Online Appendices for:
Transfer Learning, Cross Learning and Co-Learning with Operational Data Analytics (ODA)

Qi Feng[†], Lei Li[‡], and J. George Shanthikumar[†]

[†]Mitch Daniels School of Business, Purdue University, West Lafayette, IN 47907

annabellefeng@purdue.edu shanthikumar@purdue.edu

[‡]Department of Logistics and Maritime Studies, Faculty of Business, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China

leihk.li@polyu.edu.hk

A Proofs of Formal Results

Proof of Lemma 1. By the Weak Law of Large Numbers, we have $\lim_{\ell\to\infty}\hat{\mu}(\boldsymbol{X}^o) =^p \theta^o \in (0,\infty)$, where superscript p stands for "in probability." By Continuous Mapping Theorem, we have

$$\frac{1}{\hat{\mu}_{\boldsymbol{X}^o}} \xrightarrow{d} \frac{1}{\theta^o} \text{ as } n^o \to \infty.$$

By Slutsky's theorem, we deduce that $\lim_{n^o\to\infty}\frac{X_i^o}{\hat{\mu}_{\boldsymbol{X}^o}} =^d \frac{X}{\theta^o} =^d Z$. □

Proof of Theorem 2. By Slutsky's theorem, we have $\lim_{n^o\to\infty} y^o(\boldsymbol{X}^o)/(\hat{\mu}_{\boldsymbol{X}^o})^\iota =^d y^*[F_{X^o}]/(\theta^o)^\iota$. In view of Assumption 3, we obtain the first relation in (17). The second relation in (17) then follows immediately as $\hat{\mu}_{\boldsymbol{X}} \to \theta$.

Note that $\Pr\left\{\left|\hat{\mu}_{\boldsymbol{X}^o} - \theta^o\right| \ge \epsilon\right\} \le \mathcal{O}\big(1/(n^o\epsilon^2)\big)$ by hebyshev's Inequality. By Delta Method, $\Pr\left\{\left|g(\hat{\mu}_{\boldsymbol{X}^o}) - g(\theta^o)\right| \ge \epsilon\right\} \le \mathcal{O}\big(1/(n^o\epsilon^2)\big)$ holds for differentiable function $g(\cdot)$. Then,

$$\Pr\left\{\left|\frac{1}{(\hat{\mu}_{\boldsymbol{X}^o})^\iota} - \frac{1}{(\theta^o)^\iota}\right| \ge \epsilon\right\} \le \mathcal{O}\left(\frac{1}{n^o\epsilon^2}\right), \text{ and similarly} \Pr\left\{\left|(\hat{\mu}_{\boldsymbol{X}})^\iota - (\theta)^\iota\right| \ge \epsilon\right\} \le \mathcal{O}\left(\frac{1}{n\epsilon^2}\right).$$

To account for the fact that $\frac{1}{(\hat{\mu}_{\boldsymbol{X}^o})^\iota}$ may be unbounded, for any $\epsilon > 0$, we choose M to be a sufficiently large number such that $M/(\theta^o)^\iota > \epsilon$. We have

$$
\begin{aligned}
& \Pr\left\{\frac{1}{(\hat{\mu}_{\boldsymbol{X}^o})^\iota}\big|y^o(\boldsymbol{X}^o) - y^*[F_{X^o}]\big| \geq \frac{\epsilon}{2}\right\} \\
\leq\ & \Pr\left\{\frac{\big|y^o(\boldsymbol{X}^o) - y^*[F_{X^o}]\big|}{(\hat{\mu}_{\boldsymbol{X}^o})^\iota}\mathbb{I}_{\left\{\frac{1}{(\hat{\mu}_{\boldsymbol{X}^o})^\iota} > \frac{1+M}{(\theta^o)^\iota}\right\}} \geq \frac{\epsilon}{4}\right\} + \Pr\left\{\frac{1+M}{(\theta^o)^\iota}\big|y^o(\boldsymbol{X}^o) - y^*[F_{X^o}]\big| \geq \frac{\epsilon}{4}\right\} \\
\leq\ & \Pr\left\{\frac{1}{(\hat{\mu}_{\boldsymbol{X}^o})^\iota} > \frac{1+M}{(\theta^o)^\iota}\right\} + \mathcal{O}\Big(g^o(n^o,\epsilon)\Big) \\
\leq\ & \Pr\left\{\left|\frac{1}{(\hat{\mu}_{\boldsymbol{X}^o})^\iota} - \frac{1}{(\theta^o)^\iota}\right| \geq \epsilon\right\} + \mathcal{O}\Big(g^o(n^o,\epsilon)\Big) \\
\leq\ & \mathcal{O}\bigg(\frac{1}{n^o\epsilon^2}\bigg) + \mathcal{O}\Big(g^o(n^o,\epsilon)\Big).
\end{aligned}
$$

Note that $|A_nB_n - ab| \leq |A_n||B_n - b| + |b||A_n - a|$. Then, we have

$$
\begin{aligned}
& \Pr\left\{\left|\frac{y^o(\boldsymbol{X}^o)}{(\hat{\mu}_{\boldsymbol{X}^o})^\iota} - \frac{y^*[F_{X^o}]}{(\theta^o)^\iota}\right| \geq \epsilon\right\} \\
\leq\ & \Pr\left\{\frac{1}{(\hat{\mu}_{\boldsymbol{X}^o})^\iota}\big|y^o(\boldsymbol{X}^o) - y^*[F_{X^o}]\big| \geq \frac{\epsilon}{2}\right\} + \Pr\left\{y^*[F_{X^o}]\big|(\hat{\mu}_{\boldsymbol{X}^o})^\iota - (\theta^o)^\iota\big| \geq \frac{\epsilon}{2}\right\} \\
\leq\ & \mathcal{O}\Big(g^o(n^o,\epsilon)\Big) + \mathcal{O}\bigg(\frac{1}{n^o\epsilon^2}\bigg).
\end{aligned}
$$

Thus, we obtain (18). Applying a similar argument with $M(\theta)^\iota > \epsilon$, we derive

$$
\begin{aligned}
& \Pr\left\{\left|(\hat{\mu}_{\boldsymbol{X}})^\iota\frac{y^o(\boldsymbol{X}^o)}{(\hat{\mu}_{\boldsymbol{X}^o})^\iota} - (\theta)^\iota\frac{y^*[F_{X^o}]}{(\theta^o)^\iota}\right| \geq \epsilon\right\} \\
\leq\ & \Pr\left\{(\hat{\mu}_{\boldsymbol{X}})^\iota\left|\frac{y^o(\boldsymbol{X}^o)}{(\hat{\mu}_{\boldsymbol{X}^o})^\iota} - \frac{y^*[F_{X^o}]}{(\theta^o)^\iota}\right| \geq \frac{\epsilon}{2}\right\} + \Pr\left\{\frac{y^*[F_{X^o}]}{(\theta^o)^\iota}\big|(\hat{\mu}_{\boldsymbol{X}})^\iota - (\theta)^\iota\big| \geq \frac{\epsilon}{2}\right\} \\
\leq\ & \Pr\left\{\left|\frac{y^o(\boldsymbol{X}^o)}{(\hat{\mu}_{\boldsymbol{X}^o})^\iota} - \frac{y^*[F_{X^o}]}{(\theta^o)^\iota}\right|(\hat{\mu}_{\boldsymbol{X}})^\iota\mathbb{I}_{\left\{(\hat{\mu}_{\boldsymbol{X}})^\iota > (1+M)\theta^\iota\right\}} \geq \frac{\epsilon}{2}\right\} \\
& + \Pr\left\{\left|\frac{y^o(\boldsymbol{X}^o)}{(\hat{\mu}_{\boldsymbol{X}^o})^\iota} - \frac{y^*[F_{X^o}]}{(\theta^o)^\iota}\right|(1+M)\theta^\iota \geq \frac{\epsilon}{2}\right\} + \mathcal{O}\bigg(\frac{1}{n\epsilon^2}\bigg) \\
\leq\ & \Pr\left\{\big|(\hat{\mu}_{\boldsymbol{X}})^\iota - \theta^\iota\big| > M\theta^\iota\right\} + \mathcal{O}\Big(g^o(n^o,\epsilon)\Big) + \mathcal{O}\bigg(\frac{1}{n^o\epsilon^2}\bigg) + \mathcal{O}\bigg(\frac{1}{n\epsilon^2}\bigg) \\
\leq\ & \mathcal{O}\Big(g^o(n^o,\epsilon)\Big) + \mathcal{O}\bigg(\frac{1}{n^o\epsilon^2}\bigg) + \mathcal{O}\bigg(\frac{1}{n\epsilon^2}\bigg).
\end{aligned}
$$

This gives rise to (19). □

The proof of Theorem 3 uses the results in Lemmas A.3-A.9.

Lemma A.1 (Dudley's Entropy Integral) *(Theorem 5.22 in Wainwright 2019) Let X_y be a zero-mean sub-Gaussian stochastic process with respect to a distance metric d on the indexing $\mathcal{Y}$. Then we have,*

$$
\mathsf{E}[\sup_{y\in\mathcal{Y}} X_y] \leq 32\int_0^\infty \sqrt{\log N(\delta, \mathcal{Y}, d)}\mathrm{d}\delta,
$$

where $N(\epsilon, \mathcal{Y}, d)$ is the ϵ-covering number of $\mathcal{Y}$ with distance metric d.

Lemma A.2 (Uniform Weak Law of Large Numbers) *Suppose $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ are i.i.d. samples of F_X over support $\mathcal{X}$ and let $\mathcal{F} := \{\psi(y, \cdot) : y \in \mathcal{Y}\}$ be a class of uniformly bounded functions of the form $\psi : \mathcal{Y} \times \mathcal{X} \to [a, b] \subset \mathbb{R}$. Then, for any $\epsilon > 0$, we have*

$$\Pr\left\{\sup_{y\in\mathcal{Y}}\left|\frac{1}{n}\sum_{i=1}^{n}\psi(y,X_i)-\mathsf{E}[\psi(y,X)]\right|\geq 2\mathcal{R}_n(\mathcal{F})+\epsilon\right\}\leq 2\exp\Big(-\frac{2n\epsilon^2}{(b-a)^2}\Big), \tag{37}$$

where $\mathcal{R}_n(\mathcal{F}) := \mathsf{E}_{\boldsymbol{X},\boldsymbol{\sigma}}\Big[\sup_{\psi(y,\cdot)\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}\sigma_i\psi(y,X_i)\right|\Big]$ is the Rademacher complexity of $\mathcal{F}$, and $\sigma_i \in \{-1, 1\}$ with equal probability.

If, in addition, $\psi(\cdot, \cdot)$ is Lipschitz continuous and bounded, and $\mathcal{Y} \subset \mathbb{R}^q$ is compact, then $\mathcal{R}_n(\mathcal{F}) \leq C/\sqrt{n}$ for some $C > 0$, and thus for any $\epsilon > 0$

$$\Pr\left\{\sup_{y\in\mathcal{Y}}\left|\frac{1}{n}\sum_{i=1}^{n}\psi(y,X_i)-\mathsf{E}[\psi(y,X)]\right|\geq \epsilon\right\}\leq \mathcal{O}\bigg(\exp\Big(-\frac{2n\epsilon^2-4C\sqrt{n}\epsilon}{(b-a)^2}\Big)\bigg).$$

Proof. The first result can be found in Wainwright (2019). We provide the derivation below for the sake of completeness. Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ and define

$$G(\boldsymbol{x}) := \sup_{y\in\mathcal{Y}}\left|\frac{1}{n}\sum_{i=1}^{n}\psi(y,x_i)-\mathsf{E}[\psi(y,X)]\right|.$$

Define a set of n vectors $\boldsymbol{x}^\ell = (x_1^\ell, x_2^\ell, \ldots, x_n^\ell), \ell \in \{1, 2, \ldots, n\}$ such that $x_i^\ell = x_i$ for $i \neq \ell$. That is, vector $\boldsymbol{x}^\ell$ differs from $\boldsymbol{x}$ only on the ℓth element. Then

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^{n}\psi(y,x_i)-\mathsf{E}[\psi(y,X)]\right|-\sup_{q\in\mathcal{Y}}\left|\frac{1}{n}\sum_{i=1}^{n}\psi(q,x_i^\ell)-\mathsf{E}[\psi(q,X)]\right|\\
\leq\;&\left|\frac{1}{n}\sum_{i=1}^{n}\psi(y,x_i)-\mathsf{E}[\psi(y,X)]\right|-\left|\frac{1}{n}\sum_{i=1}^{n}\psi(y,x_i^\ell)-\mathsf{E}[\psi(y,X)]\right|\\
\leq\;&\frac{1}{n}\left|\psi(y,x_\ell)-\psi(y,x_\ell^\ell)\right|\leq\frac{b-a}{n}.
\end{aligned}$$

Because this inequality above holds for any $y \in \mathcal{Y}$, we deduce

$$|G(\boldsymbol{x}^\ell)-G(\boldsymbol{x})|=\left|\sup_{y\in\mathcal{Y}}\left|\frac{1}{n}\sum_{i=1}^{n}\psi(y,x_i)-\mathsf{E}[\psi(y,X)]\right|-\sup_{y\in\mathcal{Y}}\left|\frac{1}{n}\sum_{i=1}^{n}\psi(y,x_i^\ell)-\mathsf{E}[\psi(y,X)]\right|\right|\leq\frac{b-a}{n},$$

Because the inequality holds for all $\ell = 1, 2 \ldots, n$, McDiarmid's inequality implies that, for any $\varepsilon > 0$,

$$\Pr\left\{\left|G(\boldsymbol{X})-\mathsf{E}[G(\boldsymbol{X})]\right|\geq\epsilon\right\}\leq 2\exp\Big(-\frac{2n\epsilon^2}{(b-a)^2}\Big). \tag{38}$$

Next we show $\mathsf{E}[G(\boldsymbol{X})] \leq 2\mathcal{R}_n(\mathcal{F})$. Let $(Y_1, Y_2, \ldots, Y_n)$ be i.i.d. random variables independent and identical in distribution to $(X_1, X_2, \ldots, X_n)$. By symmetrization, we have

$$
\begin{aligned}
\mathsf{E}[G(\boldsymbol{X})] &= \mathsf{E}_{\boldsymbol{X}}\left[\sup_{y\in\mathcal{Y}}\left|\frac{1}{n}\sum_{i=1}^{n}\psi(y,X_i) - \mathsf{E}[\psi(y,X)]\right|\right] \\
&= \mathsf{E}_{\boldsymbol{X}}\left[\sup_{y\in\mathcal{Y}}\left|\frac{1}{n}\sum_{i=1}^{n}\big(\psi(y,X_i) - \mathsf{E}[\psi(y,Y_i)]\big)\right|\right] \\
&= \mathsf{E}_{\boldsymbol{X}}\left[\sup_{y\in\mathcal{Y}}\left|\mathsf{E}_{\boldsymbol{Y}}\left[\frac{1}{n}\sum_{i=1}^{n}\big(\psi(y,X_i) - \psi(y,Y_i)\big)\right]\right|\right] \\
&\leq \mathsf{E}_{\boldsymbol{X},\boldsymbol{Y}}\left[\sup_{y\in\mathcal{Y}}\left|\frac{1}{n}\sum_{i=1}^{n}\big(\psi(y,X_i) - \psi(y,Y_i)\big)\right|\right] \\
&= \mathsf{E}_{\boldsymbol{X},\boldsymbol{Y},\boldsymbol{\sigma}}\left[\sup_{y\in\mathcal{Y}}\left|\frac{1}{n}\sum_{i=1}^{n}\sigma_i\big(\psi(y,X_i) - \psi(y,Y_i)\big)\right|\right] \\
&\leq 2\mathsf{E}_{\boldsymbol{X},\boldsymbol{\sigma}}\left[\sup_{y\in\mathcal{Y}}\left|\frac{1}{n}\sum_{i=1}^{n}\sigma_i\psi(y,X_i)\right|\right] \\
&= 2\mathcal{R}_n(\mathcal{F}).
\end{aligned}
$$

The first inequality follows from the concavity of supremum. The fourth equality follows from the fact that $\sigma_i \in \{-1, 1\}$ with equal probability. The last inequality follows from the triangle inequality. Together with (38), we obtain (37).

Next we show $\mathcal{R}_n(\mathcal{F}) \leq C/\sqrt{n}$. Note that $\sqrt{n}\left|\frac{1}{n}\sum_{i=1}^{n}\sigma_i\psi(y,X_i)\right|$ is a zero-mean sub-Gaussian stochastic process, because $\psi(y,\cdot)$ is bounded for all $y \in \mathcal{Y}$, with respect to the Euclidean distance $||\cdot||_2$ on the indexing set $\mathcal{F}$. Thus, by Dudley's Entropy Integral (see Lemma A.1), we have

$$
\begin{aligned}
\mathcal{R}_n(\mathcal{F}) &= \frac{1}{\sqrt{n}}\mathsf{E}_{\boldsymbol{X},\boldsymbol{\sigma}}\Big[\sup_{\psi(y,\cdot)\in\mathcal{F}}\sqrt{n}\Big|\frac{1}{n}\sum_{i=1}^{n}\sigma_i\psi(y,X_i)\Big|\Big] \\
&\leq \frac{32}{\sqrt{n}}\int_0^\infty \sqrt{\log\ N(\delta,\mathcal{F},||\cdot||_2)}\mathrm{d}\delta \\
&\leq \frac{32}{\sqrt{n}}\int_0^\infty \sqrt{\log\ N(\delta/L,\mathcal{Y},||\cdot||_2)}\mathrm{d}\delta \\
&\leq \frac{32L}{\sqrt{n}}\int_0^\infty \sqrt{\log\ N(\delta,\mathcal{Y},||\cdot||_2)}\mathrm{d}\delta \\
&\leq \frac{32\sqrt{q}L}{\sqrt{n}}\int_0^{|\mathcal{Y}|} \sqrt{\log\big(1+2|\mathcal{Y}|/\delta\big)}\mathrm{d}\delta \\
&\leq \frac{C}{\sqrt{n}},
\end{aligned}
$$

where $C := (32\sqrt{q}L)\cdot|\mathcal{Y}|(\log 3 - 3\sqrt{2\pi})$ and $|\mathcal{Y}| := (\overline{y} - \underline{y})$ for $\mathcal{Y} \subseteq [\overline{y}, \underline{y}]^q$. The first and third inequalities follow from the fact that $\mathcal{N}(\delta,\mathcal{F},||\cdot||_2) \leq \mathcal{N}(\delta/L,\mathcal{Y},||\cdot||_2) \leq (1+2L|\mathcal{Y}|/\delta)^q$ because

$\psi(y,\cdot) \in \mathcal{F}, \forall y \in \mathcal{Y}$, are Lipschitz continuous. The last inequality follows from the following relation with $\delta = 3|\mathcal{Y}|e^{-a^2}$:

$$
\begin{aligned}
\int_0^{|\mathcal{Y}|} \sqrt{\log\big(1 + 2|\mathcal{Y}|/\delta\big)}\mathrm{d}\delta &\leq \int_0^{|\mathcal{Y}|} \sqrt{\log\big(3|\mathcal{Y}|/\delta\big)}\mathrm{d}\delta \\
&\leq \int_0^{\sqrt{\log 3}} a\mathrm{d}3|\mathcal{Y}|e^{-a^2} \\
&= 3|\mathcal{Y}|(ae^{-a^2})\Big|_0^{\sqrt{\log 3}} - 3|\mathcal{Y}|\int_0^{\sqrt{\log 3}} e^{-a^2}\mathrm{d}a \\
&\leq |\mathcal{Y}|(\log 3 - 3\sqrt{2\pi}).
\end{aligned}
$$

Because $\mathcal{R}_n(\mathcal{F}) \leq C/\sqrt{n}$ for the class of Lipschitz continuous and bounded functions, applying (37) with $\epsilon = \varepsilon - 2\mathcal{R}_n(\mathcal{F})$ yields

$$
\begin{aligned}
\Pr\left\{\sup_{y\in\mathcal{Y}}\left|\frac{1}{n}\sum_{i=1}^n \psi(y, X_i) - \mathsf{E}[\psi(y, X)]\right| \geq \varepsilon\right\} &\leq 2\exp\Big(-\frac{2n(\varepsilon - 2\mathcal{R}_n(\mathcal{F}))^2}{(b-a)^2}\Big) \\
&\leq 2\exp\Big(-\frac{2n(\varepsilon - 2C/\sqrt{n})^2}{(b-a)^2}\Big) \\
&= \mathcal{O}\bigg(\exp\Big(-\frac{2n\varepsilon^2 - 4C\sqrt{n}\varepsilon}{(b-a)^2}\Big)\bigg).
\end{aligned}
$$

This completes the proof. □

Lemma A.3 (Uniform Weak Law of Large Numbers) *Suppose $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ are i.i.d. draws from F_X over support $\mathcal{X}$ and let $\psi : \mathcal{Y} \times \mathcal{X} \to [a, b] \subset \mathbb{R}$ be Lipschitz for some $\mathcal{X} \subset \mathbb{R}$ and some compact $\mathcal{Y} \subset \mathbb{R}^q$. Then, for any $\epsilon > 0$,*

$$
\Pr\left\{\sup_{y\in\mathcal{Y}}\left|\frac{1}{n}\sum_{i=1}^n \psi(y, X_i) - \mathsf{E}[\psi(y, X)]\right| \geq \epsilon\right\} \leq \mathcal{O}\bigg(\frac{1}{\epsilon^q}\exp\Big(-\frac{2n\epsilon^2}{9(b-a)^2}\Big)\bigg).
$$

Proof. The convergence result was first proven by Jennrich (1969). We derive the convergence rate by ϵ-covering number and Hoeffding's inequality for the sake of completeness.

