Appendix A: Verifying Assumption 3 in Examples

In this appendix, we show that each multi-stage stochastic linear optimization problem considered in this paper satisfies Assumption 3.

A.1. Example 1 from Section 3

Consider the sample robust optimization problem

$$\begin{aligned}
\underset{x_2:\mathbb{R}\to\mathbb{R},\, x_3:\mathbb{R}^2\to\mathbb{R}}{\text{minimize}} \quad & \frac{1}{N}\sum_{j=1}^{N} \sup_{\boldsymbol{\zeta}\in\mathcal{U}_N^j} \{x_2(\zeta_1) + 2x_3(\zeta_1,\zeta_2)\} \\
\text{subject to} \quad & x_2(\zeta_1) + x_3(\zeta_1,\zeta_2) \ge \zeta_1 + \zeta_2 && \forall \boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j \\
& x_2(\zeta_1),\, x_3(\zeta_1,\zeta_2) \ge 0 && \forall \boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j.
\end{aligned}$$

We observe that the decisions must be nonnegative for every realization in the uncertainty sets. Moreover, the following constraints can be added to the above problem without affecting its optimal cost:

$$\begin{aligned}
x_2(\zeta_1) &\le \sup_{\boldsymbol{\zeta}'\in\cup_{j=1}^N \mathcal{U}_N^j} \{\zeta_1' + \zeta_2'\} \quad \forall \boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j, \\
x_3(\zeta_1,\zeta_2) &\le \sup_{\boldsymbol{\zeta}'\in\cup_{j=1}^N \mathcal{U}_N^j} \{\zeta_1' + \zeta_2'\} \quad \forall \boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j.
\end{aligned}$$

Indeed, the above constraints ensure that we are never purchasing inventory which exceeds the maximal $\zeta_1 + \zeta_2$ which can be realized in the uncertainty sets. Thus, we have shown that Assumption 3 holds.

A.2. Example 2 from Section 4.3

Consider the sample robust optimization problem

$$\begin{aligned}
\underset{x_1\in\mathbb{Z}}{\text{minimize}} \quad & x_1 \\
\text{subject to} \quad & x_1 \ge \zeta_1 \quad \forall \zeta_1 \in \cup_{j=1}^N \mathcal{U}_N^j.
\end{aligned}$$

We observe that an optimal solution to this problem is $x_1 = \max_{\zeta_1\in\cup_{j=1}^N \mathcal{U}_N^j} \lceil \zeta_1 \rceil$, and thus the constraint

$$x_1 \le \max_{\zeta_1\in\cup_{j=1}^N \mathcal{U}_N^j} \zeta_1 + 1$$

can be added to the above problem without affecting its optimal cost. We conclude that Assumption 3 holds.

A.3. Example 3 from Section 4.3

Consider the sample robust optimization problem

$$\begin{aligned}
\underset{\mathbf{x}_1\in\mathbb{R}^2}{\text{minimize}} \quad & x_{12} \\
\text{subject to} \quad & \zeta_1(1 - x_{12}) \le x_{11} \quad \forall \zeta_1 \in \cup_{j=1}^N \mathcal{U}_N^j \\
& 0 \le x_{12} \le 1.
\end{aligned}$$

We observe that an optimal solution to this problem is given by $x_{11} = \max_{\zeta_1 \in \cup_{j=1}^N \mathcal{U}_N^j} \zeta_1$ and $x_{12} = 0$. Thus, the constraint

$$x_{11} \le \max_{\zeta_1 \in \cup_{j=1}^N \mathcal{U}_N^j} \zeta_1$$

can be added to the above problem without affecting its optimal cost. We conclude that Assumption 3 holds.

A.4. Example 4 from Section 4.3

Consider the sample robust optimization problem

$$\begin{aligned} \underset{x_2:\mathbb{R}\to\mathbb{Z}}{\text{minimize}} \quad & \frac{1}{N}\sum_{j=1}^N \sup_{\zeta_1 \in \mathcal{U}_N^j} x_2(\zeta_1) \\ \text{subject to} \quad & x_2(\zeta_1) \ge \zeta_1 \quad \forall \zeta_1 \in \cup_{j=1}^N \mathcal{U}_N^j. \end{aligned}$$

Since $\Xi = [0,1]$, we observe that the constraint

$$x_2(\zeta_1) \le 1 \quad \forall \zeta_1 \in \cup_{j=1}^N \mathcal{U}_N^j$$

can be added to the above problem without affecting its optimal cost. We conclude that Assumption 3 holds.

A.5. Example from Section 7

Consider the sample robust optimization problem associated with Problem (7):

$$\begin{aligned} \underset{\mathbf{v},\,\mathbf{Q}_1\ge\mathbf{0},\mathbf{Q}_2,\,\mathbf{z}}{\text{minimize}} \quad & \frac{1}{N}\sum_{j=1}^N \max_{\boldsymbol{\zeta}\in\mathcal{U}_N^j}\left\{ c\left(Q_{10} + \sum_{r=1}^R Q_{1r}\right) + hQ_{10} + \sum_{r=1}^R v_r(\boldsymbol{\zeta}_1,\boldsymbol{\zeta}_2) + f\sum_{r=1}^R z_r(\boldsymbol{\zeta}_1)\right\} \\ \text{subject to} \quad & \sum_{r=1}^R Q_{2r}(\boldsymbol{\zeta}_1) \le Q_{10} && \forall \boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j \\ & v_r(\boldsymbol{\zeta}_1,\boldsymbol{\zeta}_2) \ge b(\zeta_{2r} + \zeta_{1r} - Q_{2r}(\boldsymbol{\zeta}_1) - Q_{1r}) - hQ_{2r}(\boldsymbol{\zeta}_1) && \forall r\in[R],\ \boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j \\ & v_r(\boldsymbol{\zeta}_1,\boldsymbol{\zeta}_2) \ge h(Q_{1r} - \zeta_{1r} - \zeta_{2r}) && \forall r\in[R],\ \boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j \\ & v_r(\boldsymbol{\zeta}_1,\boldsymbol{\zeta}_2) \ge b\,(\zeta_{1r} - Q_{1r}) - h\zeta_{2r} && \forall r\in[R],\ \boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j \\ & z_r(\boldsymbol{\zeta}_1)\mathcal{M} \ge Q_{2r}(\boldsymbol{\zeta}_1) && \forall r\in[R],\ \boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j \\ & z_r(\boldsymbol{\zeta}_1) \in \{0,1\},\ Q_{2r}(\boldsymbol{\zeta}_1) \ge 0, && \forall r\in[R],\ \boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j. \end{aligned}$$

Since $\Xi = \mathbb{R}_+^{2r}$, we observe that the constraints

$$0 \le Q_{11},\dots,Q_{1R} \le \max_{\boldsymbol{\zeta}\in\cup_{j=1}^N\mathcal{U}_N^j} (\zeta_{1r} + \zeta_{2r})$$

$$0 \le Q_{10} \le \max_{\boldsymbol{\zeta}\in\cup_{j=1}^N\mathcal{U}_N^j} \sum_{r=1}^R (\zeta_{1r} + \zeta_{2r}),$$

can be added to the above problem without affecting its optimal cost. It thus follows from the constraint $\sum_{r=1}^R Q_{2r}(\boldsymbol{\zeta}_1) \le Q_{10}$ that the constraints

$$0 \le Q_{21}(\boldsymbol{\zeta}_1),\dots,Q_{2R}(\boldsymbol{\zeta}_1) \le \max_{\boldsymbol{\zeta}'\in\cup_{j=1}^N\mathcal{U}_N^j} \sum_{r=1}^R (\zeta'_{1r} + \zeta'_{2r}), \quad \forall \boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j$$

can also be added to the above problem without affecting its optimal cost. Finally, we can without loss of generality impose the constraint for each retailer $r \in [R]$ that

$$\begin{aligned} 0 \le v_r(\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2) = \max\{ & b(\zeta_{2r} + \zeta_{1r} - Q_{2r}(\boldsymbol{\zeta}_1) - Q_{1r}) - hQ_{2r}(\boldsymbol{\zeta}_1), \\ & h(Q_{1r} - \zeta_{1r} - \zeta_{2r}), \\ & b(\zeta_{1r} - Q_{1r}) - h\zeta_{2r}\} \qquad \forall \boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j. \end{aligned}$$

Applying the aforementioned bounds on the decision rules, we conclude that Assumption 3 holds.

A.6. Example from Section 8

Consider the sample robust optimization problem

$$\begin{aligned} \underset{\mathbf{x},\mathbf{I},\mathbf{y}}{\text{minimize}} \quad & \frac{1}{N}\sum_{j=1}^{T} \sup_{\boldsymbol{\zeta}\in\mathcal{U}_N^j} \sum_{t=1}^{T} \left(c_t x_t(\zeta_1,\ldots,\zeta_{t-1}) + y_{t+1}(\zeta_1,\ldots,\zeta_t)\right) \\ \text{subject to} \quad & I_{t+1}(\zeta_1,\ldots,\zeta_t) = I_t(\zeta_1,\ldots,\zeta_{t-1}) + x_t(\zeta_1,\ldots,\zeta_{t-1}) - \zeta_t \quad \forall \boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j,\ \forall t \in [T] \\ & y_{t+1}(\zeta_1,\ldots,\zeta_t) \ge h_t I_{t+1}(\zeta_1,\ldots,\zeta_t) \quad \forall \boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j,\ \forall t \in [T] \\ & y_{t+1}(\zeta_1,\ldots,\zeta_t) \ge -b_t I_{t+1}(\zeta_1,\ldots,\zeta_t) \quad \forall \boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j,\ \forall t \in [T] \\ & 0 \le x_t(\zeta_1,\ldots,\zeta_{t-1}) \le \bar{x}_t \quad \forall \boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j,\ \forall t \in [T], \end{aligned}$$

where $I_1 = 0$ and $\Xi = \mathbb{R}_+^T$. For any feasible decision rule to the above problem and for each stage t, we observe that the following constraint is satisfied:

$$-\sup_{\boldsymbol{\zeta}'\in\cup_{j=1}^N \mathcal{U}_N^j} \sum_{s=1}^{T} \zeta'_s \le I_{t+1}(\zeta_1,\ldots,\zeta_t) \le \sum_{s=1}^{T} \bar{x}_s \quad \forall \boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j.$$

Moreover, we can without loss of generality impose the constraint that

$$0 \le y_{t+1}(\zeta_1,\ldots,\zeta_t) = \max\left\{h_t I_{t+1}(\zeta_1,\ldots,\zeta_t), -b_t I_{t+1}(\zeta_1,\ldots,\zeta_t)\right\} \quad \forall \boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j.$$

Applying the aforementioned bounds on $I_{t+1}(\zeta_1,\ldots,\zeta_t)$ over the uncertainty sets, we conclude that Assumption 3 holds.

Appendix B: On the Tightness of the Bounds from Theorem 1

In Section 4.2, we introduced a lower bound $\underline{J}$ and upper bound $\bar{J}$ on the optimal cost J^* of Problem (1). In Theorem 1, we showed under mild assumptions that these quantities also provide an asymptotic lower and upper bound on the optimal cost $\widehat{J}_N$ of Problem (2). In this appendix, we demonstrate the practical value of these bounds by establishing sufficient conditions for the lower and upper bounds to be equal and applying those sufficient conditions to applications of multi-stage stochastic linear optimization from Sections 3, 7, and 8.

B.1. Sufficient conditions for $\underline{J} = J^*$ and $\bar{J} = J^*$

We begin by developing our two primary results, Theorems EC.1 and EC.2, in which we establish sufficient conditions for the lower bound $\underline{J}$ and the upper bound $\bar{J}$ to be equal to the optimal cost J^* of Problem (1). In particular, we will show in the following Appendix B.2 that the sufficient conditions in these theorems can be verified in examples in which the underlying joint probability distribution and the support of the stochastic process are unknown. Consequently, the following two theorems can serve as practical tools for demonstrating that our proposed robust optimization approach, Problem (2), is asymptotically optimal for specific multi-stage stochastic linear optimization problems which arise in real-world applications.

We begin by developing our first primary result, Theorem EC.1, which establishes a sufficient condition for the lower bound $\underline{J}$ to be equal to the optimal cost J^* of Problem (1). Recall that S denotes the support of the joint probability distribution. Speaking intuitively, the following theorem shows that $\underline{J}$ is guaranteed to equal J^* if there exists an optimal decision rule to a stochastic problem over any restricted support $\tilde{S} \subseteq S$ that can be extended to a decision rule that is feasible for Problem (1) and has a well-behaved objective function. To see the utility of the following theorem, we refer the reader to the examples in Appendix B.2

Theorem EC.1 **(Sufficient condition for lower bound)**. *Let Assumption 1 hold, and suppose there exists an $L \geq 0$ such that, for all $\tilde{S} \subseteq S$, the optimal cost of the optimization problem*

$$\begin{aligned} \underset{\mathbf{x} \in \mathcal{X}}{\text{minimize}} \quad & \mathbb{E}\left[\sum_{t=1}^{T} \mathbf{c}_t(\boldsymbol{\xi}) \cdot \mathbf{x}_t(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{t-1}) \mathbb{I}\left\{\boldsymbol{\xi} \in \tilde{S}\right\}\right] \\ \text{subject to} \quad & \sum_{t=1}^{T} \mathbf{A}_t(\boldsymbol{\zeta}) \mathbf{x}_t(\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_{t-1}) \leq \mathbf{b}(\boldsymbol{\zeta}) \quad \forall \boldsymbol{\zeta} \in \tilde{S} \end{aligned}$$

would not change if we added the constraints

$$\begin{aligned} 0 \leq \sum_{t=1}^{T} \mathbf{c}_t(\boldsymbol{\xi}) \cdot \mathbf{x}_t(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{t-1}) \leq L\left(1 + \max\left\{\|\boldsymbol{\xi}\|, \sup_{\boldsymbol{\zeta} \in \tilde{S}} \|\boldsymbol{\zeta}\|\right\}\right) \quad & \text{a.s.} \\ \sum_{t=1}^{T} \mathbf{A}_t(\boldsymbol{\xi}) \mathbf{x}_t(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{t-1}) \leq \mathbf{b}(\boldsymbol{\xi}) \quad & \text{a.s.} \end{aligned}$$

Then, $\underline{J} = J^*$.

Proof. We recall from the definition of the lower bound that $\underline{J} := \lim_{\rho \to 0} \underline{J}_\rho$, where

$$\underline{J}_\rho := \min_{\tilde{S} \subseteq \Xi:\, \mathbb{P}(\boldsymbol{\xi} \in \tilde{S}) \geq 1-\rho} \left\{ \begin{aligned} \underset{\mathbf{x} \in \mathcal{X}}{\text{minimize}} \quad & \mathbb{E}\left[\sum_{t=1}^{T} \mathbf{c}_t(\boldsymbol{\xi}) \cdot \mathbf{x}_t(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{t-1}) \mathbb{I}\left\{\boldsymbol{\xi} \in \tilde{S}\right\}\right] \\ \text{subject to} \quad & \sum_{t=1}^{T} \mathbf{A}_t(\boldsymbol{\zeta}) \mathbf{x}_t(\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_{t-1}) \leq \mathbf{b}(\boldsymbol{\zeta}) \quad \forall \boldsymbol{\zeta} \in \tilde{S} \end{aligned} \right\}.$$

Therefore, it follows from the conditions of Theorem EC.1 that

$$\underline{J}_\rho = \min_{\tilde{S} \subseteq \Xi:\, \mathbb{P}(\boldsymbol{\xi} \in \tilde{S}) \geq 1-\rho} \left\{ \begin{aligned} \underset{\mathbf{x} \in \mathcal{X}}{\text{minimize}} \quad & \mathbb{E}\left[\sum_{t=1}^{T} \mathbf{c}_t(\boldsymbol{\xi}) \cdot \mathbf{x}_t(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{t-1}) \mathbb{I}\left\{\boldsymbol{\xi} \in \tilde{S}\right\}\right] \\ \text{subject to} \quad & \sum_{t=1}^{T} \mathbf{c}_t(\boldsymbol{\xi}) \cdot \mathbf{x}_t(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{t-1}) \geq 0 \qquad \text{a.s.} \\ & \sum_{t=1}^{T} \mathbf{A}_t(\boldsymbol{\zeta}) \mathbf{x}_t(\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_{t-1}) \leq \mathbf{b}(\boldsymbol{\zeta}) \qquad \forall \boldsymbol{\zeta} \in \tilde{S} \end{aligned} \right\}.$$

For any feasible solution to the above optimization problem, we observe that the function $\sum_{t=1}^{T} \mathbf{c}_t(\boldsymbol{\xi}) \cdot \mathbf{x}_t(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{t-1})$ is nonnegative almost surely. Therefore, for any arbitrary $r \ge 0$, a lower bound on $\underline{J}_{\rho}$ is given by

$$\underline{J}_{\rho,r} := \min_{\tilde{S}\subseteq\Xi:\, \mathbb{P}(\boldsymbol{\xi}\in\tilde{S})\ge 1-\rho} \left\{ \begin{array}{lll} \underset{\mathbf{x}\in\mathcal{X}}{\text{minimize}} & \mathbb{E}\left[\sum_{t=1}^{T} \mathbf{c}_t(\boldsymbol{\xi})\cdot \mathbf{x}_t(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{t-1}) \mathbb{I}\left\{\boldsymbol{\xi}\in\tilde{S},\ \|\boldsymbol{\xi}\|\le r\right\}\right] & \\ \text{subject to} & \sum_{t=1}^{T} \mathbf{c}_t(\boldsymbol{\xi})\cdot \mathbf{x}_t(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{t-1}) \ge 0 & \text{a.s.} \\ & \sum_{t=1}^{T} \mathbf{A}_t(\boldsymbol{\zeta})\mathbf{x}_t(\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_{t-1}) \le \mathbf{b}(\boldsymbol{\zeta}) & \forall \boldsymbol{\zeta}\in\tilde{S} \end{array} \right\} \tag{EC.1}$$

$$\ge \min_{\tilde{S}\subseteq\Xi:\, \mathbb{P}(\boldsymbol{\xi}\in\tilde{S})\ge 1-\rho} \left\{ \begin{array}{ll} \underset{\mathbf{x}\in\mathcal{X}}{\text{minimize}} & \mathbb{E}\left[\sum_{t=1}^{T} \mathbf{c}_t(\boldsymbol{\xi})\cdot \mathbf{x}_t(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{t-1}) \mathbb{I}\left\{\boldsymbol{\xi}\in\tilde{S},\ \|\boldsymbol{\xi}\|\le r\right\}\right] \\ \text{subject to} & \sum_{t=1}^{T} \mathbf{A}_t(\boldsymbol{\zeta})\mathbf{x}_t(\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_{t-1}) \le \mathbf{b}(\boldsymbol{\zeta}) \quad \forall \boldsymbol{\zeta}\in\tilde{S}: \|\boldsymbol{\zeta}\|\le r \end{array} \right\}, \tag{EC.2}$$

where the inequality follows from removing constraints from the inner minimization problem in line (EC.1). Furthermore, it follows from the conditions of Theorem EC.1 that line (EC.2) is equal to

$$\min_{\tilde{S}\subseteq\Xi:\, \mathbb{P}(\boldsymbol{\xi}\in\tilde{S})\ge 1-\rho} \left\{ \begin{array}{lll} \underset{\mathbf{x}\in\mathcal{X}}{\text{minimize}} & \mathbb{E}\left[\sum_{t=1}^{T} \mathbf{c}_t(\boldsymbol{\xi})\cdot \mathbf{x}_t(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{t-1}) \mathbb{I}\left\{\boldsymbol{\xi}\in\tilde{S},\ \|\boldsymbol{\xi}\|\le r\right\}\right] & \\ \text{subject to} & \sum_{t=1}^{T} \mathbf{A}_t(\boldsymbol{\zeta})\mathbf{x}_t(\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_{t-1}) \le \mathbf{b}(\boldsymbol{\zeta}) \quad \forall \boldsymbol{\zeta}\in\tilde{S}: \|\boldsymbol{\zeta}\|\le r & \\ & 0 \le \sum_{t=1}^{T} \mathbf{c}_t(\boldsymbol{\xi})\cdot \mathbf{x}_t(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{t-1}) \le L\left(1+\max\left\{\|\boldsymbol{\xi}\|, \sup_{\boldsymbol{\zeta}\in\tilde{S}:\|\boldsymbol{\zeta}\|\le r} \|\boldsymbol{\zeta}\|\right\}\right) & \text{a.s.} \\ & \sum_{t=1}^{T} \mathbf{A}_t(\boldsymbol{\xi})\mathbf{x}_t(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{t-1}) \le \mathbf{b}(\boldsymbol{\xi}) & \text{a.s.} \end{array} \right\} \tag{EC.3}$$

$$\ge \min_{\tilde{S}\subseteq\Xi:\, \mathbb{P}(\boldsymbol{\xi}\in\tilde{S})\ge 1-\rho} \left\{ \begin{array}{lll} \underset{\mathbf{x}\in\mathcal{X}}{\text{minimize}} & \mathbb{E}\left[\sum_{t=1}^{T} \mathbf{c}_t(\boldsymbol{\xi})\cdot \mathbf{x}_t(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{t-1}) \mathbb{I}\left\{\boldsymbol{\xi}\in\tilde{S},\ \|\boldsymbol{\xi}\|\le r\right\}\right] & \\ \text{subject to} & 0 \le \sum_{t=1}^{T} \mathbf{c}_t(\boldsymbol{\xi})\cdot \mathbf{x}_t(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{t-1}) \le L\left(1+\max\left\{\|\boldsymbol{\xi}\|, r\right\}\right) & \text{a.s.} \\ & \sum_{t=1}^{T} \mathbf{A}_t(\boldsymbol{\xi})\mathbf{x}_t(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{t-1}) \le \mathbf{b}(\boldsymbol{\xi}) & \text{a.s.} \end{array} \right\}, \tag{EC.4}$$

where the inequality follows from removing constraints from the inner minimization problem in line (EC.3) and using the fact that $\sup_{\boldsymbol{\zeta}'\in\tilde{S}:\|\boldsymbol{\zeta}'\|\le r} \|\boldsymbol{\zeta}'\| \le r$.

We now use Assumption 1 to obtain a lower bound on (EC.4). Indeed, Assumption 1 says that there exists an $a > 1$ such that $b := \mathbb{E}[\exp(\|\boldsymbol{\xi}\|^a)] < \infty$. Therefore, for any feasible solutions $\tilde{S} \subseteq \Xi$ and $\mathbf{x}\in\mathcal{X}$ to the optimization problems in (EC.4),

$$\begin{aligned} &\mathbb{E}\left[\sum_{t=1}^{T} \mathbf{c}_t(\boldsymbol{\xi})\cdot \mathbf{x}_t(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{t-1}) \mathbb{I}\left\{\boldsymbol{\xi}\notin\tilde{S} \text{ or } \|\boldsymbol{\xi}\| > r\right\}\right] \\ &\le \mathbb{E}\left[L\left(1+\max\{\|\boldsymbol{\xi}\|, r\}\right) \mathbb{I}\left\{\boldsymbol{\xi}\notin\tilde{S} \text{ or } \|\boldsymbol{\xi}\| > r\right\}\right] \end{aligned}$$

$$
\begin{aligned}
&\le \mathbb{E}\left[L\left(1+\max\left\{\|\boldsymbol{\xi}\|,r\right\}\right)\mathbb{I}\left\{\boldsymbol{\xi}\notin\tilde{S}\right\}\right]+\mathbb{E}\left[L\left(1+\max\left\{\|\boldsymbol{\xi}\|,r\right\}\right)\mathbb{I}\left\{\|\boldsymbol{\xi}\|>r\right\}\right]\\
&\le \sqrt{\mathbb{E}\left[L^2\left(1+\max\left\{\|\boldsymbol{\xi}\|,r\right\}\right)^2\right]\rho}+\sqrt{\mathbb{E}\left[L^2\left(1+\max\left\{\|\boldsymbol{\xi}\|,r\right\}\right)^2\right]\mathbb{P}(\|\boldsymbol{\xi}\|>r)}\\
&\le \underbrace{\sqrt{\mathbb{E}\left[L^2\left(1+\max\left\{\|\boldsymbol{\xi}\|,r\right\}\right)^2\right]\rho}+\sqrt{\mathbb{E}\left[L^2\left(1+\max\left\{\|\boldsymbol{\xi}\|,r\right\}\right)^2\right]\frac{b}{\exp(r^a)}}}_{h(\rho,r)}.
\end{aligned}
$$

Indeed, the first inequality follows because $0 \le \sum_{t=1}^T \mathbf{c}_t(\boldsymbol{\xi}) \cdot \mathbf{x}_t(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{t-1}) \le L(1+\max\{\|\boldsymbol{\xi}\|,r\})$ almost surely, the second inequality follows from the union bound, the third inequality follows from $\mathbb{P}(\boldsymbol{\xi}\in\tilde{S})\ge 1-\rho$ and the Cauchy-Schwartz inequality, and the fourth and final inequality follows from Markov's inequality. Therefore,

$$
\begin{aligned}
\underline{J}_{\rho,r} &\ge -h(\rho,r) + \min_{\tilde{S}\subseteq\Xi:\,\mathbb{P}(\boldsymbol{\xi}\in\tilde{S})\ge 1-\rho}\left\{\begin{array}{lll}\underset{\mathbf{x}\in\mathcal{X}}{\text{minimize}} & \mathbb{E}\left[\displaystyle\sum_{t=1}^T \mathbf{c}_t(\boldsymbol{\xi})\cdot\mathbf{x}_t(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{t-1})\right] & \\ \text{subject to} & 0\le\displaystyle\sum_{t=1}^T \mathbf{c}_t(\boldsymbol{\xi})\cdot\mathbf{x}_t(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{t-1})\le L\left(1+\max\left\{\|\boldsymbol{\xi}\|,r\right\}\right) & \text{a.s.}\\ & \displaystyle\sum_{t=1}^T \mathbf{A}_t(\boldsymbol{\xi})\mathbf{x}_t(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{t-1})\le\mathbf{b}(\boldsymbol{\xi}) & \text{a.s.}\end{array}\right\}\\
&\ge -h(\rho,r) + \left\{\begin{array}{ll}\underset{\mathbf{x}\in\mathcal{X}}{\text{minimize}} & \mathbb{E}\left[\displaystyle\sum_{t=1}^T \mathbf{c}_t(\boldsymbol{\xi})\cdot\mathbf{x}_t(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{t-1})\right]\\ \text{subject to} & \displaystyle\sum_{t=1}^T \mathbf{A}_t(\boldsymbol{\xi})\mathbf{x}_t(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{t-1})\le\mathbf{b}(\boldsymbol{\xi}) \quad \text{a.s.}\end{array}\right.\\
&= -h(\rho,r) + J^*.
\end{aligned} \tag{EC.5}
$$

The first inequality follows from the lower bound on $\underline{J}_{\rho,r}$ in line (EC.4), the definition of $h(\rho,r)$, and the law of iterated expectation. The second inequality follows from removing constraints, and the final equality follows from the definition of J^*.

