

Online Appendix of “Managing Inventory and Pricing with Contextual Robust Optimization”

Appendix A Supplement of Numerical Experiments

Appendix A.1 Parameters Extracted from den Boer (2014)

The coefficient of the covariates for all ten products is

$$\mathbf{A}^\dagger = (16.32 \ 19.57 \ 17.1 \ 17.7 \ 18.04 \ 19.13 \ 18.12 \ 15.88 \ 17.96 \ 17.45).$$

Here, $J = 1$, meaning that covariates are not available in this study of [den Boer \(2014\)](#). The coefficient of price, $\mathbf{B}^\dagger$, is formulated as $(\bar{\mathbf{B}} + \bar{\mathbf{B}}^\top)/2$, with $\bar{\mathbf{B}}$ detailed below:

$$\bar{\mathbf{B}} = \begin{pmatrix} -3.01 & 0.1 & 0.09 & 0.19 & 0.11 & 0.16 & 0.1 & 0.12 & 0.06 & 0.16 \\ 0.11 & -3.4 & 0.04 & 0.1 & 0.02 & 0.12 & 0.06 & 0.01 & 0.01 & 0.03 \\ 0.03 & 0.09 & -2.49 & 0.18 & 0.07 & 0.15 & 0.05 & 0.13 & 0.15 & 0.17 \\ 0.1 & 0.02 & 0.1 & -2.37 & 0.17 & 0.03 & 0.08 & 0.08 & 0.13 & 0.15 \\ 0.04 & 0.03 & 0.1 & 0.11 & -2.22 & 0.06 & 0.17 & 0.1 & 0.04 & 0.16 \\ 0.16 & 0.12 & 0.08 & 0.09 & 0.01 & -2.55 & 0.15 & 0.08 & 0.08 & 0.11 \\ 0.17 & 0.05 & 0.16 & 0.09 & 0.05 & 0.07 & -2.02 & 0.07 & 0.13 & 0.04 \\ 0.1 & 0.02 & 0.12 & 0.16 & 0.01 & 0.01 & 0.0 & -3.26 & 0.13 & 0.18 \\ 0.17 & 0.04 & 0.03 & 0.11 & 0.2 & 0.2 & 0.16 & 0.19 & -2.59 & 0.12 \\ 0.02 & 0.07 & 0.14 & 0.19 & 0.19 & 0.09 & 0.05 & 0.02 & 0.18 & -2.37 \end{pmatrix}.$$

The error term, $\epsilon^\dagger$, is assumed to follow a Normal distribution $\mathcal{N}(\mathbf{0}, \Sigma^\dagger)$. The covariance matrix $\Sigma^\dagger$ is diagonal and has diagonal elements captured by

$$\sigma^\dagger = (5.5 \ 6.4 \ 6.1 \ 6.4 \ 7.4 \ 7.7 \ 9.2 \ 9.9 \ 5.2 \ 6.2),$$

which is ten times higher than that in [den Boer \(2014\)](#).

Appendix A.2 Empirical Evidence for Case Study

We provide empirical evidence suggesting that the linear demand model is a reasonable and empirically adequate specification for our real-world data analysis.

First, for the following covariance matrix Σ obtained from the real-world data

$$\begin{bmatrix} 1.00 & 5.78 & 0.29 & 0.20 & 56.34 & 0.86 & 0.08 & 15.17 & 12.38 & 11.98 & 11.94 \\ 5.78 & 34.35 & 1.66 & 1.17 & 326.32 & 5.00 & 0.25 & 87.61 & 71.51 & 69.22 & 68.96 \\ 0.29 & 1.66 & 0.29 & 0.06 & 16.09 & 0.25 & 0.02 & 4.33 & 3.54 & 3.42 & 3.41 \\ 0.20 & 1.17 & 0.06 & 0.20 & 13.31 & 0.16 & 0.02 & 3.14 & 2.57 & 2.46 & 2.46 \\ 56.34 & 326.32 & 16.09 & 13.31 & 3582.99 & 52.43 & 4.01 & 857.13 & 695.38 & 673.62 & 668.16 \\ 0.86 & 5.00 & 0.25 & 0.16 & 52.43 & 0.86 & 0.06 & 13.08 & 10.64 & 10.32 & 10.25 \\ 0.08 & 0.25 & 0.02 & 0.02 & 4.01 & 0.06 & 0.08 & 1.18 & 0.97 & 0.94 & 0.94 \\ 15.17 & 87.61 & 4.33 & 3.14 & 857.13 & 13.08 & 1.18 & 230.59 & 188.19 & 182.23 & 181.58 \\ 12.38 & 71.51 & 3.54 & 2.57 & 695.38 & 10.64 & 0.97 & 188.19 & 153.75 & 148.84 & 148.41 \\ 11.98 & 69.22 & 3.42 & 2.46 & 673.62 & 10.32 & 0.94 & 182.23 & 148.84 & 144.26 & 143.74 \\ 11.94 & 68.96 & 3.41 & 2.46 & 668.16 & 10.25 & 0.94 & 181.58 & 148.41 & 143.74 & 143.42 \end{bmatrix},$$

we compute its smallest eigenvalue, which is approximately 0.0016, confirming that the positive definiteness assumption on Σ is satisfied.

Next, we estimate the parameters of the linear demand model using constrained least squares and assess the model's goodness-of-fit using R^2 and adjusted R^2 . The results, reported in Table 3, show that the R^2 values range from 0.5894 to 0.9486, with similarly high adjusted R^2 values. These results indicate that the linear model explains a substantial proportion of the variation in observed demand and provides a reasonably good fit to the empirical data.

Product	1	2	3	4
R^2	0.9486	0.7398	0.5894	0.9002
Adjusted R^2	0.9482	0.7378	0.5864	0.8994

Table 3. Goodness of linear demand model.

Appendix A.3 Experiments for Potential Negative Demand

Our paper focuses on the linear demand model with additive error that, on the one hand, is widely adopted across various domains and, on the other hand, remains arguable due to the possibility of yielding negative demand (that is, some components of $\mathbf{v} = \ell(\mathbf{\Lambda}, \mathbf{u}, \mathbf{p}) + \epsilon$ may become negative), particularly when modeling substitution effects. In this appendix, we present additional numerical experiments using both synthetic and real-world data. The proportion of negative demand resulting from our optimized pricing decisions is small, if not negligible, demonstrating that the linear demand model is a suitable choice.

Specifically, in our synthetic data experiments with the baseline setting ($N = 6$ and $S = 40$), for each $\varphi \in \{0.5, 1.0, 1.5\}$ that controls the random error's standard deviation, we generate 100 independent instances, and for each instance, we simulate demand for six products across 10,000 realizations of the error term, resulting in 60,000 demand realizations. We then compute the proportion of negative demand under the optimal pricing decisions in each instance and visualize in Figure 10 the histogram of these proportions among 100 instances.

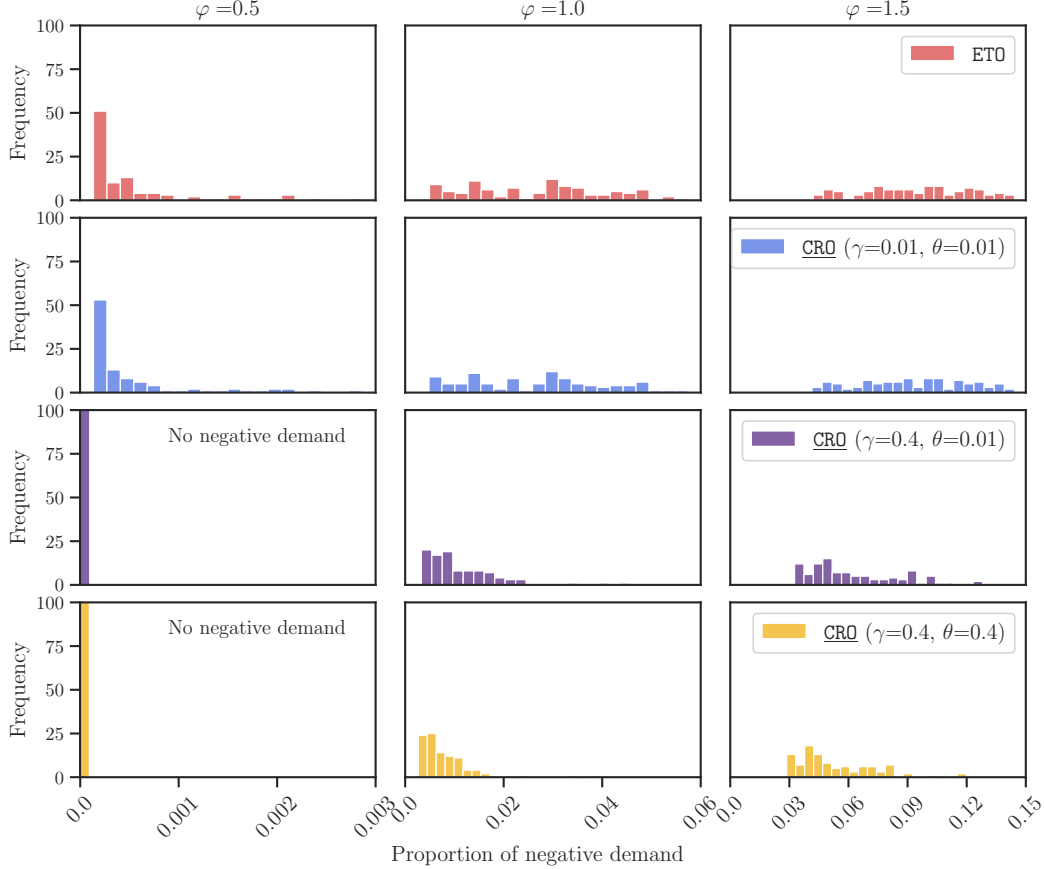


Figure 10. Histogram for proportion of negative demand under different φ for CRO and ETO.

As shown in Figure 10, the proportion of negative demand is nearly zero when $\varphi = 0.5$, and increases gradually as φ grows. Even when $\varphi = 1.5$, which corresponds to high residual variance, the largest proportion of negative demand remains below 15%. Importantly, across all scenarios, the proportion of negative demand under the CRO model is substantially lower than that under ETO. Moreover, even when we manually truncate negative demand realizations and re-evaluate the out-of-sample performance, our CRO model continues to outperform the ETO model. The observed

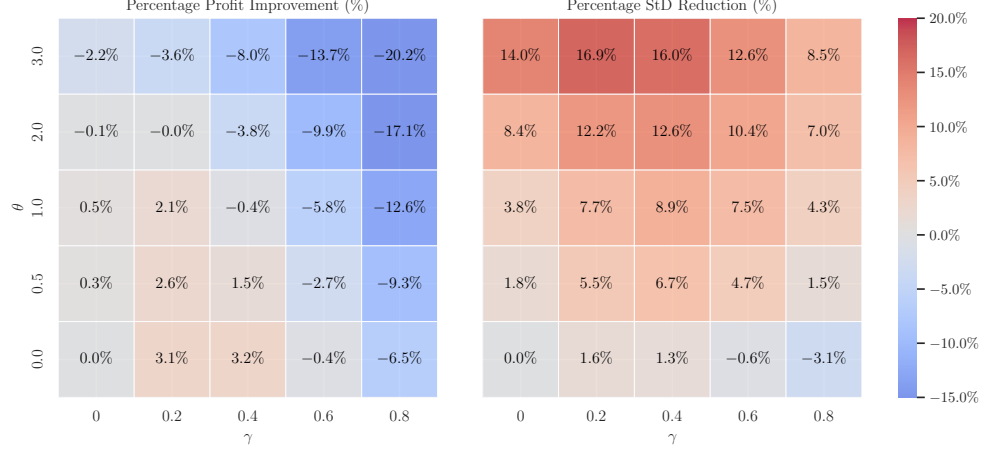


Figure 11. Heatmaps for profit improvement (left panel) and standard deviation reduction (right panel) of CRO relative to ETO under different (γ, θ) when negative demand is truncated to zero.

performance patterns remain consistent across all values of φ , and we present the representative results when $\varphi = 1.0$ in Figure 11, which demonstrates our CRO model still consistently outperforms ETO when demand is truncated.

In our real-world data experiment using sales records from Burger Cafe, we conducted an out-of-sample evaluation by applying the optimal decision obtained from the CRO model, and notably, no negative demand was observed. Based on these results, we conclude that the risk of negative demand is small and can be neglected in practice.

Appendix A.4 Comparison with Non-Parametric Residual-Based DRO

To assess whether the advantage of CRO over ETO persists when compared with a more flexible, non-parametric alternative, we conduct a simulation study that compares ETO, CRO, and a non-parametric benchmark based on random forests, which we denote by RFDR0.

Random forests are among the most widely used non-parametric prediction methods in demand estimation, as they flexibly capture non-linear relationships without imposing a prespecified functional form. One technical issue merits clarification. Non-parametric estimation generally yields a black-box demand-price mapping, rendering direct embedding into the pricing and inventory problem computationally intractable. In contrast, both ETO and CRO rely on a linear demand surrogate, allowing prices to explicitly enter the optimization problem and preserving tractability via conic reformulations. Consequently, price discretization in RFDR0 is adopted solely to ensure optimization

tractability, rather than reflecting any conceptual difference in the underlying decision problem.

Since exhaustive price enumeration becomes computationally prohibitive in the multiproduct setting, we confine this comparison to a single-product environment and assume that no contextual covariates are active beyond the intercept. Given a historical sample, we first train random forests using the demand–price observations and compute the corresponding out-of-sample residuals. Following the residual-based DRO framework of Kannan et al. (2024), we construct a Wasserstein ambiguity set centered on the empirical residual distribution. To ensure tractability, we discretize the feasible price interval $[\underline{p}, \bar{p}]$ into a grid of 100 equally spaced candidate prices. For each candidate price p , the trained random forests produce a demand prediction, and we solve the associated residual-based DRO problem to determine the optimal inventory level. The resulting **RFDR0** is the price–inventory pair that achieves the highest objective value over the discretized price grid. Similar to Qi et al. (2022), we observe that regardless of the choice of the Wasserstein radius, the resulting solution coincides with that of the standard estimate-then-optimize counterpart based on random forests. For the random forests implementation, we fix the number of trees at 200 and consider four maximum tree depths—2, 4, 8, and 16—reporting the depth that attains the highest average profit. We evaluate the three approaches with a varying sample size $S \in \{40, 80, 160\}$.

Linear data-generating process. Data are generated from the same linear demand model employed in the main simulation study (Section 4.1). All methods are evaluated under the true linear demand process, using identical performance metrics and testing protocol to ensure comparability. The comparison between **CRO** and **RFDR0** across all sample sizes is presented in Figure 12.

MNL data-generating process. Both **ETO** and **CRO** are implemented based on a fitted linear demand surrogate, while **RFDR0** is trained on the same demand–price sample and incorporated into the discretized residual-based DRO framework. All methods are evaluated under the true MNL demand process using a common testing protocol. The results are presented in Figure 13.

Two observations are particularly noteworthy. First, **RFDR0** exhibits steadily improving performance as the sample size increases under both data-generating processes. This trend is consistent with the nature of random forests, which offer substantial flexibility in capturing complex demand relationships but typically require sufficiently large samples to realize their predictive advantages.

Second, under the linear data-generating process, **RFDR0** underperforms **ETO** for all sample sizes when $\gamma = 0$ and $\theta = 0$, confirming that additional non-parametric flexibility provides no benefit when the demand model is correctly specified as linear. In contrast, under the MNL data-generating process, once the sample size reaches $S = 160$, the percentage profit improvement of **CRO** relative to **RFDR0** becomes negative. This indicates that **RFDR0** outperforms both **ETO** and **CRO** when the true demand relationship is non-linear and the available data are sufficiently abundant.

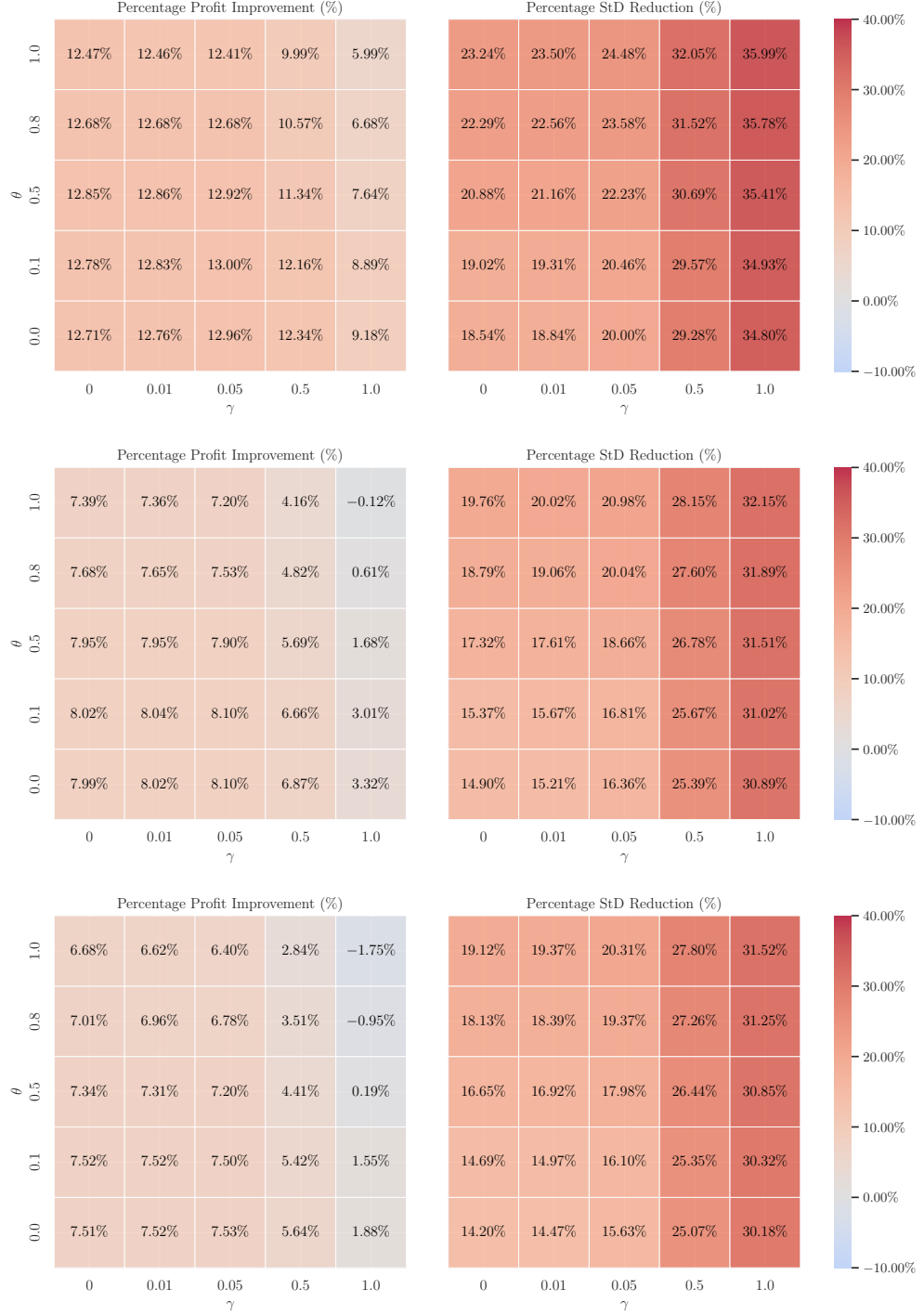


Figure 12. Heatmaps for profit improvement (left panel) and standard deviation reduction (right panel) of CRO relative to RFDR0 under different (γ, θ) and linear data-generating process: $S = 40$ (first row), $S = 80$ (second row), and $S = 160$ (third row).