When $\mathcal{Y}$ is finite, i.e., $\mathcal{Y} = \{y_1, y_2, \ldots, y_\ell\}$ for some finite integer $\ell \in \mathbb{Z}_+$. Because $\psi(y, X)$ is bounded between a and b, Hoeffding's inequality suggests that for any $\epsilon > 0$,

$$
\begin{aligned}
\Pr\left\{\sup_{y\in\mathcal{Y}}\left|\frac{1}{n}\sum_{i=1}^n \psi(y, X_i) - \mathsf{E}[\psi(y, X)]\right| \geq \epsilon\right\} &\leq \Pr\left\{\sum_{j=1}^{\ell}\left|\frac{1}{n}\sum_{i=1}^n \psi(y_j, X_i) - \mathsf{E}[\psi(y_j, X)]\right| \geq \epsilon\right\} \\
&\leq \sum_{j=1}^{\ell} 2\exp\Big(-\frac{2n\epsilon^2}{(b-a)^2}\Big) = \mathcal{O}\bigg(\exp\Big(-\frac{2n\epsilon^2}{(b-a)^2}\Big)\bigg).
\end{aligned}
$$

Now we treat the general case where $\mathcal{Y}$ is not finite. Because $\psi(\cdot, x)$ is Lipschitz continuous on some compact set $\mathcal{Y}$ for each $x \in \mathcal{X}$, there exist an ε-cover, denoted by $\{\psi(y_1,\cdot), \psi(y_2,\cdot), \ldots, \psi(y_\ell,\cdot)\}$

for some finite $\ell \in \mathbb{Z}_+$, of the function set $\{\psi(y,\cdot) : y \in \mathcal{Y}\}$, such that for every $y \in \mathcal{Y}$, there exists a $y_j \in \{y_1, y_2, \ldots, y_\ell\}$ satisfying

$$\sup_{x\in\mathcal{X}} |\psi(y,x) - \psi(y_j,x)| < \varepsilon.$$

Then, for any $y \in \mathcal{Y}$,

$$\begin{aligned}
& \left|\frac{1}{n}\sum_{i=1}^{n}\psi(y,X_i) - \mathsf{E}[\psi(y,X)]\right| \\
\leq\ & \left|\frac{1}{n}\sum_{i=1}^{n}\psi(y,X_i) - \frac{1}{n}\sum_{i=1}^{n}\psi(y_j,X_i)\right| + \left|\frac{1}{n}\sum_{i=1}^{n}\psi(y_j,X_i) - \mathsf{E}[\psi(y_j,X)]\right| + \Big|\mathsf{E}[\psi(y_j,X)] - \mathsf{E}[\psi(y,X)]\Big| \\
\leq\ & \frac{1}{n}\sum_{i=1}^{n}\big|\psi(y,X_i) - \psi(y_j,X_i)\big| + \left|\frac{1}{n}\sum_{i=1}^{n}\psi(y_j,X_i) - \mathsf{E}[\psi(y_j,X)]\right| + \mathsf{E}\big[|\psi(y_j,X)] - \psi(y,X)|\big] \\
\leq\ & \left|\frac{1}{n}\sum_{i=1}^{n}\psi(y_j,X_i) - \mathsf{E}[\psi(y_j,X)]\right| + 2\varepsilon.
\end{aligned}$$

We further note that $\ell = \mathcal{O}(\varepsilon^{-q})$ because for each $x \in \mathcal{X}$, $\psi(\cdot,x)$ is Lipschitz continuous on some compact set $\mathcal{Y} \subset \mathbb{R}^q$. Thus, by setting $\epsilon = 3\varepsilon$, we derive

$$\begin{aligned}
& \Pr\left\{\sup_{y\in\mathcal{Y}}\left|\frac{1}{n}\sum_{i=1}^{n}\psi(y,X_i) - \mathsf{E}[\psi(y,X)]\right| \geq \epsilon\right\} \\
\leq\ & \Pr\left\{\max_{j\in 1,2,\ldots,\ell}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\psi(y_j,X_i) - \mathsf{E}[\psi(y_j,X)]\right| + 2\varepsilon\right\} \geq \epsilon\right\} \\
\leq\ & \sum_{j=1}^{\ell}\Pr\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\psi(y_j,X_i) - \mathsf{E}[\psi(y_j,X)]\right| \geq \frac{1}{3}\epsilon\right\} \\
=\ & \mathcal{O}\bigg(\frac{1}{\epsilon^q}\exp\Big(-\frac{2n\epsilon^2}{9(b-a)^2}\Big)\bigg).
\end{aligned}$$

This completes the proof. □

Lemma A.4 *Let* $\hat{\boldsymbol{Z}} = (\hat{Z}_1, \hat{Z}_2, \ldots, \hat{Z}_n)$ *and* $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_n)$*, respectively, be* n *i.i.d. draws of distribution* $F_{\hat{Z}}$ *and* F_Z *over the support* $\mathbb{R}_+$*. If* $F_{\hat{Z}}(z) \to F_Z(z), \forall z \in \mathbb{R}_+$*, then*

$$F_{\hat{\mu}_{\hat{\boldsymbol{Z}}}}(z) \to F_{\hat{\mu}_{\boldsymbol{Z}}}(z), \forall z \in \mathbb{R}_+.$$

Proof. For $n = 2$, we have

$$
\begin{aligned}
& \left|F_{\hat{\mu}_{\hat{\boldsymbol{Z}}}}(z/2) - F_{\hat{\mu}_{\boldsymbol{Z}}}(z/2)\right| \\
= & \left|\int_{\xi<z} F_{\hat{Z}}(z-\xi)\mathrm{d}F_{\hat{Z}}(\xi) - \int_{\xi<z} F_Z(z-\xi)\mathrm{d}F_Z(\xi)\right| \\
\le & \left|\int_{\xi<z} F_{\hat{Z}}(z-\xi)\mathrm{d}F_{\hat{Z}}(\xi) - \int_{\xi<z} F_Z(z-\xi)\mathrm{d}F_{\hat{Z}}(\xi)\right| \\
& + \left|\int_{\xi<z} F_Z(z-\xi)\mathrm{d}F_{\hat{Z}}(\xi) - \int_{\xi<z} F_Z(z-\xi)\mathrm{d}F_Z(\xi)\right| \\
= & \left|\int_{\xi<z} \big(F_{\hat{Z}}(z-\xi) - F_Z(z-\xi)\big)\mathrm{d}F_{\hat{Z}}(\xi)\right| + \left|\int_{\xi<z} \big(F_{\hat{Z}}(z-\xi) - F_Z(z-\xi)\big)\mathrm{d}F_Z(\xi)\right| \\
\le & \int_{\xi<z} \big|F_{\hat{Z}}(z-\xi) - F_Z(z-\xi)\big|\mathrm{d}F_{\hat{Z}}(\xi) + \int_{\xi<z} \big|F_{\hat{Z}}(z-\xi) - F_Z(z-\xi)\big|\mathrm{d}F_Z(\xi).
\end{aligned}
$$

Because $F_{\hat{Z}} \in [0,1]$ and $F_Z \in [0,1]$, by the bounded convergence theorem, we have $\big|F_{\hat{\mu}_{\hat{\boldsymbol{Z}}}}(z/2) - F_{\hat{\mu}_{\boldsymbol{Z}}}(z/2)\big| \to 0, \forall z \in \mathbb{R}_+$. It follows $\big|F_{\hat{Z}_1+\hat{Z}_2}(z) - F_{Z_1+Z_2}(z)\big| \to 0, \forall z \in \mathbb{R}_+$.

Now suppose that the result is true for $(n-1)$. We derive

$$
\begin{aligned}
& \left|F_{\hat{\mu}_{\hat{\boldsymbol{Z}}}}(z/n) - F_{\hat{\mu}_{\boldsymbol{Z}}}(z/n)\right| \\
= & \left|\int_{\xi<z} F_{\sum_{i=1}^{n-1}\hat{Z}_i}(z-\xi)\mathrm{d}F_{\hat{Z}}(\xi) - \int_{\xi<z} F_{\sum_{i=1}^{n-1} Z_i}(z-\xi)\mathrm{d}F_Z(\xi)\right| \\
\le & \left|\int_{\xi<z} F_{\sum_{i=1}^{n-1}\hat{Z}_i}(z-\xi)\mathrm{d}F_{\hat{Z}}(\xi) - \int_{\xi<z} F_{\sum_{i=1}^{n-1} Z_i}(z-\xi)\mathrm{d}F_{\hat{Z}}(\xi)\right| \\
& + \left|\int_{\xi<z} F_{\sum_{i=1}^{n-1} Z_i}(z-\xi)\mathrm{d}F_{\hat{Z}}(\xi) - \int_{\xi<z} F_{\sum_{i=1}^{n-1} Z_i}(z-\xi)\mathrm{d}F_Z(\xi)\right| \\
= & \left|\int_{\xi<z} F_{\sum_{i=1}^{n-1}\hat{Z}_i}(z-\xi) - F_{\sum_{i=1}^{n-1} Z_i}(z-\xi)\mathrm{d}F_{\hat{Z}}(\xi)\right| + \left|\int_{\xi<z} F_{\hat{Z}}(z-\xi) - F_Z(z-\xi)\mathrm{d}F_{\sum_{i=1}^{n-1} Z_i}(\xi)\right| \\
\le & \int_{\xi<z} \big|F_{\sum_{i=1}^{n-1}\hat{Z}_i}(z-\xi) - F_{\sum_{i=1}^{n-1} Z_i}(z-\xi)\big|\mathrm{d}F_{\hat{Z}}(\xi) + \int_{\xi<z} \big|F_{\hat{Z}}(z-\xi) - F_Z(z-\xi)\big|\mathrm{d}F_{\sum_{i=1}^{n-1} Z_i}(\xi).
\end{aligned}
$$

We conclude the proof. □

Lemma A.5 *Suppose $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_n)$ are n i.i.d. draws of $F_Z(\cdot)$ over some support $\mathcal{Z} \in \mathbb{R}_+$ with $\mathsf{E}[Z] = 1$, and $\boldsymbol{X}^o = (X_1^o, X_2^o, \ldots, X_{n^o}^o)$ are n^o i.i.d. draws of $F_Z(\cdot/\theta^o)$ for some finite $\theta^o \in \mathbb{R}_+$. Let $\hat{F}_{Z|\boldsymbol{Z}^o}$ denote the and $\tilde{f}_{Z|\boldsymbol{Z}^o}$, respectively, denote the empirical distribution and the smoothed empirical density of $\boldsymbol{Z}^o = \boldsymbol{X}^o/\hat{\mu}_{\boldsymbol{X}^o}$, and $\hat{\boldsymbol{Z}} = (\hat{Z}_1, \hat{Z}_2, \ldots, \hat{Z}_n)$ be n i.i.d. draws of $\hat{F}_{Z|\boldsymbol{Z}^o}$ or $\tilde{f}_{Z|\boldsymbol{Z}^o}$. Then, for any continuous $g : \mathbb{R} \to \mathbb{R}$ and every sample $\boldsymbol{X}^o = \boldsymbol{x}^o$,*

$$g(\hat{Z}_1) \xrightarrow{d} g(Z) \text{ and } g(\hat{\mu}_{\hat{\boldsymbol{Z}}}) \xrightarrow{d} g(\hat{\mu}_{\boldsymbol{Z}}) \text{ as } n^o \to \infty.$$

Proof. By the strong law of large numbers, we have, along almost every sample path of $\boldsymbol{X}^o(\omega)$, $\hat{\mu}_{\boldsymbol{x}^o} \to \theta^o < \infty$, and

$$\hat{F}_{Z|\boldsymbol{Z}^o(\omega)}(z) = \frac{1}{n^o}\sum_{i=1}^{n^o} \mathbb{I}_{\{Z_i^o(\omega)\le z\}} \to F_Z(z),\ z \in \mathbb{R}.$$

Because $\tilde{f}_{Z|\boldsymbol{Z}^o(\omega)}(z) = \frac{1}{n\lambda}\sum_{i=1}^n \kappa_\lambda(\xi, Z_i) \to f_Z(z),\ z \in \mathbb{R}$ along almost every sample path of $\boldsymbol{X}^o(\omega)$, by dominated convergence theorem, we have

$$\tilde{F}_{Z|\boldsymbol{Z}^o(\omega)}(z) = \int_{-\infty}^{z} f_{Z|\boldsymbol{Z}^o(\omega)}(\xi)\mathrm{d}\xi \to F_Z(z),\ z \in \mathbb{R}.$$

Thus, $\hat{Z}_i \xrightarrow{d} Z$. By Lemma A.4, $F_{\hat{\mu}_{\hat{\boldsymbol{Z}}}} \xrightarrow{d} F_{\hat{\mu}_{\boldsymbol{Z}}}$. We deduce $\hat{\mu}_{\hat{\boldsymbol{Z}}} \xrightarrow{d} \hat{\mu}_{\boldsymbol{Z}}$. By Continuous Mapping Theorem, $g(\hat{Z}_1) \xrightarrow{d} g(Z)$ and $g(\hat{\mu}_{\hat{\boldsymbol{Z}}}) \xrightarrow{d} g(\hat{\mu}_{\boldsymbol{Z}})$ as $n^o \to \infty$ for any continuous $g : \mathbb{R} \to \mathbb{R}$. □

Lemma A.6 *Suppose* $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ *are i.i.d. samples of* F_X *over support* $\mathcal{X}$. *Define* $\Phi_n(y) := \frac{1}{n}\sum_{i=1}^n \psi(y, X_i)$ *and* $\phi(y) := \mathsf{E}[\psi(y, X)]$ *with* $\psi : \mathcal{Y} \times \mathcal{X} \to [a, b] \subset \mathbb{R}$. *Denote* Y_n^* *as a maximizer of* $\Phi_n(y)$. *If*

i) $\phi : \mathcal{Y} \to \mathbb{R}$ *has a unique maximizer* y^* *over some compact set* $\mathcal{Y} \subset \mathbb{R}^q$*;*

ii) $\psi(\cdot, x)$ *is Lipschitz continuous and bounded over* $\mathcal{Y}$ *a.s.*

then for any $\epsilon > 0$*,*

$$\lim_{n\to\infty} \Pr\left\{||Y_n^* - y^*|| \ge \epsilon\right\} = 0.$$

Proof. This result and relevant discussions can be found in van der Vaart and Wellner (1996). We provide the derivation below for the sake of completeness. Because Y_n^* and y^* are the maximizer of Φ_n and ϕ, respectively, then we must have, a.s.,

$$\begin{aligned} 0 \le \phi(y^*) - \phi(Y_n^*) &\le \phi(y^*) - \Phi_n(y^*) + \left(\Phi_n(Y_n^*) - \phi(Y_n^*)\right) \\ &\le \left|\phi(y^*) - \Phi_n(y^*)\right| + \left|\Phi_n(Y_n^*) - \phi(Y_n^*)\right| \\ &\le 2\sup_{y\in\mathcal{Y}} |\phi(y) - \Phi_n(y)|. \end{aligned}$$

Because $\phi : \mathcal{Y} \to \mathbb{R}$ is continuous with a unique maximizer y^* over some compact set $\mathcal{Y} \subset \mathbb{R}^q$, then for any $\epsilon > 0$, there exists a $\delta(\epsilon) > 0$ such that $||y^* - y|| > \epsilon$ is equivalent to

$$\sup\{\phi(y) : y \in \mathcal{Y}, ||y^* - y|| > \epsilon\} < \phi(y^*) - \delta(\epsilon).$$

Thus, the event $\{Y_n^* \in \{y : ||y^* - y|| \geq \epsilon\}\}$ implies the event $\{\phi(Y_n^*) < \phi^*(y^*) - \delta(\epsilon)\}$. We deduce

$$\begin{aligned}
\Pr\{||Y_n^* - y^*|| \geq \epsilon\} &\leq \Pr\{\phi(y^*) - \phi(Y_n^*) > \delta(\epsilon)\} \\
&\leq \Pr\Big\{2\sup_{y\in\mathcal{Y}} |\Phi_n(y) - \phi(y)| > \delta(\epsilon)\Big\}.
\end{aligned}$$

By Lemma A.2, the right-hand side goes to zero as $n \to \infty$. □

Lemma A.7 *Suppose* $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ *are i.i.d. samples of* F_X *over support* $\mathcal{X}$*. Define* $\Phi_n(y) := \frac{1}{n}\sum_{i=1}^n \psi(y, X_i)$ *and* $\phi(y) := \mathsf{E}[\psi(y, X)]$ *with* $\psi : \mathcal{Y} \times \mathcal{X} \to [a, b] \subset \mathbb{R}$*. Denote* Y_n^* *as a maximizer of* $\Phi_n(y)$*.*

i) $\phi : \mathcal{Y} \to \mathbb{R}$ *has a unique interior maximizer* y^* *over some compact set* $\mathcal{Y} \subset \mathbb{R}^q$*;*

ii) $\psi(\cdot, x)$ *is Lipschitz continuous over* $\mathcal{Y}$ *a.s. and continuously differentiable at any* y *with* $||y - y^*|| < \delta$ *for some* $\delta > 0$ *a.s.;*

iii) $\nabla\psi(\cdot, x)$ *is Lipschitz continuous and bounded over* $\mathcal{Y}$ *a.s.;*

iv) ϕ *is* α*-strongly concave for some* $\alpha > 0$*, i.e,* $\phi(y_2) \leq \phi(y_1) + \nabla\phi(y_1)||y_2 - y_1|| - \frac{\alpha}{2}||y_2 - y_1||^2$*,* $\forall y_1, y_2 \in \mathcal{Y}$*,*

then for any $\epsilon > 0$*,*

$$\Pr\Big\{||Y_n^* - y^*|| \geq \epsilon\Big\} \leq \mathcal{O}\Big(\exp\Big(-\frac{2n\epsilon^2 - 4C\sqrt{n}\epsilon}{(b-a)^2}\Big)\Big).$$

Proof. Define $g(y) := \Phi_n(y) - \phi(y)$. Note that

$$\begin{aligned}
\frac{\alpha}{2}||Y_n^* - y^*|| \leq \frac{\phi(Y_n^*) - \phi(y^*)}{||Y_n^* - y^*||} &\leq \frac{\Phi(Y_n^*) - \Phi(y^*)}{||Y_n^* - y^*||} + \frac{\phi(Y_n^*) - \phi(y^*)}{||Y_n^* - y^*||} \\
&= \frac{g(Y_n^*) - g(y^*)}{||Y_n^* - y^*||} \\
&\leq \sup_{y\in B(y^*, ||Y_n^* - y^*||)} ||\nabla g(y)||.
\end{aligned}$$

The first inequality follows from strong concavity of ϕ in Condition *iv)*. The second inequality follows from the fact that $\Phi_n(Y_n^*) - \Phi_n(y^*) \geq 0$ because Y_n^* is the maximizer of $\Phi_n(Y_n^*)$. The last inequality follows from mean value theorem for the continuity and differentiability of g.

Note that $\nabla g(y) = \nabla\Phi_n(y) - \nabla\phi(y) = \frac{1}{n}\sum_{i=1}^n \nabla_y\psi(y, X_i) - \mathsf{E}[\nabla_y\psi(y, X)]$ and y^j be the jth direction of $\mathcal{Y}$ with $j = 1, 2, \ldots, q$. Because $\nabla_{y^j}\psi(y, x)$ is Lipschitz continuous and bounded for all

$j = 1, 2, \ldots, q$, then

$$\begin{aligned}\Pr\Big\{||Y_n^* - y^*|| \geq \epsilon\Big\} &\leq \Pr\Big\{\sup_{y \in B(y^*, ||Y_n^* - y^*||)} ||\nabla g(y)|| \geq \epsilon\Big\} \\ &\leq \sum_{j=1}^{q} \Pr\Big\{\sup_{y \in \mathcal{Y}} \Big|\frac{1}{n}\sum_{i=1}^{n} \nabla_{y^j}\psi(y, X_i) - \mathsf{E}[\nabla_{y^j}\psi(y, X)]\Big| \geq \epsilon\Big\} \\ &\leq \mathcal{O}\Big(\exp\Big(-\frac{2n\epsilon^2 - 4C\sqrt{n}\epsilon}{(b-a)^2}\Big)\Big),\end{aligned}$$

The last inequality follows from Lemma A.2. □

Lemma A.8 (Scheffé's Lemma) *Suppose* $(X_1, X_2, \ldots)$ *is a sequence of integrable random variables such that* $X_n \xrightarrow{a.s.} X$ *and* $\mathsf{E}[|X|] < \infty$*. Then,*

$$\mathsf{E}[|X_n - X|] \to 0 \;\; \textit{iff} \;\; \mathsf{E}[X_n] \to \mathsf{E}[X].$$

Proof. We first show that $\mathsf{E}[X_n] \to \mathsf{E}[X]$ implies $\mathsf{E}[|X_n - X|] \to 0$. Because $X_n \xrightarrow{a.s.} X$,

$$\max\{X_n, X\} \xrightarrow{a.s.} X, \text{ and } \min\{X_n, X\} \xrightarrow{a.s.} X.$$

Since $\min\{X_n, X\} \leq_{a.s.} X$ and $\mathsf{E}[|X|] < \infty$, then by dominated convergence theorem, we have

$$\mathsf{E}[\min\{X_n, X\}] \to \mathsf{E}[X].$$

Since $\mathsf{E}[\max\{X_n, X\}] = \mathsf{E}[X + X_n - \min\{X_n, X\}] = \mathsf{E}[X] + \mathsf{E}[X_n] - \mathsf{E}[\min\{X_n, X\}]$, then

$$\mathsf{E}[\max\{X_n, X\}] - \mathsf{E}[X] = \mathsf{E}[X_n] - \mathsf{E}[\min\{X_n, X\}] \to 0.$$

Thus,

$$\mathsf{E}[|X_n - X|] = \mathsf{E}[\max\{X_n, X\}] - \mathsf{E}[\min\{X_n, X\}] \to 0.$$

Next, we show that $\mathsf{E}[|X_n - X|] \to 0$ implies $\mathsf{E}[X_n] \to \mathsf{E}[X]$. Note that

$$0 \leq |\mathsf{E}[X_n] - \mathsf{E}[X]| = |\mathsf{E}[X_n - X]| \leq \mathsf{E}[|X_n - X|] \to 0,$$

which implies that $\mathsf{E}[X_n] \to \mathsf{E}[X]$. □

Lemma A.9 *Suppose* $(X_1, X_2, \ldots)$ *is a sequence of integrable random variables such that* $X_n \xrightarrow{d} X$*,* $\mathsf{E}[X_n] \to \mathsf{E}[X]$*, and* $\mathsf{E}[|X|] < \infty$*. Then, for any Lipschitz continuous function* $g : \mathbb{R} \to \mathbb{R}$ *and compact* $\Gamma \subset \mathbb{R}$*,*

$$\sup_{\gamma \in \Gamma} |\mathsf{E}[g(\gamma X_n)] - \mathsf{E}[g(\gamma X)]| \to 0.$$

Proof. By Skorokhod representation theorem, there exist $(Y_1, Y_2, \ldots)$ and Y, defined on the common probability space, satisfy $Y_n \xrightarrow{a.s.} Y$ with $Y_n =^d X_n$ for all $n \in \{1, 2, \ldots\}$, and $Y =^d X$. Because $\mathsf{E}[X_n] \to \mathsf{E}[X]$, we have that $\mathsf{E}[Y_n] = \mathsf{E}[X_n]$ implies $\mathsf{E}[X] = \mathsf{E}[Y]$. Thus, we derive

$$\sup_{\gamma\in\Gamma} |\mathsf{E}[g(\gamma X_n)] - \mathsf{E}[g(\gamma X)]| = \sup_{\gamma\in\Gamma} |\mathsf{E}[g(\gamma Y_n)] - \mathsf{E}[g(\gamma Y)]| \le K\Big(\sup_{\gamma\in\Gamma} \gamma\Big)\mathsf{E}[|Y_n - Y|] \to 0.$$

The inequality follows from the Lipschitz continuity, where K is the Lipschitz constant. The convergence follows from Lemma A.8. □

Proof of Theorem 3. Let $\hat{Z}$ denote the random variable with distribution $\hat{F}_{Z|\boldsymbol{Z}^o}$ and $\hat{\boldsymbol{Z}} = (\hat{Z}_1, \hat{Z}_2, \ldots, \hat{Z}_n)$ be a vector of i.i.d. samples of $\tilde{f}_{Z|\boldsymbol{Z}^o}$ (or $\hat{F}_{Z|\boldsymbol{Z}^o}$). Then, given any sample of $\boldsymbol{X}^o(\omega)$, we have, by the triangle inequality,

$$\begin{aligned}
& \sup_{\gamma\in\Gamma} \big|\hat{\phi}_{\mathrm{CrL:SC}\iota}(\gamma, \mathbf{Z}^{()}) - \mathsf{E}[\psi((\hat{\mu}_{\boldsymbol{Z}})^\iota\gamma, Z)]\big| \qquad (39)\\
\le\ & \sup_{\gamma\in\Gamma} \left|\hat{\phi}_{\mathrm{CrL:SC}\iota}(\gamma, \mathbf{Z}^{()}) - \frac{1}{m}\sum_{j=1}^{m} \mathsf{E}_{\hat{Z}}[\psi((\hat{\mu}_{\boldsymbol{Z}^{(j)}})^\iota\gamma, \hat{Z})]\right| \\
& + \sup_{\gamma\in\Gamma} \left|\frac{1}{m}\sum_{j=1}^{m} \mathsf{E}_{\hat{Z}}[\psi((\hat{\mu}_{\boldsymbol{Z}^{(j)}})^\iota\gamma, \hat{Z})] - \mathsf{E}[\psi((\hat{\mu}_{\hat{\boldsymbol{Z}}})^\iota\gamma, \hat{Z})]\right| \\
& + \sup_{\gamma\in\Gamma} \big|\mathsf{E}[\psi((\hat{\mu}_{\hat{\boldsymbol{Z}}})^\iota\gamma, \hat{Z})] - \mathsf{E}[\psi((\hat{\mu}_{\boldsymbol{Z}})^\iota\gamma, Z)]\big|.
\end{aligned}$$

We will analyze each of the three terms on the right-hand side. The result would follow if we can show that for any $\epsilon > 0$, the probability that each of the first two terms is above ϵ is bounded by $\mathcal{O}\big(\exp\big(-\frac{2m\epsilon^2 - 4C_1\sqrt{m}\epsilon}{(b-a)^2}\big)\big)$, and the third term is bounded by ϵ for sufficiently large n^o, because $\Pr\{|A+B+c| > 3\epsilon\} \le \Pr\{|A|+|B|+|c| > 3\epsilon\} \le \Pr\{|A| > \epsilon\} + \Pr\{|B| > \epsilon\}$ provided that $|c| < \epsilon$.