We now combine the above results to prove the main result. Indeed,

$$
\underline{J} = \lim_{\rho\downarrow 0}\underline{J}_\rho \ge \lim_{r\to\infty}\lim_{\rho\downarrow 0}\underline{J}_{\rho,r} \ge \lim_{r\to\infty}\lim_{\rho\downarrow 0} -h(\rho,r)+J^* = J^*.
$$

The first inequality follows because $\underline{J}_\rho \ge \underline{J}_{\rho,r}$ for any arbitrary $r\ge 0$ and the quantity $\lim_{\rho\downarrow 0}\underline{J}_{\rho,r}$ is monotonically increasing in r. The second inequality follows from (EC.5). The final equality follows from the definition of $h(\rho,r)$ and Assumption 1. Since the inequality $\underline{J}\le J^*$ always holds, our proof is complete. □

We conclude the present Appendix B.1 by developing our second primary result, Theorem EC.2, which establishes a sufficient condition for the upper bound $\bar{J}$ to be equal to the optimal cost J^* of Problem (1). Speaking intuitively, the following theorem says that $\bar{J}$ is equal to J^* if there are near-optimal decision rules to Problem (1) that are feasible and result in an objective function which is both upper-semicontinuous and well-behaved on a slight extension of the support S.

Theorem EC.2 **(Sufficient condition for upper bound)**. *Let Assumption 1 hold, and suppose for all $\eta > 0$ that there exists an η-optimal decision rule $\mathbf{x}^\eta \in \mathcal{X}$ for Problem (1) and $\rho^\eta > 0$, $L^\eta \ge 0$ such that*

$$\sum_{t=1}^{T} \mathbf{A}_t(\boldsymbol{\zeta}) \mathbf{x}_t^\eta(\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_{t-1}) \le \mathbf{b}(\boldsymbol{\zeta}) \quad \forall \boldsymbol{\zeta} \in \Xi : \operatorname{dist}(\boldsymbol{\zeta}, S) \le \rho^\eta; \tag{EC.6}$$

$$0 \le \sum_{t=1}^{T} \mathbf{c}_t(\boldsymbol{\zeta}) \cdot \mathbf{x}_t^\eta(\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_{t-1}) \le L^\eta(1 + \|\boldsymbol{\zeta}\|) \quad \forall \boldsymbol{\zeta} \in \Xi : \operatorname{dist}(\boldsymbol{\zeta}, S) \le \rho^\eta; \tag{EC.7}$$

$$\lim_{\epsilon \to 0} \sup_{\boldsymbol{\zeta} \in \Xi:\, \|\boldsymbol{\zeta} - \boldsymbol{\xi}\| \le \epsilon} \left\{ \sum_{t=1}^{T} \mathbf{c}_t(\boldsymbol{\zeta}) \cdot \mathbf{x}_t^\eta(\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_{t-1}) \right\} = \sum_{t=1}^{T} \mathbf{c}_t(\boldsymbol{\xi}) \cdot \mathbf{x}_t^\eta(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{t-1}) \quad \text{a.s.} \tag{EC.8}$$

Then, $\bar{J} = J^*$.

Proof. For any arbitrary $\eta > 0$, consider an η-optimal decision rule $\mathbf{x}^\eta \in \mathcal{X}$ for Problem (1) and constants $\rho^\eta > 0$, $L^\eta \ge 0$ which satisfy properties (EC.6), (EC.7), and (EC.8). Then,

$$\begin{aligned}
\bar{J} &\le \bar{\mathbb{E}}\left[\sum_{t=1}^{T} \mathbf{c}_t(\boldsymbol{\xi}) \cdot \mathbf{x}_t^\eta(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{t-1})\right] \\
&= \lim_{\epsilon \to 0} \mathbb{E}\left[\sup_{\boldsymbol{\zeta} \in \Xi:\, \|\boldsymbol{\zeta} - \boldsymbol{\xi}\| \le \epsilon} \left\{ \sum_{t=1}^{T} \mathbf{c}_t(\boldsymbol{\zeta}) \cdot \mathbf{x}_t^\eta(\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_{t-1}) \right\}\right] \\
&= \mathbb{E}\left[\lim_{\epsilon \to 0} \sup_{\boldsymbol{\zeta} \in \Xi:\, \|\boldsymbol{\zeta} - \boldsymbol{\xi}\| \le \epsilon} \left\{ \sum_{t=1}^{T} \mathbf{c}_t(\boldsymbol{\zeta}) \cdot \mathbf{x}_t^\eta(\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_{t-1}) \right\}\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \mathbf{c}_t(\boldsymbol{\xi}) \cdot \mathbf{x}_t^\eta(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{t-1})\right] \\
&\le J^* + \eta.
\end{aligned}$$

Indeed, the first inequality follows from the definition of $\bar{J}$ and the feasiblility of $\mathbf{x}^\eta$ for the problem defining $\bar{J}$ for all sufficiently small ρ, as indicated by (EC.6). The first equality is the definition of the local upper semicontinuous envelope. The second equality follows from the dominated convergence theorem, which can be applied because of Assumption 1 and (EC.7). The third equality follows from (EC.8), and the final equality holds because $\mathbf{x}^\eta$ is an η-optimal decision rule for Problem (1). Since $\eta > 0$ was chosen arbitrarily and since the inequality $J^* \le \bar{J}$ always holds, our proof is complete. □

B.2. Applications of Theorems EC.1 and EC.2

In the previous section, we developed sufficient conditions, Theorems EC.1 and EC.2, for the proposed robust optimization approach, Problem (2), to be asymptotically optimal for multi-stage stochastic linear optimization problems. In this section, we use those sufficient conditions to show that Problem (2) is asymptotically optimal in three data-driven examples of multi-stage stochastic linear optimization based on Sections 3, 7, and 8. All together, the three examples provide evidence that the lower bound and upper bounds in Theorem 1 can be equal in applications of multi-stage stochastic linear optimization that arise in practice.

B.2.1. Example 1 from Section 3. Consider the multi-stage stochastic linear optimization problem

$$\begin{aligned} J^* = \quad & \underset{x_2:\mathbb{R}\to\mathbb{R},\, x_3:\mathbb{R}^2\to\mathbb{R}}{\text{minimize}} && \mathbb{E}\left[x_2(\xi_1) + 2x_3(\xi_1,\xi_2)\right] \\ & \text{subject to} && x_2(\xi_1) + x_3(\xi_1,\xi_2) \ge \xi_1 + \xi_2 \quad \text{a.s.} \\ &&& x_2(\xi_1),\, x_3(\xi_1,\xi_2) \ge 0 \qquad \text{a.s.,} \end{aligned} \tag{3}$$

where the random variables $\boldsymbol{\xi} = (\xi_1,\xi_2) \in \mathbb{R}^2$ denote the preorder and regular demand of a new product. We assume that this stochastic process satisfies Assumption 1 and is contained in $\Xi := \mathbb{R}^2_+$.

Proposition 1. *For Problem* (3), $\underline{J} = J^*$. *If there is an optimal* $x_2^*: \mathbb{R}\to\mathbb{R}$ *for Problem* (3) *which is continuous, then* $\bar{J} = J^*$.

Proof. Our proof is split into two parts:

- For any arbitrary $\tilde{S} \subseteq S$, we observe that the optimal cost of the optimization problem

$$\begin{aligned} & \text{minimize} && \mathbb{E}\left[(x_2(\xi_1) + 2x_3(\xi_1,\xi_2))\, \mathbb{I}\left\{\boldsymbol{\xi} \in \tilde{S}\right\}\right] \\ & \text{subject to} && x_2(\zeta_1) + x_3(\zeta_1,\zeta_2) \ge \xi_1 + \xi_2 && \forall \boldsymbol{\zeta} \in \tilde{S} \\ &&& x_2(\zeta_1),\, x_3(\zeta_1,\zeta_2) \ge 0 && \forall \boldsymbol{\zeta} \in \tilde{S} \end{aligned} \tag{EC.9}$$

  would not change if we added the constraints

$$\begin{aligned} 0 \le x_2(\xi_1) &\le \sup_{\boldsymbol{\zeta}\in\tilde{S}} \{\zeta_1 + \zeta_2\} \quad \text{a.s.} \\ 0 \le x_3(\xi_1,\xi_2) &\le \xi_1 + \xi_2 \quad \text{a.s.} \end{aligned}$$

  Moreover, any feasible solution to Problem (EC.9) which satisfies the above constraints will also satisfy

$$0 \le x_2(\xi_1) + 2x_3(\xi_1,\xi_2) \le 3\max\left\{\sup_{\boldsymbol{\zeta}\in\tilde{S}}\{\zeta_1+\zeta_2\}, \xi_1+\xi_2\right\} \quad \text{a.s.}$$

  as well as satisfy the constraints in Problem (3). Since Assumption 1 holds, we conclude from Theorem EC.1 that $\underline{J} = J^*$ for Problem (3).

- Let $x_2^*: \mathbb{R}\to\mathbb{R}$ denote an optimal second-stage decision rule to Problem (3) which is continuous. For any $M \ge 0$, define the new decision rules

$$x_2^M(\zeta_1) := \max\{0, \min\{x_2^*(\zeta_1), M\}\}, \quad x_3^M(\zeta_1,\zeta_2) := \max\left\{0, \zeta_1 + \zeta_2 - x_2^M(\zeta_1)\right\}. \tag{EC.10}$$

  We observe that the decision rules from (EC.10) satisfy the constraints of Problem (3). Moreover,

$$\begin{aligned} & \mathbb{E}\left[x_2^M(\xi_1) + 2x_3^M(\xi_1,\xi_2)\right] \\ & = \mathbb{E}\left[x_2^M(\xi_1)\right] + \mathbb{E}\left[2x_3^M(\xi_1,\xi_2)\mathbb{I}\left\{x_2^*(\xi_1) \le M\right\}\right] + \mathbb{E}\left[2x_3^M(\xi_1,\xi_2)\mathbb{I}\left\{x_2^*(\xi_1) > M\right\}\right] && \text{(EC.11)} \\ & \le \mathbb{E}\left[x_2^*(\xi_1)\right] + \mathbb{E}\left[2x_3^M(\xi_1,\xi_2)\mathbb{I}\left\{x_2^*(\xi_1) \le M\right\}\right] + \mathbb{E}\left[2x_3^M(\xi_1,\xi_2)\mathbb{I}\left\{x_2^*(\xi_1) > M\right\}\right] && \text{(EC.12)} \\ & = \mathbb{E}\left[x_2^*(\xi_1)\right] + \mathbb{E}\left[2x_3^*(\xi_1,\xi_2)\mathbb{I}\left\{x_2^*(\xi_1) \le M\right\}\right] + \mathbb{E}\left[2x_3^M(\xi_1,\xi_2)\mathbb{I}\left\{x_2^*(\xi_1) > M\right\}\right] && \text{(EC.13)} \\ & \le J^* + \mathbb{E}\left[2x_3^M(\xi_1,\xi_2)\mathbb{I}\left\{x_2^*(\xi_1) > M\right\}\right] && \text{(EC.14)} \\ & \le J^* + 2\mathbb{E}\left[(\xi_1+\xi_2)\mathbb{I}\left\{x_2^*(\xi_1) > M\right\}\right] && \text{(EC.15)} \\ & \le J^* + 2\sqrt{\mathbb{E}\left[(\xi_1+\xi_2)^2\right]}\sqrt{\mathbb{P}\left(x_2^*(\xi_1) > M\right)}. && \text{(EC.16)} \end{aligned}$$

Indeed, (EC.11) follows from the law of total probability; (EC.12) and (EC.13) follow from the definition of the decision rules from (EC.10); (EC.14) holds because the inequality $x_3^M(\xi_1, \xi_2) \geq 0$ holds almost surely; (EC.15) follows from the definition of the decision rule $x_3^M(\xi_1, \xi_2)$; (EC.16) follows from the Cauchy-Schwartz inequality.

Since $\lim_{M \to \infty} \mathbb{P}(x_2^*(\xi_1) > M) = 0$ and $\mathbb{E}[(\xi_1 + \xi_2)^2] < \infty$ (Assumption 1), we have shown, for every arbitrary $\eta > 0$, that there exists a constant $M \equiv M^\eta \geq 0$ such that the decision rules from (EC.10) are an η-optimal solution to Problem (3). Moreover, we readily observe that the decision rules from (EC.10) satisfy the properties of Theorem EC.2. Indeed, for every $M \geq 0$:

— Property (EC.6) is clearly satisfied by the decision rules (x_2^M, x_3^M).

— Property (EC.7) is satisfied by the decision rules (x_2^M, x_3^M) because the inequalities $0 \leq x_2^M(\zeta_1) + 2x_3^M(\zeta_1, \zeta_2) \leq M + 2(\zeta_1 + \zeta_2)$ hold for all $\boldsymbol{\zeta} \in \Xi$.

— Property (EC.8) is satisfied by the decision rules (x_2^M, x_3^M) because the optimal decision rules (x_2^*, x_3^*) are continuous functions, which implies that (x_2^M, x_3^M) are continuous functions as well over the domain $\boldsymbol{\zeta} \in \Xi$.

We thus conclude from Theorem EC.2 that $J^* = \bar{J}$ for Problem (3).

□

B.2.2. Example from Section 7. Consider the multi-stage stochastic linear optimization problem

$$
\begin{aligned}
J^* = \quad & \underset{\mathbf{Q} \geq \mathbf{0}, \mathbf{z} \in \{0,1\}^R, \mathbf{v}}{\text{minimize}} && \mathbb{E}\left[c\left(Q_{10} + \sum_{r=1}^R Q_{1r} \right) + hQ_{10} + \sum_{r=1}^R v_r(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) + f \sum_{r=1}^R z_r(\boldsymbol{\xi}_1) \right] \\
& \text{subject to} && \sum_{r=1}^R Q_{2r}(\boldsymbol{\xi}_1) \leq Q_{10} && \text{a.s.} \\
& && v_r(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) \geq b(\xi_{2r} + \xi_{1r} - Q_{2r}(\boldsymbol{\xi}_1) - Q_{1r}) - hQ_{2r}(\boldsymbol{\xi}_1) && \forall r \in [R], \text{ a.s.} \\
& && v_r(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) \geq h(Q_{1r} - \xi_{1r} - \xi_{2r}) && \forall r \in [R], \text{ a.s.} \\
& && v_r(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) \geq b(\xi_{1r} - Q_{1r}) - h\xi_{2r} && \forall r \in [R], \text{ a.s.} \\
& && z_r(\boldsymbol{\xi}_1)\mathcal{M} \geq Q_{2r}(\boldsymbol{\xi}_1) && \forall r \in [R], \text{ a.s.,}
\end{aligned} \tag{7}
$$

where the random variables $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) \in \mathbb{R}^{2R}$ denote the demands of a weekly magazine at different retailers. We assume that this stochastic process satisfies Assumption 1 and is contained in $\Xi := \mathbb{R}^{2R}_+$. For simplicity, we focus on the case where $f = 0$.

Proposition EC.1. *For Problem* (7), $\underline{J} = J^*$. *Moreover, if there are optimal decision rules* $Q_{21}^*, \ldots, Q_{2R}^* : \mathbb{R}^R \to \mathbb{R}$ *which are continuous, then* $\bar{J} = J^*$.

Proof. Our proof is split into two parts:

- For any arbitrary $\tilde{S} \subseteq S$, we observe that the optimal cost of the optimization problem

$$\begin{aligned} \underset{\mathbf{Q}\geq \mathbf{0}, \mathbf{z}\in\{0,1\}^R, \mathbf{v}}{\text{minimize}} \quad & \mathbb{E}\left[\left(c\left(Q_{10} + \sum_{r=1}^R Q_{1r}\right) + hQ_{10} + \sum_{r=1}^R v_r(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)\right)\mathbb{I}\left\{\boldsymbol{\xi}\in\tilde{S}\right\}\right] \\ \text{subject to} \quad & \sum_{r=1}^R Q_{2r}(\boldsymbol{\zeta}_1) \leq Q_{10} && \forall \boldsymbol{\zeta}\in\tilde{S} \\ & v_r(\boldsymbol{\zeta}_1,\boldsymbol{\zeta}_2) \geq b(\zeta_{2r}+\zeta_{1r} - Q_{2r}(\boldsymbol{\zeta}_1) - Q_{1r}) - hQ_{2r}(\boldsymbol{\zeta}_1) && \forall r\in[R],\ \forall \boldsymbol{\zeta}\in\tilde{S} \\ & v_r(\boldsymbol{\zeta}_1,\boldsymbol{\zeta}_2) \geq h(Q_{1r} - \zeta_{1r} - \zeta_{2r}) && \forall r\in[R],\ \forall \boldsymbol{\zeta}\in\tilde{S} \\ & v_r(\boldsymbol{\zeta}_1,\boldsymbol{\zeta}_2) \geq b\left(\zeta_{1r} - Q_{1r}\right) - h\zeta_{2r} && \forall r\in[R],\ \forall \boldsymbol{\zeta}\in\tilde{S} \end{aligned} \tag{EC.17}$$

would not change if we added the following constraints:

$$\begin{aligned} & Q_{10} \leq \sup_{\boldsymbol{\zeta}\in\tilde{S}}\left\{\sum_{r=1}^R (\zeta_{1r}+\zeta_{2r})\right\} \\ & Q_{1r} \leq \sup_{\boldsymbol{\zeta}\in\tilde{S}}\left\{\zeta_{1r}+\zeta_{2r}\right\} \quad \forall r\in[R] \\ & \sum_{r=1}^R Q_{2r}(\boldsymbol{\xi}_1) \leq Q_{10} && \text{a.s.} \\ & v_r(\boldsymbol{\xi}_1,\boldsymbol{\xi}_2) \leq \max\left\{b(\xi_{2r}+\xi_{1r}), hQ_{1r}, b\xi_{1r} - h\xi_{2r}\right\} && \forall r\in[R],\ \text{a.s.} \end{aligned}$$

Moreover, any feasible solution to Problem (EC.17) which satisfies the above constraints will satisfy

$$0 \leq c\left(Q_{10} + \sum_{r=1}^R Q_{1r}\right) + hQ_{10} + \sum_{r=1}^R v_r(\boldsymbol{\xi}_1,\boldsymbol{\xi}_2) \quad \text{a.s.},$$

will satisfy

$$\begin{aligned} & c\left(Q_{10} + \sum_{r=1}^R Q_{1r}\right) + hQ_{10} + \sum_{r=1}^R v_r(\boldsymbol{\xi}_1,\boldsymbol{\xi}_2) \\ & \leq c\left(\sup_{\boldsymbol{\zeta}\in\tilde{S}}\left\{\sum_{r=1}^R(\zeta_{1r}+\zeta_{2r})\right\} + \sum_{r=1}^R \sup_{\boldsymbol{\zeta}\in\tilde{S}}\left\{\zeta_{1r}+\zeta_{2r}\right\}\right) + h\sup_{\boldsymbol{\zeta}\in\tilde{S}}\left\{\sum_{r=1}^R(\zeta_{1r}+\zeta_{2r})\right\} \\ & \quad + \sum_{r=1}^R \max\left\{b(\xi_{2r}+\xi_{1r}), h\sup_{\boldsymbol{\zeta}\in\tilde{S}}\left\{\zeta_{1r}+\zeta_{2r}\right\}, b\xi_{1r} - h\xi_{2r}\right\} && \text{a.s.} \\ & \leq (2c+2h+b)\max\left\{\sum_{r=1}^R(\xi_{1r}+\xi_{2r}), \sup_{\boldsymbol{\zeta}\in\tilde{S}}\left\{\sum_{r=1}^R(\zeta_{1r}+\zeta_{2r})\right\}\right\} && \text{a.s.,} \end{aligned}$$

and will satisfy the constraints of Problem (7). Therefore, it readily follows that the conditions of Theorem EC.1 are satisfied, which implies that $\underline{J} = J^*$ for Problem (7).

- Let $Q_{21}^*, \ldots, Q_{2R}^* : \mathbb{R}^R \to \mathbb{R}$ denote optimal decision rules for Problem (7) which are continuous, and let $Q_{10}^*, \ldots, Q_{1R}^*$ denote optimal first-stage decisions for Problem (7). We define the following new decision rules for all retailers $r \in [R]$ and all realizations $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2) \in \Xi$:

$$\begin{aligned} Q'_{2r}(\boldsymbol{\zeta}_1) &:= \max\left\{0, \min\left\{Q_{21}^*(\boldsymbol{\zeta}_1), Q_{10}^* - \sum_{s=1}^{r-1} Q'_{2s}(\boldsymbol{\zeta}_1)\right\}\right\} \\ v'_r(\boldsymbol{\zeta}_1,\boldsymbol{\zeta}_2) &:= \max\left\{b(\zeta_{2r}+\zeta_{1r} - Q'_{2r}(\boldsymbol{\zeta}_1) - Q_{1r}^*) - hQ'_{2r}(\boldsymbol{\zeta}_1), h(Q_{1r}^* - \zeta_{1r} - \zeta_{2r}), b\left(\zeta_{1r} - Q_{1r}^*\right) - h\zeta_{2r}\right\}. \end{aligned}$$

We observe that $(Q_{10}^*, \ldots, Q_{1R}^*, Q'_{21}, \ldots, Q'_{2R}, v'_1, \ldots, v'_R)$ is an optimal solution to Problem (7) and satisfies the conditions of Theorem EC.2, which concludes our proof that $J^* = \bar{J}$ for Problem (7).

□

B.2.3. Example from Section 8 Consider the multi-stage stochastic optimization problem

$$
\begin{aligned}
J^* = \quad & \underset{\mathbf{x},\,\mathbf{I},\,\mathbf{y}}{\text{minimize}} && \mathbb{E}\left[\sum_{t=1}^{T}\left(c_t x_t(\xi_1,\ldots,\xi_{t-1}) + y_{t+1}(\xi_1,\ldots,\xi_t)\right)\right] \\
& \text{subject to} && I_{t+1}(\xi_1,\ldots,\xi_t) = I_t(\xi_1,\ldots,\xi_{t-1}) + x_t(\xi_1,\ldots,\xi_{t-1}) - \xi_t && \text{a.s.}, \forall t\in[T] \\
& && y_{t+1}(\xi_1,\ldots,\xi_t) \ge h_t I_{t+1}(\xi_1,\ldots,\xi_t) && \text{a.s.}, \forall t\in[T] \\
& && y_{t+1}(\xi_1,\ldots,\xi_t) \ge -b_t I_{t+1}(\xi_1,\ldots,\xi_t) && \text{a.s.}, \forall t\in[T] \\
& && 0 \le x_t(\xi_1,\ldots,\xi_{t-1}) \le \bar{x}_t && \text{a.s.}, \forall t\in[T].
\end{aligned}
\tag{10}
$$

where $I_1 = 0$ and the parameters $c_t, h_t, b_t, \bar{x}_t$ are nonnegative for all periods $t\in[T]$. We assume that the stochastic process $\boldsymbol{\xi} \equiv (\xi_1,\ldots,\xi_T) \in \mathbb{R}^T$ satisfies Assumption 1 and is contained in $\Xi = \mathbb{R}^T_+$.