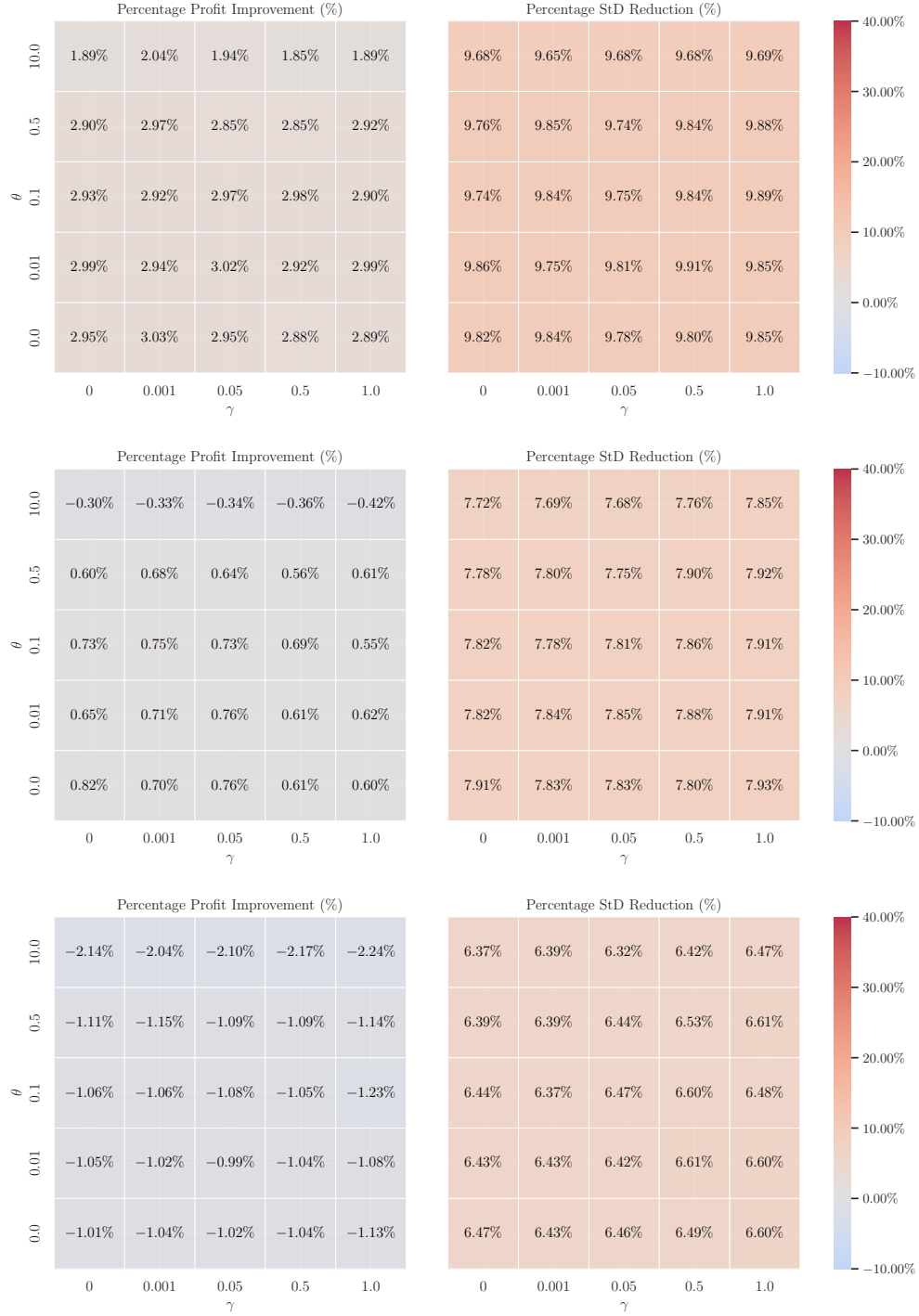


Figure 13. Heatmaps for profit improvement (left panel) and standard deviation reduction (right panel) of CRO relative to RFDR0 under different (γ, θ) and MNL data-generating process: $S = 40$ (first row), $S = 80$ (second row), and $S = 160$ (third row).

Overall, these findings suggest that while non-parametric estimation can be advantageous under structural misspecification in large-data regimes, **CRO** with a linear surrogate maintains a clear performance advantage when data are limited—precisely the regime that motivates and underpins our framework.

Appendix B Proofs

Proof of Lemma 1. Given that $\alpha \in (1, 2)$, to show $\mathbb{E}_{\mathbb{P}^*} [\exp(\|\tilde{\epsilon}\|_2^\alpha)] < \infty$, it suffices to show $\mathbb{E}_{\mathbb{P}^*} [\exp(\|\tilde{\epsilon}\|_2^\alpha) \cdot \mathbf{1}\{\|\tilde{\epsilon}\|_\infty \geq 1\}] < \infty$, where $\mathbf{1}\{\|\tilde{\epsilon}\|_\infty \geq 1\} = 1$ if $\|\tilde{\epsilon}\|_\infty \geq 1$ and 0 otherwise. Lemma 22 of Kannan et al. (2024) implies that

$$\mathbb{E}_{\mathbb{P}^*} [\exp(\|\tilde{\epsilon}\|_2^\alpha)] \leq \mathbb{E}_{\mathbb{P}^*} \left[\exp \left(\sum_{i=1}^N |\tilde{\epsilon}_i|^{1+\frac{\alpha}{2}} \right) \right] = \mathbb{E}_{\mathbb{P}^*} \left[\exp(\|\tilde{\epsilon}\|_\beta^\beta) \right],$$

where we denote $\beta = 1 + \alpha/2 \in (3/2, 2)$. Hence, it suffices to show $\mathbb{E}_{\mathbb{P}^*} [\exp(\|\tilde{\epsilon}\|_\beta^\beta)] < \infty$.

We claim that for any $k \geq 2$, $\mathbb{E}_{\mathbb{P}^*} [\|\tilde{\epsilon}\|_\beta^{\beta k}] \leq 5^N \left(4\beta k \sigma^2 N^{1-\frac{2(\beta-1)}{\beta}} e^{\frac{1}{e}} \right)^{\frac{\beta k}{2}}$. Consequently, the Fubini-Tonelli theorem, the equality

$$\mathbb{E}_{\mathbb{P}^*} [\exp(\|\tilde{\epsilon}\|_\beta^\beta)] = \mathbb{E}_{\mathbb{P}^*} \left[\sum_{k=1}^{\infty} \frac{\|\tilde{\epsilon}\|_\beta^{\beta k}}{k!} \right],$$

and the ratio test collectively imply the desired result $\mathbb{E}_{\mathbb{P}^*} [\exp(\|\tilde{\epsilon}\|_\beta^\beta)] < \infty$ as long as for $t_k = 5^N \left(4\beta k \sigma^2 N^{1-\frac{2(\beta-1)}{\beta}} e^{\frac{1}{e}} \right)^{\frac{\beta k}{2}} / k!$, it holds $\lim_{k \rightarrow \infty} t_{k+1}/t_k < 1$. Indeed, we note that since $\beta < 2$,

$$\lim_{k \rightarrow \infty} \frac{t_{k+1}}{t_k} = \lim_{k \rightarrow \infty} \frac{\left(4\beta(k+1) \sigma^2 N^{1-\frac{2(\beta-1)}{\beta}} e^{\frac{1}{e}} \right)^{\frac{\beta(k+1)}{2}} k!}{\left(4\beta k \sigma^2 N^{1-\frac{2(\beta-1)}{\beta}} e^{\frac{1}{e}} \right)^{\frac{\beta k}{2}} (k+1)!} = \lim_{k \rightarrow \infty} \frac{\left(4\beta \sigma^2 N^{1-\frac{2(\beta-1)}{\beta}} e^{\frac{1}{e}} \right)^{\frac{\beta}{2}}}{(k+1)^{1-\frac{\beta}{2}}} \cdot \left(1 + \frac{1}{k} \right)^{\frac{\beta k}{2}} = 0.$$

It remains to show $\mathbb{E}_{\mathbb{P}^*} [\|\tilde{\epsilon}\|_\beta^{\beta k}] \leq 5^N \left(4\beta k \sigma^2 N^{1-\frac{2(\beta-1)}{\beta}} e^{\frac{1}{e}} \right)^{\frac{\beta k}{2}}$ for any $k \geq 2$. To this end, we first show

$$\mathbb{P}^* [\|\tilde{\epsilon}\|_\beta \geq t] \leq 5^N \exp \left(-\frac{t^2}{8\sigma^2 N^{1-\frac{2(\beta-1)}{\beta}}} \right).$$

Let $\mathcal{K}$ be a $\frac{1}{2}$ -cover of the unit norm ball $\{\psi \in \mathbb{R}^N \mid \|\psi\|_{\frac{\beta}{\beta-1}} \leq 1\}$, whose cardinality is known to satisfy $|\mathcal{K}| \leq 5^N$ (see lemma 5.7 in Wainwright, 2019). Note that

$$\|\tilde{\epsilon}\|_\beta = \sup_{\|\psi\|_{\frac{\beta}{\beta-1}} \leq 1} \{\psi^\top \tilde{\epsilon}\} \leq \max_{\psi \in \mathcal{K}} \{\psi^\top \tilde{\epsilon}\} + \|\tilde{\epsilon}\|_\beta/2.$$

For any $\psi \in \mathcal{K}$ and $t > 0$, Assumption 1 implies that

$$\mathbb{P}^* [\psi^\top \tilde{\epsilon} \geq t] \leq \exp \left(-\frac{t^2}{2\sigma^2 \|\psi\|_2^2} \right) \leq \exp \left(-\frac{t^2}{2\sigma^2 N^{1-\frac{2(\beta-1)}{\beta}}} \right),$$

where the second inequality is due to $\|\psi\|_2^2 \leq N^{1-\frac{2(\beta-1)}{\beta}} \forall \psi \in \mathcal{K}$. Combining these two inequalities with the union bound, we obtain

$$\mathbb{P}^* [\|\tilde{\epsilon}\|_\beta \geq t] \leq \mathbb{P}^* \left[\max_{\psi \in \mathcal{K}} \{\psi^\top \tilde{\epsilon}\} \geq t/2 \right] \leq 5^N \exp \left(-\frac{t^2}{8\sigma^2 N^{1-\frac{2(\beta-1)}{\beta}}} \right).$$

It now follows that for any $k \geq 2$,

$$\begin{aligned} \mathbb{E}_{\mathbb{P}^*} [\|\tilde{\epsilon}\|_\beta^{\beta k}] &= \int_0^\infty \mathbb{P}^* [\|\tilde{\epsilon}\|_\beta^{\beta k} > t] dt \leq \int_0^\infty 5^N \exp \left(-\frac{t^{\frac{2}{\beta k}}}{8\sigma^2 N^{1-\frac{2(\beta-1)}{\beta}}} \right) dt \\ &= 5^N \left(8\sigma^2 N^{1-\frac{2(\beta-1)}{\beta}} \right)^{\frac{\beta k}{2}} \frac{\beta k}{2} \int_0^\infty e^{-u} u^{\frac{\beta k}{2}-1} du \\ &\leq 5^N \left(8\sigma^2 N^{1-\frac{2(\beta-1)}{\beta}} e^{\frac{1}{e}} \right)^{\frac{\beta k}{2}} \left(\frac{\beta k}{2} \right)^{\frac{\beta k}{2}} \\ &= 5^N \left(4\beta k \sigma^2 N^{1-\frac{2(\beta-1)}{\beta}} e^{\frac{1}{e}} \right)^{\frac{\beta k}{2}} \end{aligned}$$

where for the last inequality, we use $(\beta k/2)^{2/(\beta k)} \leq e^{1/e}$ and $\int_0^\infty e^{-u} u^{\frac{\beta k}{2}-1} du \leq (\beta k/2)^{\beta k/2}$. $\square$

Proof of Proposition 1. To ease our presentation, we denote

$$\hat{\mathbf{V}} = (\hat{\mathbf{v}}_1, \dots, \hat{\mathbf{v}}_S) \in \mathbb{R}^{N \times S}, \quad \hat{\mathbf{U}} = \left(\begin{pmatrix} \hat{\mathbf{u}}_1 \\ \hat{\mathbf{p}}_1 \end{pmatrix}, \dots, \begin{pmatrix} \hat{\mathbf{u}}_S \\ \hat{\mathbf{p}}_S \end{pmatrix} \right) \in \mathbb{R}^{(J+N) \times S}, \quad \hat{\mathbf{E}} = (\hat{\epsilon}_1, \dots, \hat{\epsilon}_S) \in \mathbb{R}^{N \times S}.$$

Under these notations, $\Sigma = S^{-1} \cdot \hat{\mathbf{U}} \hat{\mathbf{U}}^\top$ and the uncertainty set can be represented as

$$\mathcal{W}(\bar{\gamma}) = \left\{ \mathbf{\Lambda} \in \Omega \mid \sqrt{\text{tr}((\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}) \Sigma (\hat{\mathbf{\Lambda}} - \mathbf{\Lambda})^\top)} \leq \bar{\gamma} \right\} = \left\{ \mathbf{\Lambda} \in \Omega \mid \|(\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}) \hat{\mathbf{U}}\|_{\text{F}}^2 \leq \bar{\gamma}^2 S \right\},$$

where $\|\cdot\|_{\text{F}}$ is the Frobenius norm for matrices, *i.e.*, $\|\mathbf{A}\|_{\text{F}} = \sqrt{\sum_{i \in [M]} \sum_{j \in [N]} a_{ij}^2}$.

By the definition of $\hat{\mathbf{\Lambda}}$ and the fact that $\mathbf{\Lambda}^* \in \Omega$, we have

$$\sum_{s \in [S]} \|\hat{\mathbf{v}}_s - \ell(\hat{\mathbf{\Lambda}}, \hat{\mathbf{u}}_s, \hat{\mathbf{p}}_s)\|_2^2 \leq \sum_{s \in [S]} \|\hat{\mathbf{v}}_s - \ell(\mathbf{\Lambda}^*, \hat{\mathbf{u}}_s, \hat{\mathbf{p}}_s)\|_2^2,$$

which, by noting that $\hat{\mathbf{v}}_s = \ell(\mathbf{\Lambda}^*, \hat{\mathbf{u}}_s, \hat{\mathbf{p}}_s) + \hat{\epsilon}_s$ and $\ell(\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}^*, \hat{\mathbf{u}}_s, \hat{\mathbf{p}}_s) = \ell(\hat{\mathbf{\Lambda}}, \hat{\mathbf{u}}_s, \hat{\mathbf{p}}_s) - \ell(\mathbf{\Lambda}^*, \hat{\mathbf{u}}_s, \hat{\mathbf{p}}_s)$,

can be rewritten as

$$\begin{aligned}
& \sum_{s \in [S]} \|\hat{\mathbf{e}}_s - \ell(\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}^*, \hat{\mathbf{u}}_s, \hat{\mathbf{p}}_s)\|_2^2 \leq \sum_{s \in [S]} \|\hat{\mathbf{e}}_s\|_2^2 \\
\iff & \sum_{s \in [S]} \|\ell(\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}^*, \hat{\mathbf{u}}_s, \hat{\mathbf{p}}_s)\|_2^2 \leq 2 \sum_{s \in [S]} \hat{\mathbf{e}}_s^\top \ell(\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}^*, \hat{\mathbf{u}}_s, \hat{\mathbf{p}}_s) \\
\iff & \text{tr}((\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}^*) \hat{\mathbf{U}} \hat{\mathbf{U}}^\top (\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}^*)^\top) \leq 2 \cdot \text{tr}((\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}^*) \hat{\mathbf{U}} \hat{\mathbf{E}}^\top) \\
\iff & \|(\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}^*) \hat{\mathbf{U}}\|_{\text{F}}^2 \leq 2 \cdot \text{tr}((\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}^*) \hat{\mathbf{U}} \hat{\mathbf{E}}^\top) \\
\iff & \|(\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}^*) \hat{\mathbf{U}}\|_{\text{F}} \leq \frac{2 \cdot \text{tr}((\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}^*) \hat{\mathbf{U}} \hat{\mathbf{E}}^\top)}{\|(\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}^*) \hat{\mathbf{U}}\|_{\text{F}}}.
\end{aligned}$$

Here, the first equivalence follows from expanding the squares on both sides and canceling out the common terms $\|\hat{\mathbf{e}}_s\|_2^2$, $s \in [S]$, the second equivalence results from reshaping the expressions into matrices, and the final step involves dividing both sides by $\|(\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}^*) \hat{\mathbf{U}}\|_{\text{F}}$, assuming $\|(\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}^*) \hat{\mathbf{U}}\|_{\text{F}} > 0$ (otherwise, we have $\hat{\mathbf{\Lambda}} = \mathbf{\Lambda}^*$ so that $\mathbf{\Lambda}^* \in \mathcal{W}(\bar{\gamma})$ trivially holds). We then obtain the following upper bounds:

$$\begin{aligned}
\|(\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}^*) \hat{\mathbf{U}}\|_{\text{F}}^2 & \leq 4 \cdot \text{tr} \left(\frac{(\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}^*) \hat{\mathbf{U}} \hat{\mathbf{E}}^\top}{\|(\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}^*) \hat{\mathbf{U}}\|_{\text{F}}} \right)^2 \leq 4 \cdot \max_{\mathbf{\Lambda} \in \mathcal{W}(\bar{\gamma})} \left\{ \text{tr} \left(\frac{(\mathbf{\Lambda} - \mathbf{\Lambda}^*) \hat{\mathbf{U}} \hat{\mathbf{E}}^\top}{\|(\mathbf{\Lambda} - \mathbf{\Lambda}^*) \hat{\mathbf{U}}\|_{\text{F}}} \right)^2 \right\} \\
& \leq 4 \cdot \max_{\mathbf{D} \in \text{Im}(\mathbf{U}^\top), \|\mathbf{D}\|_{\text{F}}=1} \{\text{tr}(\hat{\mathbf{E}} \mathbf{D})^2\} \\
& = 4 \cdot \max_{\|\mathbf{D}\|_{\text{F}}=1} \{\text{tr}(\hat{\mathbf{E}} \mathbf{\Pi} \mathbf{D})^2\} \\
& = 4 \cdot \|\mathbf{\Pi} \hat{\mathbf{E}}\|_{\text{F}}^2 \\
& = 4 \cdot \text{tr}(\hat{\mathbf{E}} \mathbf{\Pi} \hat{\mathbf{E}}^\top).
\end{aligned}$$

Here, for the first equality, $\mathbf{\Pi} = \hat{\mathbf{U}}^\top (\hat{\mathbf{U}} \hat{\mathbf{U}}^\top)^{-1} \hat{\mathbf{U}}$ is the projection matrix onto $\text{Im}(\mathbf{U}^\top)$ —the range space of $\hat{\mathbf{U}}^\top$, the second equality is because $\|\cdot\|_{\text{F}}$ is self-dual, and the third equality follows from $\mathbf{\Pi}^2 = \mathbf{\Pi}$. That is to say,

$$4 \cdot \text{tr}(\hat{\mathbf{E}} \mathbf{\Pi} \hat{\mathbf{E}}^\top) \leq \bar{\gamma}^2 S \implies \text{tr}((\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}^*) \mathbf{\Sigma} (\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}^*)^\top) = \frac{1}{S} \cdot \|(\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}^*) \hat{\mathbf{U}}\|_{\text{F}}^2 \leq \bar{\gamma}^2,$$

implying

$$\mathbb{P}^S [\mathbf{\Lambda}^* \notin \mathcal{W}(\bar{\gamma})] \leq \mathbb{P}^S \left[\text{tr}(\hat{\mathbf{E}} \mathbf{\Pi} \hat{\mathbf{E}}^\top) > \frac{\bar{\gamma}^2 S}{4} \right].$$

It remains to investigate the right-hand side probability, *i.e.*, the concentration of $\text{tr}(\hat{\mathbf{E}} \mathbf{\Pi} \hat{\mathbf{E}}^\top)$.