To evaluate the first term on the right-hand side of (39), we take $g_j(\gamma, Z) = \psi((\hat{\mu}_{\boldsymbol{z}^{(j)}})^\iota\gamma, Z)$ for given $\boldsymbol{Z}^{(j)} = \boldsymbol{z}^{(j)}$ and apply Lemma A.2 to obtain

$$\begin{aligned}
& \Pr\left\{\sup_{\gamma\in\Gamma}\left|\frac{1}{(m-1)n} \sum_{\substack{i\in\{1,2,\ldots,n\}\\ \ell\in\{1,2,\ldots,m\}\setminus j}} \psi((\hat{\mu}_{\boldsymbol{z}^{(j)}})^\iota\gamma, Z_i^{(\ell)}) - \mathsf{E}_{\hat{Z}}[\psi((\hat{\mu}_{\boldsymbol{z}^{(j)}})^\iota\gamma, \hat{Z})]\right| \ge \epsilon\right\} \\
& \qquad \le \mathcal{O}\Big(\exp\Big(-\frac{2nm\epsilon^2 - 4C_1\sqrt{nm}\epsilon}{(b-a)^2}\Big)\Big),
\end{aligned}$$

for some $C_1 > 0$. Then, we deduce

$$\begin{aligned}
& \Pr\left\{\sup_{\gamma\in\Gamma}\left|\hat{\phi}_{\mathrm{CrL:SC}\iota}(\gamma, \mathbf{Z}^{()}) - \frac{1}{m}\sum_{j=1}^{m} \mathsf{E}_{\hat{Z}}[\psi((\hat{\mu}_{\boldsymbol{Z}^{(j)}})^\iota\gamma, \hat{Z})]\right| \ge \epsilon\right\} \\
\le\ & \Pr\left\{\sup_{\gamma\in\Gamma}\sup_{j\in\{1,2,\ldots,m\}}\left|\frac{1}{(m-1)n} \sum_{\substack{i\in\{1,2,\ldots,n\}\\ \ell\in\{1,2,\ldots,m\}\setminus j}} \psi((\hat{\mu}_{\boldsymbol{Z}^{(j)}})^\iota\gamma, Z_i^{(\ell)}) - \mathsf{E}_{\hat{Z}}[\psi((\hat{\mu}_{\boldsymbol{Z}^{(j)}})^\iota\gamma, \hat{Z})]\right| \ge \epsilon\right\} \\
\le\ & \mathcal{O}\Big(\exp\Big(-\frac{2nm\epsilon^2 - 4C_1\sqrt{nm}\epsilon}{(b-a)^2}\Big)\Big).
\end{aligned}$$

To treat the second term on the right-hand side of (39), we apply Lemma A.2 with m samples of the random vector $\hat{\boldsymbol{Z}}$ to obtain

$$\Pr\left\{\sup_{\gamma\in\Gamma}\left|\frac{1}{m}\sum_{j=1}^{m}\mathsf{E}_{\hat{Z}}[\psi((\hat{\mu}_{\boldsymbol{Z}^{(j)}})^{\iota}\gamma,\hat{Z})]-\mathsf{E}[\psi((\hat{\mu}_{\hat{\boldsymbol{Z}}})^{\iota}\gamma,\hat{Z})]\right|\geq\epsilon\right\}\leq\mathcal{O}\Big(\exp\Big(-\frac{2m\epsilon^2-4C_2\sqrt{m}\epsilon}{(b-a)^2}\Big)\Big),$$

for some $C_2>0$.

To analyze the last term on the right-hand side of (39), we note that because $\psi(\cdot,\cdot)$ is Lipschitz continuous, for any given $\boldsymbol{x}^o=(x_1^o,x_2^o,\ldots,x_{n^o}^o)$,

$$\begin{aligned}\mathsf{E}\big[|\psi((\hat{\mu}_{\hat{\boldsymbol{Z}}})^{\iota}\gamma,\hat{Z})-\psi((\hat{\mu}_{\hat{\boldsymbol{Z}}})^{\iota}\gamma,Z)|\big] &\leq K\mathsf{E}\big[|\hat{Z}-Z|\big],\text{ and}\\ \mathsf{E}\big[|\psi((\hat{\mu}_{\hat{\boldsymbol{Z}}})^{\iota}\gamma,Z)-\psi((\hat{\mu}_{\boldsymbol{Z}})^{\iota}\gamma,Z)|\big] &\leq K\mathsf{E}\big[|(\hat{\mu}_{\hat{\boldsymbol{Z}}})^{\iota}-(\hat{\mu}_{\boldsymbol{Z}})^{\iota}|\big]\gamma,\end{aligned}$$

where K is the Lipschitz constant. By Lemma A.5, we have $\psi((\hat{\mu}_{\hat{\boldsymbol{z}}})^{\iota}\gamma,\hat{Z})\to^{d}\psi((\hat{\mu}_{\hat{\boldsymbol{z}}})^{\iota}\gamma,Z),\forall\hat{\mu}_{\hat{\boldsymbol{z}}}\in\mathbb{R}_+$ and $\psi((\hat{\mu}_{\hat{\boldsymbol{Z}}})^{\iota}\gamma,z)\to^{d}\psi((\hat{\mu}_{\boldsymbol{Z}})^{\iota}\gamma,z),\forall z\in\mathbb{R}_+$. Thus, by Lemma A.9, there exists an $n^o(\epsilon)$, such that for $n^o>n^o(\epsilon)$,

$$\begin{aligned}&\sup_{\gamma\in\Gamma}|\mathsf{E}[\psi((\hat{\mu}_{\hat{\boldsymbol{Z}}})^{\iota}\gamma,\hat{Z})-\psi((\hat{\mu}_{\boldsymbol{Z}})^{\iota}\gamma,Z)]|\\ \leq\ &\sup_{\gamma\in\Gamma}|\mathsf{E}[\psi((\hat{\mu}_{\hat{\boldsymbol{Z}}})^{\iota}\gamma,\hat{Z})-\psi((\hat{\mu}_{\hat{\boldsymbol{Z}}})^{\iota}\gamma,Z)]|+\sup_{\gamma\in\Gamma}|\mathsf{E}[\psi((\hat{\mu}_{\hat{\boldsymbol{Z}}})^{\iota}\gamma,Z)-\psi((\hat{\mu}_{\boldsymbol{Z}})^{\iota}\gamma,Z)]|\\ <\ &\epsilon.\end{aligned}$$

Putting together the results of the three terms on the right-hand side of (39), we conclude (21).

Finally, the result in (22) follows immediately from Lemma A.6 and the results in (23) follows immediately from Lemma A.7. □

Proof of Lemma 2. From Chu et al. (2025),

$$\int_{\boldsymbol{x}\in\mathbb{R}_+^n}h(\boldsymbol{x})\mathrm{d}\boldsymbol{x}=\int_{\boldsymbol{z}\in\mathcal{B}^n}\int_{\alpha=0}^{\infty}h(\alpha\boldsymbol{z})\alpha^{n-1}\mathrm{d}\alpha\mathrm{d}\boldsymbol{z}.$$

For a fixed $\boldsymbol{z}_B\in\mathcal{B}^n$, let $h(\boldsymbol{x})=\prod_{i=1}^{n}\tilde{f}_{Z|\boldsymbol{z}^o}(x_i)\mathbb{I}_{\{d(\boldsymbol{x}/\hat{\mu}_{\boldsymbol{x}},\boldsymbol{z}_B)\leq\eta\}}$. Then,

$$\begin{aligned}\Pr\{d(\hat{\boldsymbol{Z}}/\hat{\mu}_{\hat{\boldsymbol{Z}}},\boldsymbol{z}_B)\leq\eta\} &= \int_{\boldsymbol{x}\in\mathbb{R}_+^n}\prod_{i=1}^{n}\tilde{f}_{Z|\boldsymbol{z}^o}(x_i)\mathbb{I}_{\{d(\boldsymbol{x}/\hat{\mu}_{\boldsymbol{x}},\boldsymbol{z}_B)\leq\eta\}}\mathrm{d}\boldsymbol{x}\\ &= \int_{\boldsymbol{z}\in\mathcal{B}^n}\Big(\mathbb{I}_{\{d(\boldsymbol{z},\boldsymbol{z}_B)\leq\eta\}}\int_{\alpha\in\mathbb{R}_+}\prod_{i=1}^{n}\tilde{f}_{Z|\boldsymbol{z}^o}(\alpha z_i)\alpha^{n-1}\mathrm{d}\alpha\Big)\mathrm{d}\boldsymbol{z}.\end{aligned}$$

Define

$$C(\alpha_1,\alpha_2,f)=\int_{\alpha\in[\alpha_1,\alpha_2]}\prod_{i=1}^{n}f(\alpha z_i)\alpha^{n-1}\mathrm{d}\alpha,\quad\text{for }\alpha_1,\alpha_2\in\mathbb{R}_+\text{ and }f:\mathbb{R}_+\to\mathbb{R}_+.$$

Note that, by $f(z) > 0$ for any $z \in \mathbb{R}_+$ (see Assumption 2), we can always choose (α_1, α_2) with $\alpha_2 > \alpha_1$ such that, given fixed $\boldsymbol{z}$, $\min\{f_Z(\alpha z_i) : \alpha \in (\alpha_1, \alpha_2), i = 1, \ldots, n\} = \epsilon > 0$ for some $\epsilon > 0$. Then for any finite n,

$$C(0, \infty, f_Z) \geq C(\alpha_1, \alpha_2, f_Z) \geq \int_{\alpha_1}^{\alpha_2} \prod_{i=1}^{n} \epsilon^n \alpha^{n-1} \mathrm{d}\alpha = \frac{\epsilon^n}{n}(\alpha_2^n - \alpha_1^n) > 0.$$

By construction, $\tilde{f}_{Z|\boldsymbol{z}^o}(z)$ is continuous and positive, and it converges to $f(z)$ for any $z \in \mathbb{R}_+$. Then, given fixed $\boldsymbol{z}$, we have $C(\alpha_1, \alpha_2, \tilde{f}_{Z|\boldsymbol{z}^o}) \to C(\alpha_1, \alpha_2, f_Z)$. Thus, for sufficiently large n^o, $C(\alpha_1, \alpha_2, \tilde{f}_{Z|\boldsymbol{z}^o}) > \epsilon^n(\alpha_2^n - \alpha_1^n)/(2n)$, implying $C(0, \infty, \tilde{f}_{Z|\boldsymbol{z}^o}(z)) \geq \underline{C}_n > 0$ for some $\underline{C}_n$ independent of n^o.

Finally, we note that $\int_{\boldsymbol{z} \in \mathcal{B}^n} \mathbb{I}_{\{d(\boldsymbol{z}, \boldsymbol{z}_B) \leq \eta\}} \mathrm{d}\boldsymbol{z} = \frac{\pi^{(n-1)/2}}{\Gamma(\frac{n-1}{2}+1)} \eta^{n-1}$ is the Lebesgue measure of an $(n-1)$-dimensional ball of radius η. We conclude the result. □

Proof of Lemma 3. Define $M_{n^o} =^d \left[\hat{\mu}_{\boldsymbol{Z}_{n^o}} \middle| \boldsymbol{Z}_{n^o}/\hat{\mu}_{\boldsymbol{Z}_{n^o}} = \boldsymbol{z}_B\right]$, $M_{n^o}^{\eta} =^d \left[\hat{\mu}_{\boldsymbol{Z}_{n^o}} \middle| d(\boldsymbol{Z}_{n^o}/\hat{\mu}_{\boldsymbol{Z}_{n^o}}, \boldsymbol{z}_B) \leq \eta\right]$ and $M =^d \left[\hat{\mu}_{\boldsymbol{Z}} \middle| \boldsymbol{Z}/\hat{\mu}_{\boldsymbol{Z}} = \boldsymbol{z}_B\right]$.

Given $\boldsymbol{z} \in \mathbb{R}_+^n$ and $\eta > 0$, let $\alpha = \frac{1}{n}\sum_{j=1}^n z_j$ and define $y_j = (z_j/\alpha - z_{B:j})/\eta, j = 1, 2, \ldots, n$. We have $z_j = \alpha(\eta y_j + z_{B:j})$ and $\sum_{j=1}^n y_j = 0$. Note that $\{d(\boldsymbol{z}/\alpha, \boldsymbol{z}_B) = \eta\}$ is equivalent to $\frac{1}{n}\sum_{j=1}^n y_j^2 = 1$. We further define $y_{n+1} = \eta$, and $z_{n+1} = \sqrt{\frac{1}{n}\sum_{j=1}^n y_j^2}$. We deduce

$$\mathbf{J} = [J_{ij}] = \left[\frac{\mathrm{d}z_i}{\mathrm{d}y_j}\right] = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

where

$$A = \alpha\eta I_n, \quad B = \begin{bmatrix} \alpha y_1 \\ \vdots \\ \alpha y_n \end{bmatrix},$$

$$C = \begin{bmatrix} \frac{y_1}{\sqrt{n\sum_{j=1}^n y_j^2}} & \frac{y_2}{n\sqrt{\sum_{j=1}^n y_j^2}} & \cdots & \frac{y_n}{\sqrt{n\sum_{j=1}^n y_j^2}} \end{bmatrix}, \quad D = [0].$$

The determinant of the Jacobian matrix $\mathbf{J}$ is

$$\begin{aligned} \det[\mathbf{J}] &= \det[A] \cdot \det[D - CA^{-1}B] \\ &= -(\alpha\eta)^n \cdot (\alpha\eta)^{-1}\alpha\left(\frac{1}{n}\sum_{j=1}^n y_j^2\right)^{\frac{1}{2}} \\ &= -\alpha^n\eta^{n-1}\left(\frac{1}{n}\sum_{j=1}^n y_j^2\right)^{\frac{1}{2}}. \end{aligned}$$

Given a function $h : \mathbb{R}_+^n \to \mathbb{R}$, let $S^\eta(\alpha) = \{\boldsymbol{z} : d(\boldsymbol{z}/\alpha, \boldsymbol{z}_B) \le \eta, \hat{\mu}_{\boldsymbol{z}} = \alpha\}$ and $T = \{\boldsymbol{y} : \frac{1}{n}\sum_{j=1}^n y_j^2 = 1\}$. We have

$$\begin{aligned}
\int_{\boldsymbol{z}\in S^\eta(\alpha)} h(\boldsymbol{z})\mathrm{d}\boldsymbol{z} &= \int_{z_{n+1}=0}^1 \int_{\boldsymbol{z}\in S^\eta(\alpha)} h(\boldsymbol{z})\mathrm{d}\boldsymbol{z}\mathrm{d}z_{n+1} \\
&= \int_{\eta_0\le\eta}\int_{\boldsymbol{y}\in T} h\big(\alpha(\eta_0\boldsymbol{y}+\boldsymbol{z}_B)\big)\cdot\left(\alpha^n\eta_0^{n-1}\left(\frac{1}{n}\sum_{j=1}^n y_j^2\right)^{\frac{1}{2}}\right)\mathrm{d}\boldsymbol{y}\mathrm{d}\eta_0 \\
&= \alpha^n\int_{\eta_0\le\eta}\int_{\boldsymbol{y}\in T} h\big(\alpha(\eta_0\boldsymbol{y}+\boldsymbol{z}_B)\big)\left(\frac{1}{n}\sum_{j=1}^n y_j^2\right)^{\frac{1}{2}}\mathrm{d}\boldsymbol{y}\mathrm{d}\eta_0^n \\
&= \alpha^n\int_{\gamma\le\eta^n}\int_{\boldsymbol{y}\in T} h\big(\alpha(\gamma^{1/n}\boldsymbol{y}+\boldsymbol{z}_B)\big)\left(\frac{1}{n}\sum_{j=1}^n y_j^2\right)^{\frac{1}{2}}\mathrm{d}\boldsymbol{y}\mathrm{d}\gamma.
\end{aligned}$$

We deduce

$$\begin{aligned}
f_{M_{n^o}^\eta}(\alpha) &\propto \int_{\boldsymbol{z}\in S^\eta(\alpha)}\prod_{j=1}^n f_{n^o}(z_j)\mathrm{d}\boldsymbol{z} \\
&= \alpha^n\int_{\gamma\le\eta^n}\int_{\boldsymbol{y}\in T}\prod_{j=1}^n f_{n^o}\big(\alpha(\gamma^{1/n}y_j+z_{B:j})\big)\left(\frac{1}{n}\sum_{j=1}^n y_j^2\right)^{\frac{1}{2}}\mathrm{d}\boldsymbol{y}\mathrm{d}\gamma.
\end{aligned}$$

Then, for almost every $\alpha_1, \alpha_2 \in \mathbb{R}_+$,

$$\begin{aligned}
\lim_{\eta\to0}\frac{f_{M_{n^o}^\eta}(\alpha_1)}{f_{M_{n^o}^\eta}(\alpha_2)} &= \lim_{\eta\to0}\frac{\alpha_1^n\int_{\gamma\le\eta^n}\int_{\boldsymbol{y}\in T}\prod_{j=1}^n f_{n^o}\big(\alpha_1(\gamma^{1/n}y_j+z_{B:j})\big)\left(\frac{1}{n}\sum_{j=1}^n y_j^2\right)^{\frac{1}{2}}\mathrm{d}\boldsymbol{y}\mathrm{d}\gamma}{\alpha_2^n\int_{\gamma\le\eta^n}\int_{\boldsymbol{y}\in T}\prod_{j=1}^n f_{n^o}\big(\alpha_2(\gamma^{1/n}y_j+z_{B:j})\big)\left(\frac{1}{n}\sum_{j=1}^n y_j^2\right)^{\frac{1}{2}}\mathrm{d}\boldsymbol{y}\mathrm{d}\gamma} \\
&= \lim_{\eta\to0}\frac{\alpha_1^n\int_{y\in T}\prod_{j=1}^n f_{n^o}\big(\alpha_1(\eta y_j+z_{B:j})\big)\left(\frac{1}{n}\sum_{j=1}^n y_j^2\right)^{\frac{1}{2}}\mathrm{d}\boldsymbol{y}}{\alpha_2^n\int_{y\in T}\prod_{j=1}^n f_{n^o}\big(\alpha_2(\eta y_j+z_{B:j})\big)\left(\frac{1}{n}\sum_{j=1}^n y_j^2\right)^{\frac{1}{2}}\mathrm{d}\boldsymbol{y}} \\
&= \frac{\alpha_1^n\prod_{j=1}^n f_{n^o}\big(\alpha_1 z_{B:j}\big)}{\alpha_2^n\prod_{j=1}^n f_{n^o}\big(\alpha_2 z_{B:j}\big)} \\
&= \frac{f_{M_{n^o}}(\alpha_1)}{f_{M_{n^o}}(\alpha_2)}.
\end{aligned}$$

The second equation follows from L'Hôpital's rule, the third equation follows from the continuity of f_{n^o}, and the last equality follows from the fact that $f_{M_{n^o}}(\alpha) \propto \alpha^n\prod_{j=1}^n f_{n^o}(\alpha z_{B:j})$. Therefore, we have

$$f_{M_{n^o}^\eta} \to f_{M_{n^o}} \text{ as } \eta \to 0.$$

We further note that $f_{M_{n^o}}(\alpha) \propto \alpha^n\prod_{j=1}^n f_{n^o}(\alpha z_{B:j})$ and $\lim_{n^o\to\infty} f_{n^o} = f_Z$, we conclude that $f_{M_{n^o}} \to f_M$ as $n^o \to \infty$. □

Proof of Theorem 4. For a given vector $\boldsymbol{Z}^o = \boldsymbol{z}^o$ of size n^o, let $\hat{Z}$ denote the random variable following smoothed empirical density $\tilde{f}_{Z|\boldsymbol{z}^o}$ and $\hat{\boldsymbol{Z}}$ denote a vector of n i.i.d. draws of $\tilde{f}_{Z|\boldsymbol{z}^o}$.

Let $T_m = \sum_{i=1}^m \mathbb{I}_{\{d(\hat{\boldsymbol{Z}}^{(j)}/\hat{\mu}_{\hat{\boldsymbol{Z}}^{(j)}}, \boldsymbol{z}_B) \leq \eta\}}$ denote the binomial random variable with success probability $q = \Pr\{d(\hat{\boldsymbol{Z}}/\hat{\mu}_{\hat{\boldsymbol{Z}}}, \boldsymbol{z}_B) \leq \eta\}$ and count m. By Lemma 2, we have, for sufficiently large m,

$$\mathsf{E}[T_m] = \sum_{j=1}^m \Pr\left\{d(\hat{\boldsymbol{Z}}^{(j)}/\hat{\mu}_{\hat{\boldsymbol{Z}}^{(j)}}, \boldsymbol{z}_B) \leq \eta\right\} \geq \underline{C}_n \frac{\pi^{(n-1)/2}}{\Gamma(\frac{n-1}{2}+1)} m\eta^{n-1} \propto m^{\frac{1}{n}}.$$

Note that $\Pr\{d(\hat{\boldsymbol{Z}}^{(j)}/\hat{\mu}_{\hat{\boldsymbol{Z}}^{(j)}}, \boldsymbol{z}_B) \leq \eta\} \to 0$ as $m \to \infty$, then it is well known that $\lim_{m\to\infty} T_m =^d \mathcal{N}(\mathcal{O}(m^{\frac{1}{n}}), \mathcal{O}(m^{\frac{1}{n}}))$, i.e., T_m converges to a normal distribution with mean and variance in the same order $\mathcal{O}(m^{\frac{1}{n}})$, as $m \to \infty$. Thus, the probability of T_m larger than $qm/2 = \frac{1}{2}\mathsf{E}[T_m] \propto \mathcal{O}(m^{\frac{1}{n}})$ converges to 1 as $m \to \infty$.