Proposition EC.2. *For Problem (10), $\underline{J} = J^*$. Moreover, if there exist optimal decision rules $x_t^* : \mathbb{R}^{t-1} \to \mathbb{R}$ which are continuous, then $\bar{J} = J^*$.*

Proof. Our proof is split into two parts:

- For any arbitrary $\tilde{S} \subseteq S$, we observe that the optimal cost of the optimization problem

$$
\begin{aligned}
& \underset{\mathbf{x},\,\mathbf{I},\,\mathbf{y}}{\text{minimize}} && \mathbb{E}\left[\sum_{t=1}^{T}\left(c_t x_t(\xi_1,\ldots,\xi_{t-1}) + y_{t+1}(\xi_1,\ldots,\xi_t)\right)\mathbb{I}\left\{\boldsymbol{\xi}\in\tilde{S}\right\}\right] \\
& \text{subject to} && I_{t+1}(\zeta_1,\ldots,\zeta_t) = I_t(\zeta_1,\ldots,\zeta_{t-1}) + x_t(\zeta_1,\ldots,\zeta_{t-1}) - \zeta_t && \forall\boldsymbol{\zeta}\in\tilde{S}, \forall t\in[T] \\
& && y_{t+1}(\zeta_1,\ldots,\zeta_t) \ge h_t I_{t+1}(\zeta_1,\ldots,\zeta_t) && \forall\boldsymbol{\zeta}\in\tilde{S}, \forall t\in[T] \\
& && y_{t+1}(\zeta_1,\ldots,\zeta_t) \ge -b_t I_{t+1}(\zeta_1,\ldots,\zeta_t) && \forall\boldsymbol{\zeta}\in\tilde{S}, \forall t\in[T] \\
& && 0 \le x_t(\zeta_1,\ldots,\zeta_{t-1}) \le \bar{x}_t && \forall\boldsymbol{\zeta}\in\tilde{S}, \forall t\in[T].
\end{aligned}
\tag{EC.18}
$$

would not change if we added the following constraints for each period $t\in[T]$:

$$
\begin{aligned}
& I_{t+1}(\xi_1,\ldots,\xi_t) = I_t(\xi_1,\ldots,\xi_{t-1}) + x_t(\xi_1,\ldots,\xi_{t-1}) - \xi_t && \text{a.s.} \\
& y_{t+1}(\xi_1,\ldots,\xi_t) = \max\left\{h_t I_{t+1}(\xi_1,\ldots,\xi_t), -b_t I_{t+1}(\xi_1,\ldots,\xi_t)\right\} && \text{a.s.} \\
& 0 \le x_t(\xi_1,\ldots,\xi_{t-1}) \le \bar{x}_t && \text{a.s.}
\end{aligned}
$$

Moreover, any feasible solution to Problem (EC.18) which satisfies the above constraints is feasible for Problem (10) and also satisfies

$$
0 \le \sum_{t=1}^{T}\left(c_t x_t(\xi_1,\ldots,\xi_{t-1}) + y_{t+1}(\xi_1,\ldots,\xi_t)\right) \le \left(\sum_{t=1}^{T} c_t + \sum_{t=1}^{T} h_t\right)\left(\sum_{t=1}^{T}\bar{x}_t\right) + \left(\sum_{t=1}^{T} b_t\right)\left(\sum_{t=1}^{T}\xi_t\right) \quad \text{a.s.}
$$

Therefore, it readily follows that the conditions of Theorem EC.1 are satisfied, which implies that $\underline{J} = J^*$ for Problem (10).

- Let $x_t^* : \mathbb{R}^{t-1} \to \mathbb{R}$ for each $t\in[T]$ denote optimal decision rules for Problem (10) which are continuous. We define the following new decision rules for each period $t\in[T]$ and all $\boldsymbol{\zeta}\in\Xi$:

$$
\begin{aligned}
x'_t(\zeta_1,\ldots,\zeta_{t-1}) &:= \max\{0, \min\{x_t^*(\zeta_1,\ldots,\zeta_{t-1}), \bar{x}_t\}\} \\
I'_{t+1}(\zeta_1,\ldots,\zeta_t) &:= I'_t(\zeta_1,\ldots,\zeta_{t-1}) + x'_t(\zeta_1,\ldots,\zeta_{t-1}) - \zeta_t \\
y'_{t+1}(\zeta_1,\ldots,\zeta_t) &:= \max\left\{h_t I'_{t+1}(\zeta_1,\ldots,\zeta_t), -b_t I'_{t+1}(\zeta_1,\ldots,\zeta_t)\right\},
\end{aligned}
$$

where $I'_1 = 0$. We observe that $(x'_1, \ldots, x'_T, y'_2, \ldots, y'_{T+1}, I'_1, \ldots, I'_{T+1})$ is an optimal solution to Problem (10) and satisfies the conditions of Theorem EC.2, which concludes our proof that $J^* = \bar{J}$ for Problem (10).

□

Appendix C: Proof of Theorem 1 from Section 4.2

In this appendix, we present the proof of Theorem 1. The theorem consists of asymptotic lower and upper bounds on the optimal cost of Problem (2), and we will address the proofs of the two bounds separately.

We first present the proof of the lower bound, which utilizes Theorem 2 from Section 4.2 and Theorem 3 from Section 4.4.

Theorem 1a. *Suppose Assumptions 1, 2, and 3 hold. Then, $\mathbb{P}^\infty$-almost surely we have*

$$\underaccent{\bar}{J} \le \liminf_{N \to \infty} \widehat{J}_N.$$

Proof. Recall from Assumption 1 that $b := \mathbb{E}[\exp(\|\boldsymbol{\xi}\|^a)] < \infty$ for some $a > 1$, and let $L \ge 0$ be the constant from Assumption 3. Then,

$$\begin{aligned}
\sum_{N=1}^{\infty} \mathbb{P}^N \left(\sup_{\boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j} L(1 + \|\boldsymbol{\zeta}\|) > \log N \right) &= \sum_{N=1}^{\infty} \mathbb{P}^N \left(\max_{j \in [N]} \left\{ L(1 + \|\hat{\boldsymbol{\xi}}^j\| + \epsilon_N) \right\} > \log N \right) && \text{(EC.19)} \\
&\le \sum_{N=1}^{\infty} N \mathbb{P} \left(L(1 + \|\boldsymbol{\xi}\| + \epsilon_N) > \log N \right) && \text{(EC.20)} \\
&= \sum_{N=1}^{\infty} N \mathbb{P} \left(\|\boldsymbol{\xi}\| > \frac{\log N}{L} - 1 - \epsilon_N \right) \\
&= \sum_{N=1}^{\infty} N \mathbb{P} \left(\exp(\|\boldsymbol{\xi}\|^a) > \exp\left(\left(\frac{\log N}{L} - 1 - \epsilon_N \right)^a \right) \right) \\
&\le \sum_{N=1}^{\infty} \frac{Nb}{\exp\left(\left(\frac{\log N}{L} - 1 - \epsilon_N \right)^a \right)} && \text{(EC.21)} \\
&< \infty, && \text{(EC.22)}
\end{aligned}$$

where (EC.19) follows from the definition of the uncertainty sets, (EC.20) follows from the union bound, (EC.21) follows from Markov's inequality, and (EC.22) follows from $a > 1$ and $\epsilon_N \to 0$. Therefore, the Borel-Cantelli lemma and Assumption 3 imply that the following equality holds for all sufficiently large $N \in \mathbb{N}$, $\mathbb{P}^\infty$-almost surely:

$$\begin{aligned}
\widehat{J}_N = \quad & \underset{\mathbf{x} \in \mathcal{X}}{\text{minimize}} && \frac{1}{N} \sum_{j=1}^{N} \sup_{\boldsymbol{\zeta} \in \mathcal{U}_N^j} \sum_{t=1}^{T} \mathbf{c}_t(\boldsymbol{\zeta}) \cdot \mathbf{x}_t(\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_{t-1}) \\
& \text{subject to} && \sum_{t=1}^{T} \mathbf{A}_t(\boldsymbol{\zeta}) \mathbf{x}_t(\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_{t-1}) \le \mathbf{b}(\boldsymbol{\zeta}) && \forall \boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j \\
& && \left\| \mathbf{x}_t(\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_{t-1}) \right\| \le \log N && \forall \boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j, \ t.
\end{aligned} \qquad \text{(EC.23)}$$

Moreover, since $\mathbf{c}_1(\boldsymbol{\zeta}) \in \mathbb{R}^{n_1}, \ldots, \mathbf{c}_T(\boldsymbol{\zeta}) \in \mathbb{R}^{n_T}$ are affine functions of the stochastic process, it follows from identical reasoning as (EC.19)-(EC.22) and the equivalence of ℓ_p-norms in finite-dimensional spaces that $\sup_{\boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j} \|\mathbf{c}_t(\boldsymbol{\zeta})\|_* \le \log N$ for all sufficiently large $N \in \mathbb{N}$, $\mathbb{P}^\infty$-almost surely.

We now apply Theorem 2 to obtain an asymptotic lower bound on the optimization problem in (EC.23). Indeed, let M_N be shorthand for $N^{-\frac{1}{(d+1)(d+2)}} \log N$. Then, for all sufficiently large $N \in \mathbb{N}$, $\mathbb{P}^\infty$-almost surely, and for any decision rule $\mathbf{x} \in \mathcal{X}$ which is feasible for the optimization problem in (EC.23),

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{t=1}^T \mathbf{c}_t(\boldsymbol{\xi}) \cdot \mathbf{x}_t(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{t-1}) \mathbb{I}\left\{\boldsymbol{\xi} \in \cup_{j=1}^N \mathcal{U}_N^j\right\}\right] \\
&\le \frac{1}{N} \sum_{j=1}^N \sup_{\boldsymbol{\zeta} \in \mathcal{U}_N^j} \sum_{t=1}^T \mathbf{c}_t(\boldsymbol{\zeta}) \cdot \mathbf{x}_t(\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_{t-1}) + M_N \sup_{\boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j} \left| \sum_{t=1}^T \mathbf{c}_t(\boldsymbol{\zeta}) \cdot \mathbf{x}_t(\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_{t-1}) \right| \\
&\le \frac{1}{N} \sum_{j=1}^N \sup_{\boldsymbol{\zeta} \in \mathcal{U}_N^j} \sum_{t=1}^T \mathbf{c}_t(\boldsymbol{\zeta}) \cdot \mathbf{x}_t(\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_{t-1}) + M_N \sum_{t=1}^T \sup_{\boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j} \|\mathbf{c}_t(\boldsymbol{\zeta})\|_* \|\mathbf{x}_t(\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_{t-1})\| \\
&\le \frac{1}{N} \sum_{j=1}^N \sup_{\boldsymbol{\zeta} \in \mathcal{U}_N^j} \sum_{t=1}^T \mathbf{c}_t(\boldsymbol{\zeta}) \cdot \mathbf{x}_t(\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_{t-1}) + T M_N (\log N)^2,
\end{aligned}
$$

where the first inequality follows from Theorem 2, the second inequality follows from the triangle inequality and the Cauchy-Schwartz inequality, and the third inequality follows because $\|\mathbf{c}_t(\boldsymbol{\zeta})\|_* \le \log N$ and $\|\mathbf{x}_t(\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_T)\| \le \log N$ for all sufficiently large $N \in \mathbb{N}$ and all realizations in the uncertainty sets. We remark that the above bound holds uniformly for all decision rules which are feasible for the optimization problem in (EC.23). Therefore, we have shown that the following inequality holds for all sufficiently large $N \in \mathbb{N}$, $\mathbb{P}^\infty$-almost surely:

$$
\begin{aligned}
\widehat{J}_N + T M_N (\log N)^2 \ge \quad \underset{\mathbf{x} \in \mathcal{X}}{\text{minimize}} \quad & \mathbb{E}\left[\sum_{t=1}^T \mathbf{c}_t(\boldsymbol{\xi}) \cdot \mathbf{x}_t(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{t-1}) \mathbb{I}\left\{\boldsymbol{\xi} \in \cup_{j=1}^N \mathcal{U}_N^j\right\}\right] \\
\text{subject to} \quad & \sum_{t=1}^T \mathbf{A}_t(\boldsymbol{\zeta}) \mathbf{x}_t(\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_{t-1}) \le \mathbf{b}(\boldsymbol{\zeta}) && \forall \boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j \\
& \left\|\mathbf{x}_t(\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_{t-1})\right\| \le \log N && \forall \boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j,\ t.
\end{aligned}
$$

Next, we obtain a lower bound on the right side of the above inequality by removing the last row of constraints and relaxing $\cup_{j=1}^N \mathcal{U}_N^j$ to any set which contains the stochastic process with sufficiently high probability:

$$
\begin{aligned}
\underset{\mathbf{x} \in \mathcal{X}, \tilde{S} \subseteq \Xi}{\text{minimize}} \quad & \mathbb{E}\left[\sum_{t=1}^T \mathbf{c}_t(\boldsymbol{\xi}) \cdot \mathbf{x}_t(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{t-1}) \mathbb{I}\left\{\boldsymbol{\xi} \in \tilde{S}\right\}\right] \\
\text{subject to} \quad & \sum_{t=1}^T \mathbf{A}_t(\boldsymbol{\zeta}) \mathbf{x}_t(\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_{t-1}) \le \mathbf{b}(\boldsymbol{\zeta}) \quad \forall \boldsymbol{\zeta} \in \tilde{S} \\
& \mathbb{P}\left(\boldsymbol{\xi} \in \tilde{S}\right) \ge \mathbb{P}\left(\boldsymbol{\xi} \in \cup_{j=1}^N \mathcal{U}_N^j\right).
\end{aligned}
$$

Finally, for any fixed $\rho \in (0,1)$, it follows from Theorem 3 that $\mathbb{P}(\boldsymbol{\xi} \in \cup_{j=1}^N \mathcal{U}_N^j \cap S) \ge 1 - \rho$ for all sufficiently large $N \in \mathbb{N}$, $\mathbb{P}^\infty$-almost surely.[3] Furthermore, we observe that $T M_N (\log N)^2$ converges to zero as the number

[3] We remark that Devroye and Wise (1980, Theorem 2) could be used here in lieu of Theorem 3. Our primary use of Theorem 3 is in the proof of Theorem 2.

of sample paths N tends to infinity. Therefore, we have shown that the following inequality holds, $\mathbb{P}^\infty$-almost surely:

$$\begin{aligned}\liminf_{N\to\infty} \widehat{J}_N \geq \quad & \underset{\mathbf{x}\in\mathcal{X},\tilde{S}\subseteq\Xi}{\text{minimize}} && \mathbb{E}\left[\sum_{t=1}^T \mathbf{c}_t(\boldsymbol{\xi})\cdot\mathbf{x}_t(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{t-1})\mathbb{I}\left\{\boldsymbol{\xi}\in\tilde{S}\right\}\right]\\ & \text{subject to} && \sum_{t=1}^T \mathbf{A}_t(\boldsymbol{\zeta})\mathbf{x}_t(\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_{t-1}) \leq \mathbf{b}(\boldsymbol{\zeta}) \quad \forall \boldsymbol{\zeta}\in\tilde{S}\\ & && \mathbb{P}\left(\boldsymbol{\xi}\in\tilde{S}\right)\geq 1-\rho.\end{aligned}$$

Since the inequality holds true for every $\rho\in(0,1)$, and since the optimal cost of the above optimization problem is monotone in ρ, we obtain the desired result. $\square$

We now conclude the proof of Theorem 1 by establishing its upper bound.

THEOREM 1B. *Suppose Assumption 2 holds. Then, $\mathbb{P}^\infty$-almost surely we have*

$$\limsup_{N\to\infty} \widehat{J}_N \leq \bar{J}.$$

Proof. Consider any $\rho>0$ such that there is a decision rule $\mathbf{x}\in\mathcal{X}$ which satisfies

$$\bar{\mathbb{E}}\left[\sum_{t=1}^T \mathbf{c}_t(\boldsymbol{\xi})\cdot\mathbf{x}_t(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{t-1})\right]<\infty, \tag{EC.24}$$

$$\sum_{t=1}^T \mathbf{A}_t(\boldsymbol{\zeta})\mathbf{x}_t(\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_{t-1})\leq\mathbf{b}(\boldsymbol{\zeta}) \quad \forall\boldsymbol{\zeta}\in\Xi:\ \operatorname{dist}(\boldsymbol{\zeta},S)\leq\rho. \tag{EC.25}$$

Indeed, if no such $\rho>0$ and $\mathbf{x}\in\mathcal{X}$ existed, then $\bar{J}=\infty$ and the desired result follows immediately. We recall from Assumption 2 that $\epsilon_N\to 0$ as $N\to\infty$. Therefore,

$$\begin{aligned}\limsup_{N\to\infty}\widehat{J}_N &\leq \limsup_{N\to\infty}\frac{1}{N}\sum_{j=1}^N \sup_{\boldsymbol{\zeta}\in\Xi:\ \|\boldsymbol{\zeta}-\hat{\boldsymbol{\xi}}^j\|\leq\epsilon_N}\sum_{t=1}^T \mathbf{c}_t(\boldsymbol{\zeta})\cdot\mathbf{x}_t(\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_{t-1})\\ &\leq \lim_{k\to\infty}\limsup_{N\to\infty}\frac{1}{N}\sum_{j=1}^N \sup_{\boldsymbol{\zeta}\in\Xi:\ \|\boldsymbol{\zeta}-\hat{\boldsymbol{\xi}}^j\|\leq\epsilon_k}\sum_{t=1}^T \mathbf{c}_t(\boldsymbol{\zeta})\cdot\mathbf{x}_t(\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_{t-1})\\ &= \lim_{k\to\infty}\mathbb{E}\left[\sup_{\boldsymbol{\zeta}\in\Xi:\ \|\boldsymbol{\zeta}-\boldsymbol{\xi}\|\leq\epsilon_k}\sum_{t=1}^T \mathbf{c}_t(\boldsymbol{\zeta})\cdot\mathbf{x}_t(\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_{t-1})\right] \quad \mathbb{P}^\infty\text{-almost surely}\\ &= \bar{\mathbb{E}}\left[\sum_{t=1}^T \mathbf{c}_t(\boldsymbol{\xi})\cdot\mathbf{x}_t(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{t-1})\right].\end{aligned} \tag{EC.26}$$

The first inequality holds because the decision rule is feasible but possibly suboptimal for Problem (2) for all $N\geq\min\{\bar{N}:\epsilon_{\bar{N}}\leq\rho\}$. The second inequality holds because $\epsilon_k\geq\epsilon_N$ for every fixed k and all $N\geq k$. The first equality follows from the strong law of large numbers (Erickson 1973), which holds since (EC.24) ensures that

$$\mathbb{E}\left[\max\left\{\sup_{\boldsymbol{\zeta}\in\Xi:\ \|\boldsymbol{\zeta}-\boldsymbol{\xi}\|\leq\epsilon_k}\sum_{t=1}^T \mathbf{c}_t(\boldsymbol{\zeta})\cdot\mathbf{x}_t(\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_{t-1}),0\right\}\right]<\infty$$

for all sufficiently large k. The final equality follows the definition of the local upper semicontinuous envelope. Since the set of decision rules which satisfy (EC.25) does not get smaller as $\rho \downarrow 0$, we conclude that the following holds $\mathbb{P}^\infty$-almost surely:

$$\limsup_{N\to\infty} \widehat{J}_N \le \lim_{\rho\downarrow 0} \begin{array}{ll} \underset{\mathbf{x}\in\mathcal{X}}{\text{minimize}} & \bar{\mathbb{E}}\left[\sum_{t=1}^T \mathbf{c}_t(\boldsymbol{\xi})\cdot\mathbf{x}_t(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{t-1})\right] \\ \text{subject to} & \sum_{t=1}^T \mathbf{A}_t(\boldsymbol{\zeta})\mathbf{x}_t(\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_{t-1}) \le \mathbf{b}(\boldsymbol{\zeta}) \quad \forall \boldsymbol{\zeta}\in\Xi:\ \mathrm{dist}(\boldsymbol{\zeta},S)\le\rho. \end{array}$$

This conclude the proof. $\square$

Appendix D: Proof of Theorem 2 from Section 4.2

In this appendix, we present the proof of Theorem 2. The proof is organized as follows. In Appendix D.1, we first develop a helpful intermediary bound (Lemma EC.2). In Appendix D.2, we use that bound to prove Theorem 2. In Appendix D.3, we provide for completeness the proofs of some miscellaneous and rather technical results that were used in Appendix D.2.

D.1. An intermediary result

Our proof of Theorem 2 relies on an intermediary result (Lemma EC.2), which establishes a relationship between sample robust optimization and distributionally robust optimization with the 1-Wasserstein ambiguity set. We begin by establishing the relationship for the specific case where there is a single data point.

Lemma EC.1. *Let $f:\mathbb{R}^d\to\mathbb{R}$ be measurable, $\mathcal{Z}\subseteq\mathbb{R}^d$, and $\hat{\boldsymbol{\xi}}\in\mathcal{Z}$. If $\theta_2\ge 2\theta_1\ge 0$, then*

$$\sup_{\mathbb{Q}\in\mathcal{P}(\mathcal{Z}):\ \mathbb{E}_\mathbb{Q}[\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}\|]\le\theta_1} \mathbb{E}_\mathbb{Q}[f(\boldsymbol{\xi})] \le \sup_{\boldsymbol{\zeta}\in\mathcal{Z}:\ \|\boldsymbol{\zeta}-\hat{\boldsymbol{\xi}}\|\le\theta_2} f(\boldsymbol{\zeta}) + \frac{2\theta_1}{\theta_2}\sup_{\boldsymbol{\zeta}\in\mathcal{Z}}|f(\boldsymbol{\zeta})|. \tag{EC.27}$$

Proof. We first apply the Richter-Rogonsinski Theorem (see Theorem 7.32 and Proposition 6.40 of Shapiro et al. (2009)), which says that a distributionally robust optimization problem with m moment constraints is equivalent to optimizing a weighted average of $m+1$ points. Thus,

$$\begin{aligned}\sup_{\mathbb{Q}\in\mathcal{P}(\mathcal{Z}):\ \mathbb{E}_\mathbb{Q}[\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}\|]\le\theta_1} \mathbb{E}_\mathbb{Q}[f(\boldsymbol{\xi})] &= \begin{cases} \underset{\boldsymbol{\zeta}^1,\boldsymbol{\zeta}^2\in\mathcal{Z},\lambda\in[0,1]}{\sup} & \lambda f(\boldsymbol{\zeta}^1)+(1-\lambda)f(\boldsymbol{\zeta}^2) \\ \text{subject to} & \lambda\left\|\boldsymbol{\zeta}^1-\hat{\boldsymbol{\xi}}\right\| + (1-\lambda)\left\|\boldsymbol{\zeta}^2-\hat{\boldsymbol{\xi}}\right\| \le \theta_1 \end{cases} \\ &\le \begin{cases} \underset{\boldsymbol{\zeta}^1,\boldsymbol{\zeta}^2\in\mathcal{Z},\lambda\in[0,1]}{\sup} & \lambda f(\boldsymbol{\zeta}^1)+(1-\lambda)f(\boldsymbol{\zeta}^2) \\ \text{subject to} & \lambda\left\|\boldsymbol{\zeta}^1-\hat{\boldsymbol{\xi}}\right\| \le \theta_1,\ (1-\lambda)\left\|\boldsymbol{\zeta}^2-\hat{\boldsymbol{\xi}}\right\| \le \theta_1, \end{cases}\end{aligned} \tag{EC.28}$$

where the inequality follows from relaxing the constraints on $\boldsymbol{\zeta}^1$ and $\boldsymbol{\zeta}^2$. Let us assume from this point onward that $\sup_{\boldsymbol{\zeta}\in\mathcal{Z}}|f(\boldsymbol{\zeta})|<\infty$; indeed, if $\sup_{\boldsymbol{\zeta}\in\mathcal{Z}}|f(\boldsymbol{\zeta})|=\infty$, then the inequality in (EC.27) would trivially hold since the right-hand side would equal infinity. Then, it follows from (EC.28) that

$$\sup_{\mathbb{Q}\in\mathcal{P}(\mathcal{Z}):\ \mathbb{E}_\mathbb{Q}[\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}\|]\le\theta_1} \mathbb{E}_\mathbb{Q}[f(\boldsymbol{\xi})] \le \sup_{0\le\lambda\le 1}\left\{\lambda\left(\sup_{\boldsymbol{\zeta}\in\mathcal{Z}:\ \|\boldsymbol{\zeta}-\hat{\boldsymbol{\xi}}\|\le\frac{\theta_1}{\lambda}} f(\boldsymbol{\zeta})\right) + (1-\lambda)\left(\sup_{\boldsymbol{\zeta}\in\mathcal{Z}:\ \|\boldsymbol{\zeta}-\hat{\boldsymbol{\xi}}\|\le\frac{\theta_1}{1-\lambda}} f(\boldsymbol{\zeta})\right)\right\}. \tag{EC.29}$$