Let $\hat{\mathbf{\xi}} = (\hat{\mathbf{e}}_1; \dots; \hat{\mathbf{e}}_S)$ be the NS -dimensional vectorization of $\hat{\mathbf{E}}$ (*i.e.*, i.i.d. samples of the sub-Gaussian random vector $\tilde{\mathbf{e}} \sim \mathbb{P}^*$ with a variance proxy σ^2) and “ $\otimes$ ” denote the Kronecker product of matrices, yielding

$$\text{tr}(\hat{\mathbf{E}} \mathbf{\Pi} \hat{\mathbf{E}}^\top) = \hat{\mathbf{\xi}}^\top (\mathbf{\Pi} \otimes \mathbf{I}_N) \hat{\mathbf{\xi}}.$$

Note that for any $\boldsymbol{\psi} = (\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_S) \in \mathbb{R}^{NS}$,

$$\mathbb{E}_{\mathbb{P}^S}[\exp(\boldsymbol{\psi}^\top \hat{\boldsymbol{\xi}})] = \prod_{s \in [S]} \mathbb{E}_{\mathbb{P}^S}[\exp(\boldsymbol{\psi}_s^\top \hat{\boldsymbol{\epsilon}}_s)] \leq \prod_{s \in [S]} \exp(\|\boldsymbol{\psi}_s\|_2^2 \sigma^2 / 2) = \exp(\|\boldsymbol{\psi}\|_2^2 \sigma^2 / 2),$$

implying that $\hat{\boldsymbol{\xi}}$ is also sub-Gaussian with a variance proxy σ^2 . Because the matrix $\boldsymbol{\Pi} \otimes \mathbf{I}_N$ is positive semidefinite, theorem 2.1 of [Hsu et al. \(2012\)](#) has established

$$\mathbb{P}^S \left[\hat{\boldsymbol{\xi}}^\top (\boldsymbol{\Pi} \otimes \mathbf{I}_N) \hat{\boldsymbol{\xi}} > \sigma^2 (\text{tr}(\boldsymbol{\Pi} \otimes \mathbf{I}_N) + 2\sqrt{\text{tr}((\boldsymbol{\Pi} \otimes \mathbf{I}_N)^2)}\psi + 2\|\boldsymbol{\Pi} \otimes \mathbf{I}_N\|\psi) \right] \leq e^{-\psi} \quad \forall \psi \geq 0.$$

We calculate that $\text{tr}(\boldsymbol{\Pi} \otimes \mathbf{I}_N) = \text{tr}(\boldsymbol{\Pi})\text{tr}(\mathbf{I}_N) = (J + N)N$, $\text{tr}((\boldsymbol{\Pi} \otimes \mathbf{I}_N)^2) = \text{tr}(\boldsymbol{\Pi}^2 \otimes \mathbf{I}_N) = \text{tr}(\boldsymbol{\Pi}^2)\text{tr}(\mathbf{I}_N) = (J + N)N$, and $\|\boldsymbol{\Pi} \otimes \mathbf{I}_N\| = \|\boldsymbol{\Pi}\| = 1$ by noticing $\text{tr}(\boldsymbol{\Pi}) = \text{tr}((\hat{\mathbf{U}}\hat{\mathbf{U}}^\top)^{-1}\hat{\mathbf{U}}\hat{\mathbf{U}}^\top) = \text{tr}(\mathbf{I}_{J+N}) = J + N$ and the eigenvalue of the projection matrix $\boldsymbol{\Pi}$ is either 0 or 1. Then, the above inequality can be simplified to

$$\begin{aligned} \mathbb{P}^S \left[\hat{\boldsymbol{\xi}}^\top (\boldsymbol{\Pi} \otimes \mathbf{I}_N) \hat{\boldsymbol{\xi}} > \sigma^2 ((J + N)N + 2\sqrt{(J + N)N\psi} + 2\psi) \right] &\leq e^{-\psi} \quad \forall \psi \geq 0 \\ \iff \mathbb{P}^S \left[\text{tr}(\hat{\mathbf{E}}\boldsymbol{\Pi}\hat{\mathbf{E}}^\top) > \sigma^2 ((J + N)N + 2\sqrt{(J + N)N\psi} + 2\psi) \right] &\leq e^{-\psi} \quad \forall \psi \geq 0. \end{aligned}$$

For any $\delta \in (0, 1)$, let $\psi = \log(1/\delta)$ and $\bar{\gamma} = \sqrt{\frac{12\sigma^2((J+N)N + \log(1/\delta))}{S}}$ that satisfy

$$\sigma^2 ((J + N)N + 2\sqrt{(J + N)N\psi} + 2\psi) \leq 3\sigma^2 ((J + N)N + \psi) = \frac{\bar{\gamma}^2 S}{4},$$

we then have

$$\mathbb{P}^S \left[\text{tr}(\hat{\mathbf{E}}\boldsymbol{\Pi}\hat{\mathbf{E}}^\top) > \frac{\bar{\gamma}^2 S}{4} \right] \leq \mathbb{P}^S [\text{tr}(\hat{\mathbf{E}}\boldsymbol{\Pi}\hat{\mathbf{E}}^\top) > \sigma^2 ((J + N)N + 2\sqrt{(J + N)N\psi} + 2\psi)] \leq e^{-\psi} = \delta,$$

concluding the proof. $\square$

Proof of Theorem 1. We focus on the case where $\hat{\boldsymbol{\Lambda}} \neq \boldsymbol{\Lambda}^*$, since the alternative case is trivial. For any $\mathbf{u} \in \mathcal{U}$ and $\mathbf{p} \in \mathcal{P}$, let $\mathbb{Q}^*(\boldsymbol{\Lambda}^*, \mathbf{u}, \mathbf{p})$ be probability distribution shifted from $\mathbb{P}^*$ by $\ell(\boldsymbol{\Lambda}^*, \mathbf{u}, \mathbf{p})$, that is, $\ell(\boldsymbol{\Lambda}^*, \mathbf{u}, \mathbf{p}) + \tilde{\epsilon} \sim \mathbb{Q}^*(\boldsymbol{\Lambda}^*, \mathbf{u}, \mathbf{p})$; and similarly, let $\mathbb{Q}_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}, \mathbf{u}, \mathbf{p})$ be the probability distribution shifted from $\mathbb{P}_{\boldsymbol{\Lambda}}$ by $\ell(\boldsymbol{\Lambda}, \mathbf{u}, \mathbf{p})$.

If $d(\mathbb{Q}^*(\boldsymbol{\Lambda}^*, \mathbf{u}, \mathbf{p}), \mathbb{Q}_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}, \mathbf{u}, \mathbf{p})) \leq \theta$ for some $\boldsymbol{\Lambda} \in \mathcal{W}(\gamma)$, then for any $(\mathbf{p}, \mathbf{q}) \in \mathcal{P} \times \mathcal{Q}$, we have

$$\begin{aligned} \mathbb{E}_{\mathbb{P}^*} [\pi(\tilde{\epsilon}, \boldsymbol{\Lambda}^*, \mathbf{u}, \mathbf{p}, \mathbf{q})] &= \mathbb{E}_{\mathbb{P}^*} [\mathbf{p}^\top (\ell(\boldsymbol{\Lambda}^*, \mathbf{u}, \mathbf{p}) + \tilde{\epsilon}) - \mathbf{g}^\top (\ell(\boldsymbol{\Lambda}^*, \mathbf{u}, \mathbf{p}) + \tilde{\epsilon} - \mathbf{q})^+ - \mathbf{c}^\top \mathbf{q}] \\ &= \mathbb{E}_{\mathbb{Q}^*(\boldsymbol{\Lambda}^*, \mathbf{u}, \mathbf{p})} [\mathbf{p}^\top \tilde{\mathbf{v}} - \mathbf{g}^\top (\tilde{\mathbf{v}} - \mathbf{q})^+ - \mathbf{c}^\top \mathbf{q}] \\ &\geq \min_{\boldsymbol{\Lambda} \in \mathcal{W}(\gamma)} \min_{\mathbb{Q} \in \mathcal{M}(\mathbb{R}^N): d(\mathbb{Q}, \mathbb{Q}_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}, \mathbf{u}, \mathbf{p})) \leq \theta} \mathbb{E}_{\mathbb{Q}} [\mathbf{p}^\top \tilde{\mathbf{v}} - \mathbf{g}^\top (\tilde{\mathbf{v}} - \mathbf{q})^+ - \mathbf{c}^\top \mathbf{q}] \\ &= \min_{\boldsymbol{\Lambda} \in \mathcal{W}(\gamma)} \min_{\mathbb{P} \in \mathcal{F}(\boldsymbol{\Lambda}, \theta)} \mathbb{E}_{\mathbb{P}} [\mathbf{p}^\top (\ell(\boldsymbol{\Lambda}, \mathbf{u}, \mathbf{p}) + \tilde{\epsilon}) - \mathbf{g}^\top (\ell(\boldsymbol{\Lambda}, \mathbf{u}, \mathbf{p}) + \tilde{\epsilon} - \mathbf{q})^+ - \mathbf{c}^\top \mathbf{q}] \\ &= \min_{\boldsymbol{\Lambda} \in \mathcal{W}(\gamma), \mathbb{P} \in \mathcal{F}(\boldsymbol{\Lambda}, \theta)} \mathbb{E}_{\mathbb{P}} [\pi(\tilde{\epsilon}, \boldsymbol{\Lambda}, \mathbf{u}, \mathbf{p}, \mathbf{q})]. \end{aligned}$$

Since the optimal solution $(\mathbf{p}^\dagger, \mathbf{q}^\dagger)$ of **CRO** satisfies

$$\mathbb{E}_{\mathbb{P}^\star} [\pi(\tilde{\epsilon}, \mathbf{\Lambda}^\star, \mathbf{u}, \mathbf{p}^\dagger, \mathbf{q}^\dagger)] \geq \min_{\mathbf{\Lambda} \in \mathcal{W}(\gamma), \mathbb{P} \in \mathcal{F}(\mathbf{\Lambda}, \theta)} \mathbb{E}_{\mathbb{P}} [\pi(\tilde{\epsilon}, \mathbf{\Lambda}, \mathbf{u}, \mathbf{p}^\dagger, \mathbf{q}^\dagger)] = Z(\gamma, \theta),$$

it thus suffices to prove that with probability at least $1 - \delta$, we have $d(\mathbb{Q}^\star(\mathbf{\Lambda}^\star, \mathbf{u}, \mathbf{p}), \mathbb{Q}_{\mathbf{\Lambda}}(\mathbf{\Lambda}, \mathbf{u}, \mathbf{p})) \leq \theta$ for some $\mathbf{\Lambda} \in \mathcal{W}(\gamma)$ and θ as specified in the theorem. To this end, we specifically construct

$$\mathbf{\Lambda} = \hat{\mathbf{\Lambda}} + \gamma \cdot \frac{\mathbf{\Lambda}^\star - \hat{\mathbf{\Lambda}}}{\sqrt{\text{tr}((\mathbf{\Lambda}^\star - \hat{\mathbf{\Lambda}})\mathbf{\Sigma}(\mathbf{\Lambda}^\star - \hat{\mathbf{\Lambda}})^\top)}},$$

which satisfies $\mathbf{\Lambda} \in \mathcal{W}(\gamma)$ as $\sqrt{\text{tr}((\mathbf{\Lambda} - \hat{\mathbf{\Lambda}})\mathbf{\Sigma}(\mathbf{\Lambda} - \hat{\mathbf{\Lambda}})^\top)} = \gamma$. We also note that

$$\begin{aligned} & d(\mathbb{Q}^\star(\mathbf{\Lambda}^\star, \mathbf{u}, \mathbf{p}), \mathbb{Q}_{\mathbf{\Lambda}}(\mathbf{\Lambda}, \mathbf{u}, \mathbf{p})) \\ \text{triangle inequality} \quad & \leq d(\mathbb{Q}^\star(\mathbf{\Lambda}^\star, \mathbf{u}, \mathbf{p}), \mathbb{Q}_{\mathbf{\Lambda}^\star}(\mathbf{\Lambda}^\star, \mathbf{u}, \mathbf{p})) + d(\mathbb{Q}_{\mathbf{\Lambda}^\star}(\mathbf{\Lambda}^\star, \mathbf{u}, \mathbf{p}), \mathbb{Q}_{\mathbf{\Lambda}}(\mathbf{\Lambda}, \mathbf{u}, \mathbf{p})) \\ & = d(\mathbb{P}^\star, \mathbb{P}_{\mathbf{\Lambda}^\star}) + d(\mathbb{Q}_{\mathbf{\Lambda}^\star}(\mathbf{\Lambda}^\star, \mathbf{u}, \mathbf{p}), \mathbb{Q}_{\mathbf{\Lambda}}(\mathbf{\Lambda}, \mathbf{u}, \mathbf{p})) \\ \text{definition of } d(\cdot, \cdot) \quad & \leq d(\mathbb{P}^\star, \mathbb{P}_{\mathbf{\Lambda}^\star}) + S^{-1} \sum_{s \in [S]} \|\ell(\mathbf{\Lambda}^\star, \mathbf{u}, \mathbf{p}) + \hat{\epsilon}_{\mathbf{\Lambda}^\star, s} - \ell(\mathbf{\Lambda}, \mathbf{u}, \mathbf{p}) - \hat{\epsilon}_{\mathbf{\Lambda}, s}\| \\ \text{triangle inequality} \quad & \leq d(\mathbb{P}^\star, \mathbb{P}_{\mathbf{\Lambda}^\star}) + \|\ell(\mathbf{\Lambda}^\star - \mathbf{\Lambda}, \mathbf{u}, \mathbf{p})\| + S^{-1} \sum_{s \in [S]} \|\ell(\mathbf{\Lambda}^\star - \mathbf{\Lambda}, \hat{\mathbf{u}}_s, \hat{\mathbf{p}}_s)\| \quad (8) \\ \text{Cauchy's inequality} \quad & \leq d(\mathbb{P}^\star, \mathbb{P}_{\mathbf{\Lambda}^\star}) + \|\ell(\mathbf{\Lambda}^\star - \mathbf{\Lambda}, \mathbf{u}, \mathbf{p})\| + \sqrt{S^{-1} \sum_{s \in [S]} \|\ell(\mathbf{\Lambda}^\star - \mathbf{\Lambda}, \hat{\mathbf{u}}_s, \hat{\mathbf{p}}_s)\|^2} \\ & \leq d(\mathbb{P}^\star, \mathbb{P}_{\mathbf{\Lambda}^\star}) + \sqrt{\text{tr}((\mathbf{\Lambda}^\star - \mathbf{\Lambda})\mathbf{\Sigma}(\mathbf{\Lambda}^\star - \mathbf{\Lambda})^\top)} \cdot (1 + \|(\mathbf{u}, \mathbf{p})\|_{\mathbf{\Sigma}^{-1}}) \\ & \leq d(\mathbb{P}^\star, \mathbb{P}_{\mathbf{\Lambda}^\star}) + \sqrt{\text{tr}((\mathbf{\Lambda}^\star - \mathbf{\Lambda})\mathbf{\Sigma}(\mathbf{\Lambda}^\star - \mathbf{\Lambda})^\top)} \cdot \left(1 + \max_{\mathbf{p} \in \mathcal{P}} \|(\mathbf{u}, \mathbf{p})\|_{\mathbf{\Sigma}^{-1}}\right) \end{aligned}$$

where we use, in the second inequality, the definition of Wasserstein distance, and in the second-to-last inequality, the fact that $\|\mathbf{A}\mathbf{a}\| \leq \sqrt{\text{tr}(\mathbf{A}\mathbf{\Sigma}\mathbf{A}^\top)}\sqrt{\mathbf{a}^\top\mathbf{\Sigma}^{-1}\mathbf{a}}$ for any matrix $\mathbf{A}$, vector $\mathbf{a}$ and positive definite $\mathbf{\Sigma}$.

Proposition 1 shows that with probability at least $1 - \frac{\delta}{2}$, $\mathbf{\Lambda}^\star \in \mathcal{W}(\bar{\gamma})$, implying that with probability at least $1 - \frac{\delta}{2}$, it holds that

$$\sqrt{\text{tr}((\mathbf{\Lambda}^\star - \mathbf{\Lambda})\mathbf{\Sigma}(\mathbf{\Lambda}^\star - \mathbf{\Lambda})^\top)} \leq \bar{\gamma} - \gamma.$$

Proposition 2, scaling k_1 by a factor of two, shows that with probability at least $1 - \frac{\delta}{2}$, it holds that $\mathbb{P}^\star \in \mathcal{F}(\mathbf{\Lambda}^\star, \bar{\theta})$, or equivalently, $d(\mathbb{P}^\star, \mathbb{P}_{\mathbf{\Lambda}^\star}) \leq \bar{\theta}$.

Applying the union bound and using the fact (8), it then follows that with probability at least $1 - \delta$,

$$d(\mathbb{Q}^\star(\mathbf{\Lambda}^\star, \mathbf{u}, \mathbf{p}), \mathbb{Q}_{\mathbf{\Lambda}}(\mathbf{\Lambda}, \mathbf{u}, \mathbf{p})) \leq \bar{\theta} + (\bar{\gamma} - \gamma) \cdot \left(\max_{\mathbf{p} \in \mathcal{P}} \|(\mathbf{u}, \mathbf{p})\|_{\mathbf{\Sigma}^{-1}} + 1\right) = \theta,$$

concluding the proof. $\square$

Proof of Proposition 3. Under Assumption 2, there exists an event $\mathcal{G}_1$ with $\mathbb{P}[\mathcal{G}_1] = 1$ such that for every $\omega \in \mathcal{G}_1$, there exist a finite index $S_1(\omega)$ and a constant $\underline{\sigma}(\omega) > 0$ satisfying

$$\sigma_{\min}(\mathbf{\Sigma}_S) \geq \underline{\sigma}(\omega) \quad \forall S \geq S_1(\omega). \quad (9)$$

By applying Proposition 1 with confidence level $\delta_S = 1/S^2$, we have

$$\mathbb{P}^S[\mathbf{\Lambda}^\star \in \mathcal{W}_S(\bar{\gamma}_S)] \geq 1 - \frac{1}{S^2} \quad \text{with} \quad \bar{\gamma}_S = \sqrt{\frac{12\sigma^2((J+N)N + 2\log S)}{S}}.$$

Since $\sum_{S=1}^{\infty} S^{-2} < \infty$, the Borel–Cantelli lemma (see theorem 2.18 of [Kallenberg 1997](#)) yields an event $\mathcal{G}_2$ with $\mathbb{P}[\mathcal{G}_2] = 1$ such that for every $\omega \in \mathcal{G}_2$, there is some $S_2(\omega) > 0$ satisfying

$$\mathbf{\Lambda}^\star \in \mathcal{W}_S(\bar{\gamma}_S) \quad \forall S \geq S_2(\omega). \quad (10)$$

For any $\omega \in \mathcal{G}_1 \cap \mathcal{G}_2$ and $S \geq \max\{S_1(\omega), S_2(\omega)\}$, (9) and (10) imply

$$\|\hat{\mathbf{\Lambda}}_S - \mathbf{\Lambda}^\star\|_{\text{F}}^2 = \text{tr}((\hat{\mathbf{\Lambda}}_S - \mathbf{\Lambda}^\star)(\hat{\mathbf{\Lambda}}_S - \mathbf{\Lambda}^\star)^\top) \leq \frac{1}{\underline{\sigma}(\omega)} \cdot \text{tr}((\hat{\mathbf{\Lambda}}_S - \mathbf{\Lambda}^\star)\mathbf{\Sigma}_S(\hat{\mathbf{\Lambda}}_S - \mathbf{\Lambda}^\star)^\top) \leq \frac{1}{\underline{\sigma}(\omega)} \cdot \bar{\gamma}_S^2.$$

Because $\bar{\gamma}_S \rightarrow 0$, we have

$$0 \leq \liminf_{S \rightarrow \infty} \|\hat{\mathbf{\Lambda}}_S - \mathbf{\Lambda}^\star\|_{\text{F}} \leq \limsup_{S \rightarrow \infty} \|\hat{\mathbf{\Lambda}}_S - \mathbf{\Lambda}^\star\|_{\text{F}} \leq \frac{1}{\sqrt{\underline{\sigma}(\omega)}} \limsup_{S \rightarrow \infty} \bar{\gamma}_S = 0$$

for every $\omega \in \mathcal{G}_1 \cap \mathcal{G}_2$. Since $\mathbb{P}[\mathcal{G}_1 \cap \mathcal{G}_2] = 1$, we obtain $\|\hat{\mathbf{\Lambda}}_S - \mathbf{\Lambda}^\star\|_{\text{F}} \xrightarrow{a.s.} 0$.

