, Xu F, Xu Z, Zhou W (2021) Sparse portfolio selection with uncertain probability distribution. *Applied Intelligence* 51(10):6665–6684.

Huber PJ (1964) Robust estimation of a location parameter. *The Annals of Mathematical Statistics* 35(1):73–101.

Kuhn D, Mohajerin Esfahani P, Nguyen VA, Shafieezadeh-Abadeh S (2019) Wasserstein distributionally robust optimization: Theory and applications in machine learning. *Operations Research & Management Science in the Age of Analytics*, 130–166 (Informs).

Lee S, Kim H, Moon I (2021) A data-driven distributionally robust newsvendor model with a Wasserstein ambiguity set. *Journal of the Operational Research Society* 72(8):1879–1897.

Liu T, Saldanha-da Gama F, Wang S, Mao Y (2022) Robust stochastic facility location: sensitivity analysis and exact solution. *INFORMS Journal on Computing* 34(5):2776–2803.

Liu Y, Pichler A, Xu H (2019) Discrete approximation and quantification in distributionally robust optimization. *Mathematics of Operations Research* 44(1):19–37.

Liu Y, Xu H (2013) Stability analysis of stochastic programs with second order dominance constraints. *Mathematical Programming* 142(1):435–460.

Mohajerin Esfahani P, Kuhn D (2018) Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming* 171(1):115–166.

Mu W, Xiong S (2023) On Huber's contaminated model. *Journal of Complexity* 77:101745.

Niculescu CP, Persson LE (2018) *Convex Functions and Their Applications: A Contemporary Approach* (Springer), second edition.

Pichler A, Xu H (2022) Quantitative stability analysis for minimax distributionally robust risk optimization. *Mathematical Programming* 191(1):47–77.

Rudin W (1976) *Principles of Mathematical Analysis*, volume 3 (McGraw-hill New York).

Shapiro A, Dentcheva D, Ruszczyński A (2014) *Lectures on Stochastic Programming: Modeling and Theory* (SIAM), second edition.

Sun H, Xu H (2016) Convergence analysis for distributionally robust optimization and equilibrium problems. *Mathematics of Operations Research* 41(2):377–401.

van der Vaart AW (2000) *Asymptotic Statistics*, volume 3 (Cambridge university press).

van der Vaart AW, Wellner JA (1996) *Weak Convergence and Empirical Processes: With Applications in Statistics* (Springer).

Wang S, Wang H, Honorio J (2023) Learning against distributional uncertainty: On the trade-off between robustness and specificity. *arXiv preprint arXiv:2301.13565* .

Yin Y, Yang Y, Yu Y, Wang D, Cheng T (2023) Robust vehicle routing with drones under uncertain demands and truck travel times in humanitarian logistics. *Transportation Research Part B: Methodological* 174:102781.