We observe that the supremum over $0 \le \lambda \le 1$ is symmetric with respect to λ, in the sense that λ can be restricted to $[0, \frac{1}{2}]$ or $[\frac{1}{2}, 1]$ without loss of generality. Moreover, under the assumption that $\theta_2 \ge 2\theta_1$, the interval $[0, 1 - \frac{\theta_1}{\theta_2}]$ is a superset of the interval $[0, \frac{1}{2}]$. Combining these arguments, we conclude that the right side of (EC.29) is equal to

$$\sup_{0 \le \lambda \le 1 - \frac{\theta_1}{\theta_2}} \left\{ \lambda \left(\sup_{\zeta \in \mathcal{Z}:\, \|\zeta - \hat{\xi}\| \le \frac{\theta_1}{\lambda}} f(\zeta) \right) + (1-\lambda) \left(\sup_{\zeta \in \mathcal{Z}:\, \|\zeta - \hat{\xi}\| \le \frac{\theta_1}{1-\lambda}} f(\zeta) \right) \right\}. \tag{EC.30}$$

Next, we observe that $\frac{\theta_1}{1-\lambda} \le \theta_2$ for every feasible λ for the above optimization problem. Using this inequality, we obtain the following upper bound:

$$\begin{aligned}
&\sup_{\mathbb{Q} \in \mathcal{P}(\mathcal{Z}):\, \mathbb{E}_{\mathbb{Q}}[\|\xi - \hat{\xi}\|] \le \theta_1} \mathbb{E}_{\mathbb{Q}}[f(\xi)] \\
&\le \sup_{0 \le \lambda \le 1 - \frac{\theta_1}{\theta_2}} \left\{ \lambda \left(\sup_{\zeta \in \mathcal{Z}:\, \|\zeta - \hat{\xi}\| \le \frac{\theta_1}{\lambda}} f(\zeta) \right) + (1-\lambda) \left(\sup_{\zeta \in \mathcal{Z}:\, \|\zeta - \hat{\xi}\| \le \theta_2} f(\zeta) \right) \right\} \\
&= \sup_{\zeta \in \mathcal{Z}:\, \|\zeta - \hat{\xi}\| \le \theta_2} f(\zeta) + \sup_{0 \le \lambda \le 1 - \frac{\theta_1}{\theta_2}} \left\{ \lambda \left(\sup_{\zeta \in \mathcal{Z}:\, \|\zeta - \hat{\xi}\| \le \frac{\theta_1}{\lambda}} f(\zeta) - \sup_{\zeta \in \mathcal{Z}:\, \|\zeta - \hat{\xi}\| \le \theta_2} f(\zeta) \right) \right\},
\end{aligned} \tag{EC.31}$$

where the above equality comes from rearranging terms. For every $\frac{\theta_1}{\theta_2} \le \lambda \le 1 - \frac{\theta_1}{\theta_2}$, it immediately follows from $\frac{\theta_1}{\lambda} \le \theta_2$ that

$$\sup_{\zeta \in \mathcal{Z}:\, \|\zeta - \hat{\xi}\| \le \frac{\theta_1}{\lambda}} f(\zeta) - \sup_{\zeta \in \mathcal{Z}:\, \|\zeta - \hat{\xi}\| \le \theta_2} f(\zeta) \le 0,$$

and the above holds at equality when $\lambda = \frac{\theta_1}{\theta_2}$. Therefore,

$$\begin{aligned}
&\sup_{0 \le \lambda \le 1 - \frac{\theta_1}{\theta_2}} \left\{ \lambda \left(\sup_{\zeta \in \mathcal{Z}:\, \|\zeta - \hat{\xi}\| \le \frac{\theta_1}{\lambda}} f(\zeta) - \sup_{\zeta \in \mathcal{Z}:\, \|\zeta - \hat{\xi}\| \le \theta_2} f(\zeta) \right) \right\} \\
&= \sup_{0 \le \lambda \le \frac{\theta_1}{\theta_2}} \left\{ \lambda \left(\sup_{\zeta \in \mathcal{Z}:\, \|\zeta - \hat{\xi}\| \le \frac{\theta_1}{\lambda}} f(\zeta) - \sup_{\zeta \in \mathcal{Z}:\, \|\zeta - \hat{\xi}\| \le \theta_2} f(\zeta) \right) \right\} && \text{(EC.32)} \\
&\le \sup_{0 \le \lambda \le \frac{\theta_1}{\theta_2}} \left\{ \lambda \left(\sup_{\zeta \in \mathcal{Z}} f(\zeta) - \inf_{\zeta \in \mathcal{Z}} f(\zeta) \right) \right\} && \text{(EC.33)} \\
&\le \frac{2\theta_1}{\theta_2} \sup_{\zeta \in \mathcal{Z}} |f(\zeta)|. && \text{(EC.34)}
\end{aligned}$$

Line (EC.32) follows because we can without loss of generality restrict λ to the interval $[0, \frac{\theta_1}{\theta_2}]$. Line (EC.33) is obtained by applying the global lower and upper bounds on $f(\zeta)$. Finally, we obtain (EC.34) since

$$0 \le \sup_{\zeta \in \mathcal{Z}} f(\zeta) - \inf_{\zeta \in \mathcal{Z}} f(\zeta) \le 2 \sup_{\zeta \in \mathcal{Z}} |f(\zeta)|.$$

Combining (EC.31) and (EC.34), we obtain the desired result. $\square$

We now extend the previous lemma to the general case with more than one data point. In the following, we let $\widehat{\mathbb{P}}_N$ denote the empirical distribution of historical data $\hat{\xi}^1, \ldots, \hat{\xi}^N \in \mathbb{R}^d$, $\mathcal{Z} \subseteq \mathbb{R}^d$ be any set that contains the historical data, and $\mathsf{d}_1(\cdot,\cdot)$ denote the 1-Wasserstein distance between two probability distributions (see Section 6).

LEMMA EC.2. *Let $f:\mathbb{R}^d\to\mathbb{R}$ be measurable, $\mathcal{Z}\subseteq\mathbb{R}^d$, and $\hat{\boldsymbol{\xi}}^1,\ldots,\hat{\boldsymbol{\xi}}^N\in\mathcal{Z}$. If $\theta_2\geq 2\theta_1\geq 0$, then*

$$\sup_{\mathbb{Q}\in\mathcal{P}(\mathcal{Z}):\,\mathsf{d}_1(\mathbb{Q},\widehat{\mathbb{P}}_N)\leq\theta_1}\mathbb{E}_{\mathbb{Q}}[f(\boldsymbol{\xi})]\leq\frac{1}{N}\sum_{j=1}^{N}\sup_{\boldsymbol{\zeta}\in\mathcal{Z}:\,\|\boldsymbol{\zeta}-\hat{\boldsymbol{\xi}}^j\|\leq\theta_2}f(\boldsymbol{\zeta})+\frac{4\theta_1}{\theta_2}\sup_{\boldsymbol{\zeta}\in\mathcal{Z}}|f(\boldsymbol{\zeta})|.$$

Proof. We recall from the proof of Mohajerin Esfahani and Kuhn (2018, Theorem 4.2) that

$$\left\{\mathbb{Q}\in\mathcal{P}(\mathcal{Z}):\,\mathsf{d}_1(\mathbb{Q},\widehat{\mathbb{P}}_N)\leq\theta_1\right\}=\left\{\frac{1}{N}\sum_{j=1}^{N}\mathbb{Q}_j:\begin{array}{r}\frac{1}{N}\sum_{j=1}^{N}\mathbb{E}_{\mathbb{Q}_j}\left[\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}^j\|\right]\leq\theta_1\\ \mathbb{Q}_1,\ldots,\mathbb{Q}_N\in\mathcal{P}(\mathcal{Z})\end{array}\right\}.$$

Therefore,

$$\sup_{\mathbb{Q}\in\mathcal{P}(\mathcal{Z}):\,\mathsf{d}_1(\mathbb{Q},\widehat{\mathbb{P}}_N)\leq\theta_1}\mathbb{E}_{\mathbb{Q}}\left[f(\boldsymbol{\xi})\right]=\sup_{\boldsymbol{\gamma}\in\mathbb{R}^N_+}\left\{\frac{1}{N}\sum_{j=1}^{N}\sup_{\mathbb{Q}\in\mathcal{P}(\mathcal{Z}):\,\mathbb{E}_{\mathbb{Q}}[\|\boldsymbol{\zeta}-\hat{\boldsymbol{\xi}}^j\|]\leq\gamma_j}\mathbb{E}_{\mathbb{Q}}\left[f(\boldsymbol{\xi})\right]:\,\frac{1}{N}\sum_{j=1}^{N}\gamma_j\leq\theta_1\right\}.$$

For any choice of $\boldsymbol{\gamma}\in\mathbb{R}^N_+$, we can partition the components γ_j into those that satisfy $2\gamma_j\leq\theta_2$ and $2\gamma_j>\theta_2$. Thus,

$$\begin{aligned}&\sup_{\mathbb{Q}\in\mathcal{P}(\mathcal{Z}):\,\mathsf{d}_1(\mathbb{Q},\widehat{\mathbb{P}}_N)\leq\theta_1}\mathbb{E}_{\mathbb{Q}}\left[f(\boldsymbol{\xi})\right]\\&\leq\sup_{\boldsymbol{\gamma}\in\mathbb{R}^N_+}\left\{\frac{1}{N}\sum_{j\in[N]:\,2\gamma_j\leq\theta_2}\sup_{\mathbb{Q}\in\mathcal{P}(\mathcal{Z}):\,\mathbb{E}_{\mathbb{Q}}[\|\boldsymbol{\zeta}-\hat{\boldsymbol{\xi}}^j\|]\leq\gamma_j}\mathbb{E}_{\mathbb{Q}}\left[f(\boldsymbol{\xi})\right]+\frac{1}{N}\sum_{j\in[N]:\,2\gamma_j>\theta_2}\sup_{\boldsymbol{\zeta}\in\mathcal{Z}}|f(\boldsymbol{\zeta})|:\,\frac{1}{N}\sum_{j=1}^{N}\gamma_j\leq\theta_1\right\},\end{aligned}\tag{EC.35}$$

where the inequality follows from upper bounding each of the inner distributionally robust optimization problems for which $2\gamma_j>\theta_2$ by $\sup_{\boldsymbol{\zeta}\in\mathcal{Z}}|f(\boldsymbol{\zeta})|$. Due to the constraints on $\boldsymbol{\gamma}$, there can be at most $\frac{2N\theta_1}{\theta_2}$ components which satisfy $2\gamma_j>\theta_2$. It thus follows from (EC.35) that

$$\begin{aligned}&\sup_{\mathbb{Q}\in\mathcal{P}(\mathcal{Z}):\,\mathsf{d}_1(\mathbb{Q},\widehat{\mathbb{P}}_N)\leq\theta_1}\mathbb{E}_{\mathbb{Q}}\left[f(\boldsymbol{\xi})\right]\\&\leq\sup_{\boldsymbol{\gamma}\in\mathbb{R}^N_+}\left\{\frac{1}{N}\sum_{j\in[N]:\,2\gamma_j\leq\theta_2}\sup_{\mathbb{Q}\in\mathcal{P}(\mathcal{Z}):\,\mathbb{E}_{\mathbb{Q}}[\|\boldsymbol{\zeta}-\hat{\boldsymbol{\xi}}^j\|]\leq\gamma_j}\mathbb{E}_{\mathbb{Q}}\left[f(\boldsymbol{\xi})\right]:\,\frac{1}{N}\sum_{j=1}^{N}\gamma_j\leq\theta_1\right\}+\frac{2\theta_1}{\theta_2}\sup_{\boldsymbol{\zeta}\in\mathcal{Z}}|f(\boldsymbol{\zeta})|.\end{aligned}\tag{EC.36}$$

To conclude the proof, we apply Lemma EC.1 to each distributionally robust optimization problem in (EC.36) to obtain the following upper bounds:

$$\sup_{\mathbb{Q}\in\mathcal{P}(\mathcal{Z}):\,\mathsf{d}_1(\mathbb{Q},\widehat{\mathbb{P}}_N)\leq\theta_1}\mathbb{E}_{\mathbb{Q}}\left[f(\boldsymbol{\xi})\right]$$

$$\leq\sup_{\boldsymbol{\gamma}\in\mathbb{R}^N_+}\left\{\frac{1}{N}\sum_{j\in[N]:\,2\gamma_j\leq\theta_2}\left(\sup_{\boldsymbol{\zeta}\in\mathcal{Z}:\,\|\boldsymbol{\zeta}-\hat{\boldsymbol{\xi}}^j\|\leq\theta_2}f(\boldsymbol{\zeta})+\frac{2\gamma_j}{\theta_2}\sup_{\boldsymbol{\zeta}\in\mathcal{Z}}|f(\boldsymbol{\zeta})|\right):\frac{1}{N}\sum_{j=1}^{N}\gamma_j\leq\theta_1\right\}+\frac{2\theta_1}{\theta_2}\sup_{\boldsymbol{\zeta}\in\mathcal{Z}}|f(\boldsymbol{\zeta})|\tag{EC.37}$$

$$\leq\sup_{\boldsymbol{\gamma}\in\mathbb{R}^N_+}\left\{\frac{1}{N}\sum_{j=1}^{N}\left(\sup_{\boldsymbol{\zeta}\in\mathcal{Z}:\,\|\boldsymbol{\zeta}-\hat{\boldsymbol{\xi}}^j\|\leq\theta_2}f(\boldsymbol{\zeta})+\frac{2\gamma_j}{\theta_2}\sup_{\boldsymbol{\zeta}\in\mathcal{Z}}|f(\boldsymbol{\zeta})|\right):\frac{1}{N}\sum_{j=1}^{N}\gamma_j\leq\theta_1\right\}+\frac{2\theta_1}{\theta_2}\sup_{\boldsymbol{\zeta}\in\mathcal{Z}}|f(\boldsymbol{\zeta})|\tag{EC.38}$$

$$=\frac{1}{N}\sum_{j=1}^{N}\sup_{\boldsymbol{\zeta}\in\mathcal{Z}:\,\|\boldsymbol{\zeta}-\hat{\boldsymbol{\xi}}^j\|\leq\theta_2}f(\boldsymbol{\zeta})+\frac{4\theta_1}{\theta_2}\sup_{\boldsymbol{\zeta}\in\mathcal{Z}}|f(\boldsymbol{\zeta})|.\tag{EC.39}$$

Line (EC.37) follows from applying Lemma EC.1 to (EC.36). Line (EC.38) follows because

$$\sup_{\boldsymbol{\zeta}\in\mathcal{Z}:\,\|\boldsymbol{\zeta}-\hat{\boldsymbol{\xi}}^j\|\leq\theta_2}f(\boldsymbol{\zeta})+\frac{2\gamma_j}{\theta_2}\sup_{\boldsymbol{\zeta}\in\mathcal{Z}}|f(\boldsymbol{\zeta})|\geq 0$$

for each component that satisfies $2\gamma_j>\theta_2$, and thus adding these quantities to (EC.37) results in an upper bound. Finally, (EC.39) follows from the constraint $\frac{1}{N}\sum_{j=1}^{N}\gamma_j\leq\theta_1$. This concludes the proof. □

D.2. Proof of Theorem 2

We have established above a deterministic bound (Lemma EC.2) between sample robust optimization and distributionally robust optimization with the 1-Wasserstein ambiguity set. We will now combine that bound with a concentration inequality of Fournier and Guillin (2015) to prove Theorem 2. We remark that the following proof will employ Theorem 3 from Section 4.4 as well as notation from Section 6. For clarity of exposition, some intermediary and rather technical details of the following proof have been relegated to Appendix D.3.

THEOREM 2. *If Assumptions 1 and 2 hold, then there exists a $\bar{N} \in \mathbb{N}$, $\mathbb{P}^\infty$-almost surely, such that*

$$\mathbb{E}\left[f(\boldsymbol{\xi})\mathbb{I}\left\{\boldsymbol{\xi} \in \cup_{j=1}^N \mathcal{U}_N^j\right\}\right] \le \frac{1}{N}\sum_{j=1}^N \sup_{\boldsymbol{\zeta} \in \mathcal{U}_N^j} f(\boldsymbol{\zeta}) + M_N \sup_{\boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j} |f(\boldsymbol{\zeta})|$$

for all $N \ge \bar{N}$ and all measurable functions $f : \mathbb{R}^d \to \mathbb{R}$, where $M_N := N^{-\frac{1}{(d+1)(d+2)}} \log N$.

Proof. Let $\kappa > 0$ be the coefficient from Assumption 2, and define $\bar{\kappa} = \kappa/8$. For each $N \in \mathbb{N}$, define

$$\delta_N := \begin{cases} \bar{\kappa} N^{-\frac{1}{2}} \log N, & \text{if } d = 1, \\ \bar{\kappa} N^{-\frac{1}{d}} (\log N)^2, & \text{if } d \ge 2. \end{cases}$$

It follows from Fournier and Guillin (2015) and Assumption 1 that $\mathsf{d}_1(\mathbb{P}, \widehat{\mathbb{P}}_N) \le \delta_N$ for all sufficiently large $N \in \mathbb{N}$, $\mathbb{P}^\infty$-almost surely (see Lemma EC.3 in Appendix D.3). Therefore, for every measurable function $f : \mathbb{R}^d \to \mathbb{R}$,

$$\begin{aligned}
&\mathbb{E}\left[f(\boldsymbol{\xi})\mathbb{I}\left\{\boldsymbol{\xi} \in \cup_{j=1}^N \mathcal{U}_N^j\right\}\right] \\
&= \mathbb{E}\left[\left(f(\boldsymbol{\xi}) + \sup_{\boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j} |f(\boldsymbol{\zeta})|\right)\mathbb{I}\left\{\boldsymbol{\xi} \in \cup_{j=1}^N \mathcal{U}_N^j\right\}\right] - \left(\sup_{\boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j} |f(\boldsymbol{\zeta})|\right)\mathbb{P}(\boldsymbol{\xi} \in \cup_{j=1}^N \mathcal{U}_N^j) \\
&\le \sup_{\mathbb{Q} \in \mathcal{P}(\Xi) :\, \mathsf{d}_1(\mathbb{Q}, \widehat{\mathbb{P}}_N) \le \delta_N} \mathbb{E}_{\mathbb{Q}} \underbrace{\left[\left(f(\boldsymbol{\xi}) + \sup_{\boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j} |f(\boldsymbol{\zeta})|\right)\mathbb{I}\left\{\boldsymbol{\xi} \in \cup_{j=1}^N \mathcal{U}_N^j\right\}\right]}_{g(\boldsymbol{\xi})} - \left(\sup_{\boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j} |f(\boldsymbol{\zeta})|\right)\mathbb{P}(\boldsymbol{\xi} \in \cup_{j=1}^N \mathcal{U}_N^j),
\end{aligned} \tag{EC.40}$$

where the inequality holds for all sufficiently large $N \in \mathbb{N}$, $\mathbb{P}^\infty$-almost surely. Next, we observe that $g(\boldsymbol{\xi})$ equals zero when $\boldsymbol{\xi}$ is not an element of $\cup_{j=1}^N \mathcal{U}_N^j$ and is nonnegative otherwise. Therefore, without loss of generality, we can restrict the supremum over the expectation of $g(\boldsymbol{\xi})$ to distributions with support contained in $\cup_{j=1}^N \mathcal{U}_N^j$ (see Lemma EC.4 in Appendix D.3). Therefore, (EC.40) is equal to

$$\begin{aligned}
&\sup_{\mathbb{Q} \in \mathcal{P}(\cup_{j=1}^N \mathcal{U}_N^j) :\, \mathsf{d}_1(\mathbb{Q}, \widehat{\mathbb{P}}_N) \le \delta_N} \mathbb{E}_{\mathbb{Q}}\left[\left(f(\boldsymbol{\xi}) + \sup_{\boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j} |f(\boldsymbol{\zeta})|\right)\mathbb{I}\left\{\boldsymbol{\xi} \in \cup_{j=1}^N \mathcal{U}_N^j\right\}\right] - \left(\sup_{\boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j} |f(\boldsymbol{\zeta})|\right)\mathbb{P}(\boldsymbol{\xi} \in \cup_{j=1}^N \mathcal{U}_N^j) \\
&= \sup_{\mathbb{Q} \in \mathcal{P}(\cup_{j=1}^N \mathcal{U}_N^j) :\, \mathsf{d}_1(\mathbb{Q}, \widehat{\mathbb{P}}_N) \le \delta_N} \mathbb{E}_{\mathbb{Q}}\left[\left(f(\boldsymbol{\xi}) + \sup_{\boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j} |f(\boldsymbol{\zeta})|\right)\right] - \left(\sup_{\boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j} |f(\boldsymbol{\zeta})|\right)\mathbb{P}(\boldsymbol{\xi} \in \cup_{j=1}^N \mathcal{U}_N^j) \\
&= \sup_{\mathbb{Q} \in \mathcal{P}(\cup_{j=1}^N \mathcal{U}_N^j) :\, \mathsf{d}_1(\mathbb{Q}, \widehat{\mathbb{P}}_N) \le \delta_N} \mathbb{E}_{\mathbb{Q}}\left[f(\boldsymbol{\xi})\right] + \left(\sup_{\boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j} |f(\boldsymbol{\zeta})|\right)\mathbb{P}(\boldsymbol{\xi} \notin \cup_{j=1}^N \mathcal{U}_N^j),
\end{aligned} \tag{EC.41}$$

where the first equality follows because the support of probability distributions in the outer-most supremum is restricted to those which assign measure only on $\cup_{j=1}^N \mathcal{U}_N^j$, and the second equality follows because $\sup_{\boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j} |f(\boldsymbol{\zeta})|$ is independent of $\mathbb{Q}$. By Assumption 2 and the construction of δ_N, we have that $\epsilon_N \geq 2\delta_N$ for all sufficiently large $N \in \mathbb{N}$. Thus, it follows from Lemma EC.2 that (EC.41) is upper bounded by

$$\frac{1}{N}\sum_{j=1}^N \sup_{\boldsymbol{\zeta}\in\mathcal{U}_N^j} f(\boldsymbol{\zeta}) + \frac{4\delta_N}{\epsilon_N} \sup_{\boldsymbol{\zeta}\in\cup_{j=1}^N\mathcal{U}_N^j} |f(\boldsymbol{\zeta})| + \left(\sup_{\boldsymbol{\zeta}\in\cup_{j=1}^N\mathcal{U}_N^j} |f(\boldsymbol{\zeta})| \right) \mathbb{P}(\boldsymbol{\xi} \notin \cup_{j=1}^N \mathcal{U}_N^j). \tag{EC.42}$$

By the definition of δ_N, and since $\epsilon_N = \kappa N^{-\frac{1}{3}}$ when $d=1$ and $\epsilon_N = \kappa N^{-\frac{1}{d+1}}$ when $d \geq 2$, we have that $\frac{4\delta_N}{\epsilon_N} \leq \frac{M_N}{2}$ for all sufficiently large N. Finally, Theorem 3 implies that $\mathbb{P}(\boldsymbol{\xi} \notin \cup_{j=1}^N \mathcal{U}_N^j) \leq \frac{M_N}{2}$ for all sufficiently large N, $\mathbb{P}^\infty$-almost surely. Combining (EC.40), (EC.41), and (EC.42), we obtain the desired result. □

D.3. Miscellaneous results

We conclude Appendix D with some intermediary and technical lemmas which were used in the proof of Theorem 2. The following lemma is a corollary of Fournier and Guillin (2015, Theorem 2) and is included for completeness.

LEMMA EC.3. *Suppose Assumption 1 holds, and let*

$$\delta_N := \begin{cases} \bar{\kappa} N^{-\frac{1}{2}} \log N, & \text{if } d = 1, \\ \bar{\kappa} N^{-\frac{1}{d}} (\log N)^2, & \text{if } d \geq 2, \end{cases}$$

for any fixed $\bar{\kappa} > 0$. Then, $\mathsf{d}_1(\mathbb{P}, \widehat{\mathbb{P}}_N) \leq \delta_N$ for all sufficiently large $N \in \mathbb{N}$, $\mathbb{P}^\infty$-almost surely.