Moreover, for any $\omega \in \mathcal{G}_1 \cap \mathcal{G}_2$ and $S \geq \max\{S_1(\omega), S_2(\omega)\}$, (9) implies that the matrix $\mathbf{\Sigma}_S$ is positive definite, and hence $\|\cdot\|_{\mathbf{\Sigma}_S}$ is a norm. Fix any $\mathbf{\Lambda} \in \mathcal{W}_S(\bar{\gamma}_S)$. Since (10) gives $\mathbf{\Lambda}^\star \in \mathcal{W}_S(\bar{\gamma}_S)$ for all sufficiently large S , the triangle inequality for $\|\cdot\|_{\mathbf{\Sigma}_S}$ yields

$$\|\mathbf{\Lambda} - \mathbf{\Lambda}^\star\|_{\mathbf{\Sigma}_S} \leq \|\mathbf{\Lambda} - \hat{\mathbf{\Lambda}}_S\|_{\mathbf{\Sigma}_S} + \|\hat{\mathbf{\Lambda}}_S - \mathbf{\Lambda}^\star\|_{\mathbf{\Sigma}_S} \leq 2\bar{\gamma}_S.$$

Combining this with (9), we obtain

$$\sup_{\mathbf{\Lambda} \in \mathcal{W}_S(\bar{\gamma}_S)} \|\mathbf{\Lambda} - \mathbf{\Lambda}^\star\|_{\text{F}} \leq \sqrt{\frac{1}{\underline{\sigma}(\omega)}} \cdot \sup_{\mathbf{\Lambda} \in \mathcal{W}_S(\bar{\gamma}_S)} \|\mathbf{\Lambda} - \mathbf{\Lambda}^\star\|_{\mathbf{\Sigma}_S} \leq \sqrt{\frac{1}{\underline{\sigma}(\omega)}} \cdot 2\bar{\gamma}_S \rightarrow 0.$$

Together with the eventual membership $\mathbf{\Lambda}^\star \in \mathcal{W}_S(\bar{\gamma}_S)$ from (10), this proves that $\mathcal{W}_S(\bar{\gamma}_S)$ shrinks to the singleton $\{\mathbf{\Lambda}^\star\}$. Since $\mathbb{P}[\mathcal{G}_1 \cap \mathcal{G}_2] = 1$, the above claims hold almost surely. $\square$

Proof of Theorem 2. Denote $\zeta^\star = \mathbb{E}_{\mathbb{P}^\star}[\pi(\tilde{\epsilon}, \mathbf{\Lambda}^\star, \mathbf{u}, \mathbf{p}^\star, \mathbf{q}^\star)]$ and $\zeta_S = \mathbb{E}_{\mathbb{P}^\star}[\pi(\tilde{\epsilon}, \mathbf{\Lambda}^\star, \mathbf{u}, \mathbf{p}_S^\dagger, \mathbf{q}_S^\dagger)]$, and let $Z_S = Z(\gamma_S, \theta_S)$ be the optimal value of **CR0** with S samples. Since $(\mathbf{p}^\star, \mathbf{q}^\star)$ solves **TRUE**,

$$\zeta_S \leq \zeta^\star \quad \forall S \geq 1. \quad (11)$$

On the other hand, by applying Theorem 1 with confidence level $\delta_S = 1/S^2$, we have

$$\mathbb{P}^S[\zeta_S \geq Z_S] \geq 1 - \frac{1}{S^2} \quad \forall S \geq 1.$$

Since $\sum_{S=1}^{\infty} S^{-2} < \infty$, the Borel–Cantelli lemma (see theorem 2.18 of [Kallenberg 1997](#)) implies that there exists an event $\mathcal{G}_1$ with $\mathbb{P}[\mathcal{G}_1] = 1$ such that for any $\omega \in \mathcal{G}_1$, there exists $S_1(\omega) > 0$ satisfying

$$\zeta_S \geq Z_S \quad \forall S \geq S_1(\omega). \quad (12)$$

Thus, to prove $\zeta_S \xrightarrow{a.s.} \zeta^*$, it suffices to show that $\liminf_{S \rightarrow \infty} Z_S \geq \zeta^*$ almost surely because it would imply that

$$\zeta^* \leq \liminf_{S \rightarrow \infty} Z_S \leq \liminf_{S \rightarrow \infty} \zeta_S \leq \limsup_{S \rightarrow \infty} \zeta_S \leq \zeta^*$$

holds almost surely.

To this end, we first note that, because $\bar{\gamma}_S \rightarrow 0$ and $\bar{\theta}_S \rightarrow 0$ as $S \rightarrow \infty$, parameters γ_S and θ_S shrink to zero, provided that the factor multiplying $\bar{\gamma}_S - \gamma_S$ remains bounded. Under Assumption 2, (9) implies that on an event $\mathcal{G}_2$ with $\mathbb{P}[\mathcal{G}_2] = 1$, such that for any $\omega \in \mathcal{G}_2$, there exists $S_2(\omega) > 0$ and $\underline{\sigma}(\omega) > 0$ satisfying

$$\sigma_{\min}(\Sigma_S) \geq \underline{\sigma}(\omega) \quad \forall S \geq S_2(\omega).$$

Hence, for any $\omega \in \mathcal{G}_2$ and $S \geq S_2(\omega)$,

$$\max_{\mathbf{p} \in \mathcal{P}} \|(\mathbf{u}, \mathbf{p})\|_{\Sigma_S^{-1}} \leq \sqrt{\frac{1}{\underline{\sigma}(\omega)}} \cdot \max_{\mathbf{p} \in \mathcal{P}} \|(\mathbf{u}, \mathbf{p})\|_2.$$

which further implies that

$$\theta_S \leq M(\omega) \bar{\gamma}_S + \bar{\theta}_S, \quad (13)$$

where $M(\omega) = 1 + \sqrt{\frac{1}{\underline{\sigma}(\omega)}} \cdot \max_{\mathbf{p} \in \mathcal{P}} \|(\mathbf{u}, \mathbf{p})\|_2$ is finite.

Next, by applying Proposition 1 with $\delta_S = 1/S^2$ and by the Borel–Cantelli lemma, there exists an event $\mathcal{G}_3$ with $\mathbb{P}[\mathcal{G}_3] = 1$ such that for any $\omega \in \mathcal{G}_3$, there exists $S_3(\omega) > 0$ satisfying

$$\mathbf{\Lambda}^* \in \mathcal{W}_S(\bar{\gamma}_S) \quad \forall S \geq S_3(\omega).$$

Therefore, fix $\omega \in \mathcal{G}_2 \cap \mathcal{G}_3$, for any $\mathbf{\Lambda} \in \mathcal{W}_S(\gamma_S)$ and $S \geq \max\{S_2(\omega), S_3(\omega)\}$, we have

$$\|\mathbf{\Lambda} - \mathbf{\Lambda}^*\|_{\Sigma_S} \leq \|\mathbf{\Lambda} - \hat{\mathbf{\Lambda}}_S\|_{\Sigma_S} + \|\hat{\mathbf{\Lambda}}_S - \mathbf{\Lambda}^*\|_{\Sigma_S} \leq \gamma_S + \bar{\gamma}_S \leq 2\bar{\gamma}_S. \quad (14)$$

Finally, by applying Proposition 2 with $\delta_S = 1/S^2$ to the empirical residual distribution $\mathbb{P}_{\mathbf{\Lambda}^*}$. Since $\sum_{S=1}^{\infty} S^{-2} < \infty$, another application of the Borel–Cantelli lemma yields an event $\mathcal{G}_4$ with $\mathbb{P}[\mathcal{G}_4] = 1$ such that for any $\omega \in \mathcal{G}_4$, there exists $S_4(\omega) > 0$ satisfying

$$d(\mathbb{P}^*, \mathbb{P}_{\mathbf{\Lambda}^*}) \leq \bar{\theta}_S \quad \forall S \geq S_4(\omega). \quad (15)$$

Before concluding $\zeta_S \xrightarrow{a.s.} \zeta^*$ by consolidating the above claims, for notation convenience we further define $\varphi_{\mathbf{p},\mathbf{q}}(\mathbf{v}) = \mathbf{p}^\top \mathbf{v} - \mathbf{g}^\top (\mathbf{v} - \mathbf{q})^+ - \mathbf{c}^\top \mathbf{q}$. For any $\mathbf{v}, \mathbf{v}' \in \mathbb{R}^N$ and any $(\mathbf{p}, \mathbf{q}) \in \mathcal{P} \times \mathcal{Q}$,

$$\begin{aligned} |\varphi_{\mathbf{p},\mathbf{q}}(\mathbf{v}) - \varphi_{\mathbf{p},\mathbf{q}}(\mathbf{v}')| &\leq \|\mathbf{p}\|_2 \|\mathbf{v} - \mathbf{v}'\|_2 + \|\mathbf{g}\|_2 \|(\mathbf{v} - \mathbf{q})^+ - (\mathbf{v}' - \mathbf{q})^+\|_2 \\ &\leq (\|\mathbf{p}\|_2 + \|\mathbf{g}\|_2) \|\mathbf{v} - \mathbf{v}'\|_2, \end{aligned}$$

where the last inequality uses the fact that the mapping $\mathbf{v} \mapsto (\mathbf{v} - \mathbf{q})^+$ is 1-Lipschitz. Let $\bar{L} = \max_{\mathbf{p} \in \mathcal{P}} (\|\mathbf{p}\|_2 + \|\mathbf{g}\|_2) < \infty$.

Fix any $\omega \in \mathcal{G}_2 \cap \mathcal{G}_3 \cap \mathcal{G}_4$, any $S \geq \max\{S_2(\omega), S_3(\omega), S_4(\omega)\}$, and choose any pair $(\mathbf{\Lambda}, \mathbb{P}) \in \mathcal{W}_S(\gamma_S) \times \mathcal{F}(\mathbf{\Lambda}, \theta_S)$, consider

$$\begin{aligned} \mathbb{Q}^* &: \ell(\mathbf{\Lambda}^*, \mathbf{u}, \mathbf{p}^*) + \tilde{\epsilon}, \quad \tilde{\epsilon} \sim \mathbb{P}^* \\ \mathbb{Q}_{\mathbf{\Lambda}} &: \ell(\mathbf{\Lambda}, \mathbf{u}, \mathbf{p}^*) + \tilde{\epsilon}, \quad \tilde{\epsilon} \sim \mathbb{P}_{\mathbf{\Lambda}} \\ \mathbb{Q} &: \ell(\mathbf{\Lambda}, \mathbf{u}, \mathbf{p}^*) + \tilde{\epsilon}, \quad \tilde{\epsilon} \sim \mathbb{P}. \end{aligned}$$

Exactly as in the proof of Theorem 1, using the triangle inequality for the Wasserstein distance, the definition of $\mathbb{P}_{\mathbf{\Lambda}}$, and the inequality $\|\mathbf{A}\mathbf{a}\|_2 \leq \sqrt{\text{tr}(\mathbf{A}\mathbf{\Sigma}_S\mathbf{A}^\top)}\sqrt{\mathbf{a}^\top \mathbf{\Sigma}_S^{-1}\mathbf{a}}$, we have

$$\begin{aligned} d(\mathbb{Q}^*, \mathbb{Q}) &\leq d(\mathbb{P}^*, \mathbb{P}_{\mathbf{\Lambda}^*}) + d(\mathbb{Q}_{\mathbf{\Lambda}^*}, \mathbb{Q}_{\mathbf{\Lambda}}) + d(\mathbb{Q}_{\mathbf{\Lambda}}, \mathbb{Q}) \\ &\leq d(\mathbb{P}^*, \mathbb{P}_{\mathbf{\Lambda}^*}) + \left(1 + \max_{\mathbf{p} \in \mathcal{P}} \|(\mathbf{u}, \mathbf{p})\|_{\mathbf{\Sigma}_S^{-1}}\right) \sqrt{\text{tr}((\mathbf{\Lambda}^* - \mathbf{\Lambda})\mathbf{\Sigma}_S(\mathbf{\Lambda}^* - \mathbf{\Lambda})^\top)} + \theta_S. \end{aligned}$$

Combining this bound with (13), (14), and (15), we obtain

$$d(\mathbb{Q}^*, \mathbb{Q}) \leq \bar{\theta}_S + M(\omega) \cdot 2\bar{\gamma}_S + \theta_S \leq 2\bar{\theta}_S + 3M(\omega)\bar{\gamma}_S.$$

Since $\varphi_{\mathbf{p}^*, \mathbf{q}^*}$ is $\bar{L}$ -Lipschitz, the Kantorovich-Rubinstein inequality (see, *e.g.*, theorem 5 of Kuhn et al. 2019) implies

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}[\pi(\tilde{\epsilon}, \mathbf{\Lambda}, \mathbf{u}, \mathbf{p}^*, \mathbf{q}^*)] &= \mathbb{E}_{\mathbb{Q}}[\varphi_{\mathbf{p}^*, \mathbf{q}^*}(\tilde{\mathbf{v}})] \geq \mathbb{E}_{\mathbb{Q}^*}[\varphi_{\mathbf{p}^*, \mathbf{q}^*}(\tilde{\mathbf{v}})] - \bar{L} \cdot d(\mathbb{Q}^*, \mathbb{Q}) \\ &= \zeta^* - \bar{L} \cdot d(\mathbb{Q}^*, \mathbb{Q}) \\ &\geq \zeta^* - \bar{L} \cdot (2\bar{\theta}_S + 3M(\omega)\bar{\gamma}_S). \end{aligned}$$

Because the above bound holds for every $(\mathbf{\Lambda}, \mathbb{P}) \in \mathcal{W}_S(\gamma_S) \times \mathcal{F}(\mathbf{\Lambda}, \theta_S)$, we conclude that

$$Z_S \geq \min_{\mathbf{\Lambda} \in \mathcal{W}_S(\gamma_S), \mathbb{P} \in \mathcal{F}(\mathbf{\Lambda}, \theta_S)} \mathbb{E}_{\mathbb{P}}[\pi(\tilde{\epsilon}, \mathbf{\Lambda}, \mathbf{u}, \mathbf{p}^*, \mathbf{q}^*)] \geq \zeta^* - \bar{L} \cdot (2\bar{\theta}_S + 3M(\omega)\bar{\gamma}_S) \quad (16)$$

for any $\omega \in \mathcal{G}_2 \cap \mathcal{G}_3 \cap \mathcal{G}_4$ and $S \geq \max\{S_2(\omega), S_3(\omega), S_4(\omega)\}$.

Fix $\omega \in \mathcal{G}_1 \cap \mathcal{G}_2 \cap \mathcal{G}_3 \cap \mathcal{G}_4$ and any $S \geq \max\{S_1(\omega), S_2(\omega), S_3(\omega), S_4(\omega)\}$, then (11), (12), and (16) imply that

$$\zeta^\star - \bar{L} \cdot (2\bar{\theta}_S + 3M(\omega)\bar{\gamma}_S) \leq Z_S \leq \zeta_S \leq \zeta^\star.$$

Because $M(\omega) < \infty$, $\bar{\gamma}_S \rightarrow 0$ and $\bar{\theta}_S \rightarrow 0$, we obtain

$$\liminf_{S \rightarrow \infty} \zeta_S \geq \zeta^\star, \quad \limsup_{S \rightarrow \infty} \zeta_S \leq \zeta^\star.$$

Therefore, we have $\lim_{S \rightarrow \infty} \zeta_S = \zeta^\star$ on the event $\mathcal{G}_1 \cap \mathcal{G}_2 \cap \mathcal{G}_3 \cap \mathcal{G}_4$. Since this event has probability one, we can conclude $\zeta_S \xrightarrow{a.s.} \zeta^\star$. $\square$

Proof of Proposition 4. Note that

$$\max_{(\mathbf{p}, \mathbf{q}) \in \mathcal{P} \times \mathcal{Q}} \min_{\mathbf{\Lambda} \in \mathcal{W}(\gamma), \mathbb{P} \in \mathcal{F}(\mathbf{\Lambda}, \theta)} \mathbb{E}_{\mathbb{P}} [\pi(\tilde{\epsilon}, \mathbf{\Lambda}, \mathbf{u}, \mathbf{p}, \mathbf{q})] = \max_{(\mathbf{p}, \mathbf{q}) \in \mathcal{P} \times \mathcal{Q}} \min_{\mathbf{\Lambda} \in \mathcal{W}(\gamma)} \min_{\mathbb{P} \in \mathcal{F}(\mathbf{\Lambda}, \theta)} \mathbb{E}_{\mathbb{P}} [\pi(\tilde{\epsilon}, \mathbf{\Lambda}, \mathbf{u}, \mathbf{p}, \mathbf{q})].$$

The inner minimization problem can be equivalently written as

$$\begin{aligned} \min_{\mathbb{Q}} \quad & \int_{\epsilon \in \mathbb{R}^N} \sum_{s \in [S]} \pi(\epsilon, \mathbf{\Lambda}, \mathbf{u}, \mathbf{p}, \mathbf{q}) d\mathbb{Q}(\epsilon, \hat{\epsilon}_{\mathbf{\Lambda}, s}) \\ \text{s.t.} \quad & \int_{\epsilon \in \mathbb{R}^N} \sum_{s \in [S]} \|\epsilon - \hat{\epsilon}_{\mathbf{\Lambda}, s}\| d\mathbb{Q}(\epsilon, \hat{\epsilon}_{\mathbf{\Lambda}, s}) \leq \theta \\ & \int_{\epsilon \in \mathbb{R}^N} d\mathbb{Q}(\epsilon, \hat{\epsilon}_{\mathbf{\Lambda}, s}) = \frac{1}{S} \quad \forall s \in [S] \\ & \mathbb{Q}(\cdot, \cdot) \geq 0, \end{aligned}$$

whose dual is given by

$$\begin{aligned} & \max_{\kappa \leq 0} \left\{ \theta\kappa + \min_{\epsilon_s \in \mathbb{R}^N} \frac{1}{S} \sum_{s \in [S]} \left\{ \pi(\epsilon_s, \mathbf{\Lambda}, \mathbf{u}, \mathbf{p}, \mathbf{q}) - \kappa \|\epsilon_s - \hat{\epsilon}_{\mathbf{\Lambda}, s}\| \right\} \right\} \\ & = \max_{\kappa \leq 0} \left\{ \theta\kappa + \frac{1}{S} \sum_{s \in [S]} \min_{\epsilon_s \in \mathbb{R}^N} \left\{ \pi(\epsilon_s, \mathbf{\Lambda}, \mathbf{u}, \mathbf{p}, \mathbf{q}) - \kappa \|\epsilon_s - \hat{\epsilon}_{\mathbf{\Lambda}, s}\| \right\} \right\}. \end{aligned}$$

Hence, the inner minimization problem can be replaced with its dual. Combining this dual formulation with the outer maximization and minimization problems yields the equivalent reformulation of **CR0** shown in (4). $\square$

Proof of Proposition 5. From the proof of Proposition 4, the **CR0** model is equivalent to

$$\max_{(\mathbf{p}, \mathbf{q}) \in \mathcal{P} \times \mathcal{Q}} \min_{\mathbf{\Lambda} \in \mathcal{W}(\gamma)} \max_{\kappa \leq 0} \left\{ \theta\kappa + \frac{1}{S} \sum_{s \in [S]} \min_{\epsilon_s \in \mathbb{R}^N} \left\{ \pi(\epsilon_s, \mathbf{\Lambda}, \mathbf{u}, \mathbf{p}, \mathbf{q}) - \kappa \|\epsilon_s - \hat{\epsilon}_{\mathbf{\Lambda}, s}\| \right\} \right\}.$$

By the minimax theorem, the above problem has the following lower bound:

$$\max_{\substack{(\mathbf{p}, \mathbf{q}) \in \mathcal{P} \times \mathcal{Q} \\ \kappa \leq 0}} \min_{\mathbf{\Lambda} \in \mathcal{W}(\gamma)} \left\{ \theta \kappa + \frac{1}{S} \sum_{s \in [S]} \min_{\boldsymbol{\epsilon}_s \in \mathbb{R}^N} \left\{ \pi(\boldsymbol{\epsilon}_s, \mathbf{\Lambda}, \mathbf{u}, \mathbf{p}, \mathbf{q}) - \kappa \|\boldsymbol{\epsilon}_s - \hat{\boldsymbol{\epsilon}}_{\mathbf{\Lambda}, s}\| \right\} \right\}. \quad (17)$$

Introducing $\mathbf{\Lambda}_s \in \mathcal{W}(\gamma)$, $s \in [S]$ such that $\mathbf{\Lambda}_1 = \mathbf{\Lambda}_s$ for $s = 2, \dots, S$, problem (17) becomes

$$\max_{\substack{(\mathbf{p}, \mathbf{q}) \in \mathcal{P} \times \mathcal{Q} \\ \kappa \leq 0}} \left\{ \theta \kappa + \min_{\substack{\mathbf{\Lambda}_1 = \mathbf{\Lambda}_s \in \mathcal{W}(\gamma) \\ s=2, \dots, S}} \frac{1}{S} \sum_{s \in [S]} \min_{\boldsymbol{\epsilon}_s \in \mathbb{R}^N} \left(\pi(\boldsymbol{\epsilon}_s, \mathbf{\Lambda}_s, \mathbf{u}, \mathbf{p}, \mathbf{q}) - \kappa \|\boldsymbol{\epsilon}_s - \hat{\boldsymbol{\epsilon}}_{\mathbf{\Lambda}_s, s}\| \right) \right\}.$$