More rigorously, because T_m is a binomial random variable, we derive

$$\begin{aligned}
1 - \Pr\{T_m \geq qm/2\} &= \sum_{k=1}^{\lfloor qm/2 \rfloor} \binom{k}{m} q^k (1-q)^{m-k} \\
&= \sum_{k=1}^{\lfloor qm/2 \rfloor} \frac{m!}{k!(m-k)!} \frac{q^k}{(1-q)^k} (1-q)^m \\
&\approx \sum_{k=1}^{\lfloor qm/2 \rfloor} \left(\frac{qm}{k}\right)^k e^{-qm} \\
&\leq qm/2 \cdot \left(\frac{qm}{qm/2}\right)^{qm/2} e^{-qm} \\
&= \frac{qm}{2e^{qm/2}} \cdot \left(\frac{2}{e}\right)^{qm/2} \to 0,
\end{aligned}$$

as $m \to \infty$ with $qm \to \mathcal{O}(m^{\frac{1}{n}})$. Now applying Lemma A.2 and Lemma 3, we have, for fixed $\boldsymbol{z}_B$ and $\boldsymbol{z}^o$,

$$\begin{aligned}
&\Pr\left\{\sup_{\gamma\in\Gamma}\left|\hat{\phi}_{\text{CrL:c}}(\gamma, y^c(\cdot), \boldsymbol{z}_B; \eta, \mathbf{Z}^{()}) - \mathsf{E}\left[\psi\left(\gamma y^c(\hat{\boldsymbol{Z}}), \hat{Z}\right) \middle| d(\hat{\boldsymbol{Z}}/\hat{\mu}_{\hat{\boldsymbol{Z}}}, \boldsymbol{z}_B) \leq \eta\right]\right| \geq \epsilon\right\} \\
\leq\ & \mathcal{O}\Big(\exp\Big(-\frac{2m^{1/n}\epsilon^2 - 4C_1\sqrt{m^{1/n}}\epsilon}{(b-a)^2}\Big)\Big),
\end{aligned}$$

for some $C_1 > 0$. Because $\hat{Z} \to^d Z$ as $n^o \to \infty$ and ψ is Lipschitz continuous, we deduce by

Lemma A.9 the existence of an $n^o(\epsilon)$ such that for any $n^o \geq n^o(\epsilon)$

$$\begin{aligned}
& \sup_{\gamma\in\Gamma}\Big|\mathsf{E}\big[\psi\big(\gamma y^c(\hat{\boldsymbol{Z}}),\hat{Z}\big)\big|d(\hat{\boldsymbol{Z}}/\hat{\mu}_{\hat{\boldsymbol{Z}}},\boldsymbol{z}_B)\leq\eta\big]-\mathsf{E}\big[\psi\big(\gamma y^c(\boldsymbol{Z}),Z\big)\big|d(\boldsymbol{Z}/\hat{\mu}_{\boldsymbol{Z}},\boldsymbol{z}_B)\leq\eta\big]\Big| \\
\leq\; & \sup_{\gamma\in\Gamma}\Big|\mathsf{E}\big[\psi\big(\gamma y^c(\hat{\boldsymbol{Z}}),\hat{Z}\big)\big|d(\hat{\boldsymbol{Z}}/\hat{\mu}_{\hat{\boldsymbol{Z}}},\boldsymbol{z}_B)\leq\eta\big]-\mathsf{E}\big[\psi\big(\gamma y^c(\hat{\boldsymbol{Z}}),Z\big)\big|d(\hat{\boldsymbol{Z}}/\hat{\mu}_{\hat{\boldsymbol{Z}}},\boldsymbol{z}_B)\leq\eta\big]\Big| \\
& +\sup_{\gamma\in\Gamma}\Big|\mathsf{E}\big[\psi\big(\gamma y^c(\hat{\boldsymbol{Z}}),Z\big)\big|d(\hat{\boldsymbol{Z}}/\hat{\mu}_{\hat{\boldsymbol{Z}}},\boldsymbol{z}_B)\leq\eta\big]-\mathsf{E}\big[\psi\big(\gamma y^c(\boldsymbol{Z}),Z\big)\big|d(\boldsymbol{Z}/\hat{\mu}_{\boldsymbol{Z}},\boldsymbol{z}_B)\leq\eta\big]\Big| \\
<\; & \frac{\epsilon}{2}+\frac{\epsilon}{2}.
\end{aligned}$$

Because $y^c \in \mathcal{H}^n_+(\iota,\pi)$ and y^c is continuous, we have, as $\eta \to 0$,

$$\begin{aligned}
\big[y^c(\boldsymbol{Z})\big|d(\boldsymbol{Z}/\hat{\mu}_{\boldsymbol{Z}},\boldsymbol{z}_B)\leq\eta\big] &= \big[(\hat{\mu}_{\boldsymbol{Z}})^\iota y^c(\boldsymbol{Z}/\hat{\mu}_{\boldsymbol{Z}})\big|d(\boldsymbol{Z}/\hat{\mu}_{\boldsymbol{Z}},\boldsymbol{z}_B)\leq\eta\big] \\
&\xrightarrow{a.s.} \big[(\hat{\mu}_{\boldsymbol{Z}})^\iota y^c(\boldsymbol{z}_B)\big|d(\boldsymbol{Z}/\hat{\mu}_{\boldsymbol{Z}},\boldsymbol{z}_B)\leq\eta\big].
\end{aligned}$$

Because ψ is Lipschitz continuous with constant K, we have

$$\begin{aligned}
& \sup_{\gamma\in\Gamma}\Big|\mathsf{E}\big[\psi\big(\gamma y^c(\boldsymbol{Z}),Z\big)\big|d(\boldsymbol{z}_B,\boldsymbol{Z}/\hat{\mu}_{\boldsymbol{Z}})\leq\eta\big]-\mathsf{E}\big[\psi\big(\gamma(\hat{\mu}_{\boldsymbol{Z}})^\iota y^c(\boldsymbol{z}_B),Z\big)\big|d(\boldsymbol{z}_B,\boldsymbol{Z}/\hat{\mu}_{\boldsymbol{Z}})\leq\eta\big]\Big| \\
\leq\; & \sup_{\gamma\in\Gamma}\Big|\mathsf{E}\big[K\gamma|y^c(\boldsymbol{Z})-(\hat{\mu}_{\boldsymbol{Z}})^\iota y^c(\boldsymbol{z}_B)|\big|d(\boldsymbol{z}_B,\boldsymbol{Z}/\hat{\mu}_{\boldsymbol{Z}})\leq\eta\big]\Big| \\
\leq\; & K\sup_{\gamma\in\Gamma}(\gamma)\mathsf{E}\big[|(\hat{\mu}_{\boldsymbol{Z}})^\iota y^c(\boldsymbol{Z}/\hat{\mu}_{\boldsymbol{Z}})-(\hat{\mu}_{\boldsymbol{Z}})^\iota y^c(\boldsymbol{z}_B)|\big|d(\boldsymbol{z}_B,\boldsymbol{Z}/\hat{\mu}_{\boldsymbol{Z}})\leq\eta\big] \\
<\; & \epsilon,
\end{aligned}$$

for any $n^o \geq n^o(\epsilon)$ with $n^o(\epsilon)$ sufficiently large. The last inequality follows because $\lim_{n^o\to\infty}\eta = 0$. By Lemma 3, we have $\big[\hat{\mu}_{\boldsymbol{Z}}\big|d(\boldsymbol{Z}/\hat{\mu}_{\boldsymbol{Z}},\boldsymbol{z}_B)\leq\eta\big] \xrightarrow{d} \big[\hat{\mu}_{\boldsymbol{Z}}\big|d(\boldsymbol{Z}/\hat{\mu}_{\boldsymbol{Z}},\boldsymbol{z}_B)=0\big]$, as $\eta \to 0$. Finally, as $n^o \to \infty$, $\eta \to 0$, and thus, for any $n^o \geq n^o(\epsilon)$,

$$\sup_{\gamma\in\Gamma}\Big|\mathsf{E}\big[\psi\big(\gamma(\hat{\mu}_{\boldsymbol{Z}})^\iota y^c(\boldsymbol{z}_B),Z\big)\big|d(\boldsymbol{z}_B,\boldsymbol{Z}/\hat{\mu}_{\boldsymbol{Z}})\leq\eta\big]-\mathsf{E}\big[\psi\big(\gamma(\hat{\mu}_{\boldsymbol{Z}})^\iota y^c(\boldsymbol{z}_B),Z\big)\big|\boldsymbol{Z}/\hat{\mu}_{\boldsymbol{Z}}=\boldsymbol{z}_B\big]\Big| < \epsilon.$$

By Equation (7) and $y^c \in \mathcal{H}^n_+(\iota,\pi)$, $\mathsf{E}\big[\psi\big(\gamma(\hat{\mu}_{\boldsymbol{Z}})^\iota y^c(\boldsymbol{z}_B),Z\big)\big|\boldsymbol{Z}/\hat{\mu}_{\boldsymbol{Z}}=\boldsymbol{z}_B\big] = \phi_B[\gamma y^c(\boldsymbol{z}_B),\boldsymbol{z}_B,F_Z]$. Putting all of the above relations together, we conclude the convergence of the boosted objective.

Now denote $\gamma^*_{\mathrm{CrL:c}}$ be the unique maximizer of $\phi_B[\gamma y^c(\boldsymbol{z}_B),\boldsymbol{z}_B,F_Z]$. We must have $\gamma^*_{\mathrm{CrL:c}}y^c(\boldsymbol{z}_B) = y^*_{\mathrm{H}\iota}(\boldsymbol{z}_B)$ based on Theorem 1 and (8). The convergence of the boosted operational statistic in (28) follows directly by applying Lemma A.6. The convergence rate of the boosted operational statistic in (29) follows immediately from Lemma A.7. □

Proof of Theorem 5. The result in (31) follows immediately by setting $y^c(\boldsymbol{x}) = (\hat{\mu}_{\boldsymbol{x}})^\iota\gamma \in \mathcal{H}^\iota_+(\iota,\pi)$ in (27). Note that in establishing (27), we do not require y^c to be consistent. Then, (32) follows by applying Lemma A.6 and (33) follows immediately from Lemma A.7. □

Proof of Theorem 6. Take $\boldsymbol{Z}$ as n i.i.d. draws from F_Z and define, for a fixed $\ell \in \{1, 2, \dots, n\}$,

$$\varphi_\ell(y(\cdot), \boldsymbol{Z}, \Theta) = \Theta^\kappa \psi(y(\boldsymbol{Z}_{\text{-}\ell}), Z_\ell).$$

It is immediate that $\varphi_\ell(y(\cdot), \boldsymbol{Z}, \Theta) = \psi(y(\boldsymbol{X}_{\text{-}\ell}), X_\ell)$ for $\boldsymbol{X} = \Theta \boldsymbol{Z}$. Also, as $\boldsymbol{X}_j = \Theta_j \boldsymbol{Z}_j$ with $\boldsymbol{Z}_j$ being n i.i.d. draws from F_Z, we have

$$\hat{\phi}(\gamma y^c(\cdot), \mathbf{X}) = \frac{1}{n} \sum_{\ell=1}^{n} \Big(\frac{1}{k} \sum_{j=1}^{k} \varphi_\ell(\gamma y^c(\cdot), \boldsymbol{Z}_j, \Theta_j) \Big).$$

Then,

$$\begin{aligned} & \sup_{\gamma \in \Gamma} \Big| \hat{\phi}(\gamma y^c(\cdot), \mathbf{X}) - \mathsf{E}[\Theta^\kappa \psi(\gamma y^c(\boldsymbol{Z}_{\text{-}\ell}), Z_\ell)] \Big| \\ = \; & \sup_{\gamma \in \Gamma} \Big| \frac{1}{n} \sum_{\ell=1}^{n} \Big(\frac{1}{k} \sum_{j=1}^{k} \varphi_\ell(\gamma y^c(\cdot), \boldsymbol{Z}_j, \Theta_j) - \mathsf{E}[\Theta^\kappa \psi(\gamma y^c(\boldsymbol{Z}_{\text{-}\ell}), Z_\ell)] \Big) \Big| \\ \leq \; & \frac{1}{n} \sum_{\ell=1}^{n} \sup_{\gamma \in \Gamma} \Big| \frac{1}{k} \sum_{j=1}^{k} \varphi_\ell(\gamma y^c(\cdot), \boldsymbol{Z}_j, \Theta_j) - \mathsf{E}[\Theta^\kappa \psi(\gamma y^c(\boldsymbol{Z}_{\text{-}\ell}), Z_\ell)] \Big|. \end{aligned}$$

Note that $\Pr\{A \geq \epsilon\} \leq \Pr\{B \geq \epsilon\}$ for $A \leq B, a.s.$ Moreover, because $\min_{j \in \{1,2,\dots,n\}} A_j \leq \frac{1}{k} \sum_{j=1}^{k} A_j$, we have $\Pr\{\frac{1}{n} \sum_{j=1}^{n} A_j \geq \epsilon\} \leq \Pr\{\min_{j=1,2,\dots,n} A_j \geq \epsilon\} = \Pr\{A_j \geq \epsilon, j = 1, 2, \dots, n\} \leq \sum_{j=1}^{n} \Pr\{A_j \geq \epsilon\}$. We deduce

$$\begin{aligned} & \Pr\Big\{ \sup_{\gamma \in \Gamma} \Big| \hat{\phi}(\gamma y^c(\cdot), \mathbf{X}) - \mathsf{E}[\Theta^\kappa \psi(\gamma y^c(\boldsymbol{Z}_{\text{-}\ell}), Z_\ell)] \Big| \geq \epsilon \Big\} \\ \leq \; & \sum_{\ell=1}^{n} \Pr\Big\{ \sup_{\gamma \in \Gamma} \Big| \frac{1}{k} \sum_{j=1}^{k} \varphi_\ell(\gamma y^c(\cdot), \boldsymbol{Z}_j, \Theta_j) - \mathsf{E}[\Theta^\kappa \psi(\gamma y^c(\boldsymbol{Z}_{\text{-}\ell}), Z_\ell)] \Big| \geq \epsilon \Big\}. \end{aligned}$$

Note that $(\Theta_1, \Theta_2, \dots, \Theta_k)$ are k i.i.d. draws from F_Θ and $\boldsymbol{Z}_j = \boldsymbol{X}_j / \Theta_j$ are k i.i.d. draws of F_Z. It is clear that $\{(\boldsymbol{Z}_j, \Theta_j), j = 1, 2, \dots, k\}$ are k draws of $(\boldsymbol{Z}, \Theta)$, which has joint density $\prod_{i=1}^{n} f_Z(z_i) f_\Theta(\theta)$. Applying Lemma A.2, we have

$$\Pr\Big\{ \sup_{\gamma \in \Gamma} \Big| \frac{1}{k} \sum_{j=1}^{k} \varphi_\ell(\gamma y^c(\cdot), \boldsymbol{Z}_j, \Theta_j) - \mathsf{E}[\Theta^\kappa \psi(\gamma y^c(\boldsymbol{Z}_{\text{-}\ell}), Z_\ell)] \Big| \geq \epsilon \Big\} \leq \mathcal{O}\Big(\exp\Big(- \frac{2k\epsilon^2 - 4C_1\sqrt{k}\epsilon}{(b-a)^2} \Big) \Big),$$

for some $C_1 > 0$. We conclude the convergence of the objective function.

To see the convergence of the boosted operational statistic, we note that if some y^* is optimal for a system with $\Theta = 1$ (i.e., maximizes $\mathsf{E}[\psi(y(\boldsymbol{Z}_{\text{-}1}), Z_1)]$ over some subclass within $\mathcal{H}_+^{n-1}(\iota, \pi)$), then the optimal base operational statistic for any sample $\boldsymbol{x}_{\text{-}1} \in \mathbb{R}_+^{n-1}$ is $y^*(\boldsymbol{x}_{\text{-}1} / \hat{\mu}_{\boldsymbol{x}_{\text{-}1}})$, regardless of the actual Θ value with which the sample is generated. Thus, when a sample of $\boldsymbol{x}_{\text{-}1}$ is observed from some system with $\Theta = \theta$, the optimal decision is $(\hat{\mu}_{\boldsymbol{x}_{\text{-}1}})^\iota y^*(\boldsymbol{x}_{\text{-}1} / \hat{\mu}_{\boldsymbol{x}_{\text{-}1}}) = y^*(\boldsymbol{x}_{\text{-}1})$. Thus, the optimal operational statistic within $\mathcal{Y}^n_{\text{CB-OS}}(y^c)$ for any system is $\gamma^*_{\text{CoL-CB:c}} y^c$. The convergence of the constant-boosting operational statistic follows from Lemma A.6. The convergence rate of the constant-boosting operational statistic follows by applying Lemma A.7. □

Proof of Theorem 7. Define

$$\varphi_\ell(y(\cdot), \boldsymbol{Z}, \Theta) = \Theta^\kappa \psi(y(\boldsymbol{Z}_{\text{-}\ell}), Z_\ell).$$

As $\boldsymbol{X}_j = \Theta_j \boldsymbol{Z}_j$ with $\boldsymbol{Z}_j$ being n i.i.d. draws from F_Z, we have, for a given $\boldsymbol{z}_B \in \mathbb{R}^{n-1}_+$,

$$\hat{\phi}(\gamma, y^c(\cdot), \boldsymbol{z}_B, \eta, \mathbf{X}) = \frac{1}{n} \sum_{\ell=1}^n \frac{\sum_{j=1}^k \mathbb{I}_{\{d(\boldsymbol{Z}_{j:\text{-}\ell}/\hat{\mu}_{\boldsymbol{Z}_{j:\text{-}\ell}}, \boldsymbol{z}_B) \le \eta\}} \varphi_\ell(\gamma y^c(\cdot), \boldsymbol{Z}_j, \Theta_j)}{\sum_{j=1}^k \mathbb{I}_{\{d(\boldsymbol{Z}_{j:\text{-}\ell}/\hat{\mu}_{\boldsymbol{Z}_{j:\text{-}\ell}}, \boldsymbol{z}_B) \le \eta\}}}.$$

Because $(\Theta_1, \Theta_2, \ldots, \Theta_k)$ are k i.i.d. draws from F_Θ and each vector $\boldsymbol{Z}_j = \boldsymbol{X}_j/\Theta_j$ are n i.i.d. draws of F_Z. It is clear that $\{(\boldsymbol{Z}_j, \Theta_j), j = 1, 2, \ldots, k\}$ are k draws of $(\boldsymbol{Z}, \Theta)$, which have joint density $\prod_{i=1}^n f_Z(z_i) f_\Theta(\theta)$. By Lemma 2, we have, for any $\ell \in \{1, 2, \ldots, n\}$, there is a finite positive $\underline{c}_{n-1}$ such that

$$\Pr\{d(\boldsymbol{Z}_{\text{-}\ell}/\hat{\mu}_{\boldsymbol{Z}_{\text{-}\ell}}, \boldsymbol{z}_B) \le \eta\} \ge \underline{c}_{n-1} \eta^{n-2}$$

Let $T_k = \sum_{j=1}^k \mathbb{I}_{\{d(\boldsymbol{Z}_{j:\text{-}\ell}/\hat{\mu}_{\boldsymbol{Z}_{j:\text{-}\ell}}, \boldsymbol{z}_B) \le \eta\}}$. Then,

$$\mathsf{E}[T_k] = \sum_{j=1}^k \Pr\left\{ d(\boldsymbol{Z}_{j:\text{-}\ell}/\hat{\mu}_{\boldsymbol{Z}_{j:\text{-}\ell}}, \boldsymbol{z}_B) \le \eta \right\} \ge \underline{c}_{n-1} k \eta^{n-2} \propto k^{\frac{1}{n-1}}.$$

An argument similar to the proof of Theorem 4 suggest that the probability of T_k larger than $\frac{1}{2}\mathsf{E}[T_k] \propto \mathcal{O}(k^{\frac{1}{n-1}})$ converges to 1 as $k \to \infty$. Applying Lemma A.2, we deduce that, for sufficiently large k, there exists some $C_1 > 0$ such that

$$\begin{aligned} \Pr\Bigg\{ \Bigg| & \frac{\sum_{j=1}^k \mathbb{I}_{\{d(\boldsymbol{Z}_{j:\text{-}\ell}/\hat{\mu}_{\boldsymbol{Z}_{j:\text{-}\ell}}, \boldsymbol{z}_B) \le \eta\}} \varphi_\ell(\gamma y^c(\cdot), \boldsymbol{Z}_j, \Theta_j)}{\sum_{j=1}^k \mathbb{I}_{\{d(\boldsymbol{Z}_{j:\text{-}\ell}/\hat{\mu}_{\boldsymbol{Z}_{j:\text{-}\ell}}, \boldsymbol{z}_B) \le \eta\}}} \\ & - \mathsf{E}\big[\Theta^\kappa \psi(\gamma y^c(\boldsymbol{Z}_{\text{-}1}), Z_1) \big| d(\boldsymbol{Z}_{\text{-}1}/\hat{\mu}_{\boldsymbol{Z}_{\text{-}1}}, \boldsymbol{z}_B) \le \eta\big] \Bigg| \ge \epsilon \Bigg\} \\ \le \mathcal{O}\Big(& \exp\Big(- \frac{2k\eta^{n-2}\epsilon^2 - 4C_1\sqrt{k\eta^{n-2}}\epsilon}{(b-a)^2} \Big)\Big). \end{aligned}$$

Because $\big[\varphi_\ell(\gamma y^c(\cdot), \boldsymbol{Z}_j, \Theta_j) \big| d(\boldsymbol{Z}_{j:\text{-}\ell}/\hat{\mu}_{\boldsymbol{Z}_{j:\text{-}\ell}}, \boldsymbol{z}_B) \le \eta\big] =^d \big[\Theta^\kappa \psi(\gamma y^c(\boldsymbol{Z}_{\text{-}1}), Z_1) \big| d(\boldsymbol{Z}_{\text{-}1}/\hat{\mu}_{\boldsymbol{Z}_{\text{-}1}}, \boldsymbol{z}_B) \le \eta\big]$.