Proof. Let $\bar{N} \in \mathbb{N}$ be any index such that $\delta_N \leq 1$ for all $N \geq \bar{N}$. It follows from Assumption 1 that there exists an $a > 1$ such that $b := \mathbb{E}[\exp(\|\boldsymbol{\xi}\|^a)] < \infty$. Thus, it follows from Fournier and Guillin (2015, Theorem 2) that there exist constants $c_1, c_2 > 0$ (which depend only a, b, and d) such that for all $N \geq \bar{N}$,

$$\mathbb{P}^N \left(\mathsf{d}_1(\mathbb{P}, \widehat{\mathbb{P}}_N) > \delta_N \right) \leq \begin{cases} c_1 \exp\left(-c_2 N \delta_N^2\right), & \text{if } d = 1, \\ c_1 \exp\left(-\frac{c_2 N \delta_N^2}{(\log(2+1/\delta_N))^2}\right), & \text{if } d = 2, \\ c_1 \exp\left(-c_2 N \delta_N^d\right), & \text{if } d \geq 3. \end{cases} \tag{EC.43}$$

First, suppose $d=1$ and $N \geq \bar{N}$. Then, it follows from the definition of $\delta_N = \bar{\kappa} N^{-\frac{1}{2}} \log N$ and (EC.43) that

$$\mathbb{P}^N \left(\mathsf{d}_1(\mathbb{P}, \widehat{\mathbb{P}}_N) > \delta_N \right) \leq c_1 \exp\left(-c_2 N \delta_N^2\right) = c_1 \exp\left(-c_2 \bar{\kappa}^2 (\log N)^2\right).$$

Second, suppose $d=2$ and $N \geq \bar{N}$. Then, it follows from the definition of $\delta_N = \bar{\kappa} N^{-\frac{1}{2}} (\log N)^2$ and (EC.43) that there exists some constant $\bar{c} > 0$ (which depends only on $\bar{\kappa}$ and c_2) such that

$$\begin{aligned} \mathbb{P}^N \left(\mathsf{d}_1(\mathbb{P}, \widehat{\mathbb{P}}_N) > \delta_N \right) &\leq c_1 \exp\left(-\frac{c_2 N \delta_N^2}{\log(2+1/\delta_N)^2} \right) \\ &= c_1 \exp\left(-\frac{c_2 \bar{\kappa}^2 (\log N)^4}{\log(2 + \bar{\kappa}^{-2} N^{\frac{1}{2}} (\log N)^{-2})^2} \right) \\ &\leq c_1 \exp\left(-\frac{c_2 \bar{\kappa}^2 (\log N)^4}{\log(2 + \bar{\kappa}^{-2} N^{\frac{1}{2}})^2} \right) \\ &\leq c_1 \exp\left(-\bar{c} (\log N)^2 \right). \end{aligned}$$

Third, suppose $d \geq 3$ and $N \geq \bar{N}$. Then, it follows from the definition of $\delta_N = \bar{\kappa} N^{-\frac{1}{d}} (\log N)^2$ and (EC.43) that

$$\mathbb{P}^N \left(\mathsf{d}_1(\mathbb{P}, \widehat{\mathbb{P}}_N) > \delta_N \right) \leq c_1 \exp\left(-c_2 N \delta_N^d\right) = c_1 \exp\left(-c_2 (\log N)^{2d}\right).$$

Therefore, for any $d \geq 1$, we have shown that

$$\sum_{N=1}^{\infty} \mathbb{P}^N \left(\mathsf{d}_1(\mathbb{P}, \widehat{\mathbb{P}}_N) > \delta_N \right) < \infty,$$

and thus the desired result follows from the Borel-Cantelli lemma. $\square$

The second lemma (Lemma EC.4) demonstrates that restrictions may be placed on the support of an ambiguity set in distributionally robust optimization without loss of generality when the objective function is nonnegative.

Lemma EC.4. *Suppose $\Xi \subseteq \mathbb{R}^d$ and $\hat{\boldsymbol{\xi}}^1, \ldots, \hat{\boldsymbol{\xi}}^N \in \mathcal{Z} \subseteq \Xi$. Let $g : \Xi \to \mathbb{R}$ be any measurable function where $g(\boldsymbol{\zeta}) \geq 0$ for all $\boldsymbol{\zeta} \in \mathcal{Z}$. Then, for all $\theta \geq 0$,*

$$\sup_{\mathbb{Q} \in \mathcal{P}(\Xi):\, \mathsf{d}_1(\mathbb{Q}, \widehat{\mathbb{P}}_N) \leq \theta} \mathbb{E}_{\mathbb{Q}} \left[g(\boldsymbol{\xi}) \mathbb{I}\{\boldsymbol{\xi} \in \mathcal{Z}\} \right] = \sup_{\mathbb{Q} \in \mathcal{P}(\mathcal{Z}):\, \mathsf{d}_1(\mathbb{Q}, \widehat{\mathbb{P}}_N) \leq \theta} \mathbb{E}_{\mathbb{Q}} \left[g(\boldsymbol{\xi}) \right].$$

Proof. For notational convenience, let $\bar{g}(\boldsymbol{\zeta}) := g(\boldsymbol{\zeta}) \mathbb{I}\{\boldsymbol{\zeta} \in \mathcal{Z}\}$ for all $\boldsymbol{\zeta} \in \Xi$. It readily follows from $\mathcal{Z} \subseteq \Xi$ that

$$\sup_{\mathbb{Q} \in \mathcal{P}(\Xi):\, \mathsf{d}_1(\mathbb{Q}, \widehat{\mathbb{P}}_N) \leq \theta} \mathbb{E}_{\mathbb{Q}} \left[\bar{g}(\boldsymbol{\xi}) \right] \geq \sup_{\mathbb{Q} \in \mathcal{P}(\mathcal{Z}):\, \mathsf{d}_1(\mathbb{Q}, \widehat{\mathbb{P}}_N) \leq \theta} \mathbb{E}_{\mathbb{Q}} \left[\bar{g}(\boldsymbol{\xi}) \right] = \sup_{\mathbb{Q} \in \mathcal{P}(\mathcal{Z}):\, \mathsf{d}_1(\mathbb{Q}, \widehat{\mathbb{P}}_N) \leq \theta} \mathbb{E}_{\mathbb{Q}} \left[g(\boldsymbol{\xi}) \right],$$

where the equality holds since $\bar{g}(\boldsymbol{\zeta}) = g(\boldsymbol{\zeta})$ for all $\boldsymbol{\zeta} \in \mathcal{Z}$.

It remains to show the other direction. By the Richter-Rogonsinski Theorem (see Theorem 7.32 and Proposition 6.40 of Shapiro et al. (2009)),

$$\sup_{\mathbb{Q} \in \mathcal{P}(\Xi):\, \mathsf{d}_1(\mathbb{Q}, \widehat{\mathbb{P}}_N) \leq \theta} \mathbb{E}_{\mathbb{Q}} \left[\bar{g}(\boldsymbol{\xi}) \right] = \begin{cases} \displaystyle \sup_{\boldsymbol{\zeta}^{j1}, \boldsymbol{\zeta}^{j2} \in \Xi,\, \lambda^j \in [0,1]} & \displaystyle \frac{1}{N} \sum_{j=1}^N \left(\lambda^j \bar{g}(\boldsymbol{\zeta}^{j1}) + (1 - \lambda^j) \bar{g}(\boldsymbol{\zeta}^{j2}) \right) \\ \text{subject to} & \displaystyle \frac{1}{N} \sum_{j=1}^N \left(\lambda^j \|\boldsymbol{\zeta}^{j1} - \hat{\boldsymbol{\xi}}^j\| + (1 - \lambda^j) \|\boldsymbol{\zeta}^{j2} - \hat{\boldsymbol{\xi}}^j\| \right) \leq \theta. \end{cases}$$

For any arbitrary $\eta > 0$, let $(\bar{\boldsymbol{\zeta}}^{j1}, \bar{\boldsymbol{\zeta}}^{j2}, \bar{\lambda}^j)_{j \in [N]}$ be an η-optimal solution to the above optimization problem. We now perform a transformation on this solution. For each $j \in [N]$, define $\breve{\lambda}^j = \bar{\lambda}^j$, and for each $* \in \{1, 2\}$, define $\breve{\boldsymbol{\zeta}}^{j*} = \bar{\boldsymbol{\zeta}}^{j*}$ if $\bar{\boldsymbol{\zeta}}^{j*} \in \mathcal{Z}$ and $\breve{\boldsymbol{\zeta}}^{j*} = \hat{\boldsymbol{\xi}}^j$ otherwise. Since $\bar{g}(\boldsymbol{\zeta}) \geq 0$ for all $\boldsymbol{\zeta} \in \Xi$ and $\bar{g}(\boldsymbol{\zeta}) = 0$ for all $\boldsymbol{\zeta} \notin \mathcal{Z}$, it follows that $\bar{g}(\breve{\boldsymbol{\zeta}}^{j*}) \geq \bar{g}(\bar{\boldsymbol{\zeta}}^{j*})$. By construction, $(\breve{\boldsymbol{\zeta}}^{j1}, \breve{\boldsymbol{\zeta}}^{j2}, \breve{\lambda}^j)_{j \in [N]}$ is a feasible solution to the above optimization problem, and is also feasible for

$$\begin{aligned} \sup_{\boldsymbol{\zeta}^{j1}, \boldsymbol{\zeta}^{j2} \in \mathcal{Z},\, \lambda^j \in [0,1]} \quad & \frac{1}{N} \sum_{j=1}^N \left(\lambda^j \bar{g}(\boldsymbol{\zeta}^{j1}) + (1 - \lambda^j) \bar{g}(\boldsymbol{\zeta}^{j2}) \right) \\ \text{subject to} \quad & \frac{1}{N} \sum_{j=1}^N \left(\lambda^j \|\boldsymbol{\zeta}^{j1} - \hat{\boldsymbol{\xi}}^j\| + (1 - \lambda^j) \|\boldsymbol{\zeta}^{j2} - \hat{\boldsymbol{\xi}}^j\| \right) \leq \theta, \end{aligned}$$

where we replaced the domain of $\boldsymbol{\zeta}^{j1}$ and $\boldsymbol{\zeta}^{j2}$ by $\mathcal{Z}$. We have thus shown that

$$\sup_{\mathbb{Q}\in\mathcal{P}(\Xi):\, \mathsf{d}_1(\mathbb{P},\widehat{\mathbb{P}}_N)\le\theta} \mathbb{E}_{\mathbb{Q}}\left[\bar{g}(\boldsymbol{\xi})\right] \le \frac{1}{N}\sum_{j=1}^{N}\left(\bar{\lambda}^j\bar{g}(\bar{\boldsymbol{\zeta}}^{j1}) + (1-\bar{\lambda}^j)\bar{g}(\bar{\boldsymbol{\zeta}}^{j2})\right) + \eta$$

$$\le \frac{1}{N}\sum_{j=1}^{N}\left(\breve{\lambda}^j\bar{g}(\breve{\boldsymbol{\zeta}}^{j1}) + (1-\breve{\lambda}^j)\bar{g}(\breve{\boldsymbol{\zeta}}^{j2})\right) + \eta \le \sup_{\mathbb{Q}\in\mathcal{P}(\mathcal{Z}):\, \mathsf{d}_1(\mathbb{Q},\widehat{\mathbb{P}}_N)\le\theta} \mathbb{E}_{\mathbb{Q}}\left[\bar{g}(\boldsymbol{\xi})\right] + \eta.$$

Since $\eta > 0$ was chosen arbitrarily, and by the equivalence of $\bar{g}$ and g on $\mathcal{Z}$, we have shown the other direction. This concludes the proof. □

Appendix E: Details for Example 2 from Section 4.3

In this appendix, we provide the omitted technical details of Example 2 from Section 4.3. For convenience, we repeat the example below.

Consider the single-stage stochastic problem

$$\begin{aligned} &\underset{x_1\in\mathbb{Z}}{\text{minimize}} && x_1 \\ &\text{subject to} && x_1 \ge \xi_1 \quad \text{a.s.,} \end{aligned}$$

where the random variable ξ_1 is governed by the probability distribution $\mathbb{P}(\xi_1 > \alpha) = (1-\alpha)^k$ for fixed $k > 0$, and $\Xi = [0,2]$. We observe that the support of the random variable is $S = [0,1]$, and thus the optimal cost of the stochastic problem is $J^* = 1$. We similarly observe that the lower bound is $\underline{J} = 1$ and the upper bound, due to the integrality of the first stage decision, is $\bar{J} = 2$. If $\epsilon_N = N^{-\frac{1}{3}}$, then we prove in Appendix E that the bounds in Theorem 1 are tight under different choices of k:

Range of k	*Result*
$k \in (0,3)$	$\mathbb{P}^\infty\left(\underline{J} < \liminf_{N\to\infty}\widehat{J}_N = \limsup_{N\to\infty}\widehat{J}_N = \bar{J}\right) = 1$
$k = 3$	$\mathbb{P}^\infty\left(\underline{J} = \liminf_{N\to\infty}\widehat{J}_N < \limsup_{N\to\infty}\widehat{J}_N = \bar{J}\right) = 1$
$k \in (3,\infty)$	$\mathbb{P}^\infty\left(\underline{J} = \liminf_{N\to\infty}\widehat{J}_N = \limsup_{N\to\infty}\widehat{J}_N < \bar{J}\right) = 1$

We now prove the above bounds. To begin, we recall that $\mathbb{P}(\xi_1 > \alpha) = (1-\alpha)^k$. Thus, for any $k > 0$,

$$\begin{aligned} \underline{J} &= \lim_{\rho\downarrow 0}\min_{x_1\in\mathbb{Z}}\left\{x_1 : \mathbb{P}(x_1 \ge \xi_1) \ge 1-\rho\right\} = 1, \text{ and} \\ \bar{J} &= \lim_{\rho\downarrow 0}\min_{x_1\in\mathbb{Z}}\left\{x_1 : x_1 \ge 1+\rho\right\} = 2. \end{aligned}$$

Furthermore, given historical data, the choice of the robustness parameter $\epsilon_N = N^{-\frac{1}{3}}$, and $\Xi = [0,2]$,

$$\widehat{J}_N = \min_{x_1\in\mathbb{Z}}\left\{x_1 : x_1 \ge \zeta_1,\ \forall \zeta_1 \in \cup_{j=1}^N \mathcal{U}_N^j\right\} = \begin{cases} 1, & \text{if } \max_{j\in[N]}\hat{\xi}_1^j \le 1 - N^{-\frac{1}{3}}, \\ 2, & \text{if } \max_{j\in[N]}\hat{\xi}_1^j > 1 - N^{-\frac{1}{3}}. \end{cases}$$

We first show that

$$\mathbb{P}^\infty\left(\limsup_{N\to\infty}\widehat{J}_N = 1\right) = \begin{cases} 0, & \text{if } 0 < k \le 3, \\ 1, & \text{if } k > 3. \end{cases} \tag{Claim 1}$$

Indeed,

$$
\begin{aligned}
&\mathbb{P}^\infty\left(\limsup_{N\to\infty} \widehat{J}_N = 1\right)\\
&= \mathbb{P}^\infty\left(\max_{j\in[N]} \hat{\xi}_1^j \le 1 - N^{-\frac{1}{3}} \text{ for all sufficiently large } N\right)\\
&= \lim_{N\to\infty} \mathbb{P}^\infty\left(\max_{j\in[n]} \hat{\xi}_1^j \le 1 - n^{-\frac{1}{3}} \text{ for all } n \ge N\right)\\
&= \lim_{N\to\infty} \mathbb{P}^\infty\left(\max_{j\in[N]} \hat{\xi}_1^j \le 1 - N^{-\frac{1}{3}} \text{ and } \max_{j\in[n]} \hat{\xi}_1^j \le 1 - n^{-\frac{1}{3}} \text{ for all } n \ge N+1\right)\\
&= \lim_{N\to\infty} \mathbb{P}^N\left(\max_{j\in[N]} \hat{\xi}_1^j \le 1 - N^{-\frac{1}{3}}\right) \prod_{n=N+1}^{\infty} \mathbb{P}\left(\max_{j\in[n]} \hat{\xi}_1^j \le 1 - n^{-\frac{1}{3}} \,\middle|\, \max_{j\in[n-1]} \hat{\xi}_1^j \le 1-(n-1)^{-\frac{1}{3}}\right) &\text{(EC.44)}\\
&= \lim_{N\to\infty} \mathbb{P}^\infty\left(\max_{j\in[N]} \hat{\xi}_1^j \le 1 - N^{-\frac{1}{3}}\right) \prod_{n=N+1}^{\infty} \mathbb{P}\left(\hat{\xi}_1^n \le 1 - n^{-\frac{1}{3}} \,\middle|\, \max_{j\in[n-1]} \hat{\xi}_1^j \le 1-(n-1)^{-\frac{1}{3}}\right) &\text{(EC.45)}\\
&= \lim_{N\to\infty} \mathbb{P}\left(\xi_1 \le 1 - N^{-\frac{1}{3}}\right)^N \prod_{n=N+1}^{\infty} \mathbb{P}\left(\xi_1 \le 1 - n^{-\frac{1}{3}}\right) &\text{(EC.46)}\\
&= \lim_{N\to\infty} \left(1 - N^{-\frac{k}{3}}\right)^N \prod_{n=N+1}^{\infty} \left(1 - n^{-\frac{k}{3}}\right). &\text{(EC.47)}
\end{aligned}
$$

Line (EC.44) follows from the law of total probability. Line (EC.45) follows because, conditional on $\max_{j\in[n-1]} \hat{\xi}_1^j \le 1-(n-1)^{-\frac{1}{3}}$, we have that $\hat{\xi}_1^j \le 1 - n^{-\frac{1}{3}}$ for all $j \in [n-1]$. Line (EC.46) follows from the independence of $\hat{\xi}^j$, $j \in \mathbb{N}$. By evaluating the limit in (EC.47), we conclude the proof of Claim 1.

Next, we show that

$$
\mathbb{P}^\infty\left(\liminf_{N\to\infty} \widehat{J}_N = 1\right) = 1 \text{ if } k \ge 3. \tag{Claim 2}
$$

Indeed,

$$
\begin{aligned}
\mathbb{P}^\infty\left(\liminf_{N\to\infty} \widehat{J}_N = 1\right) &= \mathbb{P}^\infty\left(\max_{j\in[N]} \hat{\xi}_1^j \le 1 - N^{-\frac{1}{3}} \text{ for infinitely many } N\right)\\
&= \lim_{N\to\infty} \mathbb{P}^\infty\left(\max_{j\in[n]} \hat{\xi}_1^j \le 1 - n^{-\frac{1}{3}} \text{ for some } n \ge N\right)\\
&\ge \lim_{N\to\infty} \mathbb{P}^N\left(\max_{j\in[N]} \hat{\xi}_1^j \le 1 - N^{-\frac{1}{3}}\right)\\
&= \lim_{N\to\infty} \mathbb{P}\left(\xi_1 \le 1 - N^{-\frac{1}{3}}\right)^N &\text{(EC.48)}\\
&= \lim_{N\to\infty} \left(1 - N^{-\frac{k}{3}}\right)^N. &\text{(EC.49)}
\end{aligned}
$$

Line (EC.48) follows from the independence of $\hat{\xi}^j$, $j \in \mathbb{N}$. We observe that the limit in (EC.49) is strictly positive when $k \ge 3$. It follows from the Hewitt-Savage zero-one law (see, *e.g.*, Breiman (1992), Wang and Tomkins (1992)) that the event $\{\max_{j\in[N]} \hat{\xi}_1^j \le 1 - N^{-\frac{1}{3}}$ for infinitely many $N\}$ happens with probability zero or one. Thus, (EC.49) implies that the event $\{\liminf_{N\to\infty} \widehat{J}_N = 1\}$ must occur with probability one for $k \ge 3$.

Finally, we show that

$$
\mathbb{P}^\infty\left(\liminf_{N\to\infty} \widehat{J}_N = 1\right) = 0 \text{ if } 0 < k < 3. \tag{Claim 3}
$$

Indeed, suppose that $0<k<3$. Then,

$$\sum_{N=1}^{\infty} \mathbb{P}^{\infty}\left(\widehat{J}_N = 1\right) = \sum_{N=1}^{\infty} \mathbb{P}^{N}\left(\max_{j\in[N]} \hat{\xi}_1^j \le 1 - N^{-\frac{1}{3}}\right) = \sum_{N=1}^{\infty} \left(1 - N^{-\frac{k}{3}}\right)^N < \infty.$$

Therefore, it follows from the Borel-Cantelli lemma that

$$\mathbb{P}^{\infty}\left(\liminf_{N\to\infty} \widehat{J}_N = 1\right) = \mathbb{P}^{\infty}\left(\max_{j\in[N]} \hat{\xi}_1^j > 1 - N^{-\frac{1}{3}} \text{ for all sufficiently large } N\right) = 0,$$

when $0<k<3$, which proves Claim 3.

Combining Claims 1, 2, and 3 with the definitions of $\underline{J}$ and $\bar{J}$, we have shown the desired results.

Appendix F: Proof of Theorem 3 from Section 4.4

In this appendix, we present the proof of Theorem 3. Our proof techniques follow similar reasoning to Devroye and Wise (1980) and Baíllo et al. (2000) for $S_N := \cup_{j=1}^N \mathcal{U}_N^j$, which we adapt to Assumption 1. We remark that the following theorem also provides an intermediary step in the proofs of Theorems 1 and 2, which are found in Appendices C and D.

THEOREM 3. *Suppose Assumptions 1 and 2 hold. Then, $\mathbb{P}^{\infty}$-almost surely we have*

$$\lim_{N\to\infty} \left(\frac{N^{\frac{1}{d+1}}}{(\log N)^{d+1}}\right) \mathbb{P}(\boldsymbol{\xi} \notin S_N) = 0.$$

Proof. Choose any arbitrary $\eta>0$, and let $R_N := N^{\frac{1}{d+1}}(\log N)^{-(d+1)}$. Moreover, let $a>1$ be a fixed constant such that $b := \mathbb{E}[\exp(\|\boldsymbol{\xi}\|^a)] < \infty$ (the existence of a and b follows from Assumption 1). Define

$$A_N := \left\{\boldsymbol{\zeta} \in \mathbb{R}^d : \|\boldsymbol{\zeta}\| \le (\log N)^{\frac{a+1}{2a}}\right\}.$$

We begin by showing that $R_N \mathbb{P}(\boldsymbol{\xi} \notin A_N) \le \eta$ for all sufficiently large $N \in \mathbb{N}$. Indeed,

$$R_N \mathbb{P}(\boldsymbol{\xi} \notin A_N) = R_N \mathbb{P}\left(\|\boldsymbol{\xi}\| > (\log N)^{\frac{a+1}{2a}}\right) = R_N \mathbb{P}\left(\exp(\|\boldsymbol{\xi}\|^a) > \exp((\log N)^{\frac{a+1}{2}})\right) \le \frac{bR_N}{\exp((\log N)^{\frac{a+1}{2}})} \le \eta.$$

The first inequality follows from Markov's inequality and the second inequality holds for all sufficiently large $N\in\mathbb{N}$ since $a>1$.

Next, define

$$\alpha_N := \frac{\eta}{(\log N)^{\frac{d(a+1)}{2a}} \phi R_N}, \qquad B_N := \left\{\boldsymbol{\zeta}\in\mathbb{R}^d : \mathbb{P}(\|\boldsymbol{\xi}-\boldsymbol{\zeta}\| \le \epsilon_N) > \alpha_N \epsilon_N^d\right\},$$

where $\phi>0$ is a constant which depends only on d and will be defined shortly. We now show that $R_N\mathbb{P}(\boldsymbol{\xi} \notin B_N) \le 2\eta$ for all sufficiently large N. Indeed, for all sufficiently large $N\in\mathbb{N}$,

$$\begin{aligned} R_N\mathbb{P}(\boldsymbol{\xi}\notin B_N) &= R_N\mathbb{P}(\boldsymbol{\xi}\in A_N,\, \boldsymbol{\xi}\notin B_N) + R_N\mathbb{P}(\boldsymbol{\xi}\notin A_N,\, \boldsymbol{\xi}\notin B_N) \\ &\le R_N\mathbb{P}(\boldsymbol{\xi}\in A_N,\, \boldsymbol{\xi}\notin B_N) + R_N\mathbb{P}(\boldsymbol{\xi}\notin A_N) \\ &\le R_N\mathbb{P}(\boldsymbol{\xi}\in A_N,\, \boldsymbol{\xi}\notin B_N) + \eta, \end{aligned} \tag{EC.50}$$

where the final inequality follows because $R_N\mathbb{P}(\boldsymbol{\xi}\notin A_N)\le\eta$ for all sufficiently large $N\in\mathbb{N}$. Now, choose points $\boldsymbol{\zeta}^1,\ldots,\boldsymbol{\zeta}^{K_N}\in A_N$ such that $\min_{j\in[K_N]}\|\boldsymbol{\zeta}-\boldsymbol{\zeta}^j\| \le \frac{\epsilon_N}{2}$ for all $\boldsymbol{\zeta}\in A_N$. For example, one can place the points on a grid overlaying A_N. It follows from Verger-Gaugry (2005) that this can be accomplished with a number of points K_N which satisfies

$$K_N \le \phi \left(\frac{(\log N)^{\frac{a+1}{2a}}}{\epsilon_N} \right)^d, \tag{EC.51}$$

where $\phi > 0$ is a constant that depends only on d. Then, continuing from (EC.50),

$$R_N \mathbb{P}(\boldsymbol{\xi} \notin B_N) \le R_N \mathbb{P}(\boldsymbol{\xi} \in A_N, \boldsymbol{\xi} \notin B_N) + \eta \le R_N \sum_{j=1}^{K_N} \mathbb{P}\left(\|\boldsymbol{\xi} - \boldsymbol{\zeta}^j\| \le \frac{\epsilon_N}{2}, \boldsymbol{\xi} \notin B_N \right) + \eta, \tag{EC.52}$$

where the second inequality follows from the union bound. For each $j \in [K_N]$, we have two cases to consider. First, suppose there exists a realization $\boldsymbol{\zeta} \notin B_N$ such that $\|\boldsymbol{\zeta} - \boldsymbol{\zeta}^j\| \le \frac{\epsilon_N}{2}$. Then,

$$\mathbb{P}\left(\|\boldsymbol{\xi} - \boldsymbol{\zeta}^j\| \le \frac{\epsilon_N}{2}, \boldsymbol{\xi} \notin B_N \right) \le \mathbb{P}\left(\|\boldsymbol{\xi} - \boldsymbol{\zeta}^j\| \le \frac{\epsilon_N}{2} \right) \le \mathbb{P}\left(\|\boldsymbol{\xi} - \boldsymbol{\zeta}\| \le \epsilon_N \right) \le \alpha_N \epsilon_N^d,$$

where the second inequality follows because $\|\boldsymbol{\xi} - \boldsymbol{\zeta}\| \le \|\boldsymbol{\xi} - \boldsymbol{\zeta}^j\| + \|\boldsymbol{\zeta}^j - \boldsymbol{\zeta}\| \le \epsilon_N$ whenever $\|\boldsymbol{\xi} - \boldsymbol{\zeta}^j\| \le \frac{\epsilon_N}{2}$, and the third inequality follows from $\boldsymbol{\zeta} \notin B_N$. Second, suppose there does not exist a realization $\boldsymbol{\zeta} \notin B_N$ such that $\|\boldsymbol{\zeta} - \boldsymbol{\zeta}^j\| \le \frac{\epsilon_N}{2}$. Then,

$$\mathbb{P}\left(\|\boldsymbol{\xi} - \boldsymbol{\zeta}^j\| \le \frac{\epsilon_N}{2}, \boldsymbol{\xi} \notin B_N \right) = 0.$$

In each of the two cases, we have shown that

$$\mathbb{P}\left(\|\boldsymbol{\xi} - \boldsymbol{\zeta}^j\| \le \frac{\epsilon_N}{2}, \boldsymbol{\xi} \notin B_N \right) \le \alpha_N \epsilon_N^d \tag{EC.53}$$

for each $j \in [K_N]$. Therefore, we combine (EC.52) and (EC.53) to obtain the following upper bound on $R_N \mathbb{P}(\boldsymbol{\xi} \notin B_N)$ for all sufficiently large $N \in \mathbb{N}$:

$$R_N \mathbb{P}(\boldsymbol{\xi} \notin B_N) \le R_N K_N \alpha_N \epsilon_N^d + \eta \le (\log N)^{\frac{d(a+1)}{2a}} \phi R_N \alpha_N + \eta \le 2\eta. \tag{EC.54}$$

The first inequality follows from (EC.52) and (EC.53), the second inequality follows from (EC.51), and the third inequality follows from the definition of α_N.