By Lagrange relaxation, we have

$$\max_{\substack{(\mathbf{p}, \mathbf{q}) \in \mathcal{P} \times \mathcal{Q} \\ \kappa \leq 0}} \left\{ \theta \kappa + \min_{\substack{\mathbf{\Lambda}_s \in \mathcal{W}(\gamma) \\ s \in [S]}} \max_{\sum_{s \in [S]} \mathbf{D}_s = \mathbf{0}} \frac{1}{S} \sum_{s \in [S]} \min_{\boldsymbol{\epsilon}_s \in \mathbb{R}^N} \left(\pi(\boldsymbol{\epsilon}_s, \mathbf{\Lambda}_s, \mathbf{u}, \mathbf{p}, \mathbf{q}) - \kappa \|\boldsymbol{\epsilon}_s - \hat{\boldsymbol{\epsilon}}_{\mathbf{\Lambda}_s, s}\| \right) + \mathbf{D}_s \bullet \mathbf{\Lambda}_s \right\}.$$

Applying the minimax theorem again then yields the following lower bound:

$$\begin{aligned} \max \quad & \theta \kappa + \frac{1}{S} \cdot \mathbf{e}^\top \mathbf{t} \\ \text{s.t.} \quad & t_s \leq \pi(\boldsymbol{\epsilon}_s, \mathbf{\Lambda}_s, \mathbf{u}, \mathbf{p}, \mathbf{q}) - \kappa \|\boldsymbol{\epsilon}_s - \hat{\boldsymbol{\epsilon}}_{\mathbf{\Lambda}_s, s}\| + \mathbf{D}_s \bullet \mathbf{\Lambda}_s \quad \forall \mathbf{\Lambda}_s \in \mathcal{W}(\gamma), \boldsymbol{\epsilon}_s \in \mathbb{R}^N, s \in [S] \\ & \sum_{s \in [S]} \mathbf{D}_s = \mathbf{0} \\ & \mathbf{D}_s \in \mathbb{R}^{N \times (J+N)} \quad \forall s \in [S] \\ & (\mathbf{p}, \mathbf{q}) \in \mathcal{P} \times \mathcal{Q}, \kappa \leq 0, \mathbf{t} \in \mathbb{R}^S. \end{aligned} \quad (18)$$

Defining uncertainty sets $\mathcal{E}_s(\gamma)$, $s \in [S]$, we can equivalently represent model (18) as

$$\begin{aligned} \max \quad & \theta \kappa + S^{-1} \cdot \mathbf{e}^\top \mathbf{t} \\ \text{s.t.} \quad & t_s \leq \pi(\boldsymbol{\epsilon}_s, \mathbf{\Lambda}_s, \mathbf{u}, \mathbf{p}, \mathbf{q}) - \kappa \phi_s + \mathbf{D}_s \bullet \mathbf{\Lambda}_s \quad \forall (\mathbf{\Lambda}_s, \boldsymbol{\epsilon}_s, \phi_s) \in \mathcal{E}_s(\gamma), s \in [S] \\ & \sum_{s \in [S]} \mathbf{D}_s = \mathbf{0} \\ & \mathbf{D}_s \in \mathbb{R}^{N \times (J+N)} \quad \forall s \in [S] \\ & (\mathbf{p}, \mathbf{q}) \in \mathcal{P} \times \mathcal{Q}, \kappa \leq 0, \mathbf{t} \in \mathbb{R}^S. \end{aligned}$$

Finally, applying the enhanced affine recourse adaptation (6) to approximate $(\ell(\mathbf{\Lambda}_s, \mathbf{u}, \mathbf{p}) + \boldsymbol{\epsilon}_s - \mathbf{q})^+$ then yields the [CRO](#) model. The lower bound $Z(\gamma, \theta) \geq \underline{Z}(\gamma, \theta)$ follows straightforwardly. $\square$

Proof of Proposition 6. When $\gamma = 0$, $\mathcal{W}(\gamma) = \{\hat{\mathbf{\Lambda}}\}$. Hence, for each $s \in [S]$, the affine function

$\mathbf{y}_s \in \mathcal{L}$ degenerates to $\mathbf{y}_s \in \underline{\mathcal{L}}$. Letting $\mathbf{D}_s = \mathbf{0}$ for all $s \in [S]$, the **CR0** model now becomes

$$\underline{Z}(0, \theta) = \left[\begin{array}{ll} \max & \mathbf{p}^\top \ell(\hat{\Lambda}, \mathbf{u}, \mathbf{p}) - \mathbf{c}^\top \mathbf{q} + \theta \kappa + S^{-1} \cdot \mathbf{e}^\top \mathbf{t} \\ \text{s.t.} & t_s \leq \mathbf{p}^\top \boldsymbol{\epsilon}_s - \mathbf{g}^\top \mathbf{y}_s(\boldsymbol{\epsilon}_s, \phi_s) - \kappa \phi_s \quad \forall (\boldsymbol{\epsilon}_s, \phi_s) \in \underline{\mathcal{E}}_s, s \in [S] \\ & \mathbf{y}_s(\boldsymbol{\epsilon}_s, \phi_s) \geq \ell(\hat{\Lambda}, \mathbf{u}, \mathbf{p}) + \boldsymbol{\epsilon}_s - \mathbf{q} \quad \forall (\boldsymbol{\epsilon}_s, \phi_s) \in \underline{\mathcal{E}}_s, s \in [S] \\ & \mathbf{y}_s(\boldsymbol{\epsilon}_s, \phi_s) \geq \mathbf{0} \quad \forall (\boldsymbol{\epsilon}_s, \phi_s) \in \underline{\mathcal{E}}_s, s \in [S] \\ & \mathbf{y}_s \in \underline{\mathcal{L}} \quad \forall s \in [S] \\ & (\mathbf{p}, \mathbf{q}) \in \mathcal{P} \times \mathcal{Q}, \mathbf{t} \in \mathbb{R}^S, \kappa \leq 0 \end{array} \right],$$

which is exactly the **RBO** model. Suppose $\theta = 0$ further, we must have $\hat{\boldsymbol{\epsilon}}_s = \hat{\boldsymbol{\epsilon}}_{\hat{\Lambda},s}$ for each $s \in [S]$. Note that the optimal value of the **ETO** model is

$$Z(0, 0) = \left[\begin{array}{ll} \max & \mathbf{p}^\top \ell(\hat{\Lambda}, \mathbf{u}, \mathbf{p}) - \mathbf{c}^\top \mathbf{q} + S^{-1} \cdot \mathbf{e}^\top \mathbf{t} \\ \text{s.t.} & t_s \leq \mathbf{p}^\top \hat{\boldsymbol{\epsilon}}_{\hat{\Lambda},s} - \mathbf{g}^\top \mathbf{y}_s \quad \forall s \in [S] \\ & \mathbf{y}_s \geq \ell(\hat{\Lambda}, \mathbf{u}, \mathbf{p}) + \hat{\boldsymbol{\epsilon}}_{\hat{\Lambda},s} - \mathbf{q} \quad \forall s \in [S] \\ & \mathbf{y}_s \geq \mathbf{0} \quad \forall s \in [S] \\ & \mathbf{y}_s \in \mathbb{R}^N \quad \forall s \in [S] \\ & (\mathbf{p}, \mathbf{q}) \in \mathcal{P} \times \mathcal{Q}, \mathbf{t} \in \mathbb{R}^S \end{array} \right]. \quad (19)$$

We next show that for any feasible solution $(\hat{\mathbf{p}}, \hat{\mathbf{q}}, \hat{\mathbf{t}}, \hat{\mathbf{y}}_s)$ of problem (19), one can construct a feasible solution of **RBO** that achieves the same objective value, thus establishing $\underline{Z}(0, 0) = Z(0, 0)$. To this end, we construct the following solution: $\mathbf{p} = \hat{\mathbf{p}}, \mathbf{q} = \hat{\mathbf{q}}, \mathbf{t} = \hat{\mathbf{t}}, \kappa = -(\|\hat{\mathbf{p}}\|_* + \mathbf{g}^\top \mathbf{e})$ and $\mathbf{y}_s(\boldsymbol{\epsilon}_s, \phi_s) = \hat{\mathbf{y}}_s + \phi_s \mathbf{e}$. Then for any $(\boldsymbol{\epsilon}_s, \phi_s) \in \underline{\mathcal{E}}_s$, we can easily verify that $\mathbf{y}_s(\boldsymbol{\epsilon}_s, \phi_s) \geq \mathbf{0}$,

$$\begin{aligned} \mathbf{y}_s(\boldsymbol{\epsilon}_s, \phi_s) - (\ell(\hat{\Lambda}, \mathbf{u}, \mathbf{p}) + \boldsymbol{\epsilon}_s - \mathbf{q}) &= \hat{\mathbf{y}}_s + \phi_s \mathbf{e} - (\ell(\hat{\Lambda}, \mathbf{u}, \mathbf{p}) + \boldsymbol{\epsilon}_s - \mathbf{q}) \\ &\geq \mathbf{e} \|\boldsymbol{\epsilon}_s - \hat{\boldsymbol{\epsilon}}_{\hat{\Lambda},s}\| - (\boldsymbol{\epsilon}_s - \hat{\boldsymbol{\epsilon}}_{\hat{\Lambda},s}) \\ &\geq \mathbf{e} \|\boldsymbol{\epsilon}_s - \hat{\boldsymbol{\epsilon}}_{\hat{\Lambda},s}\|_\infty - (\boldsymbol{\epsilon}_s - \hat{\boldsymbol{\epsilon}}_{\hat{\Lambda},s}) \geq \mathbf{0}, \end{aligned}$$

and

$$\begin{aligned} t_s - (\mathbf{p}^\top \boldsymbol{\epsilon}_s - \mathbf{g}^\top \mathbf{y}_s(\boldsymbol{\epsilon}_s, \phi_s) - \kappa \phi_s) &\leq \hat{\mathbf{p}}^\top \hat{\boldsymbol{\epsilon}}_{\hat{\Lambda},s} - \mathbf{g}^\top \hat{\mathbf{y}}_s - (\mathbf{p}^\top \boldsymbol{\epsilon}_s - \mathbf{g}^\top \mathbf{y}_s(\boldsymbol{\epsilon}_s, \phi_s) - \kappa \phi_s) \\ &= \hat{\mathbf{p}}^\top (\hat{\boldsymbol{\epsilon}}_{\hat{\Lambda},s} - \boldsymbol{\epsilon}_s) + \mathbf{g}^\top \phi_s \mathbf{e} + \kappa \phi_s \\ &\leq \|\hat{\mathbf{p}}\|_* \phi_s + \mathbf{g}^\top \phi_s \mathbf{e} + \kappa \phi_s \leq 0. \end{aligned}$$

Additionally, the objective value of the **RBO** model is

$$\mathbf{p}^\top \ell(\hat{\Lambda}, \mathbf{u}, \mathbf{p}) - \mathbf{c}^\top \mathbf{q} + \theta \kappa + S^{-1} \cdot \mathbf{e}^\top \mathbf{t} = \hat{\mathbf{p}}^\top \ell(\hat{\Lambda}, \mathbf{u}, \hat{\mathbf{p}}) - \mathbf{c}^\top \hat{\mathbf{q}} + S^{-1} \cdot \mathbf{e}^\top \hat{\mathbf{t}},$$

which is exactly the objective value of problem (19). Hence, we can conclude the proof. $\square$

Appendix C SDP Reformulation of [CRO](#)

Proof of Theorem 3. We show that the [CRO](#) model is equivalent to the semidefinite program:

$$\begin{aligned}
& \max \quad \frac{1}{S} \sum_{s \in [S]} \left(\hat{\mathbf{v}}_s^\top \boldsymbol{\alpha}_s - \mathbf{g}^\top \mathbf{y}_s^0 + (\hat{\mathbf{A}} \hat{\mathbf{U}}^{\mathbf{A}} + \hat{\mathbf{B}} \hat{\mathbf{U}}^{\mathbf{B}}) \bullet \mathbf{R}_s - \gamma \sqrt{S} \eta_s \right) - \mathbf{c}^\top \mathbf{q} + \theta \kappa \\
& \text{s.t.} \quad \boldsymbol{\alpha}_s = \mathbf{p} - \mathbf{Y}_s^\epsilon \mathbf{g} \quad \forall s \in [S] \\
& \quad \mathbf{g}(\mathbf{y}_s^{\mathbf{A}})^\top - \mathbf{p} \mathbf{u}^\top + \boldsymbol{\alpha}_s \hat{\mathbf{u}}_s^\top + \mathbf{R}_s (\hat{\mathbf{U}}^{\mathbf{A}})^\top - \mathbf{D}_s^{\mathbf{A}} = \mathbf{0} \quad \forall s \in [S] \\
& \quad \begin{pmatrix} \mathbf{g}(\mathbf{y}_s^{\mathbf{B}})^\top + \boldsymbol{\alpha}_s \hat{\mathbf{p}}_s^\top + \mathbf{R}_s (\hat{\mathbf{U}}^{\mathbf{B}})^\top - \mathbf{D}_s^{\mathbf{B}} & \mathbf{p} \\ \mathbf{p}^\top & 1 \end{pmatrix} \succeq \mathbf{0} \quad \forall s \in [S] \\
& \quad \|\boldsymbol{\alpha}_s\|_* \leq -\kappa - \mathbf{g}^\top \mathbf{y}_s^1, \|\mathbf{R}_s\|_{\mathbf{F}} \leq \eta_s \quad \forall s \in [S] \\
& \quad \mathbf{y}_{s,n}^0 + (\boldsymbol{\alpha}_{s,n}^1)^\top \hat{\mathbf{v}}_s - \eta_{s,n}^1 \gamma \sqrt{S} + (\hat{\mathbf{A}} \hat{\mathbf{U}}^{\mathbf{A}} + \hat{\mathbf{B}} \hat{\mathbf{U}}^{\mathbf{B}}) \bullet \mathbf{R}_{s,n}^1 + q_n \geq 0 \quad \forall n \in [N], s \in [S] \\
& \quad \boldsymbol{\alpha}_{s,n}^1 = \mathbf{Y}_s^\epsilon \mathbf{e}_n - \mathbf{e}_n \quad \forall n \in [N], s \in [S] \\
& \quad \boldsymbol{\alpha}_{s,n}^1 \hat{\mathbf{u}}_s^\top + \mathbf{R}_{s,n}^1 (\hat{\mathbf{U}}^{\mathbf{A}})^\top = \mathbf{e}_n (\mathbf{y}_s^{\mathbf{A}} - \mathbf{u})^\top \quad \forall n \in [N], s \in [S] \\
& \quad \boldsymbol{\alpha}_{s,n}^1 \hat{\mathbf{p}}_s^\top + \mathbf{R}_{s,n}^1 (\hat{\mathbf{U}}^{\mathbf{B}})^\top - \mathbf{Q}_{s,n}^1 = \mathbf{e}_n (\mathbf{y}_s^{\mathbf{B}} - \mathbf{p})^\top \quad \forall n \in [N], s \in [S] \\
& \quad \|\boldsymbol{\alpha}_{s,n}^1\|_* \leq y_{s,n}^1 \quad \forall n \in [N], s \in [S] \\
& \quad \|\mathbf{R}_{s,n}^1\|_{\mathbf{F}} \leq \eta_{s,n}^1 \quad \forall n \in [N], s \in [S] \quad (20) \\
& \quad \mathbf{y}_{s,n}^0 + (\boldsymbol{\alpha}_{s,n}^2)^\top \hat{\mathbf{v}}_s - \eta_{s,n}^2 \gamma \sqrt{S} + \mathbf{R}_{s,n}^2 \bullet (\hat{\mathbf{A}} \hat{\mathbf{U}}^{\mathbf{A}} + \hat{\mathbf{B}} \hat{\mathbf{U}}^{\mathbf{B}}) \geq 0 \quad \forall n \in [N], s \in [S] \\
& \quad \boldsymbol{\alpha}_{s,n}^2 = \mathbf{Y}_s^\epsilon \mathbf{e}_n \quad \forall n \in [N], s \in [S] \\
& \quad \boldsymbol{\alpha}_{s,n}^2 \hat{\mathbf{u}}_s^\top + \mathbf{R}_{s,n}^2 (\hat{\mathbf{U}}^{\mathbf{A}})^\top = \mathbf{e}_n (\mathbf{y}_s^{\mathbf{A}})^\top \quad \forall n \in [N], s \in [S] \\
& \quad \boldsymbol{\alpha}_{s,n}^2 \hat{\mathbf{p}}_s^\top + \mathbf{R}_{s,n}^2 (\hat{\mathbf{U}}^{\mathbf{B}})^\top - \mathbf{Q}_{s,n}^2 = \mathbf{e}_n (\mathbf{y}_s^{\mathbf{B}})^\top \quad \forall n \in [N], s \in [S] \\
& \quad \|\boldsymbol{\alpha}_{s,n}^2\|_* \leq y_{s,n}^1 \quad \forall n \in [N], s \in [S] \\
& \quad \|\mathbf{R}_{s,n}^2\|_{\mathbf{F}} \leq \eta_{s,n}^2 \quad \forall n \in [N], s \in [S] \\
& \quad \sum_{s \in [S]} \mathbf{D}_s = \mathbf{0}, (\mathbf{p}, \mathbf{q}) \in \mathcal{P} \times \mathcal{Q}, \kappa \leq 0 \\
& \quad \mathbf{y}_s^0 \in \mathbb{R}^N, \mathbf{y}_s^{\mathbf{A}} \in \mathbb{R}^J, \mathbf{y}_s^{\mathbf{B}} \in \mathbb{R}^N, \mathbf{Y}_s^\epsilon \in \mathbb{R}^{N \times N}, \mathbf{y}_s^1 \in \mathbb{R}^N \quad \forall s \in [S] \\
& \quad \mathbf{D}_s \in \mathbb{R}^{N \times (J+N)}, \mathbf{R}_s \in \mathbb{R}^{N \times S}, \boldsymbol{\alpha}_s \in \mathbb{R}^N, \eta_s \in \mathbb{R}_+ \quad \forall s \in [S] \\
& \quad \mathbf{Q}_{s,n}^1 \in \mathbb{S}_+^N, \mathbf{R}_{s,n}^1 \in \mathbb{R}^{N \times S}, \boldsymbol{\alpha}_{s,n}^1 \in \mathbb{R}^N, \eta_{s,n}^1 \in \mathbb{R}_+ \quad \forall n \in [N], s \in [S] \\
& \quad \mathbf{Q}_{s,n}^2 \in \mathbb{S}_+^N, \mathbf{R}_{s,n}^2 \in \mathbb{R}^{N \times S}, \boldsymbol{\alpha}_{s,n}^2 \in \mathbb{R}^N, \eta_{s,n}^2 \in \mathbb{R}_+ \quad \forall n \in [N], s \in [S].
\end{aligned}$$

Here, we use $\|\cdot\|_*$ as the dual norm of norm $\|\cdot\|$.