Because $y^c \in \mathcal{H}^{n-1}_+(\iota, \pi)$, y^c is continuous. We deduce

$$\begin{aligned} \big[y^c(\boldsymbol{Z}_{\text{-}1}) \big| d(\boldsymbol{Z}_{\text{-}1}/\hat{\mu}_{\boldsymbol{Z}_{\text{-}1}}, \boldsymbol{z}_B) \le \eta\big] &= \big[(\hat{\mu}_{\boldsymbol{Z}_{\text{-}1}})^\iota y^c(\boldsymbol{Z}_{\text{-}1}/\hat{\mu}_{\boldsymbol{Z}_{\text{-}1}}) \big| d(\boldsymbol{Z}_{\text{-}1}/\hat{\mu}_{\boldsymbol{Z}_{\text{-}1}}, \boldsymbol{z}_B) \le \eta\big] \\ &\xrightarrow{a.s.} \big[(\hat{\mu}_{\boldsymbol{Z}_{\text{-}1}})^\iota y^c(\boldsymbol{z}_B) \big| d(\boldsymbol{Z}_{\text{-}1}/\hat{\mu}_{\boldsymbol{Z}_{\text{-}1}}, \boldsymbol{z}_B) \le \eta\big], \end{aligned}$$

as $\eta \to 0$. Together with the Lipschitz continuity of ψ, we deduce

$$\Big|\mathsf{E}\big[\Theta^\kappa \psi(\gamma y^c(\boldsymbol{Z}_{\text{-}1}), Z_1)\big| d(\boldsymbol{Z}_{\text{-}1}/\hat{\mu}_{\boldsymbol{Z}_{\text{-}1}}, \boldsymbol{z}_B) \le \eta\big] \\ - \mathsf{E}\big[\Theta^\kappa \psi(\gamma(\hat{\mu}_{\boldsymbol{Z}_{\text{-}1}})^\iota y^c(\boldsymbol{z}_B), Z_1)\big| d(\boldsymbol{Z}_{\text{-}1}/\hat{\mu}_{\boldsymbol{Z}_{\text{-}1}}, \boldsymbol{z}_B) \le \eta\big]\Big| \\ \to 0, \quad \text{as } \eta \to 0.$$

By Lemma 3, we have

$$\big[\hat{\mu}_{\boldsymbol{Z}_{\text{-}1}} | d(\boldsymbol{Z}_{\text{-}1}/\hat{\mu}_{\boldsymbol{Z}_{\text{-}1}}, \boldsymbol{z}_B) \le \eta\big] \xrightarrow{d} \big[\hat{\mu}_{\boldsymbol{Z}_{\text{-}1}} | d(\boldsymbol{Z}_{\text{-}1}/\hat{\mu}_{\boldsymbol{Z}_{\text{-}1}}, \boldsymbol{z}_B) = 0\big], \text{ as } \eta \to 0.$$

Thus, as $\eta \to 0$,

$$\Big|\mathsf{E}\big[\Theta^\kappa \psi(\gamma(\hat{\mu}_{\boldsymbol{Z}_{\text{-}1}})^\iota y^c(\boldsymbol{z}_B), Z_1)\big| d(\boldsymbol{Z}_{\text{-}1}/\hat{\mu}_{\boldsymbol{Z}_{\text{-}1}}, \boldsymbol{z}_B) \le \eta\big] \\ - \mathsf{E}\big[\Theta^\kappa \psi(\gamma(\hat{\mu}_{\boldsymbol{Z}_{\text{-}1}})^\iota y^c(\boldsymbol{z}_B), Z_1)\big| \boldsymbol{Z}_{\text{-}1}/\hat{\mu}_{\boldsymbol{Z}_{\text{-}1}} = \boldsymbol{z}_B\big]\Big| \to 0.$$

Because $\mathsf{E}\big[\Theta^\kappa \psi(\gamma(\hat{\mu}_{\boldsymbol{Z}_{\text{-}1}})^\iota y^c(\boldsymbol{z}_B), Z_1)\big| \boldsymbol{Z}_{\text{-}1}/\hat{\mu}_{\boldsymbol{Z}_{\text{-}1}} = \boldsymbol{z}_B\big] = \mathsf{E}[\Theta^\kappa]\phi_B[\gamma y^c(\boldsymbol{z}_B), \boldsymbol{z}_B, F_Z]$ by Equation (7) and Assumption 3, we conclude the convergence of targeted-boosting objective for any $\epsilon > 0$ with a sufficiently large k.

Now denote $\gamma^*_{\text{CoL-TB:c}}$ be the unique maximizer of $\phi_B[\gamma y^c(\boldsymbol{z}_B), \boldsymbol{z}_B, F_Z]$. Then we must have $\gamma^*_{\text{CoL-TB:c}} y^c(\boldsymbol{z}_B) = y^*_{\text{H}\iota}(\boldsymbol{z}_B)$ based on Theorem 1 and Equation (8). The convergence of the boosted operational statistic follows directly by applying Lemma A.6. The convergence rate of the targeted boosting operational statistic follows by applying Lemma A.7. □

B Examples to Demonstrate Theorem 1

In the newsvendor application, we have $\psi(y, x) = p\min\{y, x\} - cy$ satisfying Assumption 1 with $\kappa = \iota = 1$. By Lemma 5 of Chu et al. (2025), the function ϕ_B in (7) can be explicitly written as

$$\phi_B[y, \boldsymbol{z}_B, F_Z] = \int_0^\infty \frac{1}{\eta}\mathsf{E}\big[p\min(\eta y \wedge Z) - c\eta y\big]\eta^n \prod_{i=1}^n f_Z(\eta z_{B:i})\mathrm{d}\eta. \tag{40}$$

Based on Equation (8), the derivation of the optimal operational statistic boils down to computing the base operational statistic y_B^{OS} from the above equation. Next, we apply Theorem 1 to specific distribution families.

B.1 Gamma Distribution

The random variable Z follows a gamma distribution with mean one. Let k_Z denote the shape parameter, then the probability density function of Z is

$$f_Z(z) = \frac{z^{k_Z-1}e^{-z}}{\Gamma(k_Z)},\ \forall z \in \mathbb{R}_+,$$

where $\Gamma(\cdot)$ is a Gamma function. Note that when $k_Z = 1$, Z becomes an exponential random variable with mean one.

Substituting the density function into Equation (40) and differentiating with respect to y gives

$$\begin{aligned}
\frac{\partial \phi_B}{\partial y} &= \int_0^\infty \left(p\bar{F}_Z(\eta y) - c\right)\eta^n \prod_{i=1}^n f_Z\Big(\eta z_{B:i}\Big)\mathrm{d}\eta \\
&= \int_0^\infty \left(p\int_y^\infty \frac{(\eta x)^{k_Z-1}e^{-\eta x}}{\Gamma(k_Z)}\eta dx - c\right)\eta^n \prod_{i=1}^n \frac{(\eta z_{B:i})^{k_Z-1}e^{-\eta z_{B:i}}}{\Gamma(k_Z)} d\eta \\
&= \big(\Gamma(k_Z)\big)^{-n}\Big(\prod_{i=1}^n z_{B:i}^{k_Z-1}\Big)\cdot \int_0^\infty \left(p\int_y^\infty \frac{\eta^{k_Z}x^{k_Z-1}e^{-\eta x}}{\Gamma(k_Z)}dx - c\right)\eta^{nk_Z}e^{-\eta\sum_{i=1}^n z_{B:i}}d\eta \\
&= \big(\Gamma(k_Z)\big)^{-(n+1)}\Big(\prod_{i=1}^n z_{B:i}^{k_Z-1}\Big)\cdot \int_0^\infty \left(p\int_y^\infty \eta^{k_Z}x^{k_Z-1}e^{-\eta x}dx - c\Gamma(k_Z)\right)\eta^{nk_Z}e^{-\eta n\hat{\mu}_{\boldsymbol{z}_B}}d\eta.
\end{aligned}$$

Setting the right-hand side to zero, we deduce

$$\begin{aligned}
\frac{c}{p} &= \frac{\int_0^\infty\int_y^\infty \eta^{k_Z}x^{k_Z-1}e^{-\eta x}\mathrm{d}x\cdot \eta^{nk_Z}e^{-\eta n\hat{\mu}_{\boldsymbol{z}_B}}\mathrm{d}\eta}{\Gamma(k_Z)\int_0^\infty \eta^{nk_Z}e^{-\eta n\hat{\mu}_{\boldsymbol{z}_B}}\mathrm{d}\eta} \\
&= \frac{\int_y^\infty x^{k_Z-1}\int_0^\infty \eta^{(n+1)k_Z}e^{-\eta(x+n\hat{\mu}_{\boldsymbol{z}_B})}\mathrm{d}\eta\mathrm{d}x}{\Gamma(k_Z)\int_0^\infty \eta^{nk_Z}e^{-\eta n\hat{\mu}_{\boldsymbol{z}_B}}\mathrm{d}\eta} \\
&= \frac{(n\hat{\mu}_{\boldsymbol{z}_B})^{nk_Z+1}\int_y^\infty \frac{x^{k_Z-1}}{(x+n\hat{\mu}_{\boldsymbol{z}_B})^{(n+1)k_Z+1}}\int_0^\infty \eta^{(n+1)k_Z}(x+n\hat{\mu}_{\boldsymbol{z}_B})^{(n+1)k_Z+1}\cdot e^{-\eta(x+n\hat{\mu}_{\boldsymbol{z}_B})}\mathrm{d}\eta\mathrm{d}x}{\Gamma(k_Z)\int_0^\infty (\eta n\hat{\mu}_{\boldsymbol{z}_B})^{nk_Z}e^{-\eta n\hat{\mu}_{\boldsymbol{z}_B}}\mathrm{d}(\eta n\hat{\mu}_{\boldsymbol{z}_B})} \\
&= \frac{\Gamma\big((n+1)k_Z+1\big)}{\Gamma(k_Z)\Gamma(nk_Z+1)}\cdot\int_y^\infty \frac{\big(n\hat{\mu}_{\boldsymbol{z}_B}\big)^{nk_Z+1}x^{k_Z-1}}{(x+n\hat{\mu}_{\boldsymbol{z}_B})^{(n+1)k_Z+1}}\mathrm{d}x \\
&= \frac{\Gamma\big((n+1)k_Z+1\big)}{\Gamma(k_Z)\Gamma(nk_Z+1)}\cdot\int_y^\infty \frac{n\cdot n^{nk_Z+1}x^{k_Z-1}}{(x+n)^{nk_Z+2}(x+n)^{k_Z-1}}\mathrm{d}\Big(\frac{x}{n}\Big) \\
&= \frac{\Gamma\big((n+1)k_Z+1\big)}{\Gamma(k_Z)\Gamma(nk_Z+1)}\cdot\int_{\frac{y}{n}}^\infty \Big(\frac{1}{1+x}\Big)^{nk_Z+2}\Big(\frac{x}{1+x}\Big)^{k_Z-1}\mathrm{d}x. \qquad (41)
\end{aligned}$$

The solution to this equation gives the optimal base operational statistic, $y_B^{\mathrm{OS}}(\boldsymbol{z}_B)$. We observe that $y_B^{\mathrm{OS}}(\boldsymbol{z}_B) \in \mathcal{SC}_+^n(1,\pi)$. In other words, $y^*_{\mathrm{H}\iota} = y^*_{\mathrm{SC}\iota}$.

When Z follows an exponential distribution (i.e., $k_Z = 1$), (41) reduces to $c/p = (1+y/n)^{-(n+1)}$, In this case, the optimal base operational statistic becomes

$$y_B^{OS}(\boldsymbol{z}_B) = y^*_{\mathrm{SC1}}[F_Z] = n\Big(\Big(\frac{p}{c}\Big)^{\frac{1}{n+1}} - 1\Big),\ \forall \boldsymbol{z}_B \in \mathcal{B}^n.$$

B.2 Beta Distribution

The random variable Z follows a beta distribution with mean one and shape parameters (α, β). Thus, Z can be expressed as $Z =^d \alpha_0 Z_0$, where $\alpha_0 = \frac{\alpha+\beta}{\alpha}$, with the density of Z_0 being

$$f_{Z_0}(z) = \frac{z^{\alpha-1}(1-z)^{\beta-1}}{B(\alpha,\beta)}, \ \forall z \in [0,1],$$

where $B(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\mathrm{d}t$ is a beta function. Note that when $\alpha = \beta = 1$, Z becomes a uniform random variable over $[0,2]$.

Given that the distribution is bounded, we need to examine the values of $\eta z_{B:i}$ to determine the range of integration in (40). Let $z_{B[i]}$ be the ith largest element in $\boldsymbol{z}_B$. Then, we should evaluate the integration in (40) for $\eta z_{B[n]} \leq \alpha_0$, or $\eta \leq \alpha_0 / z_{B[n]}$. Now substituting the density function into (40) and differentiating with respect to y, we derive

$$\begin{aligned}
\frac{\partial \phi_B}{\partial y} &= \int_0^{\alpha_0/z_{B[n]}} (p\bar{F}_Z(\eta y) - c)\eta^n \prod_{i=1}^n f_Z\Big(\eta z_{B:i}\Big)\mathrm{d}\eta = \int_0^{\alpha_0/z_{B[n]}} (p\bar{F}_{Z_0}\Big(\frac{\eta y}{\alpha_0}\Big) - c)\eta^n \prod_{i=1}^n \frac{1}{\alpha_0} f_{Z_0}\Big(\frac{\eta z_{B:i}}{\alpha_0}\Big)\mathrm{d}\eta \\
&= \int_0^{\alpha_0/z_{B[n]}} \left(p\int_{\min\{\eta y/\alpha_0, 1\}}^1 \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}\mathrm{d}x - c\right)\eta^n \prod_{i=1}^n \frac{(\frac{\eta z_{B:i}}{\alpha_0})^{\alpha-1}(1-(\frac{\eta z_{B:i}}{\alpha_0}))^{\beta-1}}{\alpha_0 B(\alpha,\beta)}\mathrm{d}\eta \\
&= \frac{\alpha_0}{\big(B(\alpha,\beta)\big)^n}\int_0^{1/z_{B[n]}} \left(p\int_{\min\{\tilde{\eta} y, 1\}}^1 \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}\mathrm{d}x - c\right)\tilde{\eta}^n \prod_{i=1}^n \big(\tilde{\eta} z_{B:i}\big)^{\alpha-1}\big(1-\tilde{\eta} z_{B:i}\big)^{\beta-1}\mathrm{d}\tilde{\eta}.
\end{aligned}$$

The fourth equation is derived by replacing η/α_0 with $\tilde{\eta}$. Setting the right-hand side to zero, we deduce

$$B(\alpha,\beta)\frac{c}{p} = \frac{\int_0^{1/z_{B[n]}} \left(\int_{\min\{\tilde{\eta} y,1\}}^1 x^{\alpha-1}(1-x)^{\beta-1}\mathrm{d}x\right)\tilde{\eta}^{\alpha n}\prod_{i=1}^n \big(1-\tilde{\eta} z_{B:i}\big)^{\beta-1}\mathrm{d}\tilde{\eta}}{\int_0^{1/z_{B[n]}} \tilde{\eta}^{\alpha n}\prod_{i=1}^n \big(1-\tilde{\eta} z_{B:i}\big)^{\beta-1}\mathrm{d}\tilde{\eta}}. \tag{42}$$

The solution to this equation gives the optimal base operational statistic, $y_B^{\mathrm{OS}}(\boldsymbol{z}_B)$. We observe that $y_B^{\mathrm{OS}}(\cdot) \notin \mathcal{SC}_+^n(1,\pi)$.

When Z follows a uniform distribution (i.e., $\alpha = \beta = 1$), (42) reduces to

$$\begin{aligned}
\frac{c}{p} &= \frac{\int_0^{1/z_{B[n]}} \max\big\{1-\tilde{\eta} y, 0\big\}\tilde{\eta}^n \mathrm{d}\tilde{\eta}}{\int_0^{1/z_{B[n]}} \tilde{\eta}^n \mathrm{d}\tilde{\eta}} = \frac{\int_0^{\min\{1/z_{B[n]}, 1/y\}} (1-\tilde{\eta} y)\tilde{\eta}^n \mathrm{d}\tilde{\eta}}{\frac{1}{n+1}(z_{B[n]})^{-(n+1)}} \\
&= \begin{cases} \frac{1}{n+2}\big(\frac{y}{z_{B[n]}}\big)^{-(n+1)} & \text{if } z_{B[n]} \leq y; \\ 1 - \frac{n+1}{n+2}\frac{y}{z_{B[n]}} & \text{if } z_{B[n]} > y. \end{cases}
\end{aligned}$$

The solution to this equation gives the optimal base operational statistic

$$y_B^{OS}(\boldsymbol{z}_B) = \begin{cases} \Big(\frac{p}{(n+2)c}\Big)^{\frac{1}{n+1}} z_{B[n]}, & \text{if } n \leq \frac{p}{c} - 2; \\ \frac{n+2}{n+1}\Big(1 - \frac{c}{p}\Big) z_{B[n]}, & \text{if } n > \frac{p}{c} - 2. \end{cases} \tag{43}$$

To derive the optimal scaled operational statistic for uniformly-distributed Z, we substitute the density function $f_Z(z) = 1/2, z \in [0, 2]$, into (12) to obtain

$$\begin{aligned}
\mathsf{E}[\psi(\hat{\mu}_{\boldsymbol{Z}}\gamma, Z)] &= \int_{\boldsymbol{z}\in[0,2]^n} \int_0^1 \psi(\hat{\mu}_{\boldsymbol{z}}\gamma, z_0) f_Z(z_0)\mathrm{d}z_0 \prod_{i=1}^n f_Z(z_i)\mathrm{d}\boldsymbol{z} \\
&= \frac{1}{2^{n+1}}\Big(\int_{\boldsymbol{z}\in[0,2]^n} p\int_0^1 [(\hat{\mu}_{\boldsymbol{z}}\gamma)\wedge z_0]\mathrm{d}z_0\mathrm{d}\boldsymbol{z} - c\int_{\boldsymbol{z}\in[0,2]^n} \hat{\mu}_{\boldsymbol{z}}\gamma\mathrm{d}\boldsymbol{z}\Big) \\
&= \frac{p}{2^{n+1}}\Big(\int_{\boldsymbol{z}\in[0,2]^n\cap\hat{\mu}_{\boldsymbol{z}}\gamma\le 1}\Big[(\hat{\mu}_{\boldsymbol{z}}\gamma) - \frac{1}{2}(\hat{\mu}_{\boldsymbol{z}}\gamma)^2\Big]\mathrm{d}\boldsymbol{z} + \int_{\boldsymbol{z}\in[0,2]^n\cap\hat{\mu}_{\boldsymbol{z}}\gamma>1}\frac{1}{2}\mathrm{d}\boldsymbol{z} - \frac{1}{2}\frac{c}{p}\gamma\Big) \\
&= \frac{p}{2^{n+2}}\Big(\Big(1-\frac{c}{p}\Big)\gamma - \frac{3n+1}{12n}\gamma^2 + \mathbb{I}_{\{\gamma>1\}}\int_{\boldsymbol{z}\in[0,2]^n\cap\hat{\mu}_{\boldsymbol{z}}\gamma>1}(\hat{\mu}_{\boldsymbol{z}}\gamma-1)^2\mathrm{d}\boldsymbol{z}\Big).
\end{aligned}$$

The third and forth equations follow from the fact that $\int_{\boldsymbol{z}\in[0,2]^n}\hat{\mu}_{\boldsymbol{z}}\mathrm{d}\boldsymbol{z} = 1/2$ and $\int_{\boldsymbol{z}\in[0,2]^n}(\hat{\mu}_{\boldsymbol{z}})^2\mathrm{d}\boldsymbol{z} = \frac{3n+1}{12n}$, respectively. Differentiating the right-hand side with respect to γ gives

$$\frac{\partial\mathsf{E}[\psi(\hat{\mu}_{\boldsymbol{Z}}\gamma, Z)]}{\partial\gamma} = \frac{p}{2^{n+2}}\Big(\Big(1-\frac{c}{p}\Big) - \frac{3n+1}{6n}\gamma + \mathbb{I}_{\{\gamma>1\}}\frac{\partial\int_{\boldsymbol{z}\in[0,2]^n\cap\hat{\mu}_{\boldsymbol{z}}\gamma>1}(\hat{\mu}_{\boldsymbol{z}}\gamma-1)^2\mathrm{d}\boldsymbol{z}}{\partial\gamma}\Big).$$

Setting the right-hand side to zero gives the optimal scale base operational statistic

$$y^*_{\mathrm{SC1}}[F_Z] = \begin{cases} \frac{6n}{3n+1}\Big(1-\frac{c}{p}\Big) & \text{if } \frac{2}{3n-1} \ge \frac{p}{c} - 2; \\ \arg\max\Big\{\big(1-\frac{c}{p}\big)\gamma - \frac{3n+1}{12n}\gamma^2 & \\ +\int_{\boldsymbol{z}\in[0,2]^n\cap\hat{\mu}_{\boldsymbol{z}}\gamma>1}(\hat{\mu}_{\boldsymbol{z}}\gamma-1)^2\mathrm{d}\boldsymbol{z} : \gamma \ge 1\Big\} & \text{if } \frac{2}{3n-1} < \frac{p}{c} - 2. \end{cases} \tag{44}$$

It is clear that the solutions in (43) and (10) are not equivalent.

B.3 Log-normal Distribution

The random variable Z follows a log-normal distribution with mean one and shape parameter σ. Thus, Z can be expressed as $Z =^d \alpha_0 Z_0$, where $\alpha_0 = 1/\mathsf{E}[Z_0]$ and $\ln Z_0 \sim \mathcal{N}(0, \sigma^2)$ with the density of $\ln Z_0$ being

$$f_{\ln Z_0}(z) = \frac{1}{\sqrt{2\pi}\sigma}\exp\Big(-\frac{z^2}{2\sigma^2}\Big),\ \forall z \in \mathbb{R}_+.$$

Note that $f_Z(z) = \frac{1}{\alpha_0}f_{Z_0}(z/\alpha_0) = \frac{1}{z}f_{\ln Z_0}(\ln z - \ln\alpha_0)$ and $\bar{F}_Z(z) = \bar{F}_{Z_0}(z/\alpha_0) = \bar{F}_{\ln Z_0}(\ln z - \ln\alpha_0)$.