We now prove the main result. Indeed, for all sufficiently large $N \in \mathbb{N}$,

$$R_N \mathbb{P}(\boldsymbol{\xi} \notin S_N) = R_N \mathbb{P}(\boldsymbol{\xi} \notin S_N, \boldsymbol{\xi} \in B_N) + R_N \mathbb{P}(\boldsymbol{\xi} \notin S_N, \boldsymbol{\xi} \notin B_N) \le R_N \mathbb{P}(\boldsymbol{\xi} \notin S_N, \boldsymbol{\xi} \in B_N) + 2\eta, \tag{EC.55}$$

where the equality follows from the law of total probability and the inequality follows from (EC.54). Let $\rho := \frac{d(a-1)}{2a} > 0$. Then, for all sufficiently large $N \in \mathbb{N}$:

$$\begin{aligned}
\mathbb{P}^N \left(R_N \mathbb{P}(\boldsymbol{\xi} \notin S_N) > 3\eta \right) &\le \mathbb{P}^N \left(R_N \mathbb{P}(\boldsymbol{\xi} \notin S_N, \boldsymbol{\xi} \in B_N) > \eta \right) && \text{(EC.56)}\\
&\le \eta^{-1} R_N \mathbb{E}_{\mathbb{P}^N} \left[\mathbb{P}\left(\boldsymbol{\xi} \notin S_N, \boldsymbol{\xi} \in B_N \right) \right] && \text{(EC.57)}\\
&= \eta^{-1} R_N \mathbb{E} \left[\mathbb{I}\left\{ \boldsymbol{\xi} \in B_N \right\} \mathbb{P}^N \left(\boldsymbol{\xi} \notin S_N \right) \right] && \text{(EC.58)}\\
&= \eta^{-1} R_N \mathbb{E} \left[\mathbb{I}\left\{ \boldsymbol{\xi} \in B_N \right\} \mathbb{P}^N \left(\|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^1\| > \epsilon_N, \ldots, \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^N\| > \epsilon_N \right) \right] && \text{(EC.59)}\\
&= \eta^{-1} R_N \mathbb{E} \left[\mathbb{I}\left\{ \boldsymbol{\xi} \in B_N \right\} \mathbb{P}^1 \left(\|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^1\| > \epsilon_N \right)^N \right] && \text{(EC.60)}\\
&\le \eta^{-1} R_N (1 - \alpha_N \epsilon_N^d)^N && \text{(EC.61)}\\
&\le \eta^{-1} R_N \exp\left(-N \alpha_N \epsilon_N^d \right) && \text{(EC.62)}\\
&\le \eta^{-1} R_N \exp\left(-\kappa^d N^{\frac{1}{d+1}} \alpha_N \right) && \text{(EC.63)}\\
&= \eta^{-1} R_N \exp\left(-\kappa^d \eta \phi^{-1} (\log N)^{1+\rho} \right). && \text{(EC.64)}
\end{aligned}$$

Line (EC.56) follows from (EC.55), (EC.57) follows from Markov's inequality, (EC.58) follows from Fubini's thoerem, and (EC.59) follows from the definition of S_N. Line (EC.60) follows because, given any fixed realization of $\boldsymbol{\xi}$, the random variables $\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}^1\|,\dots,\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}^N\|$ are independent. Note that the random vector $\hat{\boldsymbol{\xi}}^1$ is drawn from the measure $\mathbb{P}^1$. Line (EC.61) follows from the definition of B_N and the fact that the function $\boldsymbol{\xi}\mapsto\mathbb{I}\{\boldsymbol{\xi}\in B_N\}\,\mathbb{P}^1\left(\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}^1\|>\epsilon_N\right)^N$ is equal to zero if $\boldsymbol{\xi}\notin B_N$. Line (EC.62) follows from the mean value theorem. Line (EC.63) holds since Assumption 2 implies that $\epsilon_N\ge\kappa N^{-\frac{1}{d+1}}$. Line (EC.64) follows from the definitions of α_N, R_N, and ρ. Since $\rho>0$, it follows from (EC.64) and the definition of R_N that

$$\sum_{N=1}^{\infty}\mathbb{P}^N\left(R_N\mathbb{P}(\boldsymbol{\xi}\notin S_N)>3\eta\right)<\infty,\quad\forall\eta>0,$$

and thus the Borel-Cantelli lemma implies that $R_N\mathbb{P}(\boldsymbol{\xi}\notin S_N)\to 0$ as $N\to\infty$, $\mathbb{P}^\infty$-almost surely. □

Appendix G: Proof of Proposition 3 from Section 6

In this appendix, we present the proof of Proposition 3. We begin with the following intermediary lemma.

Lemma EC.5. *The ∞-Wasserstein ambiguity set is equivalent to*

$$\left\{\frac{1}{N}\sum_{j=1}^N\mathbb{Q}_j:\begin{array}{l}\mathbb{Q}_j\left(\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}^j\|\le\epsilon_N\right)=1\quad\forall j\in[N],\\ \mathbb{Q}_1,\dots,\mathbb{Q}_N\in\mathcal{P}(\Xi)\end{array}\right\}.$$

Proof. By the definition of the ∞-Wasserstein distance from Section 6,

$$\left\{\mathbb{Q}\in\mathcal{P}(\Xi):\mathsf{d}_\infty\left(\mathbb{Q},\widehat{\mathbb{P}}_N\right)\le\epsilon_N\right\}=\left\{\mathbb{Q}\in\mathcal{P}(\Xi):\begin{array}{l}\Pi\in\mathcal{P}(\Xi\times\Xi),\\ \Pi\left(\|\boldsymbol{\xi}-\boldsymbol{\xi}'\|\le\epsilon_N\right)=1,\text{ and}\\ \Pi\text{ is a joint distribution of }\boldsymbol{\xi}\text{ and }\boldsymbol{\xi}'\\ \text{with marginals }\mathbb{Q}\text{ and }\widehat{\mathbb{P}}_N\text{, respectively}\end{array}\right\}.\tag{EC.65}$$

Let $\bar{\boldsymbol{\xi}}^1,\dots,\bar{\boldsymbol{\xi}}^L$ be the distinct vectors among $\hat{\boldsymbol{\xi}}^1,\dots,\hat{\boldsymbol{\xi}}^N$, and let $I_1,\dots,I_L$ be index sets defined as

$$I_\ell:=\{j\in[N]:\hat{\boldsymbol{\xi}}^j=\bar{\boldsymbol{\xi}}^\ell\}.$$

For any joint distribution Π that satisfies the constraints in the ambiguity set in (EC.65), let $\mathbb{Q}_\ell$ be the conditional distribution of $\boldsymbol{\xi}$ given $\boldsymbol{\xi}'=\bar{\boldsymbol{\xi}}^\ell$. Then, for every Borel set $A\subseteq\Xi$,

$$\mathbb{Q}(\boldsymbol{\xi}\in A)=\Pi\left((\boldsymbol{\xi},\boldsymbol{\xi}')\in A\times\Xi\right)=\sum_{\ell=1}^L\Pi\left(\boldsymbol{\xi}\in A\mid\boldsymbol{\xi}'=\bar{\boldsymbol{\xi}}^\ell\right)\widehat{\mathbb{P}}_N(\boldsymbol{\xi}'=\bar{\boldsymbol{\xi}}^\ell)=\sum_{\ell=1}^L\mathbb{Q}_\ell(\boldsymbol{\xi}\in A)\frac{|I_\ell|}{N}.$$

The first equality follows because Π is a joint distribution of $\boldsymbol{\xi}$ and $\boldsymbol{\xi}'$ with marginals $\mathbb{Q}$ and $\widehat{\mathbb{P}}_N$, respectively. The second equality follows from the law of total probability. The final equality follows from the definitions of $\mathbb{Q}_\ell$ and $\widehat{\mathbb{P}}_N$. Since the above equalities holds for every Borel set, we have shown that

$$\mathbb{Q}=\sum_{\ell=1}^L\frac{|I_\ell|}{N}\mathbb{Q}_\ell.$$

Furthermore, by using similar reasoning as above, we observe that

$$\Pi\left(\|\boldsymbol{\xi}-\boldsymbol{\xi}'\|\le\epsilon_N\right)=\sum_{\ell=1}^L\Pi\left(\|\boldsymbol{\xi}-\boldsymbol{\xi}'\|\le\epsilon_N\mid\boldsymbol{\xi}'=\bar{\boldsymbol{\xi}}^\ell\right)\widehat{\mathbb{P}}_N(\boldsymbol{\xi}'=\bar{\boldsymbol{\xi}}^\ell)=\sum_{\ell=1}^L\mathbb{Q}_\ell(\|\boldsymbol{\xi}-\bar{\boldsymbol{\xi}}^\ell\|\le\epsilon_N)\frac{|I_\ell|}{N}.$$

Combining the above results, the ambiguity set from (EC.65) can be rewritten as

$$\left\{ \sum_{\ell=1}^{L} \frac{|I_\ell|}{N} \mathbb{Q}_\ell : \begin{array}{l} \sum_{\ell=1}^{L} \mathbb{Q}_\ell(\|\boldsymbol{\xi} - \bar{\boldsymbol{\xi}}^\ell\| \le \epsilon_N) \frac{|I_\ell|}{N} = 1, \\ \mathbb{Q}_1, \ldots, \mathbb{Q}_L \in \mathcal{P}(\Xi) \end{array} \right\} = \left\{ \sum_{\ell=1}^{L} \frac{|I_\ell|}{N} \mathbb{Q}_\ell : \begin{array}{ll} \mathbb{Q}_\ell(\|\boldsymbol{\xi} - \bar{\boldsymbol{\xi}}^\ell\| \le \epsilon_N) = 1 & \forall \ell \in [L], \\ \mathbb{Q}_1, \ldots, \mathbb{Q}_L \in \mathcal{P}(\Xi) & \end{array} \right\}$$
$$= \left\{ \frac{1}{N} \sum_{j=1}^{N} \mathbb{Q}_j : \begin{array}{ll} \mathbb{Q}_j(\|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^j\| \le \epsilon_N) = 1 & \forall j \in [N], \\ \mathbb{Q}_1, \ldots, \mathbb{Q}_N \in \mathcal{P}(\Xi) & \end{array} \right\}.$$

The first equality follows because $\mathbb{Q}_\ell(\|\boldsymbol{\xi} - \bar{\boldsymbol{\xi}}^\ell\| \le \epsilon_N) \le 1$ for each $\ell \in [L]$. The second equality follows because $\mathbb{Q}_\ell(\|\boldsymbol{\xi} - \bar{\boldsymbol{\xi}}^\ell\| \le \epsilon_N) = 1$ if and only if there exists $\mathbb{Q}_j \in \mathcal{P}(\Xi)$ for each $j \in I_\ell$ such that $\mathbb{Q}_j(\|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^j\| \le \epsilon_N) = 1$ and $\sum_{j \in I_\ell} \frac{1}{|I_\ell|} \mathbb{Q}_j = \mathbb{Q}_\ell$. This concludes the proof. □

We now present the proof of Proposition 3.

PROPOSITION 3. *Problem* (2) *with uncertainty sets of the form*

$$\mathcal{U}_N^j := \left\{ \boldsymbol{\zeta} \equiv (\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_T) \in \Xi : \|\boldsymbol{\zeta} - \hat{\boldsymbol{\xi}}^j\| \le \epsilon_N \right\}$$

is equivalent to ∞-*WDRO.*

Proof. It follows from Lemma EC.5 that the ∞-Wasserstein ambiguity set can be decomposed into separate distributions, each having a support that is contained in $\{\boldsymbol{\zeta} \in \Xi : \|\boldsymbol{\zeta} - \hat{\boldsymbol{\xi}}^j\| \le \epsilon_N\}$ for $j \in [N]$. Of course, these sets are exactly equal to the uncertainty sets from Section 3, and thus Lemma EC.5 implies that the ∞-Wasserstein ambiguity set is equivalent to

$$\left\{ \frac{1}{N} \sum_{j=1}^{N} \mathbb{Q}_j : \ \mathbb{Q}_j \in \mathcal{P}(\mathcal{U}_N^j) \text{ for each } j \in [N] \right\}.$$

Therefore, when $\mathcal{A}_N$ is the ∞-Wasserstein ambiguity set and each $\mathcal{U}_N^j$ is a closed balls around $\hat{\boldsymbol{\xi}}^j$ which is intersected with Ξ,

$$\sup_{\mathbb{Q} \in \mathcal{A}_N} \mathbb{E}_{\mathbb{Q}} \left[\sum_{t=1}^{T} \mathbf{c}_t(\boldsymbol{\xi}) \cdot \mathbf{x}_t(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{t-1}) \right] = \frac{1}{N} \sum_{j=1}^{N} \sup_{\mathbb{Q} \in \mathcal{P}(\mathcal{U}_N^j)} \mathbb{E}_{\mathbb{Q}} \left[\sum_{t=1}^{T} \mathbf{c}_t(\boldsymbol{\xi}) \cdot \mathbf{x}_t(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{t-1}) \right]$$
$$= \frac{1}{N} \sum_{j=1}^{N} \sup_{\boldsymbol{\zeta} \in \mathcal{U}_N^j} \sum_{t=1}^{T} \mathbf{c}_t(\boldsymbol{\zeta}) \cdot \mathbf{x}_t(\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_{t-1}).$$

Moreover, it similarly follows from Lemma EC.5 that the following inequalities are equivalent:

$$\mathbb{Q}\left(\sum_{t=1}^{T} \mathbf{A}_t(\boldsymbol{\xi}) \mathbf{x}_t(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{t-1}) \le \mathbf{b}(\boldsymbol{\xi}) \right) = 1 \quad \forall \mathbb{Q} \in \mathcal{A}_N$$
$$\frac{1}{N} \sum_{j=1}^{N} \mathbb{Q}_j \left(\sum_{t=1}^{T} \mathbf{A}_t(\boldsymbol{\xi}) \mathbf{x}_t(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{t-1}) \le \mathbf{b}(\boldsymbol{\xi}) \right) = 1 \quad \forall \mathbb{Q}_j \in \mathcal{P}(\mathcal{U}_N^j),\ j \in [N]$$
$$\mathbb{Q}_j \left(\sum_{t=1}^{T} \mathbf{A}_t(\boldsymbol{\xi}) \mathbf{x}_t(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{t-1}) \le \mathbf{b}(\boldsymbol{\xi}) \right) = 1 \quad \forall \mathbb{Q}_j \in \mathcal{P}(\mathcal{U}_N^j),\ j \in [N]$$
$$\sum_{t=1}^{T} \mathbf{A}_t(\boldsymbol{\zeta}) \mathbf{x}_t(\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_{t-1}) \le \mathbf{b}(\boldsymbol{\zeta}) \qquad \forall \boldsymbol{\zeta} \in \mathcal{U}_N^j,\ j \in [N].$$

We have thus shown that Problem (2) and Problem (6) have equivalent objective functions and constraints under the specified constructions of the uncertainty sets and ambiguity set. This concludes the proof. □

Appendix H: Proof of Proposition 4 from Section 6

In this appendix, we present the proof of Proposition 4.

PROPOSITION 4. *If $p \in [1,\infty)$ and $\epsilon_N > 0$, then a decision rule is feasible for p-WDRO only if*

$$\sum_{t=1}^{T} \mathbf{A}_t(\boldsymbol{\zeta})\mathbf{x}_t(\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_{t-1}) \le \mathbf{b}(\boldsymbol{\zeta}) \quad \forall \boldsymbol{\zeta} \in \Xi.$$

Proof. Consider any arbitrary $\bar{\boldsymbol{\xi}} \in \Xi$ such that $\bar{\boldsymbol{\xi}} \neq \hat{\boldsymbol{\xi}}^j$ for each $j \in [N]$. Let $\delta_{\bar{\boldsymbol{\xi}}}$ denote the Dirac delta distribution which satisfies $\delta_{\bar{\boldsymbol{\xi}}}(\boldsymbol{\xi} = \bar{\boldsymbol{\xi}}) = 1$, and let $\widehat{\mathbb{P}}_N := \frac{1}{N}\sum_{j=1}^N \delta_{\hat{\boldsymbol{\xi}}^j}$ be the empirical distribution of the sample paths. For any $\lambda \in (0,1)$, let the convex combination of the two distributions be given by

$$\mathbb{Q}^{\lambda}_{\bar{\boldsymbol{\xi}}} := (1-\lambda)\widehat{\mathbb{P}}_N + \lambda \delta_{\bar{\boldsymbol{\xi}}}.$$

We recall the definition of the p-Wasserstein distance between $\widehat{\mathbb{P}}_N$ and $\mathbb{Q}^{\lambda}_{\bar{\boldsymbol{\xi}}}$:

$$\mathsf{d}_p\left(\widehat{\mathbb{P}}_N, \mathbb{Q}^{\lambda}_{\bar{\boldsymbol{\xi}}}\right) = \inf\left\{ \left(\int_{\Xi\times\Xi} \|\boldsymbol{\xi} - \boldsymbol{\xi}'\|^p \, d\Pi(\boldsymbol{\xi},\boldsymbol{\xi}') \right)^{\frac{1}{p}} : \begin{array}{l} \Pi \text{ is a joint distribution of } \boldsymbol{\xi} \text{ and } \boldsymbol{\xi}' \\ \text{with marginals } \widehat{\mathbb{P}}_N \text{ and } \mathbb{Q}^{\lambda}_{\bar{\boldsymbol{\xi}}}, \text{ respectively} \end{array} \right\}. \tag{EC.66}$$

Consider a feasible joint distribution $\bar{\Pi}$ for the above optimization problem in which $\boldsymbol{\xi}' \sim \mathbb{Q}^{\lambda}_{\bar{\boldsymbol{\xi}}}$, $\boldsymbol{\xi}'' \sim \widehat{\mathbb{P}}_N$, and

$$\boldsymbol{\xi} = \begin{cases} \boldsymbol{\xi}', & \text{if } \boldsymbol{\xi}' = \hat{\boldsymbol{\xi}}^j \text{ for some } j \in [N], \\ \boldsymbol{\xi}'', & \text{otherwise.} \end{cases}$$

Indeed, we readily verify that the marginal distributions of $\boldsymbol{\xi}$ and $\boldsymbol{\xi}'$ are $\widehat{\mathbb{P}}_N$ and $\mathbb{Q}^{\lambda}_{\bar{\boldsymbol{\xi}}}$, respectively, and thus this joint distribution is feasible for the optimization problem in (EC.66). Moreover,

$$\begin{aligned}
\mathsf{d}_p\left(\widehat{\mathbb{P}}_N, \mathbb{Q}^{\lambda}_{\bar{\boldsymbol{\xi}}}\right) &\le \left(\int_{\Xi\times\Xi} \|\boldsymbol{\xi} - \boldsymbol{\xi}'\|^p \, d\bar{\Pi}(\boldsymbol{\xi},\boldsymbol{\xi}') \right)^{\frac{1}{p}} \\
&= \left(\int_{\Xi\times\Xi} \|\boldsymbol{\xi} - \boldsymbol{\xi}'\|^p \, \mathbb{I}\left\{\boldsymbol{\xi}' = \bar{\boldsymbol{\xi}}\right\} d\bar{\Pi}(\boldsymbol{\xi},\boldsymbol{\xi}') + \underbrace{\int_{\Xi\times\Xi} \|\boldsymbol{\xi} - \boldsymbol{\xi}'\|^p \, \mathbb{I}\left\{\boldsymbol{\xi}' \neq \bar{\boldsymbol{\xi}}\right\} d\bar{\Pi}(\boldsymbol{\xi},\boldsymbol{\xi}')}_{=0} \right)^{\frac{1}{p}} \\
&= \left(\frac{1}{N} \sum_{j=1}^N \lambda \left\| \hat{\boldsymbol{\xi}}^j - \bar{\boldsymbol{\xi}} \right\|^p \right)^{\frac{1}{p}}.
\end{aligned}$$

The inequality follows since $\bar{\Pi}$ is a feasible but possibly suboptimal joint distribution for the optimization problem in (EC.66). The first equality follows from splitting the integral into two cases, and observing that the second case equals zero since $\boldsymbol{\xi} = \boldsymbol{\xi}'$ whenever $\boldsymbol{\xi}' \neq \bar{\boldsymbol{\xi}}$. The final equality follows because $\boldsymbol{\xi} = \boldsymbol{\xi}''$ whenever $\boldsymbol{\xi}' = \bar{\boldsymbol{\xi}}$, and $\boldsymbol{\xi}''$ is distributed uniformly over the historical sample paths. Thus, for any arbitrary choice of $\bar{\boldsymbol{\xi}} \in \Xi$, we have shown that $\mathbb{Q}^{\lambda}_{\bar{\boldsymbol{\xi}}}$ is contained in the p-Wasserstein ambiguity set whenever $\lambda \in (0,1)$ satisfies

$$\begin{aligned}
\left(\frac{1}{N} \sum_{j=1}^N \lambda \left\| \hat{\boldsymbol{\xi}}^j - \bar{\boldsymbol{\xi}} \right\|^p \right)^{\frac{1}{p}} &\le \epsilon_N \\
\frac{1}{N} \sum_{j=1}^N \lambda \left\| \hat{\boldsymbol{\xi}}^j - \bar{\boldsymbol{\xi}} \right\|^p &\le \epsilon_N^p \\
\lambda &\le \frac{\epsilon_N^p}{\frac{1}{N}\sum_{j=1}^N \left\| \hat{\boldsymbol{\xi}}^j - \bar{\boldsymbol{\xi}} \right\|^p}.
\end{aligned}$$

Now, consider any feasible decision rule for Problem (6), *i.e.*, a decision rule $\mathbf{x} \in \mathcal{X}$ which satisfies

$$\sum_{t=1}^{T} \mathbf{A}_t(\boldsymbol{\xi})\mathbf{x}_t(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{t-1}) \le \mathbf{b}(\boldsymbol{\xi}) \quad \mathbb{Q}\text{-a.s., } \forall \mathbb{Q} \in \mathcal{A}_N. \tag{EC.67}$$

Let $\mathcal{A}_N$ be the p-Wasserstein ambiguity set for $1 \le p < \infty$ and $\epsilon_N > 0$. Then, for any arbitrary $\bar{\boldsymbol{\xi}} \in \Xi$, there exists a $\lambda \in (0,1)$ such that $\mathbb{Q}^{\lambda}_{\bar{\boldsymbol{\xi}}}$ is contained in $\mathcal{A}_N$, and so it follows from (EC.67) that the decision rule must satisfy

$$\sum_{t=1}^{T} \mathbf{A}_t(\bar{\boldsymbol{\xi}})\mathbf{x}_t(\bar{\boldsymbol{\xi}}_1,\ldots,\bar{\boldsymbol{\xi}}_{t-1}) \le \mathbf{b}(\bar{\boldsymbol{\xi}}).$$

Since $\bar{\boldsymbol{\xi}} \in \Xi$ was chosen arbitrarily, we conclude that the decision rule must satisfy

$$\sum_{t=1}^{T} \mathbf{A}_t(\boldsymbol{\zeta})\mathbf{x}_t(\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_{t-1}) \le \mathbf{b}(\boldsymbol{\zeta}) \quad \forall \boldsymbol{\zeta} \in \Xi,$$

which is what we wished to show. $\square$

Appendix I: Linear Decision Rules for Problem (6) with 1-Wasserstein Ambiguity Sets

In this appendix, we present a reformulation of linear decision rules for Problem (6) using the 1-Wasserstein ambiguity set. The performance of this data-driven approach is illustrated in Section 8.