Given each $s \in [S]$, let us first consider the following problem,

$$\min_{(\boldsymbol{\Lambda}_s, \boldsymbol{\epsilon}_s, \phi_s) \in \mathcal{E}_s(\gamma)} \mathbf{p}^\top (\ell(\boldsymbol{\Lambda}_s, \mathbf{u}, \mathbf{p}) + \boldsymbol{\epsilon}_s) - \mathbf{g}^\top \mathbf{y}_s(\boldsymbol{\epsilon}_s, \mathbf{A}_s, \mathbf{B}_s, \phi_s) - \kappa \phi_s + \mathbf{D}_s \bullet \boldsymbol{\Lambda}_s. \quad (21)$$

Recall the definition of $\mathcal{E}_s(\gamma)$, $\mathbf{A} = (\mathbf{A}, \mathbf{B})$, $\mathbf{\Sigma} = S^{-1} \cdot \hat{\mathbf{U}}\hat{\mathbf{U}}^\top$, and

$$\hat{\mathbf{U}} = \left(\begin{pmatrix} \hat{\mathbf{u}}_1 \\ \hat{\mathbf{p}}_1 \end{pmatrix}, \dots, \begin{pmatrix} \hat{\mathbf{u}}_S \\ \hat{\mathbf{p}}_S \end{pmatrix} \right) \in \mathbb{R}^{(J+N) \times S}.$$

In what follows, we directly use $\mathbf{A}$ and $\mathbf{B}$ and denote $\hat{\mathbf{U}}^A = (\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, \dots, \hat{\mathbf{u}}_S) \in \mathbb{R}^{J \times S}$, $\hat{\mathbf{U}}^B = (\hat{\mathbf{p}}_1, \hat{\mathbf{p}}_2, \dots, \hat{\mathbf{p}}_S) \in \mathbb{R}^{N \times S}$. We now can express problem (21) as

$$\begin{aligned} \min \quad & \mathbf{p}^\top (\mathbf{A}_s \mathbf{u} + \mathbf{B}_s \mathbf{p} + \boldsymbol{\epsilon}_s) - \mathbf{g}^\top (\mathbf{y}_s^0 + \mathbf{A}_s \mathbf{y}_s^A + \mathbf{B}_s \mathbf{y}_s^B + \mathbf{Y}_s^\epsilon \boldsymbol{\epsilon}_s + \phi_s \mathbf{y}_s^1) - \kappa \phi_s + \mathbf{D}_s^A \bullet \mathbf{A}_s + \mathbf{D}_s^B \bullet \mathbf{B}_s \\ \text{s.t.} \quad & \|\boldsymbol{\epsilon}_s - \hat{\mathbf{v}}_s + \mathbf{A}_s \hat{\mathbf{u}}_s + \mathbf{B}_s \hat{\mathbf{p}}_s\| \leq \phi_s \\ & \|(\mathbf{A}_s - \hat{\mathbf{A}})\hat{\mathbf{U}}^A + (\mathbf{B}_s - \hat{\mathbf{B}})\hat{\mathbf{U}}^B\|_F \leq \gamma \sqrt{S} \\ & -\mathbf{B}_s \geq \mathbf{0} \\ & \mathbf{A}_s \in \mathbb{R}^{N \times J}, \mathbf{B}_s \in \mathbb{R}^{N \times N}, \boldsymbol{\epsilon}_s \in \mathbb{R}^N, \phi_s \in \mathbb{R}_+, \end{aligned}$$

whose dual can be derived as follows:

$$\begin{aligned} \max \quad & \hat{\mathbf{v}}_s^\top \boldsymbol{\alpha}_s - \mathbf{g}^\top \mathbf{y}_s^0 + (\hat{\mathbf{A}}\hat{\mathbf{U}}^A + \hat{\mathbf{B}}\hat{\mathbf{U}}^B) \bullet \mathbf{R}_s - \gamma \sqrt{S} \eta_s \\ \text{s.t.} \quad & \boldsymbol{\alpha}_s = \mathbf{p} - \mathbf{Y}_s^\epsilon \mathbf{g} \\ & \mathbf{g}(\mathbf{y}_s^A)^\top - \mathbf{p} \mathbf{u}^\top + \boldsymbol{\alpha}_s \hat{\mathbf{u}}_s^\top + \mathbf{R}_s (\hat{\mathbf{U}}^A)^\top - \mathbf{D}_s^A = \mathbf{0} \\ & \mathbf{g}(\mathbf{y}_s^B)^\top - \mathbf{p} \mathbf{p}^\top + \boldsymbol{\alpha}_s \hat{\mathbf{p}}_s^\top + \mathbf{R}_s (\hat{\mathbf{U}}^B)^\top - \mathbf{Q}_s - \mathbf{D}_s^B = \mathbf{0} \\ & \|\boldsymbol{\alpha}_s\|_* \leq -\kappa - \mathbf{g}^\top \mathbf{y}_s^1 \\ & \|\mathbf{R}_s\|_F \leq \eta_s \\ & \mathbf{Q}_s \in \mathbb{S}_+^N, \mathbf{R}_s \in \mathbb{R}^{N \times S}, \boldsymbol{\alpha}_s \in \mathbb{R}^N, \eta_s \in \mathbb{R}_+. \end{aligned}$$

Here, the only non-linear constraint $\mathbf{g}(\mathbf{y}_s^B)^\top - \mathbf{p} \mathbf{p}^\top + \boldsymbol{\alpha}_s \hat{\mathbf{p}}_s^\top + \mathbf{R}_s (\hat{\mathbf{U}}^B)^\top - \mathbf{Q}_s - \mathbf{D}_s^B = \mathbf{0}$ can be eliminated by Schur complement using $\mathbf{Q}_s \in \mathbb{S}_+^N$:

$$\begin{pmatrix} \mathbf{g}(\mathbf{y}_s^B)^\top + \boldsymbol{\alpha}_s \hat{\mathbf{p}}_s^\top + \mathbf{R}_s (\hat{\mathbf{U}}^B)^\top - \mathbf{D}_s^B & \mathbf{p} \\ \mathbf{p}^\top & 1 \end{pmatrix} \in \mathbb{S}_+^{N+1}.$$

Now, let us consider the following batch of constraints:

$$\mathbf{y}_s(\mathbf{A}_s, \mathbf{B}_s, \boldsymbol{\epsilon}_s, \phi_s) \geq \mathbf{A}_s \mathbf{u} + \mathbf{B}_s \mathbf{p} + \boldsymbol{\epsilon}_s - \mathbf{q} \quad \forall (\mathbf{A}_s, \mathbf{B}_s, \boldsymbol{\epsilon}_s, \phi_s) \in \mathcal{E}_s(\gamma), s \in [S].$$

Given any $n \in [N]$ and $s \in [S]$, the associated constraint can be written as

$$\min_{(\mathbf{A}_s, \mathbf{B}_s, \boldsymbol{\epsilon}_s, \phi_s) \in \mathcal{E}_s(\gamma)} \mathbf{e}_n^\top (\mathbf{A}_s (\mathbf{y}_s^A - \mathbf{u}) + \mathbf{B}_s (\mathbf{y}_s^B - \mathbf{p}) + (\mathbf{Y}_s^\epsilon \boldsymbol{\epsilon}_s - \boldsymbol{\epsilon}_s) + \phi_s \mathbf{y}_s^1) \geq -q_n - y_{s,n}^0.$$

For the left-hand side problem of this inequality, we can express it as

$$\begin{aligned}
\min \quad & \mathbf{e}_n^\top (\mathbf{A}_s(\mathbf{y}_s^A - \mathbf{u}) + \mathbf{B}_s(\mathbf{y}_s^B - \mathbf{p}) + (\mathbf{Y}_s^\epsilon - \mathbf{I})\boldsymbol{\epsilon}_s + \mathbf{y}_s^1 \phi_s) \\
\text{s.t.} \quad & \|\boldsymbol{\epsilon}_s - \hat{\mathbf{v}}_s + \mathbf{A}_s \hat{\mathbf{u}}_s + \mathbf{B}_s \hat{\mathbf{p}}_s\| \leq \phi_s \\
& \|(\mathbf{A}_s - \hat{\mathbf{A}})\hat{\mathbf{U}}^A + (\mathbf{B}_s - \hat{\mathbf{B}})\hat{\mathbf{U}}^B\|_F \leq \gamma\sqrt{S} \\
& -\mathbf{B}_s \geq \mathbf{0} \\
& \mathbf{A}_s \in \mathbb{R}^{N \times J}, \mathbf{B}_s \in \mathbb{R}^{N \times N}, \boldsymbol{\epsilon}_s \in \mathbb{R}^N, \phi_s \in \mathbb{R}_+.
\end{aligned} \tag{22}$$

The dual of problem (22) is as follows:

$$\begin{aligned}
\max \quad & \hat{\mathbf{v}}_s^\top \boldsymbol{\alpha}_{s,n}^1 - \gamma\sqrt{S}\eta_{s,n}^1 + (\hat{\mathbf{A}}\hat{\mathbf{U}}^A + \hat{\mathbf{B}}\hat{\mathbf{U}}^B) \bullet \mathbf{R}_{s,n}^1 \\
\text{s.t.} \quad & \boldsymbol{\alpha}_{s,n}^1 = \mathbf{Y}_s^{\epsilon\top} \mathbf{e}_n - \mathbf{e}_n \\
& \boldsymbol{\alpha}_{s,n}^1 \hat{\mathbf{u}}_s^\top + \mathbf{R}_{s,n}^1 (\hat{\mathbf{U}}^A)^\top = \mathbf{e}_n (\mathbf{y}_s^A - \mathbf{u})^\top \\
& \boldsymbol{\alpha}_{s,n}^1 \hat{\mathbf{p}}_s^\top + \mathbf{R}_{s,n}^1 (\hat{\mathbf{U}}^B)^\top - \mathbf{Q}_{s,n}^1 = \mathbf{e}_n (\mathbf{y}_s^B - \mathbf{p})^\top \\
& \|\boldsymbol{\alpha}_{s,n}^1\|_* \leq y_{s,n}^1 \\
& \|\mathbf{R}_{s,n}^1\|_F \leq \eta_{s,n}^1 \\
& \mathbf{Q}_{s,n}^1 \in \mathbb{S}_+^N, \mathbf{R}_{s,n}^1 \in \mathbb{R}^{N \times S}, \boldsymbol{\alpha}_{s,n}^1 \in \mathbb{R}^N, \eta_{s,n}^1 \in \mathbb{R}_+.
\end{aligned}$$

Similarly, for the batch of constraints

$$\mathbf{y}_s(\mathbf{A}_s, \mathbf{B}_s, \boldsymbol{\epsilon}_s, \phi_s) \geq \mathbf{0} \quad \forall (\mathbf{A}_s, \mathbf{B}_s, \boldsymbol{\epsilon}_s, \phi_s) \in \mathcal{E}_s(\gamma), s \in [S],$$

given $n \in [N], s \in [S]$, the corresponding left-hand side problem is

$$\begin{aligned}
\min \quad & \mathbf{e}_n^\top (\mathbf{A}_s \mathbf{y}_s^A + \mathbf{B}_s \mathbf{y}_s^B + \mathbf{Y}_s^\epsilon \boldsymbol{\epsilon}_s + \mathbf{y}_s^1 \phi_s) \\
\text{s.t.} \quad & \|\boldsymbol{\epsilon}_s - \hat{\mathbf{v}}_s + \mathbf{A}_s \hat{\mathbf{u}}_s + \mathbf{B}_s \hat{\mathbf{p}}_s\| \leq \phi_s \\
& \|(\mathbf{A}_s - \hat{\mathbf{A}})\hat{\mathbf{U}}^A + (\mathbf{B}_s - \hat{\mathbf{B}})\hat{\mathbf{U}}^B\|_F \leq \gamma\sqrt{S} \\
& -\mathbf{B}_s \geq \mathbf{0} \\
& \mathbf{A}_s \in \mathbb{R}^{N \times J}, \mathbf{B}_s \in \mathbb{R}^{N \times N}, \boldsymbol{\epsilon}_s \in \mathbb{R}_+, \phi_s \in \mathbb{R}_+.
\end{aligned}$$

We can derive its dual problem as

$$\begin{aligned}
\max \quad & \hat{\mathbf{v}}_s^\top \boldsymbol{\alpha}_{s,n}^2 - \gamma\sqrt{S}\eta_{s,n}^2 + (\hat{\mathbf{A}}\hat{\mathbf{U}}^A + \hat{\mathbf{B}}\hat{\mathbf{U}}^B) \bullet \mathbf{R}_{s,n}^2 \\
\text{s.t.} \quad & \boldsymbol{\alpha}_{s,n}^2 = \mathbf{Y}_s^{\epsilon\top} \mathbf{e}_n \\
& \boldsymbol{\alpha}_{s,n}^2 \hat{\mathbf{u}}_s^\top + \mathbf{R}_{s,n}^2 (\hat{\mathbf{U}}^A)^\top = \mathbf{e}_n (\mathbf{y}_s^A)^\top \\
& \boldsymbol{\alpha}_{s,n}^2 \hat{\mathbf{p}}_s^\top + \mathbf{R}_{s,n}^2 (\hat{\mathbf{U}}^B)^\top - \mathbf{Q}_{s,n}^2 = \mathbf{e}_n (\mathbf{y}_s^B)^\top \\
& \|\boldsymbol{\alpha}_{s,n}^2\|_* \leq y_{s,n}^1 \\
& \|\mathbf{R}_{s,n}^2\|_F \leq \eta_{s,n}^2 \\
& \mathbf{Q}_{s,n}^2 \in \mathbb{S}_+^N, \mathbf{R}_{s,n}^2 \in \mathbb{R}^{N \times S}, \boldsymbol{\alpha}_{s,n}^2 \in \mathbb{R}^N, \eta_{s,n}^2 \in \mathbb{R}_+.
\end{aligned}$$

Therefore, we can finally obtain problem (20) as a reformulation of the **CR0** model. $\square$

Appendix D Details for The **G-CR0** Model

In this appendix, we show that for the out-of-sample performance of any solution $(\mathbf{p}^\dagger, \mathbf{q}^\dagger) \in \mathcal{P} \times \mathcal{Q}$, the **G-CR0** model provides a $(1 - \delta)$ -confident lower bound as specified in the form (2). Subsequently, in Appendix D.2, we establish conditions under which **G-CR0** admits a tractable lower bound.

Appendix D.1 Performance Guarantees and Optimality

We make the common assumption about the residual distribution.

Assumption 3. *There exist some $\alpha > 1$ and $K > 0$ such that $\mathbb{E}_{\mathbb{P}^\star} [\exp(\|\tilde{\epsilon}\|^\alpha)] \leq K < \infty$.*

We use the following assumption to characterize a finite-sample confidence region of the estimated parameter, thereby allowing the prediction model (7) in **G-CR0** to extend beyond linear models.

Assumption 4. *For any $\delta \in (0, 1)$, it holds that*

$$\mathbb{P}^S[\mathbf{\Lambda}^\star \in \mathcal{W}(\gamma_*)] \geq 1 - \delta \quad \text{for some } \gamma_* = \mathcal{O}\left(\left(\frac{\log(1/\delta)}{S}\right)^\nu\right) \text{ with } \nu > 0.$$

Moreover, there exists a constant $L > 0$ such that

$$\|\ell(\mathbf{\Lambda}, \mathbf{u}, \mathbf{p}) - \ell(\mathbf{\Lambda}^\star, \mathbf{u}, \mathbf{p})\| \leq L \cdot \rho(\mathbf{\Lambda} - \mathbf{\Lambda}^\star) \quad \forall \mathbf{\Lambda} \in \Omega, \mathbf{u} \in \mathcal{U}, \mathbf{p} \in \mathcal{P}.$$

We have the following generalized finite-sample performance guarantees.

Theorem 4. *Suppose $N \neq 2$, and Assumptions 3 and 4 hold. Then there exist constants $k_1, k_2 > 0$ depending on $\mathbb{P}^\star$ only through α, K , and N such that for any $\mathbf{u} \in \mathcal{U}$ and $\delta \in (0, 1)$, with probability at least $1 - \delta$, the optimal value $Z(\gamma, \theta)$ of the **G-CR0** model with*

$$\gamma \in [0, \gamma_*] \quad \text{and} \quad \theta = (\gamma_* - \gamma) \cdot 2L + \begin{cases} \left(\frac{\log(k_1/\delta)}{k_2 S}\right)^{\min\{1/N, 1/2\}} & \text{if } S \geq \frac{\log(k_1/\delta)}{k_2} \\ \left(\frac{\log(k_1/\delta)}{k_2 S}\right)^{1/\alpha} & \text{if } S < \frac{\log(k_1/\delta)}{k_2} \end{cases}$$

is a lower bound of the out-of-sample performance of the optimal solution $(\mathbf{p}^\dagger, \mathbf{q}^\dagger)$ of **G-CR0**. That is,

$$\mathbb{P}^S\left[\mathbb{E}_{\mathbb{P}^\star}[\pi(\tilde{\epsilon}, \mathbf{\Lambda}^\star, \mathbf{u}, \mathbf{p}^\dagger, \mathbf{q}^\dagger)] \geq Z(\gamma, \theta)\right] \geq 1 - \delta.$$

Proof of Theorem 4. We use the same notations and arguments as in the proof of Theorem 1. It suffices to prove that with probability at least $1 - \delta$, we have $d(\mathbb{Q}^*(\Lambda^*, \mathbf{u}, \mathbf{p}), \mathbb{Q}_\Lambda(\Lambda, \mathbf{u}, \mathbf{p})) \leq \theta$ for some $\Lambda \in \mathcal{W}(\gamma)$ and θ as specified in the theorem. To this end, we construct

$$\Lambda = \hat{\Lambda} + \gamma \cdot \frac{\Lambda^* - \hat{\Lambda}}{\rho(\Lambda^* - \hat{\Lambda})},$$

which satisfies $\Lambda \in \mathcal{W}(\gamma)$ as $\rho(\Lambda - \hat{\Lambda}) = \gamma$. Equipped with Assumption 4, we can establish

$$d(\mathbb{Q}^*(\Lambda^*, \mathbf{u}, \mathbf{p}), \mathbb{Q}_\Lambda(\Lambda, \mathbf{u}, \mathbf{p})) \leq d(\mathbb{P}^*, \mathbb{P}_{\Lambda^*}) + 2L \cdot \rho(\Lambda^* - \Lambda) \quad (23)$$

Assumption 4 implies that $\rho(\Lambda^* - \Lambda) \leq \gamma_* - \gamma$ with probability at least $1 - \frac{\delta}{2}$. Proposition 2 shows that under Assumption 3, $d(\mathbb{P}^*, \mathbb{P}_{\Lambda^*}) \leq \bar{\theta}$ with probability at least $1 - \frac{\delta}{2}$. Applying the union bound and using (23) then completes the proof. $\square$

Based on the finite-sample performance guarantee, we can further establish the asymptotic optimality of the **G-CRO** model under additional assumptions that the norm used to construct the parameter uncertainty set is sample-independent and that the reward function π is uniformly Lipschitz continuous with respect to the uncertain input parameter.

Assumption 5. *The norm ρ used in the **G-CRO** model is sample-independent. Moreover, for $\|\cdot\|$ used in the Wasserstein distance $d(\cdot, \cdot)$, there exists a finite constant $L_\pi > 0$ such that*

$$|\pi(\mathbf{p}, \mathbf{q}, \mathbf{v}) - \pi(\mathbf{p}, \mathbf{q}, \mathbf{v}')| \leq L_\pi \cdot \|\mathbf{v} - \mathbf{v}'\| \quad \forall (\mathbf{p}, \mathbf{q}) \in \mathcal{P} \times \mathcal{Q}, \mathbf{v}, \mathbf{v}' \in \mathcal{V}.$$

Theorem 5. *Suppose $N \neq 2$, and Assumptions 3, 4, and 5 hold. Fix any covariate $\mathbf{u} \in \mathcal{U}$. For each sample size S , set $\delta_S = S^{-2}$, and choose a size $\gamma_{*,S}$ satisfying Assumption 4 at confidence level δ_S and a radius $\bar{\theta}_S$ as in Theorem 4 evaluated at δ_S such that $\gamma_{*,S} = \mathcal{O}((\frac{\log S}{S})^\nu)$ and $\bar{\theta}_S \rightarrow 0$. Define $\mathcal{W}_S(\gamma) = \{\Lambda \in \Omega \mid \rho(\hat{\Lambda}_S - \Lambda) \leq \gamma\}$ and $\mathcal{F}_S(\Lambda, \theta) = \{\mathbb{P} \in \mathcal{M}(\Xi) \mid d(\mathbb{P}, \mathbb{P}_\Lambda) \leq \theta\}$. Consider the optimal solution $(\mathbf{p}_S^\dagger, \mathbf{q}_S^\dagger)$ of the **G-CRO** with $\mathcal{W}_S(\gamma_S)$ and $\mathcal{F}_S(\Lambda, \theta_S)$ for some $\gamma_S \in [0, \gamma_{*,S}]$ and $\theta_S = 2L(\gamma_{*,S} - \gamma_S) + \bar{\theta}_S$, where L is the Lipschitz constant in Assumption 4. Let $(\mathbf{p}^*, \mathbf{q}^*)$ be the optimal solution of the true contextual stochastic optimization problem. It holds that*

$$\mathbb{E}_{\mathbb{P}^*}[\pi(\tilde{\epsilon}, \Lambda^*, \mathbf{u}, \mathbf{p}_S^\dagger, \mathbf{q}_S^\dagger)] \xrightarrow{a.s.} \mathbb{E}_{\mathbb{P}^*}[\pi(\tilde{\epsilon}, \Lambda^*, \mathbf{u}, \mathbf{p}^*, \mathbf{q}^*)].$$

Proof of Theorem 5. The proof is similar to the proof of Theorem 2. For completeness, we include the details below.