Substituting the density function into (40) and differentiating with respect to y gives

$$\begin{aligned}
\frac{\partial \phi_B}{\partial y} &= \int_0^\infty (p\bar{F}_Z(\eta y) - c)\eta^n \prod_{i=1}^n f_Z\Big(\eta z_{B:i}\Big)\mathrm{d}\eta = \int_0^\infty \Big(p\bar{F}_{Z_0}\Big(\frac{\eta y}{\alpha_0}\Big) - c\Big)\eta^n \prod_{i=1}^n \frac{1}{\alpha_0} f_{Z_0}\Big(\frac{\eta z_{B:i}}{\alpha_0}\Big)\mathrm{d}\eta \\
&= (\alpha_0)^n \int_0^\infty \Big(p\bar{F}_{\ln Z_0}\big(\ln(\eta/\alpha_0) + \ln y\big) - c\Big)\frac{\eta^n}{(\alpha_0)^n}\prod_{i=1}^n \frac{\alpha_0}{\eta z_{B:i}} f_{\ln Z_0}\big(\ln(\eta/\alpha_0) + \ln z_{B:i}\big)\mathrm{d}(\eta/\alpha_0) \\
&= (\alpha_0)^n \int_0^\infty \Big(p\int_{\ln\tilde{\eta}+\ln y}^\infty f_{\ln Z_0}(x)\mathrm{d}x - c\Big)\tilde{\eta}^n \prod_{i=1}^n \frac{1}{\tilde{\eta} z_{B:i}} f_{\ln Z_0}\big(\ln\tilde{\eta} + \ln z_{B:i}\big)\mathrm{d}\tilde{\eta} \\
&= (\alpha_0)^n \int_0^\infty \Big(p\int_{\ln y}^\infty \frac{1}{\sqrt{2\pi}\sigma}\exp\Big(\frac{-(x+\ln\tilde{\eta})^2}{2\sigma^2}\Big)\mathrm{d}x - c\Big)\prod_{i=1}^n \frac{1}{z_{B:i}}\frac{1}{\sqrt{2\pi}\sigma}\exp\Big(\frac{-(\ln z_{B:i} + \ln\tilde{\eta})^2}{2\sigma^2}\Big)\mathrm{d}\tilde{\eta} \\
&= C_0 \int_{-\infty}^\infty \Big(p\int_{\ln y}^\infty \frac{1}{\sqrt{2\pi}\sigma}\exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big)\mathrm{d}x - c\Big)\exp(-\check{\eta})\prod_{i=1}^n \exp\Big(-\frac{(\ln z_{B:i} - \check{\eta})^2}{2\sigma^2}\Big)\mathrm{d}\check{\eta} \\
&= C_0 \int_{-\infty}^\infty \Big(p\int_{\ln y}^\infty \frac{1}{\sqrt{2\pi}\sigma}\exp\Big(-\frac{(x-\check{\eta})^2}{2\sigma^2}\Big)\mathrm{d}x - c\Big)\exp\Big(-\frac{\sum_{i=1}^n(\ln z_{B:i} - \check{\eta})^2 + 2\sigma^2\check{\eta}}{2\sigma^2}\Big)\mathrm{d}\check{\eta} \\
&= C_0C_1 \int_{-\infty}^\infty \Big(p\int_{\ln y}^\infty \frac{1}{\sqrt{2\pi}\sigma}\exp\Big(-\frac{(x-\check{\eta})^2}{2\sigma^2}\Big)dx - c\Big)\exp\Big(-\frac{(\check{\eta} - \bar{z}_{\ln} + \sigma^2/n)^2}{2\sigma^2/n}\Big)\mathrm{d}\check{\eta}.
\end{aligned}$$

In the above derivation, we have changed variables with $\check{\eta} = -\ln\tilde{\eta}$, $C_0 = \Big(\frac{\alpha_0}{\sqrt{2\pi}\sigma}\Big)^n \frac{1}{\prod_{i=1}^n z_{B:i}}$, $\bar{z}_{\ln} = \frac{1}{n}\sum_{i=1}^n \ln z_{B:i}$, and $C_1 = \exp\Big(-\frac{\sum_{i=1}^n(\ln z_{B:i})^2}{2\sigma^2} + \frac{(\bar{z}_{\ln} - \sigma^2/n)^2}{2\sigma^2/n}\Big)$. The fourth and sixth equations follow from changing of variable η/α_0 with $\tilde{\eta}$ and $-\ln\tilde{\eta}$ with $\check{\eta}$, respectively. Setting the right-hand side to zero, we deduce

$$\begin{aligned}
\frac{c}{p} &= \frac{\int_{-\infty}^\infty \Big(\int_{\ln y}^\infty \frac{1}{\sqrt{2\pi}\sigma}\exp\Big(-\frac{(x-\check{\eta})^2}{2\sigma^2}\Big)\mathrm{d}x\Big)\exp\Big(-\frac{(\check{\eta} - \bar{z}_{\ln} + \sigma^2/n)^2}{2\sigma^2/n}\Big)\mathrm{d}\check{\eta}}{\int_{-\infty}^\infty \exp\Big(-\frac{(\check{\eta} - \bar{z}_{\ln} + \sigma^2/n)^2}{2\sigma^2/n}\Big)\mathrm{d}\check{\eta}} \\
&= \frac{\sqrt{n}}{\sqrt{2\pi}\sigma}\int_{-\infty}^\infty \Big(\int_{\ln y}^\infty \frac{1}{\sqrt{2\pi}\sigma}\exp\Big(-\frac{(x-\check{\eta})^2}{2\sigma^2}\Big)\mathrm{d}x\Big)\exp\Big(-\frac{(\check{\eta} - \bar{z}_{\ln} + \sigma^2/n)^2}{2\sigma^2/n}\Big)\mathrm{d}\check{\eta} \\
&= \frac{\sqrt{n}}{\sqrt{2\pi}\sigma}\int_{\ln y}^\infty \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}\sigma}\exp\Big(-\frac{(n+1)}{2\sigma^2}\Big[\check{\eta} - \frac{x + n\bar{z}_{\ln} - \sigma^2/n}{n+1}\Big]^2\Big)\mathrm{d}\check{\eta}\exp\Big(-\frac{n(z - \bar{z}_{\ln} + \sigma^2/n)^2}{2\sigma^2(n+1)}\Big)\mathrm{d}x \\
&= \frac{\sqrt{n}}{\sqrt{2\pi}\sigma\sqrt{n+1}}\int_{\ln y}^\infty \exp\Big(-\frac{(x - \bar{z}_{\ln} + \sigma^2/n)^2}{2\sigma^2(n+1)/n}\Big)\mathrm{d}x \\
&= 1 - \Phi\Big(\frac{\ln y - \bar{z}_{\ln} + \sigma^2/n}{\sigma}\sqrt{\frac{n}{n+1}}\Big),
\end{aligned}$$

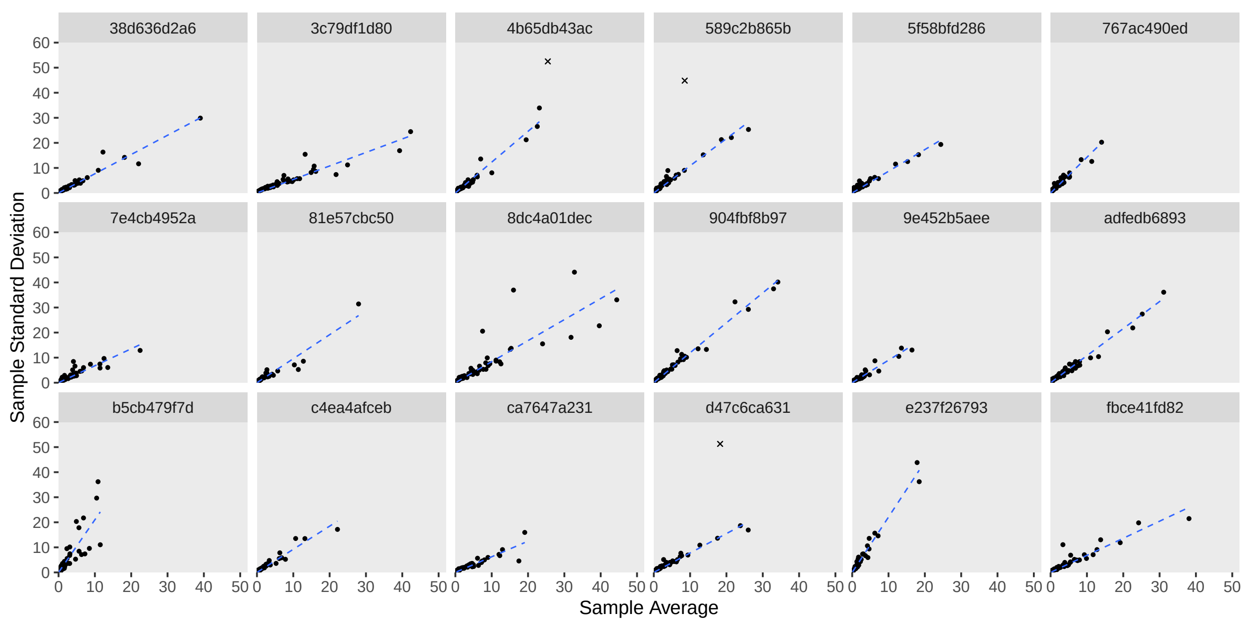
where $\Phi(\cdot)$ is the standard normal distribution function. The second and the fourth equations follow from the fact that $\int_{-\infty}^\infty e^{-\frac{(z-\check{\eta})^2}{2\sigma^2}}dz = \sqrt{2\pi}\sigma$ for any $\check{\eta} \in \mathbb{R}$. The solution to this equation gives the optimal base operational statistic

$$y_B^{\mathrm{OS}}(\boldsymbol{z}_B) = \Big(\prod_{i=1}^n z_{B:i}\Big)^{1/n}\exp\Big\{\sigma\Big[\Phi^{\mathrm{inv}}\Big(1 - \frac{c}{p}\Big)\sqrt{\frac{n+1}{n}} - \frac{\sigma}{n}\Big]\Big\},\ \forall \boldsymbol{z}_B \in \mathcal{B}^n, \tag{45}$$

where $\Phi^{\text{inv}}(\cdot)$ is the inverse of the standard normal distribution function. We observe that the right-hand side does depend on the elements of $\boldsymbol{z}_B$, and thus $y_B^{\text{OS}}(\cdot) \notin \mathcal{SC}_+^n(1, \pi)$.

C Additional Evidence to Justify Assumption 2

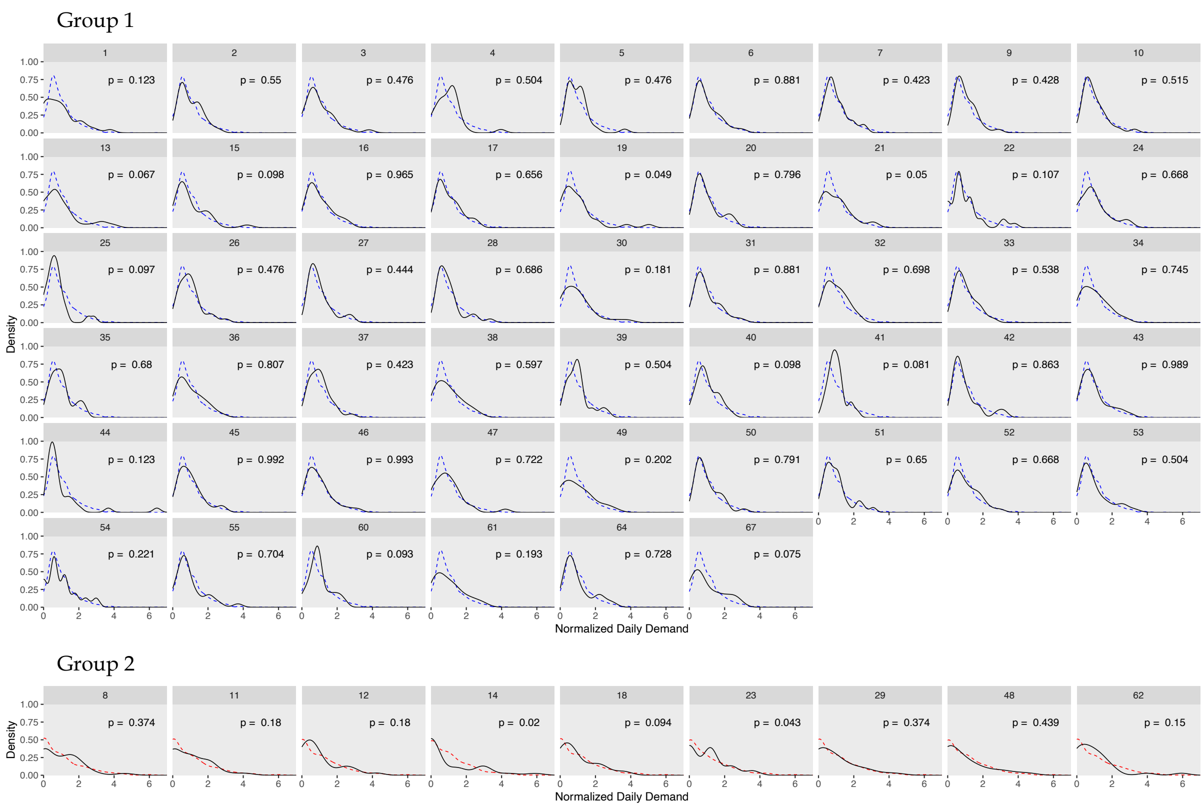
To confirm the observation of the demand pattern in Figure 1, we randomly selected, out of several hundreds, 18 products from the JD.com data set. For each product, we compute the sample average and the sample standard deviation of the daily demands at each of the 60 distribution centers. As we can see from Figure 2, the ratio of the sample average and the sample standard deviation stays close to constant across these products.



Note. Each subfigure corresponds to one randomly selected product from the JD.com data set available at `https://connect.informs.org/msom/events/datadriven2020`. Each subfigure plots the sample average and sample standard deviation of the daily demands for that product in the 60 distribution centers over 31 days.

Figure 2: The daily product demands at different distribution centers of JD.com

There are, indeed, rare exceptions. In particular, for each SKUs 4b65db43ac, 589c2b865b, and d47c6ca631, there is one distribution center (marked with an "x" in the figure), whose demand variability is significantly larger than the rest of the 59 distribution centers. Unfortunately, the available data set does not offer enough information to identify the unique characteristics of these distribution centers. In such situations, we cannot directly utilize the data from such distribution centers for learning. Nevertheless, one can still cross learn or co-learn among the rest of the 59 distribution centers.



Note. Each subfigure corresponds to one distribution center for the daily demands of SKU 068f4481b3 from the JD.com data set available at `https://connect.informs.org/msom/events/datadriven2020`. Each subfigure plots the normalized daily demand density of each distribution center (black solid line) over 31 days and the density of pooled normalized demand over all other distribution centers in the corresponding group (blue and red dashed line) which are identical in either group. The number at top-right corner of each subfigure is the p-value of two-sample Kolmogorov-Smirnov test.

Figure 3: The normalized demand curves of each distribution centers of JD.com

To further validate the Assumption 2 about the scaled demand distributions following a common normalized distribution Z, we analyze the data of SKU 068f4481b3 and adopt two-sample Kolmogorov-Smirnov test for the samples of each normalized demand distribution Z_{Bj}. We normalize the daily demands $\boldsymbol{X}_j$ (with sample size $n_j = n = 31$) by sample average $\hat{\mu}_{\boldsymbol{X}_j}$ for each distribution center j to obtain $\boldsymbol{Z}_{Bj} = \boldsymbol{X}_j/\hat{\mu}_{\boldsymbol{X}_j}$. We start with grouping the distribution centers exhibiting similar empirical shapes of $\boldsymbol{Z}_{Bj}$ and identifying two groups, one with 35 distribution centers and the other with 9. Next, we examine the distribution centers within each group. We take any center j in a group and pool the data of the remaining centers in the same group to form $\boldsymbol{Z}_{B\text{-}j} = \{\boldsymbol{Z}_{B1}, \ldots, \boldsymbol{Z}_{B(j-1)}, \boldsymbol{Z}_{B(j+1)}, \ldots, \boldsymbol{Z}_{Bk}\}$ with $n_{\text{-}j} = |\boldsymbol{Z}_{B\text{-}j}|$. We run the two-sample Kolmogorov-Smirnov test for $\boldsymbol{Z}_{Bj}$ and $\boldsymbol{Z}_{B\text{-}j}$ and report the p-value in Figure 3. Specifically, the two-sample Kolmogorov-Smirnov statistic (left side of the following inequality) follows

$$\sup_{x\in\mathbb{R}} |\hat{F}_{Z|\boldsymbol{Z}_{B\text{-}j}}(x) - \hat{F}_{Z_j|\boldsymbol{Z}_{Bj}}(x)| \leq \sqrt{\ln\left(\frac{2}{\alpha}\right)\frac{n_{\text{-}j}+n_j}{2n_{\text{-}j}n_j}},$$

with probability $1-\alpha$ for sample size $n_{\text{-}j}$ and n_j.

The empirical density curves and corresponding test results are plotted in Figure 3. For Group 1 (2) with 51 (9) distribution centers, all of the hypotheses that $\boldsymbol{Z}_j$ is from the same distribution of the "true" normalized distribution cannot be rejected with significance level of 96% (98%). With the conventional statistical standard of 5% threshold, the result suggests that we cannot reject the null hypothesis that the tested distributions are the same. Thus, we can pool the data within each group to apply co-learning.

Naturally, given the limited sample sizes, there is no guarantee that all pooled distribution centers are truly drawn from the same underlying distribution family with a p-value threshold of 0.05. In practice, different criteria may be used to decide whether data from multiple distribution centers should be pooled. For example, one may begin by pooling centers whose empirical distributions exhibit very similar shapes to form an initial benchmark sample. Then, pairwise tests can be performed between this benchmark and each remaining center, adding the center with the highest p-value to the pooled benchmark. This process can be repeated iteratively until the p-values of all remaining centers fall below a predefined threshold.

The application of the co-learning solutions for this product is reported in Section D.

D Additional Numerical Results

D.1 Synthetic Data

In this section, we introduce the details on how we implement each oracle solution, transfer-learning solution, cross-learning solution, and co-learning solution.

Directly Trained Solutions. Given i.i.d. observations $\boldsymbol{X} = \{X_1, X_2, \ldots, X_n\}$.

- **Sample average approximation (SAA):** Define $\varphi(y, \boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^{n}\psi(y, X_i)$ as the sample average profit. In general, the sample average approximation solution is $\text{argmax}\{\varphi(y, \boldsymbol{X}), y \in \mathbb{R}_+\}$. However, under small samples, this solution can perform extremely badly. Thus, given the continuity of Assumption 2-(2), we use the interpolated SAA newsvendor problem,

$$y_{\text{SAA}} = (r + 1 - n + n \cdot c/s)X_{[r]} + (n - n \cdot c/s - r)X_{[r+1]},$$

where $r = \lfloor n - n \cdot c/s \rfloor$ and $\boldsymbol{X}_{[]}$ is the ascending sequence of X (i.e., $X_{[1]} \le X_{[2]} \le \ldots \le X_{[n]}$).

- **Regularized sample average (Re-Var & Re-Std):** Define $\mathsf{Var}[\varphi(y, \boldsymbol{X})] = \frac{1}{n}\sum_{i=1}^{n}\big(\psi(y, X_i)\big)^2 - \big(\varphi(y, \boldsymbol{X})\big)^2$ as the variance of the sample average profit. The regularized solutions are

$$\begin{aligned} y_{\text{Re-Var}}(\boldsymbol{X}) = y_{\text{Re-Var-}\beta}(\boldsymbol{X}, \beta_{\text{Re-Var}}(\boldsymbol{X})) &= \arg\max \big\{\varphi(y, \boldsymbol{X}) - \beta_{\text{Re-Var}}(\boldsymbol{X})\mathsf{Var}[\varphi(y, \boldsymbol{X})]\big\}, \\ y_{\text{Re-Std}}(\boldsymbol{X}) = y_{\text{Re-Std-}\beta}(\boldsymbol{X}, \beta_{\text{Re-Std}}(\boldsymbol{X})) &= \arg\max \big\{\varphi(y, \boldsymbol{X}) - \beta_{\text{Re-Std}}(\boldsymbol{X})\sqrt{\mathsf{Var}[\varphi(y, \boldsymbol{X})]}\big\}, \end{aligned}$$

where the parameters $\beta_{\text{Re-Var}}(\boldsymbol{X})$ and $\beta_{\text{Re-Std}}(\boldsymbol{X})$ are obtained through five-fold cross validation. Define $\{(\boldsymbol{X}_{\text{-}1}, \boldsymbol{X}_1), (\boldsymbol{X}_{\text{-}2}, \boldsymbol{X}_2), \ldots, (\boldsymbol{X}_{\text{-}5}, \boldsymbol{X}_5)\}$ as five pairs of partition of $\boldsymbol{X}$, which satisfies $\bigcup_{\ell=1}^{5}\boldsymbol{X}_\ell = \boldsymbol{X}$, $\boldsymbol{X}_\ell \bigcap \boldsymbol{X}_{\ell'} = \emptyset$, $\forall \ell \ne \ell' \in \{1, 2, \ldots, 5\}$, and $\boldsymbol{X}_{\text{-}\ell} \bigcap \boldsymbol{X}_\ell = \emptyset$, $\boldsymbol{X}_{\text{-}\ell} \bigcup \boldsymbol{X}_\ell = \boldsymbol{X}$, $\forall \ell \in \{1, 2 \ldots, 5\}$. We derive

$$\beta_{\text{Re-Var}}(\boldsymbol{X}) = \arg\max \Big\{\frac{1}{5}\sum_{\ell=1}^{5}\psi(y_{\text{Re-Var}}(\boldsymbol{X}_{\text{-}\ell}, \beta), \boldsymbol{X}_\ell) : \beta \in \{0, 0.02, 0.04, \ldots, 2\}\Big\}.$$

and

$$\beta_{\text{Re-Std}}(\boldsymbol{X}) = \arg\max \Big\{\frac{1}{5}\sum_{\ell=1}^{5}\psi(y_{\text{Re-Std}}(\boldsymbol{X}_{\text{-}\ell}, \beta), \boldsymbol{X}_\ell) : \beta \in \{0, 0.02, 0.04, \ldots, 2\}\Big\}.$$

In almost all of the tested instances, the β value is strictly below 2.

- **Order statistics (OrS):** The order statistics of $\boldsymbol{X}$ is

$$y_{\text{OrS}}(\boldsymbol{X}) = \sum_{i=1}^{n} w_i X_{[i]},$$

where the weight vector $\boldsymbol{w} = (w_1, w_2, \ldots, w_n)$ is chosen through a robust optimization proposed by Besbes and Mouchtaki (2023) to maximize the following competitive ratio

$$\boldsymbol{w} = \arg\max \Big\{\max_{\mu \in [0,1]} \sum_{i=1}^{n} w_i \frac{(1 - B_{i,n}(1 - \mu)(c/s - \mu) + (1 - c/s)\mu}{\min\{c/s(1 - \mu), (1 - c/s)\mu\}} : \boldsymbol{w} \in \mathbb{R}_+, \sum_{i=1}^{n} w_i = 1\Big\}.$$

- **ODA based on empirical distribution (SC1-Emp & H1-Emp):** Let $\boldsymbol{Z} = \boldsymbol{X}/\hat{\mu}_{\boldsymbol{X}}$ be the normalized sample and $\boldsymbol{Z}_{[]}$ be the ascending sequence of Z. Let $m_0 = 500$ and $m = 2n \times 10^4$.

We sample from the empirical density function with Gaussian kernel:

$$\tilde{f}_{Z|\boldsymbol{Z}}(\xi) = \frac{1}{n\lambda}\sum_{i=1}^{n}\frac{1}{\sqrt{2\pi}}e^{\frac{(\xi - Z_i)^2}{2\lambda^2}}.$$

Then, the smoothed empirical distribution function is

$$\tilde{F}_{Z|\boldsymbol{Z}}(z) = \int_{\xi=0}^{z} \tilde{f}_{Z|\boldsymbol{Z}}(\xi)\mathrm{d}\xi. \tag{46}$$

Using $\tilde{f}_{Z|\boldsymbol{Z}}$ to replace f_Z in Theorem 1, the ODA solutions $y_{\text{H1-Emp}}$ and $y_{\text{SC1-Emp}}$ can be derived. We calculate $y_{\text{H}\iota\text{-Emp}}$ and $y_{\text{SC}\iota\text{-Emp}}$ by simulation. From the same distribution, we also sample $\tilde{Z}_i = \tilde{F}^{\text{inv}}_{Z|\boldsymbol{Z}}(U), i \in \{1, 2, \ldots, m_0\}$, with U following uniform distribution over $[0, 1]$. The simulated sample consists of $\mathbf{Z}^{()} = (\boldsymbol{Z}^{(j)} \in \mathbb{R}^n : j = 1, 2, \ldots, m)$, where $\boldsymbol{Z}^{(j)} = (Z_1^{(j)}, Z_2^{(j)}, \ldots, Z_n^{(j)})$. Then,

$$y_{\text{SC1-Emp}}(\boldsymbol{X}) = \hat{\mu}_{\boldsymbol{X}} \cdot \arg\max \Big\{ \frac{1}{m_0^2} \sum_{j=1}^{m_0} \sum_{\ell=1}^{m_0} \psi(\hat{\mu}_{\boldsymbol{Z}^{(j)}}\gamma, \tilde{Z}_\ell) : \gamma \geq 0 \Big\}.$$

The computation for $y_{\text{H1-Emp}}$ is similar, with one key difference: we require each sample $\boldsymbol{Z}^{(j)}$ to satisfy the condition $d(\boldsymbol{Z}, \boldsymbol{Z}^{(j)}/\hat{\mu}_{\boldsymbol{Z}^{(j)}} < \eta$. To achieve this, any sample that fails the condition is discarded, and we resample. This process is repeated until m_0 valid samples have been obtained for the computation.