We first review the necessary notation. Following Section 5, we focus on a specific case of Problem (6) of the form

$$\begin{aligned} &\underset{\mathbf{x} \in \mathcal{X}}{\text{minimize}} && \sup_{\mathbb{Q} \in \mathcal{A}_N} \mathbb{E}_{\mathbb{Q}} \left[\sum_{t=1}^{T} \mathbf{c}_t \cdot \mathbf{x}_t(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{t-1}) \right] \\ &\text{subject to} && \sum_{t=1}^{T} \mathbf{A}_t \mathbf{x}_t(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{t-1}) \le \mathbf{b}(\boldsymbol{\xi}) \quad \mathbb{Q}\text{-a.s., } \forall \mathbb{Q} \in \mathcal{A}_N, \end{aligned} \tag{EC.68}$$

in which $\mathbf{A}_t(\boldsymbol{\xi})$ and $\mathbf{c}_t(\boldsymbol{\xi})$ do not depend on the stochastic process. The ambiguity set is constructed as

$$\mathcal{A}_N = \left\{ \mathbb{Q} \in \mathcal{P}(\Xi) : \mathsf{d}_1(\mathbb{Q}, \widehat{\mathbb{P}}_N) \le \epsilon_N \right\},$$

where $\widehat{\mathbb{P}}_N$ is the empirical distribution of the historical data, $\epsilon_N \ge 0$ is the robustness parameter, and the 1-Wasserstein distance between two distributions is given by

$$\mathsf{d}_1(\mathbb{Q}, \mathbb{Q}') = \inf \left\{ \int_{\Xi \times \Xi} \|\boldsymbol{\xi} - \boldsymbol{\xi}'\| \, d\Pi(\boldsymbol{\xi}, \boldsymbol{\xi}') : \begin{array}{l} \Pi \text{ is a joint distribution of } \boldsymbol{\xi} \text{ and } \boldsymbol{\xi}' \\ \text{with marginals } \mathbb{Q} \text{ and } \mathbb{Q}', \text{ respectively} \end{array} \right\}.$$

We refer to Section 6 for more details on the 1-Wasserstein ambiguity set. We assume that the robustness parameter satisfies $\epsilon_N > 0$, in which case it follows from Proposition 4 in Section 6 that Problem (EC.68) is equivalent to

$$\begin{aligned} &\underset{\mathbf{x} \in \mathcal{X}}{\text{minimize}} && \sup_{\mathbb{Q} \in \mathcal{A}_N} \mathbb{E}_{\mathbb{Q}} \left[\sum_{t=1}^{T} \mathbf{c}_t \cdot \mathbf{x}_t(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{t-1}) \right] \\ &\text{subject to} && \sum_{t=1}^{T} \mathbf{A}_t \mathbf{x}_t(\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_{t-1}) \le \mathbf{b}(\boldsymbol{\zeta}) \quad \forall \boldsymbol{\zeta} \in \Xi. \end{aligned} \tag{EC.69}$$

We next present an extension of the linear decision rule approach to Problem (EC.69), in which we restrict the space of decision rules to those of the form

$$\mathbf{x}_t(\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_{t-1}) = \mathbf{x}_{t,0} + \sum_{s=1}^{t-1} \mathbf{X}_{t,s}\boldsymbol{\zeta}_s.$$

The resulting approximation of Problem (EC.69) is given by

$$\begin{aligned} &\text{minimize} && \sup_{\mathbb{Q}\in\mathcal{A}_N} \mathbb{E}_{\mathbb{Q}}\left[\sum_{t=1}^{T} \mathbf{c}_t \cdot \left(\mathbf{x}_{t,0} + \sum_{s=1}^{t-1} \mathbf{X}_{t,s}\boldsymbol{\xi}_s\right)\right] \\ &\text{subject to} && \sum_{t=1}^{T} \mathbf{A}_t \left(\mathbf{x}_{t,0} + \sum_{s=1}^{t-1} \mathbf{X}_{t,s}\boldsymbol{\zeta}_s\right) \le \mathbf{b}(\boldsymbol{\zeta}) \quad \forall \boldsymbol{\zeta}\in\Xi, \end{aligned} \tag{EC.70}$$

where the decision variables are $\mathbf{x}_{t,0}\in\mathbb{R}^{n_t}$ and $\mathbf{X}_{t,s}\in\mathbb{R}^{n_t\times d_s}$ for all $1\le s<t$.

In the remainder of this appendix, we develop a tractable reformulation of Problem (EC.70). Our reformulation, which will use similar duality techniques to those presented in Section 5, is presented as Theorem EC.3. Our reformulation requires the following assumption:

Assumption EC.1. *The set $\Xi\subseteq\mathbb{R}^d$ is a nonempty multi-dimensional box of the form $[\boldsymbol{\ell},\mathbf{u}]$, where any component of $\boldsymbol{\ell}$ might be $-\infty$ and any component of $\mathbf{u}$ may be ∞. Moreover, the norm in the 1-Wasserstein distance is equal to $\|\cdot\|_1$.*

We now present the reformulation of Problem (EC.70) given Assumption EC.1.

Theorem EC.3. *If Assumption EC.1 holds, then Problem* (EC.70) *can be reformulated by adding at most $O(md)$ additional continuous decision variables and $O(md)$ additional linear constraints. The reformulation is given by*

$$\begin{aligned} &\text{minimize} && \lambda\epsilon_N + \frac{1}{N}\sum_{j=1}^{N}\left(\sum_{t=1}^{T}\mathbf{c}_t\cdot\left(\mathbf{x}_{t,0}+\sum_{s=1}^{t-1}\mathbf{X}_{t,s}\hat{\boldsymbol{\xi}}_s^j\right)+\boldsymbol{\alpha}\cdot(\mathbf{u}-\hat{\boldsymbol{\xi}}^j)+\boldsymbol{\beta}(\hat{\boldsymbol{\xi}}^j-\boldsymbol{\ell})\right) \\ &\text{subject to} && \left\|\sum_{s=t+1}^{T}(\mathbf{X}_{s,t})^\intercal\mathbf{c}_s-\boldsymbol{\alpha}_t+\boldsymbol{\beta}_t\right\|_\infty\le\lambda && t\in[T] \\ &&& \mathbf{M}_t-\boldsymbol{\Lambda}_t=-\mathbf{B}_t+\sum_{s=t+1}^{T}\mathbf{A}_s\mathbf{X}_{s,t} && t\in[T] \\ &&& \sum_{t=1}^{T}(\mathbf{M}_t\mathbf{u}_t-\boldsymbol{\Lambda}_t\boldsymbol{\ell}_t+\mathbf{A}_t\mathbf{x}_{t,0})\le\mathbf{b}^0, \end{aligned}$$

where the auxilary decision variables are $\boldsymbol{\alpha}:=(\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_T),\boldsymbol{\beta}:=(\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_T)\in\mathbb{R}^d_+$, as well as $\mathbf{M}:=(\mathbf{M}_1,\ldots,\mathbf{M}_T),\boldsymbol{\Lambda}:=(\boldsymbol{\Lambda}_1,\ldots,\boldsymbol{\Lambda}_T)\in\mathbb{R}^{m\times d}_+$.

Proof. Following similar reasoning to Theorem 4, the constraints

$$\sum_{t=1}^{T}\mathbf{A}_t\left(\mathbf{x}_{t,0}+\sum_{s=1}^{t-1}\mathbf{X}_{t,s}\boldsymbol{\zeta}_s\right)\le\mathbf{b}^0+\sum_{t=1}^{T}\mathbf{B}_t\boldsymbol{\zeta}_t \quad \forall\boldsymbol{\zeta}\in\Xi$$

are satisfied if and only if there exist $\mathbf{M} := (\mathbf{M}_1, \ldots, \mathbf{M}_T), \mathbf{\Lambda} := (\mathbf{\Lambda}_1, \ldots, \mathbf{\Lambda}_T) \in \mathbb{R}_+^{m \times d}$ which satisfy

$$\begin{aligned} \sum_{t=1}^{T} (\mathbf{M}_t \mathbf{u}_t - \mathbf{\Lambda}_t \boldsymbol{\ell}_t + \mathbf{A}_t \mathbf{x}_{t,0}) &\le \mathbf{b}^0, \\ \mathbf{M}_t - \mathbf{\Lambda}_t &= \sum_{s=t+1}^{T} \mathbf{A}_s \mathbf{X}_{s,t} - \mathbf{B}_t, \qquad t \in [T]. \end{aligned}$$

The remainder of the proof focuses on the objective function. Note that for any fixed solution to Problem (EC.70) one can define a function $f : \Xi \to \mathbb{R}$ as follows

$$f(\boldsymbol{\zeta}) = \sum_{t=1}^{T} \mathbf{c}_t \cdot \left(\mathbf{x}_{t,0} + \sum_{s=1}^{t-1} \mathbf{X}_{t,s} \boldsymbol{\zeta}_s \right).$$

It follows from Assumption EC.1 that $\Xi \subseteq \mathbb{R}^d$ is nonempty, convex, and closed, and $-f(\cdot)$ is proper, convex, and lower semicontinuous on Ξ, thus satisfying Mohajerin Esfahani and Kuhn (2018, Assumption 4.1). Therefore, we conclude from Mohajerin Esfahani and Kuhn (2018, Equation 12b) that

$$\begin{aligned} \sup_{\mathbb{Q} \in \mathcal{A}_N} \mathbb{E}_{\mathbb{Q}} \left[\sum_{t=1}^{T} \mathbf{c}_t \cdot \left(\mathbf{x}_{t,0} + \sum_{s=1}^{t-1} \mathbf{X}_{t,s} \boldsymbol{\xi}_s \right) \right] &= \sup_{\mathbb{Q} \in \mathcal{A}_N} \mathbb{E}_{\mathbb{Q}} \left[f(\boldsymbol{\xi}) \right] \\ &= \inf_{\lambda \ge 0} \lambda \epsilon_N + \frac{1}{N} \sum_{j=1}^{N} \sup_{\boldsymbol{\zeta} \in \Xi} \left\{ f(\boldsymbol{\zeta}) - \lambda \| \boldsymbol{\zeta} - \hat{\boldsymbol{\xi}}^j \|_1 \right\} \\ &= \inf_{\lambda \ge 0} \lambda \epsilon_N + \frac{1}{N} \sum_{j=1}^{N} \underbrace{\sup_{\boldsymbol{\zeta} \in \Xi} \left\{ \sum_{t=1}^{T} \mathbf{c}_t \cdot \left(\mathbf{x}_{t,0} + \sum_{s=1}^{t-1} \mathbf{X}_{t,s} \boldsymbol{\zeta}_s \right) - \lambda \| \boldsymbol{\zeta} - \hat{\boldsymbol{\xi}}^j \|_1 \right\}}_{\gamma_j}. \end{aligned} \tag{EC.71}$$

We now reformulate the expression γ_j for each $j \in [N]$. Indeed, it follows from strong duality for linear programming that

$$\begin{aligned} \gamma_j = \quad & \underset{\boldsymbol{\alpha},\, \boldsymbol{\beta} \in \mathbb{R}_+^d}{\text{minimize}} && \sum_{t=1}^{T} \Big(\mathbf{c}_t \cdot \mathbf{x}_{t,0} + \boldsymbol{\alpha}_t \cdot (\mathbf{u}_t - \hat{\boldsymbol{\xi}}_t^j) + \boldsymbol{\beta}_t \cdot (\hat{\boldsymbol{\xi}}_t^j - \boldsymbol{\ell}_t) \Big) \\ & \text{subject to} && \left\| \sum_{s=t+1}^{T} (\mathbf{X}_{s,t})^\intercal \mathbf{c}_s - \boldsymbol{\alpha}_t + \boldsymbol{\beta}_t \right\|_\infty \le \lambda, \quad t \in [T]. \end{aligned} \tag{EC.72}$$

Remark: For any index l such that $u_l = \infty$ (alternatively, $\ell_l = -\infty$), the corresponding decision variable α_l (alternatively, β_l) should be set to zero and the term $\alpha_l(u_l - \hat{\xi}_l^j)$ (alternatively, $\beta_l(\hat{\xi}_l^j - \ell_l)$) should be dropped from the objective.

Note that problem (EC.72) is component-wise separable to d problems of the form

$$\begin{aligned} & \underset{\alpha_l,\, \beta_l \in \mathbb{R}_+}{\text{minimize}} && \alpha_l(u_l^k - \hat{\xi}_l^j) + \beta_l(\hat{\xi}_l^j - \ell_l) \\ & \text{subject to} && |g_l - \alpha_l + \beta_l| \le \lambda, \end{aligned} \tag{EC.73}$$

where $\mathbf{g} := \left(\sum_{s=2}^{T} (\mathbf{X}_{s,1})^\intercal \mathbf{c}_s, \sum_{s=3}^{T} (\mathbf{X}_{s,2})^\intercal \mathbf{c}_s, \ldots, (\mathbf{X}_{T,T-1})^\intercal \mathbf{c}_T, 0 \right) \in \mathbb{R}^d$. Moreover, $\hat{\boldsymbol{\xi}}^j \in \Xi$ implies that both $(u_l - \hat{\xi}_l^j)$ and $(\hat{\xi}_l^j - \ell_l)$ are nonnegative, and so for any fixed λ and g_l, an optimal solution of (EC.73) is given by $\alpha_l = \max\{g_l - \lambda, 0\}$ and $\beta_l = \max\{-g_l - \lambda, 0\}$ (their corresponding minimal values). This solution is independent of the value of $\hat{\xi}_l^j$, and therefore, the same variables $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ can be used in (EC.72) for all values of $j \in [N]$. Combining (EC.71) and (EC.72) and plugging the result to the objective function of (EC.70), we obtain the desired formulation. □

Appendix J: Supplement to Section 7

J.1. Reformulation of the Three-Stage Inventory Replenishment Problem

Let $I^-_{2r}(\boldsymbol{\xi}_1) = \max\{0, \xi_{1r} - Q_{1r}\}$ denote the demand at retailer r from the first half of the week that exceeded the initial inventory of the retailer. With these auxiliary decision rules, we observe that the three-stage stochastic nonlinear optimization problem from Section 7.1 is equivalent to

$$
\begin{aligned}
\underset{\mathbf{Q}\geq \mathbf{0}, \mathbf{z}\in\{0,1\}^R, \mathbf{I}}{\text{minimize}} \quad & \mathbb{E}\Bigg[c\left(Q_{10} + \sum_{r=1}^R Q_{1r}\right) + h\left(Q_{10} - \sum_{r=1}^R Q_{2r}(\boldsymbol{\xi}_1)\right) + b\left(\sum_{r=1}^R I^-_{2r}(\boldsymbol{\xi}_1)\right) \\
& \quad + f\left(\sum_{r=1}^R z_r(\boldsymbol{\xi}_1)\right) + b\left(\sum_{r=1}^R \max\{0, -I_{3r}(\boldsymbol{\xi}_1,\boldsymbol{\xi}_2)\}\right) + h\left(\sum_{r=1}^R \max\{0, I_{3r}(\boldsymbol{\xi}_1,\boldsymbol{\xi}_2)\}\right)\Bigg] \\
\text{subject to} \quad & \sum_{r=1}^R Q_{2r}(\boldsymbol{\xi}_1) \leq Q_{10} && \text{a.s.} \\
& I^-_{2r}(\boldsymbol{\xi}_1) = \max\{0, \xi_{1r} - Q_{1r}\} && \forall r \in [R], \text{ a.s.} \\
& I_{3r}(\boldsymbol{\xi}_1,\boldsymbol{\xi}_2) = Q_{1r} - \xi_{1r} + I^-_{2r}(\boldsymbol{\xi}_1,\boldsymbol{\xi}_2) + Q_{2r}(\boldsymbol{\xi}_1) - \xi_{2r} && \forall r \in [R], \text{ a.s.} \\
& z_r(\boldsymbol{\xi}_1)\mathcal{M} \geq Q_{2r}(\boldsymbol{\xi}_1) && \forall r \in [R], \text{ a.s.}
\end{aligned}
$$

After substituting the equality of $I_{3r}(\boldsymbol{\xi}_1,\boldsymbol{\xi}_2)$ into the objective function, and after adding epigraph decision rules, we obtain the following simplified formulation:

$$
\begin{aligned}
\underset{\mathbf{Q}\geq \mathbf{0}, \mathbf{z}\in\{0,1\}^R, \mathbf{I}^-, \mathbf{v}}{\text{minimize}} \quad & \mathbb{E}\left[c\left(Q_{10} + \sum_{r=1}^R Q_{1r}\right) + hQ_{10} + \sum_{r=1}^R v_r(\boldsymbol{\xi}_1,\boldsymbol{\xi}_2) + f\sum_{r=1}^R z_r(\boldsymbol{\xi}_1)\right] \\
\text{subject to} \quad & \sum_{r=1}^R Q_{2r}(\boldsymbol{\xi}_1) \leq Q_{10} && \text{a.s.} \\
& I^-_{2r}(\boldsymbol{\xi}_1) = \max\{0, \xi_{1r} - Q_{1r}\} && \forall r \in [R], \text{ a.s.} \\
& v_r(\boldsymbol{\xi}_1,\boldsymbol{\xi}_2) \geq b(\xi_{2r} + \xi_{1r} - Q_{2r}(\boldsymbol{\xi}_1) - Q_{1r}) - hQ_{2r}(\boldsymbol{\xi}_1) && \forall r \in [R], \text{ a.s.} \\
& v_r(\boldsymbol{\xi}_1,\boldsymbol{\xi}_2) \geq (h+b)I^-_{2r}(\boldsymbol{\xi}_1) + h(Q_{1r} - \xi_{1r} - \xi_{2r}) && \forall r \in [R], \text{ a.s.} \\
& z_r(\boldsymbol{\xi}_1)\mathcal{M} \geq Q_{2r}(\boldsymbol{\xi}_1) && \forall r \in [R], \text{ a.s.}
\end{aligned}
$$

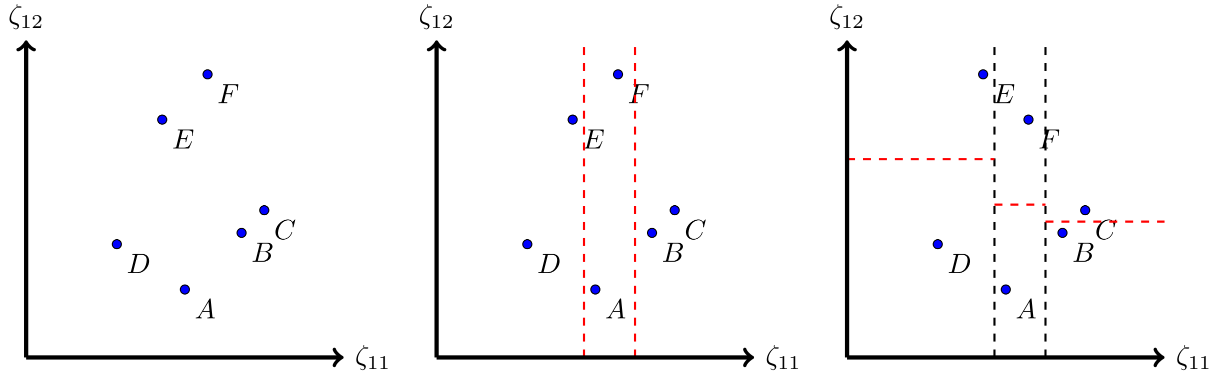
Plugging in the equality of the auxiliary decision rules $I^-_{2r}(\boldsymbol{\xi}_1)$ into the remaining constraints, and eliminating the maximization by splitting the relevant constraint, we conclude our derivation of Problem (7).

J.2. Heuristic partitioning algorithm

In this appendix, we describe the heuristic partitioning algorithm for finite adaptability which is used in our numerical experiments in Section 7. The goal of the algorithm is to construct a partition comprised of hyper-rectangular regions of the form $P^1 := [\boldsymbol{\ell}^1, \mathbf{u}^1] \times \mathbb{R}^R_+, \ldots, P^N := [\boldsymbol{\ell}^N, \mathbf{u}^N] \times \mathbb{R}^R_+$ such that the inclusion $\hat{\boldsymbol{\xi}}^j \in P^j$ is satisfied for each sample path $j \in [N]$. The output of the partitioning algorithm from this section are thus the vectors $\boldsymbol{\ell}^1, \mathbf{u}^1, \ldots, \boldsymbol{\ell}^N, \mathbf{u}^N \in [0,\infty]^R$ which define the partition of the set $\Xi := \mathbb{R}^{2R}_+$.

Our algorithm is comprised of the following steps, which are formalized in Algorithm 1 and visualized in Figure EC.1. We first define $M := \lceil N^{\frac{1}{R}} \rceil$ as the smallest integer which satisfies the inequality $M^R \geq N$.

Figure EC.1 Illustration of the partitioning algorithm



Note. **Left**: This shows the $N = 6$ historical sample paths in an example with $R = 2$ retailers. **Center**: In the first iteration ($r = 1$) of the heuristic partitioning algorithm, there are $N = 6$ historical sample paths. Since there are $R = 2$ retailers, we subdivide the historical sample paths into $M = 3$ regions along the demand of the first retailer. **Right**: In the second iteration ($r = 2$), we subdivide each of the 3 regions along the demand of the second retailer. The regions comprise the partition which is returned by the algorithm.

We then iterate over the indices $r = 1, 2, \ldots, R$. In each iteration r, we start with a partition of Ξ and then subdivide each region in that partition into at most M smaller regions by adding cuts along the first-stage demand ζ_{1r}. The borders of the new regions are determined by the historical data points such that each region contains approximately the same number of data points and the region borders are the furthest possible from the closest data point in the ζ_{1r} dimension. The progression of the algorithm is illustrated in Figure EC.1.

When the algorithm concludes its Rth iteration, we have obtained a partition $\mathcal{P} = \{P^1, \ldots, P^N\}$ of Ξ with exactly N regions such that each $\hat{\boldsymbol{\xi}}_1^j$ lies in its own region. Moreover, since Ξ is a hyperrectangle, each region P^j is also a hyperrectangle. This concludes the description of our heuristic partitioning algorithm. We note that, for simplicity, the version of the algorithm presented here ignores cases where there exist $j_1 \neq j_2 \in [N]$ and r such that $\hat{\boldsymbol{\xi}}_{1r}^{j_1} = \hat{\boldsymbol{\xi}}_{1r}^{j_2}$, which may happen if the distribution is not continuous; however, this is not the case in our numerical experiments, and, furthermore, our algorithm can be easily adjusted to address such ties.

J.3. Reformulation of the robust optimization problem

In this appendix, we present the derivation of Problem (9) from Section 7.2. Indeed, we recall from Section 7.2 that the uncertainty sets $\mathcal{U}_N^1, \ldots, \mathcal{U}_N^N$ are hyperrectangles, and it follows from Appendix J.2 that the regions $P^1, \ldots, P^K \subseteq \Xi$ that are obtained from our heuristic partitioning algorithm are hyperrectangles as well.

Algorithm 1 Partitioning algorithm

Input: Ξ, N, $\{\hat{\boldsymbol{\xi}}^1, \ldots, \hat{\boldsymbol{\xi}}^N\}$

Output: Partition $\mathcal{P}$

Initialize $M = \lceil N^{\frac{1}{R}} \rceil$, $\mathcal{P} = \{\Xi\}$

for $r := 1, \ldots, R$ **do**

  for all $P \in \mathcal{P}$ **do**

    Find $J = \{j : \hat{\boldsymbol{\xi}}^j \in P\}$

    Let $\{j_{(k)}\}_{k=1,\ldots,|J|}$ be an ordering of the indexes in J such that

$$\hat{\xi}_{1r}^{j_{(k)}} \leq \hat{\xi}_{1r}^{j_{(k+1)}}, k \in [|J|-1].$$

    Update $\mathcal{P} = \mathcal{P} \setminus \{P\}$.