First, we define the true optimal value and the out-of-sample value of the **G-CRO** solution by

$$\zeta^* = \mathbb{E}_{\mathbb{P}^*}[\pi(\tilde{\epsilon}, \Lambda^*, \mathbf{u}, \mathbf{p}^*, \mathbf{q}^*)] \quad \text{and} \quad \zeta_S = \mathbb{E}_{\mathbb{P}^*}[\pi(\tilde{\epsilon}, \Lambda^*, \mathbf{u}, \mathbf{p}_S^\dagger, \mathbf{q}_S^\dagger)].$$

Let Z_S be the optimal value of **G-CRO** with (γ_S, θ_S) . Since $(\mathbf{p}^*, \mathbf{q}^*)$ is optimal for the true problem,

$$\zeta_S \leq \zeta^* \quad \forall S \geq 1. \quad (24)$$

On the other hand, Theorem 4, applied with confidence level $\delta_S = S^{-2}$, gives $\mathbb{P}^S[\zeta_S \geq Z_S] \geq 1 - S^{-2}$. Since $\sum_{S=1}^{\infty} S^{-2} < \infty$, the Borel–Cantelli lemma yields an event $\mathcal{G}_1$ with $\mathbb{P}[\mathcal{G}_1] = 1$ such that for any $\omega \in \mathcal{G}_1$, there exists $S_1(\omega) > 0$ satisfying

$$\zeta_S \geq Z_S \quad \forall S \geq S_1(\omega). \quad (25)$$

Therefore, it remains to prove that $\liminf_{S \rightarrow \infty} Z_S \geq \zeta^*$ almost surely.

By Assumption 4 with $\delta_S = S^{-2}$, $\mathbb{P}^S[\mathbf{\Lambda}^* \in \mathcal{W}_S(\gamma_{*,S})] \geq 1 - S^{-2}$. The Borel–Cantelli lemma then gives an event $\mathcal{G}_2$ with $\mathbb{P}[\mathcal{G}_2] = 1$ such that, for any $\omega \in \mathcal{G}_2$, there exists $S_2(\omega) > 0$ satisfying

$$\mathbf{\Lambda}^* \in \mathcal{W}_S(\gamma_{*,S}) \quad \forall S \geq S_2(\omega). \quad (26)$$

Similarly, the Borel–Cantelli lemma and Theorem 4 with $\delta_S = S^{-2}$ imply an event $\mathcal{G}_3$ with $\mathbb{P}[\mathcal{G}_3] = 1$ such that, for any $\omega \in \mathcal{G}_3$, there exists $S_3(\omega) > 0$ satisfying

$$d(\mathbb{P}^*, \mathbb{P}_{\mathbf{\Lambda}^*}) \leq \bar{\theta}_S \quad \forall S \geq S_3(\omega). \quad (27)$$

Fix any $\omega \in \mathcal{G}_2 \cap \mathcal{G}_3$ and any $S \geq \max\{S_2(\omega), S_3(\omega)\}$, we have (26) and (27). For any $\mathbf{\Lambda} \in \mathcal{W}_S(\gamma_S)$ and any $\mathbb{P} \in \mathcal{F}_S(\mathbf{\Lambda}, \theta_S)$, by the triangle inequality for the fixed norm ρ , we have

$$\rho(\mathbf{\Lambda}^* - \mathbf{\Lambda}) \leq \rho(\mathbf{\Lambda}^* - \hat{\mathbf{\Lambda}}_S) + \rho(\hat{\mathbf{\Lambda}}_S - \mathbf{\Lambda}) \leq \gamma_{*,S} + \gamma_S.$$

Let $\mathbb{Q}^*(\mathbf{\Lambda}^*, \mathbf{u}, \mathbf{p}^*)$ be the distribution of $\ell(\mathbf{\Lambda}^*, \mathbf{u}, \mathbf{p}^*) + \tilde{\epsilon}$ under $\tilde{\epsilon} \sim \mathbb{P}^*$, $\mathbb{Q}_{\mathbf{\Lambda}}(\mathbf{\Lambda}, \mathbf{u}, \mathbf{p}^*)$ be the distribution of $\ell(\mathbf{\Lambda}, \mathbf{u}, \mathbf{p}^*) + \tilde{\epsilon}$ under $\tilde{\epsilon} \sim \mathbb{P}_{\mathbf{\Lambda}}$, and $\mathbb{Q}$ be the distribution of $\ell(\mathbf{\Lambda}, \mathbf{u}, \mathbf{p}^*) + \tilde{\epsilon}$ under $\tilde{\epsilon} \sim \mathbb{P}$. Using the same shifted-distribution comparison as in the proof of Theorem 4, followed by the definition of $\mathcal{F}_S(\mathbf{\Lambda}, \theta_S)$, we obtain

$$\begin{aligned} d(\mathbb{Q}^*(\mathbf{\Lambda}^*, \mathbf{u}, \mathbf{p}^*), \mathbb{Q}) &\leq d(\mathbb{P}^*, \mathbb{P}_{\mathbf{\Lambda}^*}) + 2L \cdot \rho(\mathbf{\Lambda}^* - \mathbf{\Lambda}) + \theta_S \\ &\leq \bar{\theta}_S + 2L \cdot (\gamma_{*,S} + \gamma_S) + 2L \cdot (\gamma_{*,S} - \gamma_S) + \bar{\theta}_S \\ &= 2\bar{\theta}_S + 4L \cdot \gamma_{*,S}. \end{aligned}$$

As $\mathbf{v} \mapsto \pi(\mathbf{p}^*, \mathbf{q}^*, \mathbf{v})$ is L_π -Lipschitz (Assumption 5), the Kantorovich–Rubinstein inequality implies

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}[\pi(\tilde{\epsilon}, \mathbf{\Lambda}, \mathbf{u}, \mathbf{p}^*, \mathbf{q}^*)] &= \mathbb{E}_{\mathbb{Q}}[\pi(\mathbf{p}^*, \mathbf{q}^*, \tilde{\mathbf{v}})] \\ &\geq \mathbb{E}_{\mathbb{Q}^*(\mathbf{\Lambda}^*, \mathbf{u}, \mathbf{p}^*)}[\pi(\mathbf{p}^*, \mathbf{q}^*, \tilde{\mathbf{v}})] - L_\pi \cdot d(\mathbb{Q}^*(\mathbf{\Lambda}^*, \mathbf{u}, \mathbf{p}^*), \mathbb{Q}) \\ &\geq \zeta^* - L_\pi \cdot (2\bar{\theta}_S + 4L\gamma_{*,S}), \end{aligned}$$

which holds uniformly for all $\mathbf{\Lambda} \in \mathcal{W}_S(\gamma_S)$ and $\mathbb{P} \in \mathcal{F}_S(\mathbf{\Lambda}, \theta_S)$. Consequently, we obtain

$$Z_S \geq \min_{\mathbf{\Lambda} \in \mathcal{W}_S(\gamma_S), \mathbb{P} \in \mathcal{F}_S(\mathbf{\Lambda}, \theta_S)} \mathbb{E}_{\mathbb{P}}[\pi(\tilde{\epsilon}, \mathbf{\Lambda}, \mathbf{u}, \mathbf{p}^*, \mathbf{q}^*)] \geq \zeta^* - L_{\pi} \cdot (2\bar{\theta}_S + 4L\gamma_{*,S}). \quad (28)$$

Finally, on the event $\mathcal{G}_1 \cap \mathcal{G}_2 \cap \mathcal{G}_3$, equations (24), (25), and (28) imply that, for any $S \geq \max\{S_1(\omega), S_2(\omega), S_3(\omega)\}$, we have $\zeta^* - L_{\pi} \cdot (2\bar{\theta}_S + 4L\gamma_{*,S}) \leq Z_S \leq \zeta_S \leq \zeta^*$. Because $\gamma_{*,S} \rightarrow 0$ and $\bar{\theta}_S \rightarrow 0$, it holds that

$$\liminf_{S \rightarrow \infty} \zeta_S \geq \zeta^* \quad \text{and} \quad \limsup_{S \rightarrow \infty} \zeta_S \leq \zeta^*.$$

Therefore, $\zeta_S \rightarrow \zeta^*$ on $\mathcal{G}_1 \cap \mathcal{G}_2 \cap \mathcal{G}_3$. Since $\mathbb{P}[\mathcal{G}_1 \cap \mathcal{G}_2 \cap \mathcal{G}_3] = 1$, we have $\zeta_S \xrightarrow{a.s.} \zeta^*$. $\square$

Appendix D.2 Tractable Approximation

First of all, by the same argument as in Proposition 4, we can reformulate the **G-CRO** model as

$$Z(\gamma, \theta) = \max_{(\mathbf{p}, \mathbf{q}) \in \mathcal{P} \times \mathcal{Q}} \min_{\mathbf{\Lambda} \in \mathcal{W}(\gamma)} \max_{\kappa \leq 0} \left\{ \theta\kappa + S^{-1} \cdot \sum_{s \in [S]} \min_{\epsilon_s \in \Xi} (\pi(\epsilon_s, \mathbf{\Lambda}, \mathbf{u}, \mathbf{p}, \mathbf{q}) - \kappa \|\epsilon_s - \hat{\epsilon}_{\mathbf{\Lambda}, s}\|) \right\}.$$

Next, by the same argument as in Proposition 5, we have the following lower bound:

$$\begin{aligned} & \max \quad \theta\kappa + S^{-1} \cdot \mathbf{e}^\top \mathbf{t} \\ \text{s.t.} \quad & t_s \leq \pi(\epsilon_s, \mathbf{\Lambda}_s, \mathbf{u}, \mathbf{p}, \mathbf{q}) - \kappa \|\epsilon_s - \hat{\epsilon}_{\mathbf{\Lambda}_s, s}\| + \mathbf{D}_s \bullet \mathbf{\Lambda}_s \quad \forall \mathbf{\Lambda}_s \in \mathcal{W}(\gamma), \epsilon_s \in \Xi, s \in [S] \\ & \sum_{s \in [S]} \mathbf{D}_s = \mathbf{0} \\ & (\mathbf{p}, \mathbf{q}) \in \mathcal{P} \times \mathcal{Q}, \kappa \leq 0, \mathbf{t} \in \mathbb{R}^S, \mathbf{D}_s \in \mathbb{R}^{N \times (J+N)} \quad \forall s \in [S] \end{aligned}$$

Indeed, such a lower bound is exact if for any $(\epsilon, \mathbf{u}, \mathbf{p}, \mathbf{q}) \in \Xi \times \mathcal{U} \times \mathcal{P} \times \mathcal{Q}$, $\pi(\epsilon, \mathbf{\Lambda}, \mathbf{u}, \mathbf{p}, \mathbf{q})$ is convex in $\mathbf{\Lambda} \in \mathcal{W}(\gamma)$. This is the case when considering a multiproduct pricing problem without inventory, where the reward function, expressed as

$$\pi(\epsilon, \mathbf{\Lambda}, \mathbf{u}, \mathbf{p}, \mathbf{q}) = \mathbf{p}^\top (\ell(\mathbf{\Lambda}, \mathbf{u}, \mathbf{p}) + \epsilon),$$

is actually affine in $\mathbf{\Lambda}$.

Finally, we can invoke the enhanced affine recourse adaptation to obtain a robust optimization

problem as follows:

$$\underline{Z}(\gamma, \theta) = \left[\begin{array}{l} \max \quad \theta \kappa + S^{-1} \cdot \mathbf{e}^\top \mathbf{t} \\ \text{s.t.} \quad t_s \leq \mathbf{d}^\top \mathbf{y}_s(\boldsymbol{\Lambda}_s, \boldsymbol{\epsilon}_s, \phi_s) - \kappa \phi_s + \mathbf{D}_s \bullet \boldsymbol{\Lambda}_s \quad \forall (\boldsymbol{\Lambda}_s, \boldsymbol{\epsilon}_s, \phi_s) \in \mathcal{E}_s(\gamma), s \in [S] \\ \quad \mathbf{W} \mathbf{y}_s(\boldsymbol{\Lambda}_s, \boldsymbol{\epsilon}_s, \phi_s) \leq \mathbf{h}(\boldsymbol{\epsilon}_s, \boldsymbol{\Lambda}_s, \mathbf{u}, \mathbf{p}, \mathbf{q}) \quad \forall (\boldsymbol{\Lambda}_s, \boldsymbol{\epsilon}_s, \phi_s) \in \mathcal{E}_s(\gamma), s \in [S] \\ \quad \sum_{s \in [S]} \mathbf{D}_s = \mathbf{0} \\ \quad \mathbf{D}_s \in \mathbb{R}^{N \times (J+N)}, \mathbf{y}_s \in \mathcal{L} \quad \forall s \in [S] \\ \quad (\mathbf{p}, \mathbf{q}) \in \mathcal{P} \times \mathcal{Q}, \mathbf{t} \in \mathbb{R}^S, \kappa \leq 0 \end{array} \right], \quad (\underline{\mathbf{G}}\text{-}\underline{\mathbf{CRQ}})$$

where for each $s \in [S]$, the uncertainty set $\mathcal{E}_s(\gamma)$ is represented as

$$\mathcal{E}_s(\gamma) = \{(\boldsymbol{\Lambda}_s, \boldsymbol{\epsilon}_s, \phi_s) \in \mathcal{W}(\gamma) \times \Xi \times \mathbb{R} \mid \|\boldsymbol{\epsilon}_s - \hat{\mathbf{v}}_s + \ell(\boldsymbol{\Lambda}_s, \hat{\mathbf{u}}_s, \hat{\mathbf{p}}_s)\| \leq \phi_s\}.$$

Note that the G-CRQ model involves only robust linear constraints with convex uncertainty sets $\mathcal{E}_s(\gamma)$. Moreover, because each component of $\mathbf{h}$ is concave in $(\mathbf{p}, \mathbf{q})$ for any $(\boldsymbol{\Lambda}_s, \boldsymbol{\epsilon}_s, \phi_s) \in \mathcal{E}_s(\gamma)$, the G-CRQ model is a convex optimization problem that, if conic representable using second-order cones, power cones, positive-semidefinite cones, and exponential cones, can be efficiently solved by off-the-shelf solvers such as COPT (Ge et al., 2023) and MOSEK (ApS, 2019).

Appendix E Contextual Robust Satisficing Models

This appendix revisits the multiproduct inventory and pricing problem through the lens of contextual robust satisficing. In the first part, we directly adopt the modeling framework and approximation strategy in Sim et al. (2025), wherein uncertainty is modeled on the random demand $\tilde{\mathbf{v}}$ and recourse decisions are defined as affine functions of the demand realization. This would result in a non-convex problem with bilinear terms that cannot be efficiently solved. In contrast, the second part adopts our modeling framework, which captures ambiguity in the random residual $\tilde{\epsilon}$, and applies our approximation strategy based on the enhanced affine recourse adaptation. This then yields a tractable approximation that can be efficiently solved as a semidefinite program.

Appendix E.1 The Contextual Robust Satisficing Model of Sim et al. (2025)

We first present a residual-based robust satisficing model as follows:

$$\kappa_\tau^* = \left[\begin{array}{l} \min \quad \kappa \\ \text{s.t.} \quad \tau - \mathbb{E}_{\mathbb{P}} [\pi(\tilde{\mathbf{v}}, \mathbf{u}, \mathbf{p}, \mathbf{q})] \leq \kappa \cdot d(\mathbb{P}, \mathbb{P}_{\hat{\boldsymbol{\Lambda}}}) \quad \forall \mathbb{P} \in \mathcal{M}(\mathcal{Z}) \\ \quad (\mathbf{p}, \mathbf{q}) \in \mathcal{P} \times \mathcal{Q}, \kappa \in \mathbb{R}_+ \end{array} \right]. \quad (29)$$

Here, τ is a specified target profit, $\pi(\mathbf{v}, \mathbf{u}, \mathbf{p}, \mathbf{q}) = \mathbf{p}^\top \mathbf{v} - \mathbf{g}^\top (\mathbf{v} - \mathbf{q})^+ - \mathbf{c}^\top \mathbf{q}$ is the profit function, $\mathcal{Z}$ is the support for demand $\tilde{\mathbf{v}}$, and $\mathbb{P}_\Lambda$ is the empirical distribution over S predicted demand according to historical samples, *i.e.*, $\hat{\mathbf{v}}(\mathbf{A}, \mathbf{B}, \mathbf{u}, \mathbf{p}, s) = \hat{\mathbf{v}}_s + \mathbf{A}(\mathbf{u} - \hat{\mathbf{u}}_s) + \mathbf{B}(\mathbf{p} - \hat{\mathbf{p}}_s)$ for each $s \in [S]$.

Sim et al. (2025) propose the following contextual robust satisficing model,

$$\begin{aligned} \min \quad & \theta \\ \text{s.t.} \quad & \underline{\tau} - \mathbb{E}_{\mathbb{P}}[\pi(\tilde{\mathbf{v}}, \mathbf{u}, \mathbf{p}, \mathbf{q})] \leq \kappa_\tau^* \cdot (d(\mathbb{P}, \mathbb{P}_\Lambda) + \theta \Upsilon(\mathbf{A}, \mathbf{B})) \quad \forall \mathbb{P} \in \mathcal{M}(\mathcal{Z}), \mathbf{B} \leq \mathbf{0} \\ & (\mathbf{p}, \mathbf{q}) \in \mathcal{P} \times \mathcal{Q}, \theta \in \mathbb{R}_+, \end{aligned} \quad (30)$$

to account for parameter uncertainty. Here, $\underline{\tau}$ is another target specified by the retailer, and $\Upsilon(\mathbf{A}, \mathbf{B})$ is the estimation distance metric, defined as

$$\Upsilon(\mathbf{A}, \mathbf{B}) = \frac{1}{S} \sum_{s \in [S]} \|(\mathbf{A} - \hat{\mathbf{A}})\hat{\mathbf{u}}_s + (\mathbf{B} - \hat{\mathbf{B}})\hat{\mathbf{p}}_s\|_2.$$

Problem (30) can be equivalently reformulated as the following robust optimization problem,

$$\begin{aligned} \min \quad & \theta \\ \text{s.t.} \quad & \underline{\tau} \leq \frac{1}{S} \sum_{s \in [S]} (\pi(\mathbf{v}_s, \mathbf{u}, \mathbf{p}, \mathbf{q}) + \kappa_\tau^* \|\mathbf{v}_s - \hat{\mathbf{v}}(\mathbf{A}, \mathbf{B}, \mathbf{u}, \mathbf{p}, s)\|) + \theta \kappa_\tau^* \Upsilon(\mathbf{A}, \mathbf{B}) \\ & \forall \mathbf{B} \leq \mathbf{0}, \mathbf{v}_s \in \mathcal{Z}, s \in [S] \\ & (\mathbf{p}, \mathbf{q}) \in \mathcal{P} \times \mathcal{Q}, \theta \in \mathbb{R}_+. \end{aligned} \quad (31)$$

Similar to the CRO model, problem (31) is also a two-stage optimization problem that needs to be approximated. Sim et al. (2025) consider the conventional affine recourse adaptation that approximates the recourse decision $\mathbf{y}_s \in \mathbb{R}^N$ as an affine function of the demand $\mathbf{v}_s$ and the auxiliary variable ϕ_s . That is, $\mathbf{y}_s(\cdot) \in \underline{\mathcal{L}}$ with

$$\underline{\mathcal{L}} = \left\{ \mathbf{y} : \mathbb{R}^N \times \mathbb{R} \rightarrow \mathbb{R}^N \mid \mathbf{y}(\mathbf{v}, \phi) = \mathbf{y}^0 + \mathbf{Y}^v \mathbf{v} + \mathbf{y}^1 \phi \text{ for some } \mathbf{y}^0, \mathbf{y}^1 \in \mathbb{R}^N \text{ and } \mathbf{Y}^v \in \mathbb{R}^{N \times N} \right\}.$$