Transfer-Learning Solutions. Given i.i.d. observations $\boldsymbol{X} = \{X_1, X_2, \ldots, X_n\}$ from a focal system and i.i.d. observations $\boldsymbol{X}^o = \{X_1^o, X_2^o, \ldots, X_{n^o}^o\}$ from an old system. First calculate the oracle solution $y^o(\boldsymbol{X}^o)$, o $\in$ {SAA, Re-Var, Re-Std, OrS, SC1-Emp, H1-Emp}, according to each directly trained solution define above, then calculate the corresponding transfer-learning solution by (15), i.e.,

$$y_{\text{TrL:o}}(\boldsymbol{X}) = \frac{\hat{\mu}_{\boldsymbol{X}}}{\hat{\mu}_{\boldsymbol{X}^o}} y^o(\boldsymbol{X}^o).$$

Cross-Learning Solutions. Given i.i.d. observations $\boldsymbol{X} = \{X_1, X_2, \ldots, X_n\}$ from a focal system and i.i.d. observations $\boldsymbol{X}^o = \{X_1^o, X_2^o, \ldots, X_{n^o}^o\}$ from an old system. Let $\boldsymbol{Z}^o = \boldsymbol{X}^o/\hat{\mu}_{\boldsymbol{X}^o} = (Z_1^o, Z_2^o, \ldots, Z_{n^o}^o)$ be the normalized sample of old system and $\boldsymbol{Z}_B = \boldsymbol{X}/\hat{\mu}_{\boldsymbol{X}}$ be the base point of the focal system sample. Let $m_0 = 500$, $m = 1 \times 10^{5+n/5}$.

- **Scaled operational statistic (CrL:SC1):** Randomly sample $Z_i^{(j)}$ with replacement from $\boldsymbol{Z}^o$ such that $\mathbf{Z}^{()} = (\boldsymbol{Z}^{(j)} \in \mathbb{R}^n : j = 1, 2, \ldots, m_0)$, where $\boldsymbol{Z}^{(j)} = (Z_1^{(j)}, Z_2^{(j)}, \ldots, Z_n^{(j)})$. Then,

$$y_{\text{CrL:SC1-Emp}}(\boldsymbol{X}) = \hat{\mu}_{\boldsymbol{X}} \cdot \arg\max \Big\{ \frac{1}{m_0 n^o} \sum_{j=1}^{m_0} \sum_{\ell=1}^{n^o} \psi(\hat{\mu}_{\boldsymbol{Z}^{(j)}}\gamma, Z_\ell^o) : \gamma \geq 0 \Big\}.$$

- **Targeted boosting (CrL:c):** First construct $\tilde{F}_{Z|\boldsymbol{Z}^o}$ according to (46), then randomly sample from $\tilde{F}_{Z|\boldsymbol{Z}^o}(\cdot)$ to form $\mathbf{Z}^{()} = (\boldsymbol{Z}^{(j)} \in \mathbb{R}^n : j = 1, 2, \ldots, m)$, where $\boldsymbol{Z}^{(j)} = (Z_1^{(j)}, Z_2^{(j)}, \ldots, Z_n^{(j)})$ satisfies $d(\boldsymbol{Z}, \boldsymbol{Z}^{(j)}/\hat{\mu}_{\boldsymbol{Z}^{(j)}}) \leq \eta$. Then, the targeted-boosting solution is

$$y_{\text{CrL:c}}(\boldsymbol{X}) = \hat{\mu}_{\boldsymbol{X}} \cdot \arg\max \Big\{ \frac{1}{m_0 n^o} \sum_{j \in \mathcal{M}(\boldsymbol{Z}, \mathbf{Z}^{(j)})} \sum_{\ell=1}^{n^o} \psi(y_{\text{c}}(\boldsymbol{Z}^{(j)})\gamma, Z_\ell^o) : \gamma \geq 0 \Big\}$$

  for c $\in$ {SAA, Re-Var, Re-Std, OrS, SC1-Emp, H1-Emp}. We use $m_0 = 500$ and $m = 5,000$ in the reported experiments.

- **Homogeneous operational statistic (CrL:H1):**

$$y_{\text{CrL:H1}}(\boldsymbol{X}) = \hat{\mu}_{\boldsymbol{X}} \cdot \arg\max \Big\{ \frac{1}{m_0 n^o} \sum_{j \in \mathcal{M}(\boldsymbol{Z}, \mathbf{Z}^{(j)})} \sum_{\ell=1}^{n^o} \psi(\hat{\mu}_{\boldsymbol{Z}^{(j)}}\gamma, Z_\ell^o) : \gamma \geq 0 \Big\}.$$

Co-Learning Solutions. Given i.i.d. observations $\boldsymbol{X}_j = \{X_{j:1}, X_{j:2}, \ldots, X_{j:n}\}$ from each system $j = 1, 2, \ldots, k$ with $\mathbf{X} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_k\}$. Normalize the sample to $\boldsymbol{Z}_j = \boldsymbol{X}_j / \hat{\mu}_{\boldsymbol{X}_j}$ and denote $\mathbf{Z} = \{\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_k\}$.

First calculate $y_{\text{c}}(\boldsymbol{X}_{j:\text{-}i})$, $\forall j \in \{1, 2, \ldots, k\}$ and $i \in \{1, 2, \ldots, n\}$.

- **Constant boosting (CoL-CB:c):** For some system j,

$$y_{\text{CoL-CB:c}}(\boldsymbol{X}_j) = y^c(\boldsymbol{X}_j) \cdot \arg\max \Big\{ \frac{1}{kn} \sum_{j=1}^{k} \sum_{\ell=1}^{n} \psi(y^c(\boldsymbol{X}_{j:\text{-}\ell})\gamma, X_\ell) : \gamma \geq 0 \Big\}.$$

- **Targeted boosting (CoL-TB:c):** For some system j, first we use K-means clustering approach with n_K number of clusters to construct a partition $\{\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_{n_K}\}$ of $\{\boldsymbol{Z}_{j:\text{-}\ell} : j = 1, 2, \ldots, k, \ell = 1, 2, \ldots, n\}$. Then, if $\boldsymbol{Z}_{j:\text{-}1} \in \mathcal{M}_q$ for some $q \in \{1, 2, \ldots, n_K\}$,

$$y_{\text{CoL-TB:c}}(\boldsymbol{X}_j; n_K) = y^c(\boldsymbol{X}_j) \cdot \arg\max \Big\{ \frac{1}{|\mathcal{M}_q| n} \sum_{(j,i) \in \mathcal{M}_q} \sum_{\ell=1}^{n} \psi(y^c(\boldsymbol{X}_{j:\text{-}\ell})\gamma, X_\ell) : \gamma \geq 0 \Big\}.$$

  To determine the value of n_K, we use the following cross validation

$$y_{\text{CoL-TB:c}}(\boldsymbol{X}_j) = \arg\max \Big\{ \frac{1}{kn} \sum_{j=1}^{k} \sum_{\ell=1}^{n} \psi(y_{\text{CoL-TB:c}}(\boldsymbol{X}_j; n_K), X_\ell) : n_K = 1, 2, 3, 4, 5 \Big\}.$$

Table 3: Performance of Transfer/Cross-Learning Solutions Against Directly Trained Solutions 2

		Directly Trained Solutions				Transfer-Learning Solutions				Cross-Learning Solutions			
			Ave	Stdev	Min		Ave	Stdev	Min		Ave	Stdev	Min
$n=5$	Benchmark	Ψ_{H1}	0.624	0.04	0.1								
	$n^o=200$	Ψ_{SAA}	0.298	0.50	-5.4	$\Psi_{\text{TrL:SAA}}$	0.575	0.09	-0.4	$\Psi_{\text{CrL:SAA}}$	0.601	0.07	-0.2
		$\Psi_{\text{Re-Var}}$	0.010	0.97	-6.0	$\Psi_{\text{TrL:Re-Var}}$	0.574	0.08	-0.5	$\Psi_{\text{CrL:Re-Var}}$	0.603	0.07	0.0
		$\Psi_{\text{Re-Std}}$	0.006	0.97	-6.0	$\Psi_{\text{TrL:Re-Std}}$	0.575	0.08	-0.5	$\Psi_{\text{CrL:Re-Std}}$	0.601	0.07	-0.2
		Ψ_{OrS}	0.018	0.96	-5.9	$\Psi_{\text{TrL:OrS}}$	0.575	0.09	-0.5	$\Psi_{\text{CrL:OrS}}$	0.601	0.07	0.0
		$\Psi_{\text{SC1-Emp}}$	0.248	0.61	-5.7	$\Psi_{\text{TrL:SC1-Emp}}$	0.574	0.09	-0.5	$\Psi_{\text{CrL:SC1-Emp}}$	0.601	0.07	-0.1
		$\Psi_{\text{H1-Emp}}$	0.296	0.51	-5.4	$\Psi_{\text{TrL:H1-Emp}}$	0.574	0.09	-0.5	$\Psi_{\text{CrL:H1-Emp}}$	0.601	0.07	0.0
										$\Psi_{\text{CrL:SC1}}$	0.581	0.07	-0.1
										$\Psi_{\text{CrL:H1}}$	0.610	0.04	0.0
	$n^o=1000$	Ψ_{SAA}	0.298	0.50	-5.4	$\Psi_{\text{TrL:SAA}}$	0.584	0.07	0.0	$\Psi_{\text{CrL:SAA}}$	0.611	0.07	0.0
		$\Psi_{\text{Re-Var}}$	0.010	0.97	-6.0	$\Psi_{\text{TrL:Re-Var}}$	0.583	0.07	0.0	$\Psi_{\text{CrL:Re-Var}}$	0.612	0.07	0.0
		$\Psi_{\text{Re-Std}}$	0.006	0.97	-6.0	$\Psi_{\text{TrL:Re-Std}}$	0.585	0.07	0.0	$\Psi_{\text{CrL:Re-Std}}$	0.610	0.07	0.0
		Ψ_{OrS}	0.018	0.96	-5.9	$\Psi_{\text{TrL:OrS}}$	0.584	0.07	0.0	$\Psi_{\text{CrL:OrS}}$	0.610	0.07	0.0
		$\Psi_{\text{SC1-Emp}}$	0.248	0.61	-5.7	$\Psi_{\text{TrL:SC1-Emp}}$	0.584	0.07	0.0	$\Psi_{\text{CrL:SC1-Emp}}$	0.611	0.07	0.0
		$\Psi_{\text{H1-Emp}}$	0.296	0.51	-5.4	$\Psi_{\text{TrL:H1-Emp}}$	0.584	0.07	0.0	$\Psi_{\text{CrL:H1-Emp}}$	0.610	0.07	0.0
										$\Psi_{\text{CrL:SC1}}$	0.587	0.07	0.0
										$\Psi_{\text{CrL:H1}}$	0.621	0.04	0.0
$n=10$	Benchmark	Ψ_{H1}	0.635	0.01	0.2								
	$n^o=200$	Ψ_{SAA}	0.353	0.42	-5.1	$\Psi_{\text{TrL:SAA}}$	0.602	0.05	-0.2	$\Psi_{\text{CrL:SAA}}$	0.607	0.06	-0.2
		$\Psi_{\text{Re-Var}}$	0.319	0.46	-5.1	$\Psi_{\text{TrL:Re-Var}}$	0.598	0.05	-0.1	$\Psi_{\text{CrL:Re-Var}}$	0.609	0.05	-0.2
		$\Psi_{\text{Re-Std}}$	0.322	0.47	-5.1	$\Psi_{\text{TrL:Re-Std}}$	0.600	0.05	-0.1	$\Psi_{\text{CrL:Re-Std}}$	0.607	0.06	-0.2
		Ψ_{OrS}	0.318	0.52	-5.1	$\Psi_{\text{TrL:OrS}}$	0.601	0.05	-0.2	$\Psi_{\text{CrL:OrS}}$	0.619	0.03	0.0
		$\Psi_{\text{SC1-Emp}}$	0.364	0.40	-4.5	$\Psi_{\text{TrL:SC1-Emp}}$	0.601	0.05	-0.2	$\Psi_{\text{CrL:SC1-Emp}}$	0.612	0.05	0.0
		$\Psi_{\text{H1-Emp}}$	0.356	0.41	-5.0	$\Psi_{\text{TrL:H1-Emp}}$	0.601	0.05	-0.2	$\Psi_{\text{CrL:H1-Emp}}$	0.608	0.06	0.0
										$\Psi_{\text{CrL:SC1}}$	0.604	0.05	-0.1
										$\Psi_{\text{CrL:H1}}$	0.622	0.02	0.3
	$n^o=1000$	Ψ_{SAA}	0.353	0.42	-5.1	$\Psi_{\text{TrL:SAA}}$	0.609	0.04	0.1	$\Psi_{\text{CrL:SAA}}$	0.616	0.06	0.0
		$\Psi_{\text{Re-Var}}$	0.319	0.46	-5.1	$\Psi_{\text{TrL:Re-Var}}$	0.609	0.04	0.1	$\Psi_{\text{CrL:Re-Var}}$	0.619	0.05	0.0
		$\Psi_{\text{Re-Std}}$	0.322	0.47	-5.1	$\Psi_{\text{TrL:Re-Std}}$	0.610	0.04	0.1	$\Psi_{\text{CrL:Re-Std}}$	0.616	0.06	0.0
		Ψ_{OrS}	0.318	0.52	-5.1	$\Psi_{\text{TrL:OrS}}$	0.609	0.04	0.1	$\Psi_{\text{CrL:OrS}}$	0.628	0.03	0.0
		$\Psi_{\text{SC1-Emp}}$	0.364	0.40	-4.5	$\Psi_{\text{TrL:SC1-Emp}}$	0.610	0.04	0.1	$\Psi_{\text{CrL:SC1-Emp}}$	0.621	0.05	0.0
		$\Psi_{\text{H1-Emp}}$	0.356	0.41	-5.0	$\Psi_{\text{TrL:H1-Emp}}$	0.609	0.04	0.1	$\Psi_{\text{CrL:H1-Emp}}$	0.616	0.06	0.0
										$\Psi_{\text{CrL:SC1}}$	0.611	0.04	0.2
										$\Psi_{\text{CrL:H1}}$	0.632	0.01	0.3
$n=15$	Benchmark	Ψ_{H1}	0.637	0.00	0.4								
	$n^o=200$	Ψ_{SAA}	0.444	0.29	-4.0	$\Psi_{\text{TrL:SAA}}$	0.611	0.04	0.0	$\Psi_{\text{CrL:SAA}}$	0.620	0.03	0.0
		$\Psi_{\text{Re-Var}}$	0.389	0.32	-4.4	$\Psi_{\text{TrL:Re-Var}}$	0.606	0.04	0.0	$\Psi_{\text{CrL:Re-Var}}$	0.616	0.03	0.0
		$\Psi_{\text{Re-Std}}$	0.404	0.32	-4.4	$\Psi_{\text{TrL:Re-Std}}$	0.609	0.04	0.0	$\Psi_{\text{CrL:Re-Std}}$	0.612	0.05	0.0
		Ψ_{OrS}	0.393	0.39	-4.3	$\Psi_{\text{TrL:OrS}}$	0.610	0.04	0.0	$\Psi_{\text{CrL:OrS}}$	0.619	0.03	0.0
		$\Psi_{\text{SC1-Emp}}$	0.439	0.30	-3.9	$\Psi_{\text{TrL:SC1-Emp}}$	0.610	0.04	0.0	$\Psi_{\text{CrL:SC1-Emp}}$	0.619	0.03	0.0
		$\Psi_{\text{H1-Emp}}$	0.444	0.29	-4.0	$\Psi_{\text{TrL:H1-Emp}}$	0.610	0.04	0.0	$\Psi_{\text{CrL:H1-Emp}}$	0.619	0.03	0.0
										$\Psi_{\text{CrL:SC1}}$	0.611	0.04	0.0
										$\Psi_{\text{CrL:H1}}$	0.623	0.02	0.4
	$n^o=1000$	Ψ_{SAA}	0.444	0.29	-4.0	$\Psi_{\text{TrL:SAA}}$	0.618	0.03	0.3	$\Psi_{\text{CrL:SAA}}$	0.629	0.02	0.0
		$\Psi_{\text{Re-Var}}$	0.389	0.32	-4.3	$\Psi_{\text{TrL:Re-Var}}$	0.617	0.03	0.3	$\Psi_{\text{CrL:Re-Var}}$	0.625	0.03	0.0
		$\Psi_{\text{Re-Std}}$	0.403	0.32	-4.3	$\Psi_{\text{TrL:Re-Std}}$	0.618	0.03	0.3	$\Psi_{\text{CrL:Re-Std}}$	0.621	0.05	0.0
		Ψ_{OrS}	0.393	0.39	-4.3	$\Psi_{\text{TrL:OrS}}$	0.618	0.03	0.3	$\Psi_{\text{CrL:OrS}}$	0.629	0.02	0.0
		$\Psi_{\text{SC1-Emp}}$	0.439	0.30	-3.9	$\Psi_{\text{TrL:SC1-Emp}}$	0.618	0.03	0.3	$\Psi_{\text{CrL:SC1-Emp}}$	0.629	0.03	0.0
		$\Psi_{\text{H1-Emp}}$	0.443	0.29	-4.0	$\Psi_{\text{TrL:H1-Emp}}$	0.618	0.03	0.3	$\Psi_{\text{CrL:H1-Emp}}$	0.628	0.03	0.0
										$\Psi_{\text{CrL:SC1}}$	0.619	0.03	0.3
										$\Psi_{\text{CrL:H1}}$	0.633	0.01	0.4

Notes. $Z \sim 2 \times Beta(0.5, 0.5)$ (implying $\mathsf{Cv}[Z] = 0.71$), $\mathsf{E}[X] = 1.5$, $\mathsf{E}[X^o] = 15$, and $n = \{5, 10, 15\}$. The profit function is $\psi(y, x) = 10\min\{y, x\} - 7y$. The subscripts TrL:$j$ and CrL:j stand for, respectively, transfer-learning and cross-learning solutions, where j = SAA, Re-Var, Re-Std, OrS, SC1-Emp, or H1-Emp stands for the sample-average approximation, regularization with variance, regularization with standard deviation, order statistics, scaled operational statistic with empirical distribution, or homogeneous operational statistic with empirical distribution, respectively. The result is generated by computing the average, standard deviation, and minimum of the actual profit Ψ_j out of 100,000 randomly generated instances.

Table 4: Performance of Co-Learning Solutions Against Directly Trained Solutions 2