    Set $K = \min\{M, |J|\}$ and $k_0 = 0$

    for all $l := 1, \ldots, K$ **do**

      Set $k_l = \max\{\lceil |J| l / K \rceil, k_{l-1} + 1\}$

      if $k < |J|$ **then**

        $\bar{\zeta}^l = \frac{\hat{\xi}_{1r}^{j_{(k)}} + \hat{\xi}_{1r}^{j_{(k+1)}}}{2}$

        Set $P^l = \{\boldsymbol{\zeta} \in P : \zeta_{1r} \leq \bar{\zeta}^l\}$ and update $P = P \setminus P^l$.

      else

        Set $K = l$ and $P^K = P$

      end if

    end for

    Update $\mathcal{P} = \mathcal{P} \cup \{P^1, \ldots, P^K\}$.

  end for

end for

Consequently, the approximation of the robust optimization problem using finite adaptability is given by

$$\begin{aligned}
\underset{\substack{\mathbf{v},\,\mathbf{Q}_1 \geq \mathbf{0},\\ \mathbf{Q}_2^k \geq 0, \mathbf{z}^k \in \{0,1\}^R}}{\text{minimize}} \quad & c\left(Q_{10} + \sum_{r=1}^R Q_{1r}\right) + hQ_{10} + \frac{1}{N}\sum_{j=1}^N \max_{k \in \mathcal{K}_j, \boldsymbol{\zeta} \in \mathcal{U}_N^j \cap P^k} \left\{ \sum_{r=1}^R v_r(\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2) + f\sum_{r=1}^R z_r^k \right\} \\
\text{subject to} \quad & \sum_{r=1}^R Q_{2r}^k \leq Q_{10} && \forall k \in [K] \\
& v_r(\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2) \geq b(\zeta_{2r} + \zeta_{1r} - Q_{2r}^k - Q_{1r}) - hQ_{2r}^k && \forall r \in [R],\, k \in [K],\, \boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j \cap P^k \\
& v_r(\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2) \geq h(Q_{1r} - \zeta_{1r} - \zeta_{2r}) && \forall r \in [R],\, k \in [K],\, \boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j \cap P^k \\
& v_r(\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2) \geq b(\zeta_{1r} - Q_{1r}) - h\zeta_{2r} && \forall r \in [R],\, k \in [K],\, \boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j \cap P^k \\
& z_r^k \mathcal{M} \geq Q_{2r}^k && \forall r \in [R],\, k \in [K],
\end{aligned} \tag{EC.74}$$

where $\mathcal{K}_j := \{k \in [N] : \mathcal{U}_N^j \cap P^k \neq \emptyset\}$ contains the indices of regions $P^1, \ldots, P^K$ that intersect the uncertainty set $\mathcal{U}_N^j$. In the remainder of this appendix, we show that the above optimization problem is equivalent to Problem (9).

We first observe that there is an optimal solution to Problem (EC.74) in which each decision rule $v_r(\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2)$ is equivalent to a function that depends only on ζ_{1r} and ζ_{2r}. Therefore, since each $\mathcal{U}_N^j \cap P^k$ is a hyperrectangle, we observe that

$$\max_{k \in \mathcal{K}_j, \boldsymbol{\zeta} \in \mathcal{U}_N^j \cap P^k} \left\{ \sum_{r=1}^R v_r(\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2) + f\sum_{r=1}^R z_r^k \right\} = \max_{k \in \mathcal{K}_j} \left\{ \sum_{r=1}^R \left(\max_{\boldsymbol{\zeta} \in \mathcal{U}_N^j \cap P^k} v_r(\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2) + f z_r^k \right) \right\}.$$

It follows from the above observation that Problem (EC.74) is equivalent to

$$\begin{aligned}
\underset{\substack{\mathbf{v},\,\mathbf{Q}_1 \geq \mathbf{0},\\ \mathbf{Q}_2^k \geq 0, \mathbf{z}^k \in \{0,1\}^R\\ \mathbf{u}^{j,k}, \mathbf{v}^j \in \mathbb{R}^R}}{\text{minimize}} \quad & c\left(Q_{10} + \sum_{r=1}^R Q_{1r}\right) + hQ_{10} + \frac{1}{N}\sum_{j=1}^N v_r^j \\
\text{subject to} \quad & \sum_{r=1}^R Q_{2r}^k \leq Q_{10} && \forall k \in [K] \\
& v_r^j \geq \sum_{r=1}^R \left(u_r^{j,k} + f z_r^k\right) && \forall r \in [R],\, j \in [N],\, k \in \mathcal{K}_j \\
& u_r^{j,k} \geq b(\zeta_{2r} + \zeta_{1r} - Q_{2r}^k - Q_{1r}) - hQ_{2r}^k && \forall r \in [R],\, k \in [K],\, \boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j \cap P^k \\
& u_r^{j,k} \geq h(Q_{1r} - \zeta_{1r} - \zeta_{2r}) && \forall r \in [R],\, k \in [K],\, \boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j \cap P^k \\
& u_r^{j,k} \geq b(\zeta_{1r} - Q_{1r}) - h\zeta_{2r} && \forall r \in [R],\, k \in [K],\, \boldsymbol{\zeta} \in \cup_{j=1}^N \mathcal{U}_N^j \cap P^k \\
& z_r^k \mathcal{M} \geq Q_{2r}^k && \forall r \in [R],\, k \in [K],
\end{aligned}$$

where at optimality each auxiliary decision variable $u_r^{j,k}$ will satisfy the equality $u_r^{j,k} = \max_{\boldsymbol{\zeta} \in \mathcal{U}^j \cap P^k} v_r(\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2)$. Applying the standard 'robust counterpart' reformulation technique, we observe that the above optimization problem can be rewritten as

$$\begin{array}{lll}
\underset{\substack{\mathbf{v}^j,\,\mathbf{Q}_1 \geq \mathbf{0},\\ \mathbf{Q}_2^k \geq 0, \mathbf{z}^k \in \{0,1\}^R\\ \mathbf{u}^{j,k}, \mathbf{v}^j \in \mathbb{R}^R}}{\text{minimize}} & c\left(Q_{10} + \displaystyle\sum_{r=1}^R Q_{1r}\right) + hQ_{10} + \dfrac{1}{N}\displaystyle\sum_{j=1}^N \sum_{r=1}^R v_r^j & \\
\text{subject to} & \displaystyle\sum_{r=1}^R Q_{2r}^k \leq Q_{10} & \forall k \in [K] \\
& v_r^j \geq \displaystyle\sum_{r=1}^R \left(u_r^{j,k} + f z_r^k\right) & \forall r \in [R],\, j \in [N],\, k \in \mathcal{K}_j \\
& u_r^{j,k} \geq \displaystyle\max_{\boldsymbol{\zeta} \in \mathcal{U}_N^j \cap P^k} b(\zeta_{2r} + \zeta_{1r} - Q_{2r}^k - Q_{1r}) - hQ_{2r}^k & \forall r \in [R],\, j \in [N],\, k \in \mathcal{K}_j \\
& u_r^{j,k} \geq \displaystyle\max_{\boldsymbol{\zeta} \in \mathcal{U}_N^j \cap P^k} h(Q_{1r} - \zeta_{1r} - \zeta_{2r}) & \forall r \in [R],\, j \in [N],\, k \in \mathcal{K}_j \\
& u_r^{j,k} \geq \displaystyle\max_{\boldsymbol{\zeta} \in \mathcal{U}_N^j \cap P^k} b\left(\zeta_{1r} - Q_{1r}\right) - h\zeta_{2r} & \forall r \in [R],\, j \in [N],\, k \in \mathcal{K}_j \\
& z_r^k \mathcal{M} \geq Q_{2r}^k & \forall r \in [R],\, k \in [K].
\end{array}$$

Let us define a lower bound $\underline{\zeta}_{tr}^{jk} := \min_{\boldsymbol{\zeta} \in \mathcal{U}_N^j \cap P^k} \zeta_{tr}$ and an upper bound $\bar{\zeta}_{tr}^{jk} := \max_{\boldsymbol{\zeta} \in \mathcal{U}_N^j \cap P^k} \zeta_{tr}$ for each period $t \in \{1, 2\}$, retailer $r \in [R]$, sample path $j \in [N]$, and region $k \in \mathcal{K}_j$. Since each set $\mathcal{U}_N^j \cap P^k$ is a hyperrectangle, and since the holding costs and backlogging costs are nonnegative ($h, b \geq 0$), we see that the above optimization problem is equivalent to Problem (9).

J.4. Supplement to Section 7.3

We conclude the present Appendix J with additional numerical results which were omitted from Section 7. In Figure EC.2, we show the impact of having a fixed cost of $f = 0.1$ versus not having a fixed cost on the replenishment decision rules obtained from SRO-FA. In Figures EC.3 and EC.4, we show the impact of the robustness parameter on SRO-FA for various numbers of retailers and sizes of training datasets with and without fixed costs.

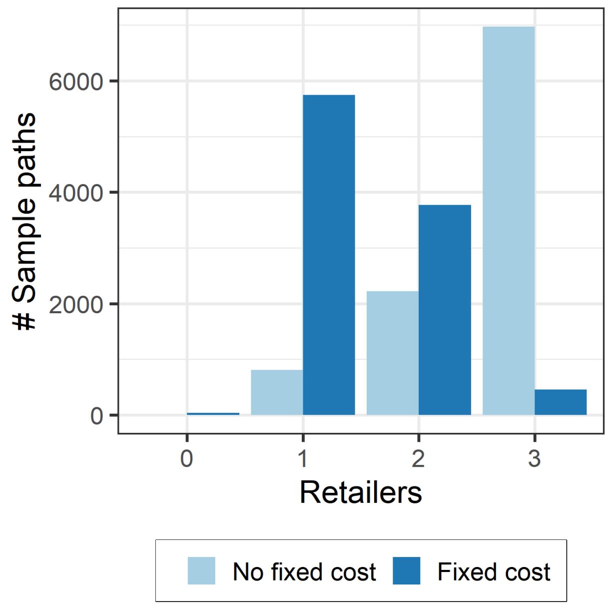
Appendix K: Supplement to Section 8.3

In this appendix, we provide supplemental numerical results for the multi-stage stochastic inventory management problem from Section 8. Specifically, the aim of this appendix is to evaluate the impact of the projection procedure, described at the end of Section 8.2, on the out-of-sample costs of SRO-LDR and SAA-LDR reported in Table 1.

Following the same notation in Section 8.2, let $\mathbf{x}^{\mathcal{A},i,\ell} = (x_1^{\mathcal{A},i,\ell}, \ldots, x_T^{\mathcal{A},i,\ell})$ be the production quantities obtained when the decision rule from approach $\mathcal{A}$ on training dataset ℓ is applied to the ith sample path in the testing dataset. For each approach $\mathcal{A}$ and training dataset ℓ, the probability that the resulting decision rule is *feasible* is approximated by

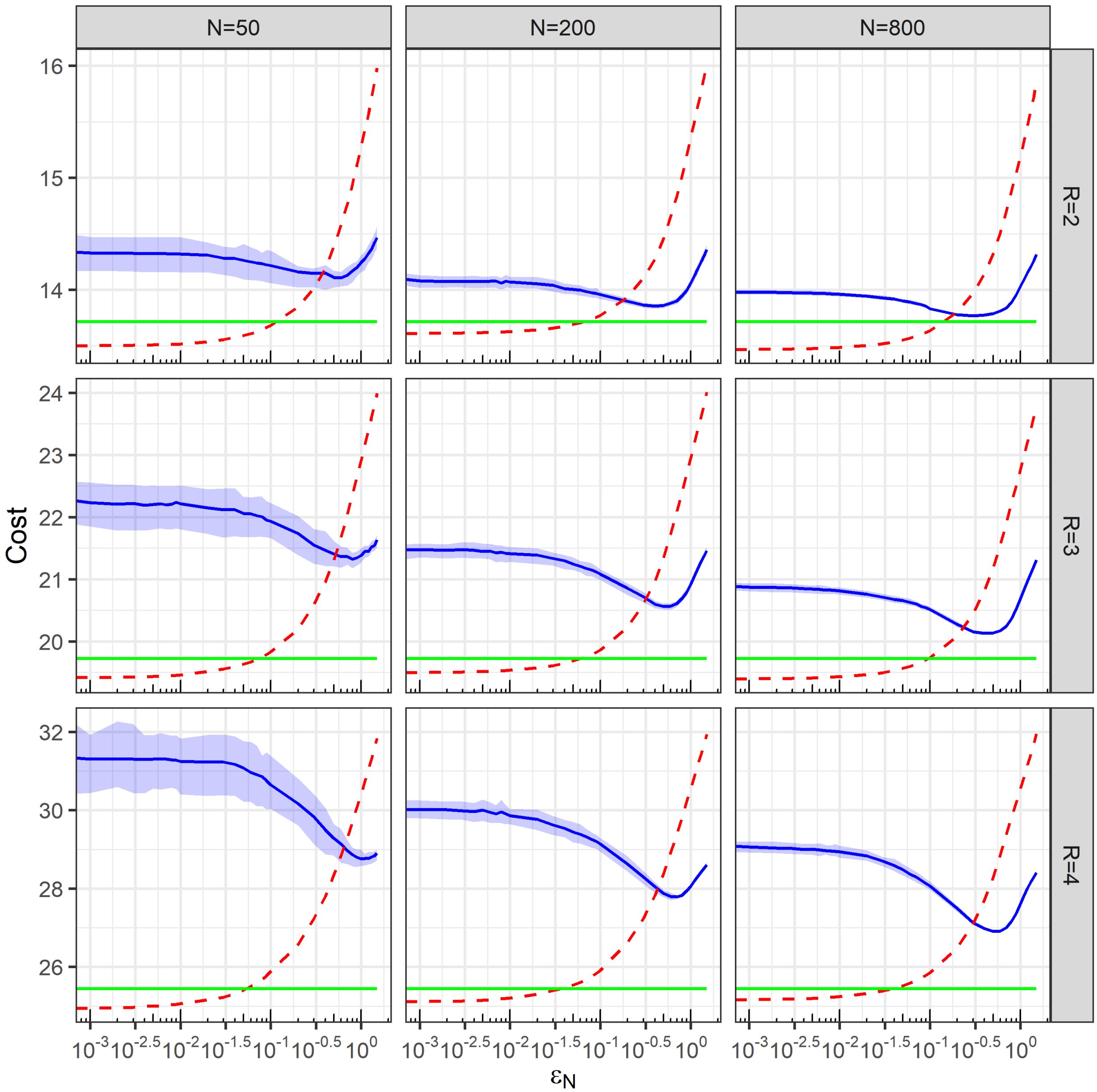
$$P^{\mathcal{A},\ell} = \frac{1}{10000} \sum_{i=1}^{10000} \mathbb{1}_{\{\mathbf{x}^{\mathcal{A},i,\ell} \in [0, \bar{\mathbf{x}}]\}},$$

Figure EC.2 Three-stage inventory replenishment problem: Histogram of replenishment decision rules for SRO-FA



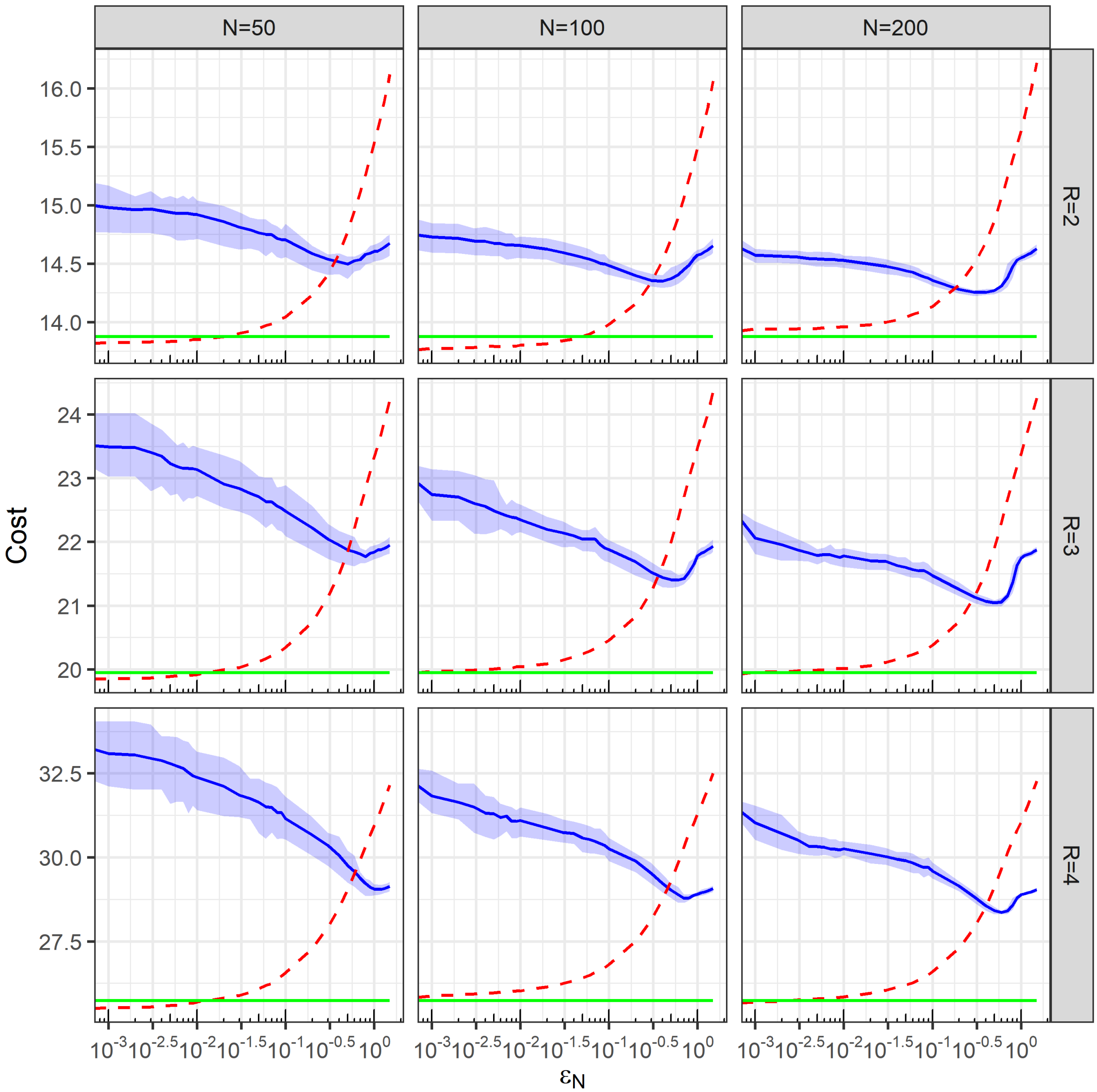
Note. The histogram corresponds to the replenishment decision rules obtained by SRO-FA in experiments where $N = 200$, $R = 3$, and $\epsilon_N = 0.7$. The light-black bars correspond to experiments in which there was no fixed cost ($f = 0$) and the dark-black bars correspond to experiments in which there was fixed cost ($f = 0.1$). For each number of retailers $0, 1, 2, 3$, the histogram shows the number of sample paths in the test set for which the replenishment policies sent a nonzero quantity of magazines to the corresponding number of retailers.

Figure EC.3 Three-stage inventory replenishment problem:
Impact of robustness parameter on SRO-FA, no fixed cost ($f = 0$)



Note. The solid black lines are the average out-of-sample cost of decision rules produced by SRO-FA, and the shaded regions are the 20th and 80th percentiles over the 50 training datasets. The dotted red lines are the average in-sample cost for SRO-FA. The green line is the benchmark cost of Problem (7). Results are shown for $N \in \{50, 200, 800\}$.

Figure EC.4 Three-stage inventory replenishment problem: Impact of robustness parameter on SRO-FA, fixed cost ($f = 0.1$)



Note. The solid black lines are the average out-of-sample cost of decision rules produced by SRO-FA, and the shaded regions are the 20th and 80th percentiles over the 50 training datasets. The dotted red lines are the average in-sample cost for SRO-FA. The green line is the benchmark cost of Problem (7). Results are shown for $N \in \{50, 100, 200\}$.

and the *infeasibility magnitude* is approximated by

$$C^{\mathcal{A},\ell} = \frac{1}{10000} \sum_{i=1}^{10000} \min_{\mathbf{y} \in [0, \bar{\mathbf{x}}]} \left\| \mathbf{x}^{\mathcal{A},i,\ell} - \mathbf{y} \right\|_1 .$$

In other words, $P^{\mathcal{A},\ell}$ tells us how frequently the projection procedure needs to be applied, and $C^{\mathcal{A},\ell}$ captures the average number of production units which are changed due to the projection procedure.

In Tables EC.1 and EC.2, for each experiment in Section 8.3 and for each approach $\mathcal{A} \in \{\text{SRO-LDR, SAA-LDR}\}$, we report the average and standard deviations for $P^{\mathcal{A},\ell}$ and $C^{\mathcal{A},\ell}$ over the 100 training datasets. For almost all choices of T, α, and N, the decision rules produced by SRO-LDR are feasible for over 93% of the sample paths in testing dataset, and their *infeasibility magnitude* is below 2 units. These results imply that the projection procedure does not significantly impact the out-of-sample cost of SRO-LDR reported in Table 1. In contrast, SAA-LDR produces decision rules which have low feasibility and high *infeasibility magnitude* when N is small. This shows, for small training datasets, that the decision rules obtained by SAA-LDR can be unreliable and require significant corrections to obtain feasible production quantities.

Table EC.1 Multi-stage stochastic inventory management: out-of-sample feasibility.

			Size of training dataset (N)			
T	α	Approach	10	25	50	100
5	0	SRO-LDR	96.3(6.6)	98.9(2.1)	99.7(0.8)	100.0(0.1)
		SAA-LDR	83.0(13.4)	96.5(4.3)	99.5(1.3)	100.0(0.1)
	0.25	SRO-LDR	93.8(7.3)	95.9(3.5)	97.3(2.3)	98.1(1.1)
		SAA-LDR	79.2(12.9)	92.3(5.3)	96.5(2.8)	98.1(1.1)
	0.5	SRO-LDR	89.7(8.6)	91.0(4.9)	91.1(3.7)	94.1(2.4)
		SAA-LDR	73.4(11.3)	85.4(4.9)	89.9(3.5)	94.0(2.3)
10	0	SRO-LDR	99.6(1.0)	100.0(0.1)	100.0(0.0)	100.0(0.0)
		SAA-LDR	61.5(24.6)	99.0(1.6)	100.0(0.1)	100.0(0.0)
	0.25	SRO-LDR	99.4(1.8)	99.9(0.3)	100.0(0.1)	100.0(0.0)
		SAA-LDR	60.2(23.9)	97.8(2.2)	99.8(0.4)	100.0(0.0)
	0.5	SRO-LDR	96.7(2.9)	97.7(1.4)	98.6(0.7)	98.9(0.3)
		SAA-LDR	57.6(22.4)	93.9(3.0)	97.7(1.2)	98.9(0.3)

Mean (standard deviation) of the percentage of the 10,000 sample paths in the testing dataset for which the linear decision rule resulted in feasible production quantities ($P^{\mathcal{A},i}$). In other words, 100% minus the above values indicates the percentage of sample paths in the testing dataset for which the production quantities needed correction. The mean and standard deviation are computed over 100 training datasets for each value of N, T, α.

Table EC.2 Multi-stage stochastic inventory management: infeasibility magnitude.

			Size of training dataset (N)			
T	α	Approach	10	25	50	100
5	0	SRO-LDR	0.5(1.6)	0.1(0.3)	0.0(0.0)	0.0(0.0)
		SAA-LDR	4.6(6.5)	0.4(0.9)	0.0(0.1)	0.0(0.0)
	0.25	SRO-LDR	0.8(1.4)	0.4(0.6)	0.2(0.3)	0.1(0.1)
		SAA-LDR	5.7(6.6)	0.9(1.1)	0.2(0.3)	0.1(0.1)
	0.5	SRO-LDR	1.7(2.1)	1.1(0.8)	1.0(0.6)	0.6(0.4)
		SAA-LDR	7.8(7.3)	2.0(1.2)	1.1(0.7)	0.6(0.4)
10	0	SRO-LDR	0.0(0.1)	0.0(0.0)	0.0(0.0)	0.0(0.0)
		SAA-LDR	218.2(1417.0)	0.1(0.2)	0.0(0.0)	0.0(0.0)
	0.25	SRO-LDR	0.0(0.2)	0.0(0.0)	0.0(0.0)	0.0(0.0)
		SAA-LDR	218.8(1417.2)	0.2(0.3)	0.0(0.0)	0.0(0.0)
	0.5	SRO-LDR	0.4(0.4)	0.2(0.2)	0.1(0.1)	0.1(0.0)
		SAA-LDR	220.3(1417.4)	0.7(0.5)	0.2(0.2)	0.1(0.0)

Mean (standard deviation) of the average *infeasibility magnitude* on the testing dataset resulting from applying the projecting procedure on the production quantity produced by the linear decision rule ($C^{\mathcal{A},i}$). The mean and standard deviation are computed over 100 training datasets for each value of N, T, α.