		Directly Trained Solutions				Constant-Boosting Solutions				Targeted-Boosting Solutions			
			Ave	Stdev	Min		Ave	Stdev	Min		Ave	Stdev	Min
$n = 5$	Benchmark	Ψ_{H1}	0.624	0.14	0.3								
	$k = 20$	Ψ_{SAA}	0.296	0.17	-1.3	$\Psi_{\text{CoL-CB:SAA}}$	0.337	0.12	-0.7	$\Psi_{\text{CoL-TB:SAA}}$	0.348	0.12	-0.7
		$\Psi_{\text{Re-Var}}$	0.017	0.31	-2.8	$\Psi_{\text{CoL-CB:Re-Var}}$	0.349	0.12	-0.4	$\Psi_{\text{CoL-TB:Re-Var}}$	0.359	0.12	-0.4
		$\Psi_{\text{Re-Std}}$	0.000	0.31	-2.8	$\Psi_{\text{CoL-CB:Re-Std}}$	0.343	0.12	-0.4	$\Psi_{\text{CoL-TB:Re-Std}}$	0.353	0.12	-0.4
		Ψ_{OrS}	0.015	0.30	-2.8	$\Psi_{\text{CoL-CB:OrS}}$	0.362	0.12	-0.6	$\Psi_{\text{CoL-TB:OrS}}$	0.373	0.12	-0.6
		$\Psi_{\text{SC1-Emp}}$	0.244	0.20	-1.5	$\Psi_{\text{CoL-CB:SC1-Emp}}$	0.339	0.12	-0.7	$\Psi_{\text{CoL-TB:SC1-Emp}}$	0.350	0.12	-0.7
		$\Psi_{\text{H1-Emp}}$	0.293	0.18	-1.3	$\Psi_{\text{CoL-CB:H1-Emp}}$	0.347	0.12	-0.6	$\Psi_{\text{CoL-TB:H1-Emp}}$	0.357	0.12	-0.6
						$\Psi_{\text{CoL-CB:SC1}}$	0.538	0.14	0.0	$\Psi_{\text{CoL-TB:H1}}$	0.531	0.14	0.0
	Benchmark	Ψ_{H1}	0.624	0.03	0.5								
	$k = 500$	Ψ_{SAA}	0.297	0.03	0.1	$\Psi_{\text{CoL-CB:SAA}}$	0.378	0.02	0.3	$\Psi_{\text{CoL-TB:SAA}}$	0.391	0.03	0.3
		$\Psi_{\text{Re-Var}}$	0.018	0.06	-0.3	$\Psi_{\text{CoL-CB:Re-Var}}$	0.371	0.03	0.3	$\Psi_{\text{CoL-TB:Re-Var}}$	0.388	0.03	0.3
		$\Psi_{\text{Re-Std}}$	0.002	0.06	-0.3	$\Psi_{\text{CoL-CB:Re-Std}}$	0.366	0.03	0.3	$\Psi_{\text{CoL-TB:Re-Std}}$	0.383	0.03	0.3
		Ψ_{OrS}	0.016	0.06	-0.3	$\Psi_{\text{CoL-CB:OrS}}$	0.393	0.03	0.3	$\Psi_{\text{CoL-TB:OrS}}$	0.410	0.03	0.3
		$\Psi_{\text{SC1-Emp}}$	0.246	0.04	0.0	$\Psi_{\text{CoL-CB:SC1-Emp}}$	0.380	0.02	0.3	$\Psi_{\text{CoL-TB:SC1-Emp}}$	0.394	0.03	0.3
		$\Psi_{\text{H1-Emp}}$	0.295	0.04	0.1	$\Psi_{\text{CoL-CB:H1-Emp}}$	0.387	0.02	0.3	$\Psi_{\text{CoL-TB:H1-Emp}}$	0.402	0.03	0.3
						$\Psi_{\text{CoL-CB:SC1}}$	0.563	0.03	0.4	$\Psi_{\text{CoL-TB:H1}}$	0.564	0.03	0.4
$n = 10$	Benchmark	Ψ_{H1}	0.635	0.15	0.2								
	$k = 20$	Ψ_{SAA}	0.351	0.15	-1.0	$\Psi_{\text{CoL-CB:SAA}}$	0.394	0.12	-0.8	$\Psi_{\text{CoL-TB:SAA}}$	0.411	0.12	-0.8
		$\Psi_{\text{Re-Var}}$	0.350	0.15	-1.0	$\Psi_{\text{CoL-CB:Re-Var}}$	0.378	0.11	-0.6	$\Psi_{\text{CoL-TB:Re-Var}}$	0.392	0.12	-0.2
		$\Psi_{\text{Re-Std}}$	0.316	0.15	-1.0	$\Psi_{\text{CoL-CB:Re-Std}}$	0.359	0.11	-0.3	$\Psi_{\text{CoL-TB:Re-Std}}$	0.373	0.12	-0.2
		Ψ_{OrS}	0.316	0.18	-1.2	$\Psi_{\text{CoL-CB:OrS}}$	0.449	0.12	-0.2	$\Psi_{\text{CoL-TB:OrS}}$	0.462	0.12	-0.1
		$\Psi_{\text{SC1-Emp}}$	0.362	0.15	-1.1	$\Psi_{\text{CoL-CB:SC1-Emp}}$	0.410	0.12	-0.7	$\Psi_{\text{CoL-TB:SC1-Emp}}$	0.425	0.12	-0.3
		$\Psi_{\text{H1-Emp}}$	0.354	0.15	-1.0	$\Psi_{\text{CoL-CB:H1-Emp}}$	0.399	0.12	-0.7	$\Psi_{\text{CoL-TB:H1-Emp}}$	0.414	0.12	-0.7
						$\Psi_{\text{CoL-CB:SC1}}$	0.590	0.14	0.1	$\Psi_{\text{CoL-TB:H1}}$	0.584	0.14	0.1
	Benchmark	Ψ_{H1}	0.635	0.03	0.6								
	$k = 500$	Ψ_{SAA}	0.351	0.03	0.2	$\Psi_{\text{CoL-CB:SAA}}$	0.418	0.02	0.3	$\Psi_{\text{CoL-TB:SAA}}$	0.435	0.03	0.3
		$\Psi_{\text{Re-Var}}$	0.351	0.03	0.2	$\Psi_{\text{CoL-CB:Re-Var}}$	0.403	0.02	0.3	$\Psi_{\text{CoL-TB:Re-Var}}$	0.420	0.03	0.3
		$\Psi_{\text{Re-Std}}$	0.317	0.03	0.2	$\Psi_{\text{CoL-CB:Re-Std}}$	0.380	0.02	0.3	$\Psi_{\text{CoL-TB:Re-Std}}$	0.398	0.03	0.3
		Ψ_{OrS}	0.317	0.04	0.2	$\Psi_{\text{CoL-CB:OrS}}$	0.471	0.02	0.4	$\Psi_{\text{CoL-TB:OrS}}$	0.486	0.03	0.4
		$\Psi_{\text{SC1-Emp}}$	0.362	0.03	0.2	$\Psi_{\text{CoL-CB:SC1-Emp}}$	0.433	0.02	0.3	$\Psi_{\text{CoL-TB:SC1-Emp}}$	0.449	0.03	0.3
		$\Psi_{\text{H1-Emp}}$	0.355	0.03	0.2	$\Psi_{\text{CoL-CB:H1-Emp}}$	0.422	0.02	0.3	$\Psi_{\text{CoL-TB:H1-Emp}}$	0.439	0.03	0.3
						$\Psi_{\text{CoL-CB:SC1}}$	0.606	0.03	0.5	$\Psi_{\text{CoL-TB:H1}}$	0.609	0.03	0.5
$n = 15$	Benchmark	Ψ_{H1}	0.637	0.14	0.2								
	$k = 20$	Ψ_{SAA}	0.443	0.13	-0.8	$\Psi_{\text{CoL-CB:SAA}}$	0.464	0.12	-0.2	$\Psi_{\text{CoL-TB:SAA}}$	0.478	0.12	-0.2
		$\Psi_{\text{Re-Var}}$	0.380	0.14	-1.0	$\Psi_{\text{CoL-CB:Re-Var}}$	0.418	0.12	-0.4	$\Psi_{\text{CoL-TB:Re-Var}}$	0.435	0.12	-0.4
		$\Psi_{\text{Re-Std}}$	0.374	0.15	-1.0	$\Psi_{\text{CoL-CB:Re-Std}}$	0.426	0.12	-0.4	$\Psi_{\text{CoL-TB:Re-Std}}$	0.441	0.12	-0.4
		Ψ_{OrS}	0.393	0.15	-0.9	$\Psi_{\text{CoL-CB:OrS}}$	0.474	0.12	-0.4	$\Psi_{\text{CoL-TB:OrS}}$	0.488	0.12	-0.4
		$\Psi_{\text{SC1-Emp}}$	0.438	0.14	-0.9	$\Psi_{\text{CoL-CB:SC1-Emp}}$	0.468	0.12	-0.3	$\Psi_{\text{CoL-TB:SC1-Emp}}$	0.482	0.12	-0.3
		$\Psi_{\text{H1-Emp}}$	0.443	0.13	-0.8	$\Psi_{\text{CoL-CB:H1-Emp}}$	0.465	0.12	-0.3	$\Psi_{\text{CoL-TB:H1-Emp}}$	0.479	0.12	-0.3
						$\Psi_{\text{CoL-CB:SC1}}$	0.606	0.14	0.2	$\Psi_{\text{CoL-TB:H1}}$	0.601	0.14	0.2
	Benchmark	Ψ_{H1}	0.637	0.03	0.6								
	$k = 500$	Ψ_{SAA}	0.443	0.03	0.3	$\Psi_{\text{CoL-CB:SAA}}$	0.478	0.02	0.4	$\Psi_{\text{CoL-TB:SAA}}$	0.494	0.03	0.4
		$\Psi_{\text{Re-Var}}$	0.380	0.03	0.3	$\Psi_{\text{CoL-CB:Re-Var}}$	0.431	0.02	0.3	$\Psi_{\text{CoL-TB:Re-Var}}$	0.446	0.03	0.4
		$\Psi_{\text{Re-Std}}$	0.373	0.03	0.3	$\Psi_{\text{CoL-CB:Re-Std}}$	0.442	0.02	0.4	$\Psi_{\text{CoL-TB:Re-Std}}$	0.458	0.03	0.4
		Ψ_{OrS}	0.393	0.03	0.3	$\Psi_{\text{CoL-CB:OrS}}$	0.488	0.02	0.4	$\Psi_{\text{CoL-TB:OrS}}$	0.506	0.03	0.4
		$\Psi_{\text{SC1-Emp}}$	0.438	0.03	0.3	$\Psi_{\text{CoL-CB:SC1-Emp}}$	0.480	0.02	0.4	$\Psi_{\text{CoL-TB:SC1-Emp}}$	0.497	0.03	0.4
		$\Psi_{\text{H1-Emp}}$	0.443	0.03	0.3	$\Psi_{\text{CoL-CB:H1-Emp}}$	0.478	0.02	0.4	$\Psi_{\text{CoL-TB:H1-Emp}}$	0.495	0.03	0.4
						$\Psi_{\text{CoL-CB:SC1}}$	0.618	0.03	0.5	$\Psi_{\text{CoL-TB:H1}}$	0.621	0.03	0.5

Notes. $Z \sim 2 \times Beta(0.5, 0.5)$ (implying $\mathsf{Cv}[Z] = 0.71$), $\Theta \sim 1.5 \times Exp(1)$, and $n = \{5, 10, 15\}$. In each system, the profit function is $\psi(y, x) = 10 \min\{y, x\} - 7y$. The average profits over k systems $\Psi_{\text{CoL-TB:}j}$ and $\Psi_{\text{CoL-CB:}j}$ are derived by, respectively, constant-boosting and targeted-boosting candidate solution j, where $j =$ SAA, Re-Var, Re-Std, OrS, SC1-Emp, or H1-Emp stands for the sample-average approximation, regularization with variance, regularization with standard deviation, order statistic, scaled operational statistic with empirical distribution, or homogeneous operational statistic with empirical distribution, respectively. The result is generated by computing the average, standard deviation, and minimum of Ψ_{j} out of $30,000$ randomly generated problem instances.

D.2 Real Data

Next, we present a case study using the data analyzed in Section C for SKU 068f4481b3. Initially, we randomly generate a sample of size n from the 31 observed samples for each distribution center to serve as training data. We then compute the co-learning solutions for the 51 (9) distribution centers in Group 1 (Group 2) depicted in Figure 3. To assess the performance of each solution, we use the remaining $(31 - n)$ samples from each distribution center as validation data, allowing us to approximate the out-of-sample performance of the solutions derived from the training data. This process is repeated 10,000 times, and the average profit results for the 51 (9) distribution centers in Group 1 (Group 2) are presented in Table 5 (Table 6).

Table 5: Case Study 1: Performance of Co-Learning Solutions Against Directly Trained Solutions

		Directly Trained Solutions				Constant-Boosting Solutions				Targeted-Boosting Solutions			
			Ave	Stdev	Min		Ave	Stdev	Min		Ave	Stdev	Min
$k = 51$	$n = 5$	Ψ_{SAA}	13.928	1.31	3.2	$\Psi_{\text{CoL-CB:SAA}}$	14.408	0.88	7.6	$\Psi_{\text{CoL-TB:SAA}}$	14.411	0.84	7.7
		$\Psi_{\text{Re-Var}}$	10.129	2.95	-8.3	$\Psi_{\text{CoL-CB:Re-Var}}$	12.480	1.05	8.5	$\Psi_{\text{CoL-TB:Re-Var}}$	12.422	1.08	7.1
		$\Psi_{\text{Re-Std}}$	10.129	2.95	-8.3	$\Psi_{\text{CoL-CB:Re-Std}}$	13.204	0.96	9.5	$\Psi_{\text{CoL-TB:Re-Std}}$	13.141	0.98	7.5
		Ψ_{OrS}	12.002	2.25	-3.0	$\Psi_{\text{CoL-CB:OrS}}$	13.814	0.83	10.2	$\Psi_{\text{CoL-TB:OrS}}$	13.802	0.82	10.2
		$\Psi_{\text{SC1-Emp}}$	13.601	1.62	-0.5	$\Psi_{\text{CoL-CB:SC1-Emp}}$	14.433	0.87	7.3	$\Psi_{\text{CoL-TB:SC1-Emp}}$	14.441	0.84	7.8
		$\Psi_{\text{H1-Emp}}$	13.910	1.45	1.4	$\Psi_{\text{CoL-CB:H1-Emp}}$	14.455	0.82	7.8	$\Psi_{\text{CoL-TB:H1-Emp}}$	14.449	0.79	7.8
						$\Psi_{\text{CoL-CB:SC1}}$	15.331	0.62	12.1	$\Psi_{\text{CoL-TB:H1}}$	15.270	0.64	12.1
	$n = 10$	Ψ_{SAA}	15.516	0.99	7.4	$\Psi_{\text{CoL-CB:SAA}}$	15.615	0.98	7.0	$\Psi_{\text{CoL-TB:SAA}}$	15.596	0.92	7.5
		$\Psi_{\text{Re-Var}}$	14.702	1.25	5.8	$\Psi_{\text{CoL-CB:Re-Var}}$	15.019	0.97	6.8	$\Psi_{\text{CoL-TB:Re-Var}}$	14.909	0.91	7.3
		$\Psi_{\text{Re-Std}}$	15.307	1.32	6.1	$\Psi_{\text{CoL-CB:Re-Std}}$	15.309	0.92	7.6	$\Psi_{\text{CoL-TB:Re-Std}}$	15.237	0.84	8.2
		Ψ_{OrS}	15.181	1.41	5.7	$\Psi_{\text{CoL-CB:OrS}}$	15.698	0.83	8.3	$\Psi_{\text{CoL-TB:OrS}}$	15.712	0.78	9.1
		$\Psi_{\text{SC1-Emp}}$	15.366	1.20	6.0	$\Psi_{\text{CoL-CB:SC1-Emp}}$	15.740	0.81	8.6	$\Psi_{\text{CoL-TB:SC1-Emp}}$	15.742	0.77	9.2
		$\Psi_{\text{H1-Emp}}$	15.482	1.10	7.1	$\Psi_{\text{CoL-CB:H1-Emp}}$	15.646	0.86	8.2	$\Psi_{\text{CoL-TB:H1-Emp}}$	15.628	0.82	8.5
						$\Psi_{\text{CoL-CB:SC1}}$	16.419	0.49	13.9	$\Psi_{\text{CoL-TB:H1}}$	16.391	0.49	13.0
	$n = 15$	Ψ_{SAA}	16.278	0.92	7.9	$\Psi_{\text{CoL-CB:SAA}}$	16.382	0.85	8.6	$\Psi_{\text{CoL-TB:SAA}}$	16.377	0.82	8.6
		$\Psi_{\text{Re-Var}}$	15.641	1.04	6.2	$\Psi_{\text{CoL-CB:Re-Var}}$	15.762	1.10	7.7	$\Psi_{\text{CoL-TB:Re-Var}}$	15.749	1.09	7.7
		$\Psi_{\text{Re-Std}}$	16.177	0.91	8.2	$\Psi_{\text{CoL-CB:Re-Std}}$	16.189	1.08	8.5	$\Psi_{\text{CoL-TB:Re-Std}}$	16.176	1.07	8.5
		Ψ_{OrS}	16.102	1.17	7.0	$\Psi_{\text{CoL-CB:OrS}}$	16.263	0.98	9.4	$\Psi_{\text{CoL-TB:OrS}}$	16.288	0.91	9.4
		$\Psi_{\text{SC1-Emp}}$	16.136	1.07	6.8	$\Psi_{\text{CoL-CB:SC1-Emp}}$	16.368	0.81	8.6	$\Psi_{\text{CoL-TB:SC1-Emp}}$	16.371	0.79	8.6
		$\Psi_{\text{H1-Emp}}$	16.237	0.99	7.2	$\Psi_{\text{CoL-CB:H1-Emp}}$	16.366	0.79	8.9	$\Psi_{\text{CoL-TB:H1-Emp}}$	16.359	0.77	8.9
						$\Psi_{\text{CoL-CB:SC1}}$	16.832	0.59	14.2	$\Psi_{\text{CoL-TB:H1}}$	16.815	0.59	13.8

Notes. The case study utilizes the data of SKU 068f4481b3 including 31 daily demand samples in each of 51 distribution centers. within Group 1 of Figure 3. Training data includes randomly generated samples with size $n = \{5, 10, 15\}$ from 31 samples in each of 51 distribution centers. Validation data contain the remaining $31 - n$ samples excluding each size-n sample in the training data. In each system, the profit function is $\psi(y, x) = 10 \min\{y, x\} - 7y$. The average profits over 51 systems $\Psi_{\text{CoL-TB:}j}$ and $\Psi_{\text{CoL-CB:}j}$ are derived by, respectively, constant-boosting and targeted-boosting candidate solution j, where $j = \text{SAA}, \text{Re-Var}, \text{Re-Std}, \text{OrS}, \text{SC1-Emp}$, or H1-Emp stands for the sample-average approximation, regularization with variance, regularization with standard deviation, order statistic, scaled operational statistic with empirical distribution, or homogeneous operational statistic with empirical distribution, respectively. The result is generated by computing the average, standard deviation, and minimum of Ψ_{j} out of 10,000 randomly generated samples from the data.

As observed in both Table 5 and Table 6, constant-boosting solutions and targeted-boosting solutions both outperform their directly trained counterparts. Notably, most constant-boosting solutions outperform the targeted-boosting solutions, as the empirical distributions in both Group 1 and Group 2 closely resemble some gamma distributions. Furthermore, the advantage of co-learning solutions over the directly trained solutions is more significant in Group 2, which has fewer distribution centers compared with Group 1. This is because the empirical densities in Group 2 exhibit a decreasing trend with greater variability.

Table 6: Case Study 2: Performance of Co-Learning Solutions Against Directly Trained Solutions

		Directly Trained Solutions				Constant-Boosting Solutions				Targeted-Boosting Solutions			
			Ave	Stdev	Min		Ave	Stdev	Min		Ave	Stdev	Min
$k=9$	$n=5$	Ψ_{SAA}	-0.309	0.29	-2.9	$\Psi_{\text{CoL-CB:SAA}}$	-0.063	0.17	-2.1	$\Psi_{\text{CoL-TB:SAA}}$	-0.066	0.17	-2.1
		$\Psi_{\text{Re-Var}}$	-1.015	0.72	-6.1	$\Psi_{\text{CoL-CB:Re-Var}}$	-0.068	0.22	-3.6	$\Psi_{\text{CoL-TB:Re-Var}}$	-0.072	0.22	-3.6
		$\Psi_{\text{Re-Std}}$	-1.015	0.72	-6.1	$\Psi_{\text{CoL-CB:Re-Std}}$	-0.068	0.22	-3.6	$\Psi_{\text{CoL-TB:Re-Std}}$	-0.072	0.22	-3.6
		Ψ_{OrS}	-0.794	0.55	-4.9	$\Psi_{\text{CoL-CB:OrS}}$	-0.070	0.23	-3.6	$\Psi_{\text{CoL-TB:OrS}}$	-0.074	0.23	-3.6
		$\Psi_{\text{SC1-Emp}}$	-0.364	0.34	-3.3	$\Psi_{\text{CoL-CB:SC1-Emp}}$	-0.065	0.18	-2.5	$\Psi_{\text{CoL-TB:SC1-Emp}}$	-0.068	0.18	-2.5
		$\Psi_{\text{H1-Emp}}$	-0.335	0.31	-3.0	$\Psi_{\text{CoL-CB:H1-Emp}}$	-0.064	0.17	-2.2	$\Psi_{\text{CoL-TB:H1-Emp}}$	-0.067	0.17	-2.2
						$\Psi_{\text{CoL-CB:SC1}}$	-0.007	0.08	-2.0	$\Psi_{\text{CoL-TB:H1}}$	-0.028	0.13	-2.0
	$n=10$	Ψ_{SAA}	-0.226	0.21	-1.8	$\Psi_{\text{CoL-CB:SAA}}$	-0.079	0.21	-2.3	$\Psi_{\text{CoL-TB:SAA}}$	-0.079	0.21	-2.3
		$\Psi_{\text{Re-Var}}$	-0.261	0.26	-1.8	$\Psi_{\text{CoL-CB:Re-Var}}$	-0.073	0.18	-1.8	$\Psi_{\text{CoL-TB:Re-Var}}$	-0.074	0.18	-1.8
		$\Psi_{\text{Re-Std}}$	-0.288	0.27	-1.8	$\Psi_{\text{CoL-CB:Re-Std}}$	-0.072	0.18	-1.8	$\Psi_{\text{CoL-TB:Re-Std}}$	-0.073	0.18	-1.8
		Ψ_{OrS}	-0.356	0.28	-2.0	$\Psi_{\text{CoL-CB:OrS}}$	-0.050	0.17	-1.9	$\Psi_{\text{CoL-TB:OrS}}$	-0.053	0.17	-1.9
		$\Psi_{\text{SC1-Emp}}$	-0.237	0.23	-1.8	$\Psi_{\text{CoL-CB:SC1-Emp}}$	-0.063	0.17	-1.8	$\Psi_{\text{CoL-TB:SC1-Emp}}$	-0.064	0.17	-1.8
		$\Psi_{\text{H1-Emp}}$	-0.230	0.22	-1.8	$\Psi_{\text{CoL-CB:H1-Emp}}$	-0.075	0.20	-2.0	$\Psi_{\text{CoL-TB:H1-Emp}}$	-0.075	0.19	-2.0
						$\Psi_{\text{CoL-CB:SC1}}$	-0.001	0.03	-1.5	$\Psi_{\text{CoL-TB:H1}}$	-0.010	0.07	-1.2
	$n=15$	Ψ_{SAA}	-0.180	0.18	-1.3	$\Psi_{\text{CoL-CB:SAA}}$	-0.045	0.13	-1.3	$\Psi_{\text{CoL-TB:SAA}}$	-0.045	0.13	-1.3
		$\Psi_{\text{Re-Var}}$	-0.147	0.18	-1.3	$\Psi_{\text{CoL-CB:Re-Var}}$	-0.058	0.15	-1.3	$\Psi_{\text{CoL-TB:Re-Var}}$	-0.057	0.14	-1.3
		$\Psi_{\text{Re-Std}}$	-0.153	0.19	-1.5	$\Psi_{\text{CoL-CB:Re-Std}}$	-0.059	0.15	-1.5	$\Psi_{\text{CoL-TB:Re-Std}}$	-0.057	0.14	-1.5
		Ψ_{OrS}	-0.267	0.23	-1.7	$\Psi_{\text{CoL-CB:OrS}}$	-0.043	0.14	-1.5	$\Psi_{\text{CoL-TB:OrS}}$	-0.042	0.14	-1.5
		$\Psi_{\text{SC1-Emp}}$	-0.169	0.17	-1.2	$\Psi_{\text{CoL-CB:SC1-Emp}}$	-0.038	0.12	-1.6	$\Psi_{\text{CoL-TB:SC1-Emp}}$	-0.039	0.12	-1.6
		$\Psi_{\text{H1-Emp}}$	-0.169	0.18	-1.2	$\Psi_{\text{CoL-CB:H1-Emp}}$	-0.048	0.16	-1.4	$\Psi_{\text{CoL-TB:H1-Emp}}$	-0.048	0.13	-1.4
						$\Psi_{\text{CoL-CB:SC1}}$	0.000	0.00	0.0	$\Psi_{\text{CoL-TB:H1}}$	-0.002	0.03	-0.8

Notes. The case study utilizes the data of SKU 068f4481b3 including 31 daily demand samples in each of 9 distribution centers. (within Group 2 of Figure 3). Training data includes randomly generated samples with size $n = \{5, 10, 15\}$ from 31 samples in each of 9 distribution centers. Validation data contain the remaining $31 - n$ samples excluding each size-n sample in the training data. In each system, the profit function is $\psi(y, x) = 10 \min\{y, x\} - 7y$. The average profits over 9 systems $\Psi_{\text{CoL-TB}:j}$ and $\Psi_{\text{CoL-CB}:j}$ are derived by, respectively, constant-boosting and targeted-boosting candidate solution j, where $j = \text{SAA}, \text{Re-Var}, \text{Re-Std}, \text{OrS}, \text{SC1-Emp}$, or H1-Emp stands for the sample-average approximation, regularization with variance, regularization with standard deviation, order statistic, scaled operational statistic with empirical distribution, or homogeneous operational statistic with empirical distribution, respectively. The result is generated by computing the average, standard deviation, and minimum of Ψ_{j} out of 10,000 randomly generated samples from the data.

References

Besbes O, Mouchtaki O (2023) How big should your data really be? Data-driven newsvendor and the transient of learning. *Management Science* 69(10):5695–6415.

Chu LY, Feng Q, Shanthikumar JG, Shen ZJM, Wu J (2025) Solving the price-setting newsvendor problem with parametric operational data analytics (oda). *Management Science* 71(8):6627–6646.

Jennrich RI (1969) Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics* 40(2):633–643.

van der Vaart AW, Wellner JA (1996) *Weak convergence and Empirical Process* (Springer).

Wainwright MJ (2019) *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48 (Cambridge university press).