Subsequently, Sim et al. (2025) propose a conservatively approximated formulation as follows:

$$\begin{aligned} \min \quad & \theta \\ \text{s.t.} \quad & \underline{\tau} + \mathbf{c}^\top \mathbf{q} \leq \frac{1}{S} \sum_{s \in [S]} (\mathbf{p}^\top \mathbf{v}_s - \mathbf{g}^\top \mathbf{y}_s(\mathbf{v}_s, \phi_s) + \kappa_\tau^* \eta) \quad \forall (\eta, (\mathbf{v}_s, \phi_s)_{s \in [S]}) \in \bar{\mathcal{Z}}(\theta, \mathbf{p}) \\ & \mathbf{y}_s(\mathbf{v}_s, \phi_s) \geq \mathbf{v}_s - \mathbf{q} \quad \forall (\mathbf{v}_s, \phi_s) \in \bar{\mathcal{Z}}_s(\mathbf{p}), s \in [S] \\ & \mathbf{y}_s(\mathbf{v}_s, \phi_s) \geq \mathbf{0} \quad \forall (\mathbf{v}_s, \phi_s) \in \bar{\mathcal{Z}}_s(\mathbf{p}), s \in [S] \\ & \mathbf{y}_s \in \underline{\mathcal{L}} \quad \forall s \in [S] \\ & (\mathbf{p}, \mathbf{q}) \in \mathcal{P} \times \mathcal{Q}, \theta \in \mathbb{R}_+, \end{aligned} \quad (32)$$

where

$$\bar{\mathcal{Z}}(\theta, \mathbf{p}) = \left\{ (\eta, (\mathbf{v}_s, \phi_s)_{s \in [S]}) \left| \begin{array}{l} \eta \in \mathbb{R}, (\mathbf{v}_s, \phi_s) \in \bar{\mathcal{Z}}_s(\mathbf{p}) \quad \forall s \in [S] \\ \eta \geq \frac{1}{S} \sum_{s \in [S]} (\|\mathbf{v}_s - \hat{\mathbf{v}}_s - \mathbf{A}(\mathbf{u} - \hat{\mathbf{u}}_s) - \mathbf{B}(\mathbf{p} - \hat{\mathbf{p}}_s)\|) + \theta \Upsilon(\mathbf{A}, \mathbf{B}) \\ \eta \geq \frac{\theta}{\vartheta' S} \sum_{s \in [S]} \phi_s \\ \text{for some } \mathbf{A} \in \mathbb{R}^{N \times J} \text{ and } -\mathbf{B} \geq \mathbf{0} \end{array} \right. \right\}$$

for some fixed ϑ' such that $\vartheta' \geq \bar{\theta}$ and

$$\bar{\mathcal{Z}}_s(\mathbf{p}) = \{(\mathbf{v}_s, \phi_s) \in \mathcal{Z} \times \mathbb{R} \mid \phi_s \geq \|\mathbf{v}_s - \hat{\mathbf{v}}_s - \hat{\mathbf{A}}(\mathbf{u} - \hat{\mathbf{u}}_s) - \hat{\mathbf{B}}(\mathbf{p} - \hat{\mathbf{p}}_s)\|\}.$$

Here, $\bar{\theta}$ is some constant that satisfies $|d(\mathbb{P}, \mathbb{P}_{\hat{\Lambda}}) - d(\mathbb{P}, \mathbb{P}_{\Lambda})| \leq \bar{\theta} \cdot \Upsilon(\mathbf{A}, \mathbf{B})$.

To reformulate the approximation (32), we first consider the following problem arising from the the first collection of constraints in (32):

$$\begin{aligned} \min \quad & \frac{1}{S} \sum_{s \in [S]} \left(\mathbf{p}^\top \mathbf{v}_s - \mathbf{g}^\top (\mathbf{y}_s^0 + \mathbf{Y}_s^v \mathbf{v}_s + \mathbf{y}_s^1 \phi_s) \right) + \kappa_\tau^\star \eta \\ \text{s.t.} \quad & \eta \geq \frac{1}{S} \sum_{s \in [S]} \left(\|\mathbf{v}_s - \mathbf{A}(\mathbf{u} - \hat{\mathbf{u}}_s) - \mathbf{B}(\mathbf{p} - \hat{\mathbf{p}}_s) - \hat{\mathbf{v}}_s\| + \theta \|(\mathbf{A} - \hat{\mathbf{A}})\hat{\mathbf{u}}_s + (\mathbf{B} - \hat{\mathbf{B}})\hat{\mathbf{p}}_s\|_2 \right) \\ & \phi_s \geq \|\mathbf{v}_s - \hat{\mathbf{A}}(\mathbf{u} - \hat{\mathbf{u}}_s) - \hat{\mathbf{B}}(\mathbf{p} - \hat{\mathbf{p}}_s) - \hat{\mathbf{v}}_s\| \quad \forall s \in [S] \\ & \eta - \frac{\theta}{\vartheta' S} \sum_{s \in [S]} \phi_s \geq 0, \quad \mathbf{B} \leq \mathbf{0} \\ & \phi_s \in \mathbb{R}_+, \quad \mathbf{v}_s \in \mathbb{R}^N \quad \forall s \in [S] \\ & \eta \in \mathbb{R}_+, \quad \mathbf{A} \in \mathbb{R}^{N \times J}, \end{aligned}$$

whose dual can be derived as

$$\begin{aligned} \max \quad & \frac{1}{S} \sum_{s \in [S]} \left(-\mathbf{g}^\top \mathbf{y}_s^0 + \boldsymbol{\alpha}_s^\top \hat{\mathbf{v}}_s + \boldsymbol{\beta}_s^\top (\hat{\mathbf{A}}\hat{\mathbf{u}}_s + \hat{\mathbf{B}}\hat{\mathbf{p}}_s) + \mathbf{t}_s^\top (\hat{\mathbf{v}}_s + \hat{\mathbf{A}}(\mathbf{u} - \hat{\mathbf{u}}_s) + \hat{\mathbf{B}}(\mathbf{p} - \hat{\mathbf{p}}_s)) \right) \\ \text{s.t.} \quad & \frac{1}{S} \sum_{s \in [S]} (\boldsymbol{\alpha}_s(\mathbf{u} - \hat{\mathbf{u}}_s)^\top - \boldsymbol{\beta}_s \hat{\mathbf{u}}_s^\top) = \mathbf{0} \\ & \frac{1}{S} \sum_{s \in [S]} (\boldsymbol{\alpha}_s(\mathbf{p} - \hat{\mathbf{p}}_s)^\top - \boldsymbol{\beta}_s \hat{\mathbf{p}}_s^\top) \geq \mathbf{0} \\ & \boldsymbol{\alpha}_s + \mathbf{t}_s = \mathbf{p} - \mathbf{Y}_s^{v\top} \mathbf{g} \quad \forall s \in [S] \\ & \|\boldsymbol{\alpha}_s\|_* \leq \gamma, \quad \|\mathbf{t}_s\|_* \leq \mu_s \quad \forall s \in [S] \\ & \|\boldsymbol{\beta}_s\|_2 \leq \theta \gamma \quad \forall s \in [S] \\ & -\delta \frac{\theta}{\vartheta'} + \mu_s \leq -\mathbf{g}^\top \mathbf{y}_s^1 \quad \forall s \in [S] \\ & \gamma + \delta \leq \kappa_\tau^\star, \quad \gamma, \delta \in \mathbb{R}_+ \\ & \mu_s \in \mathbb{R}_+, \quad \boldsymbol{\alpha}_s, \boldsymbol{\beta}_s, \mathbf{t}_s \in \mathbb{R}^N \quad \forall s \in [S]. \end{aligned}$$

Then for each product $n \in [N]$ and any $s \in [S]$, note that the constraint

$$\mathbf{e}_n^\top (\mathbf{y}_s^0 + \mathbf{Y}_s^v \mathbf{v}_s + \mathbf{y}_s^1 \phi_s) \geq \mathbf{e}_n^\top (\mathbf{v}_s - \mathbf{q}) \quad \forall (\mathbf{v}_s, \phi_s) \in \bar{\mathcal{Z}}_s(\mathbf{p})$$

arising from the second collection can be equivalently reformulated as

$$\min_{(\mathbf{v}_s, \phi_s) \in \bar{\mathcal{Z}}_s(\mathbf{p})} \mathbf{e}_n^\top (\mathbf{Y}_s^v \mathbf{v}_s + \mathbf{y}_s^1 \phi_s - \mathbf{v}_s) \geq -\mathbf{e}_n^\top (\mathbf{y}_s^0 + \mathbf{q}).$$

For the left-hand-side problem,

$$\begin{aligned} \min \quad & \mathbf{e}_n^\top (\mathbf{Y}_s^v \mathbf{v}_s + \mathbf{y}_s^1 \phi_s - \mathbf{v}_s) \\ \text{s.t.} \quad & \phi_s \geq \|\mathbf{d}_s\| \\ & \mathbf{d}_s = \mathbf{v}_s - \hat{\mathbf{A}}(\mathbf{u} - \hat{\mathbf{u}}_s) - \hat{\mathbf{B}}(\mathbf{p} - \hat{\mathbf{p}}_s) - \hat{\mathbf{v}}_s \\ & \mathbf{v}_s, \mathbf{d}_s \in \mathbb{R}^N, \phi_s \in \mathbb{R}_+, \end{aligned}$$

we can also obtain its dual problem as

$$\begin{aligned} \max \quad & \mathbf{r}_{n,s}^{1\top} (\hat{\mathbf{A}}(\mathbf{u} - \hat{\mathbf{u}}_s) + \hat{\mathbf{B}}(\mathbf{p} - \hat{\mathbf{p}}_s) + \hat{\mathbf{v}}_s) \\ \text{s.t.} \quad & \|\mathbf{r}_{n,s}^1\|_* \leq \mathbf{e}_n^\top \mathbf{y}_s^1 \\ & \mathbf{r}_{n,s}^1 = \mathbf{Y}_s^{v\top} \mathbf{e}_n - \mathbf{e}_n \\ & \mathbf{r}_{n,s}^1 \in \mathbb{R}^N. \end{aligned}$$

In other words, the constraint

$$\mathbf{e}_n^\top (\mathbf{y}_s^0 + \mathbf{Y}_s^v \mathbf{v}_s + \mathbf{y}_s^1 \phi_s) \geq \mathbf{e}_n^\top (\mathbf{v}_s - \mathbf{q}) \quad \forall (\mathbf{v}_s, \phi_s) \in \bar{\mathcal{Z}}_s(\mathbf{p})$$

can be equivalently reformulated as

$$\begin{aligned} \mathbf{r}_{n,s}^{1\top} (\hat{\mathbf{A}}(\mathbf{u} - \hat{\mathbf{u}}_s) + \hat{\mathbf{B}}(\mathbf{p} - \hat{\mathbf{p}}_s) + \hat{\mathbf{v}}_s) &\geq -\mathbf{e}_n^\top (\mathbf{y}_s^0 + \mathbf{q}) \\ \|\mathbf{r}_{n,s}^1\|_* &\leq \mathbf{e}_n^\top \mathbf{y}_s^1 \\ \mathbf{r}_{n,s}^1 &= \mathbf{Y}_s^{v\top} \mathbf{e}_n - \mathbf{e}_n \\ \mathbf{r}_{n,s}^1 &\in \mathbb{R}^N. \end{aligned}$$

We can also reformulate the constraint

$$\mathbf{e}_n^\top (\mathbf{y}_s^0 + \mathbf{Y}_s^v \mathbf{v}_s + \mathbf{y}_s^1 \phi_s) \geq 0 \quad \forall (\mathbf{v}_s, \phi_s) \in \bar{\mathcal{Z}}_s(\mathbf{p})$$

arising from the third collection, in a similar manner.

Now we can conclude that the approximation (32) can be equivalently reformulated as the following deterministic optimization problem:

$$\begin{aligned}
& \min \quad \theta \\
& \text{s.t.} \quad \frac{1}{S} \sum_{s \in [S]} \left(-\mathbf{g}^\top \mathbf{y}_s^0 + \boldsymbol{\alpha}_s^\top \hat{\mathbf{v}}_s + \boldsymbol{\beta}_s^\top (\hat{\mathbf{A}} \hat{\mathbf{u}}_s + \hat{\mathbf{B}} \hat{\mathbf{p}}_s) + \mathbf{t}_s^\top (\hat{\mathbf{v}}_s + \hat{\mathbf{A}}(\mathbf{u} - \hat{\mathbf{u}}_s) + \hat{\mathbf{B}}(\mathbf{p} - \hat{\mathbf{p}}_s)) \right) \geq \underline{\tau} + \mathbf{c}^\top \mathbf{q} \\
& \quad \frac{1}{S} \sum_{s \in [S]} (\boldsymbol{\alpha}_s(\mathbf{u} - \hat{\mathbf{u}}_s)^\top - \boldsymbol{\beta}_s \hat{\mathbf{u}}_s^\top) = \mathbf{0} \\
& \quad \frac{1}{S} \sum_{s \in [S]} (\boldsymbol{\alpha}_s(\mathbf{p} - \hat{\mathbf{p}}_s)^\top - \boldsymbol{\beta}_s \hat{\mathbf{p}}_s^\top) \geq \mathbf{0} \\
& \quad \boldsymbol{\alpha}_s + \mathbf{t}_s = \mathbf{p} - \mathbf{Y}_s^{v\top} \mathbf{g} \quad \forall s \in [S] \\
& \quad \|\boldsymbol{\alpha}_s\|_* \leq \gamma, \quad \forall s \in [S] \\
& \quad \|\boldsymbol{\beta}_s\|_2 \leq \gamma \theta \quad \forall s \in [S] \\
& \quad \|\mathbf{t}_s\|_* \leq \delta \frac{\theta}{\vartheta'} - \mathbf{g}^\top \mathbf{y}_s^1 \quad \forall s \in [S] \\
& \quad \gamma + \delta \leq \kappa_\tau^* \\
& \quad \mathbf{e}_n^\top \mathbf{y}_s^0 + \mathbf{r}_{n,s}^{1\top} (\hat{\mathbf{A}}(\mathbf{u} - \hat{\mathbf{u}}_s) + \hat{\mathbf{B}}(\mathbf{p} - \hat{\mathbf{p}}_s) + \hat{\mathbf{v}}_s) \geq -\mathbf{e}_n^\top \mathbf{q} \quad \forall n \in [N], s \in [S] \\
& \quad \|\mathbf{r}_{n,s}^1\|_* \leq \mathbf{e}_n^\top \mathbf{y}_s^1 \quad \forall n \in [N], s \in [S] \\
& \quad \mathbf{r}_{n,s}^1 = \mathbf{Y}_s^{v\top} \mathbf{e}_n - \mathbf{e}_n \quad \forall n \in [N], s \in [S] \\
& \quad \mathbf{e}_n^\top \mathbf{y}_s^0 + \mathbf{r}_{n,s}^{2\top} (\hat{\mathbf{A}}(\mathbf{u} - \hat{\mathbf{u}}_s) + \hat{\mathbf{B}}(\mathbf{p} - \hat{\mathbf{p}}_s) + \hat{\mathbf{v}}_s) \geq 0, \quad \forall n \in [N], s \in [S] \\
& \quad \|\mathbf{r}_{n,s}^2\|_* \leq \mathbf{e}_n^\top \mathbf{y}_s^1 \quad \forall n \in [N], s \in [S] \\
& \quad \mathbf{r}_{n,s}^2 = \mathbf{Y}_s^{v\top} \mathbf{e}_n \quad \forall n \in [N], s \in [S] \\
& \quad \mathbf{r}_{n,s}^1, \mathbf{r}_{n,s}^2 \in \mathbb{R}^N \quad \forall n \in [N], s \in [S] \\
& \quad \boldsymbol{\alpha}_s, \boldsymbol{\beta}_s, \mathbf{t}_s, \mathbf{y}_s^0, \mathbf{y}_s^1 \in \mathbb{R}^N, \mathbf{Y}_s^v \in \mathbb{R}^{N \times N} \quad \forall s \in [S] \\
& \quad (\mathbf{p}, \mathbf{q}) \in \mathcal{P} \times \mathcal{Q}, \theta, \gamma, \delta \in \mathbb{R}_+
\end{aligned} \tag{33}$$

It is evident that even reformulation (33)—the approximation of problem (30) via affine recourse adaptation—involves some bilinear terms such as $\mathbf{t}_s^\top \hat{\mathbf{B}}(\mathbf{p} - \hat{\mathbf{p}}_s)$ and $\boldsymbol{\alpha}_s(\mathbf{p} - \hat{\mathbf{p}}_s)^\top$ in both the objective and constraints, making it a non-convex optimization problem that is challenging to solve to optimality. We also remark that even when adopting the residual $\tilde{\epsilon}$ as the source of the uncertainty, similar bilinear terms persist due to the corresponding decision-dependent uncertainty set:

$$\bar{\mathcal{Z}}_s(\mathbf{p}) = \{(\boldsymbol{\epsilon}_s, \phi_s) \in \mathbb{R}^N \times \mathbb{R} \mid \phi_s \geq \|\boldsymbol{\epsilon}_s - \hat{\mathbf{v}}_s + \hat{\mathbf{A}}\mathbf{u} + \hat{\mathbf{B}}\mathbf{p}\|\}.$$

Appendix E.2 Our Contextual Robust Satisficing Model

Modeling the random residual $\tilde{\epsilon}$ as the underlying uncertainty, the corresponding residual-based robust satisficing model becomes

$$\kappa_\tau^\star = \left[\begin{array}{l} \min \quad \kappa \\ \text{s.t.} \quad \tau - \mathbb{E}_{\mathbb{P}}[\pi(\tilde{\epsilon}, \hat{\mathbf{\Lambda}}, \mathbf{u}, \mathbf{p}, \mathbf{q})] \leq \kappa \cdot d(\mathbb{P}, \mathbb{P}_{\hat{\mathbf{\Lambda}}}) \quad \forall \mathbb{P} \in \mathcal{M}(\mathbb{R}^N) \\ (\mathbf{p}, \mathbf{q}) \in \mathcal{P} \times \mathcal{Q}, \kappa \in \mathbb{R}_+ \end{array} \right],$$

where the profit function is $\pi(\epsilon, \mathbf{\Lambda}, \mathbf{u}, \mathbf{p}, \mathbf{q}) = \mathbf{p}^\top(\ell(\mathbf{\Lambda}, \mathbf{u}, \mathbf{p}) + \epsilon) - \mathbf{g}^\top(\ell(\mathbf{\Lambda}, \mathbf{u}, \mathbf{p}) + \epsilon - \mathbf{q})^+ - \mathbf{c}^\top \mathbf{q}$.

The above model can be equivalently reformulated as the following robust optimization problem:

$$\kappa_\tau^\star = \left[\begin{array}{l} \min \quad \kappa \\ \text{s.t.} \quad \tau - S^{-1} \cdot \sum_{s \in [S]} [\pi(\epsilon_s, \hat{\mathbf{\Lambda}}, \mathbf{u}, \mathbf{p}, \mathbf{q}) + \kappa \eta_s] \leq 0 \quad \forall (\epsilon_s, \eta_s) \in \bar{\mathcal{E}}_s, s \in [S] \\ (\mathbf{p}, \mathbf{q}) \in \mathcal{P} \times \mathcal{Q}, \kappa \in \mathbb{R}_+ \end{array} \right],$$

where $\bar{\mathcal{E}}_s = \{(\epsilon_s, \eta_s) \in \mathbb{R}^N \times \mathbb{R}_+ : \eta_s \geq \|\epsilon_s - \hat{\epsilon}_{\hat{\mathbf{\Lambda}}, s}\|\}$.

Corresponding to problem (30), we propose a contextual robust satisficing model considering parameter uncertainty as follows:

$$\begin{array}{ll} \min & \theta \\ \text{s.t.} & \underline{\tau} - \mathbb{E}_{\mathbb{P}}[\pi(\mathbf{p}, \mathbf{q}, \tilde{\epsilon}, \mathbf{u}, \mathbf{\Lambda})] \leq \kappa_\tau^\star \cdot (d(\mathbb{P}, \mathbb{P}_{\mathbf{\Lambda}}) + \theta \|\mathbf{\Lambda} - \hat{\mathbf{\Lambda}}\|_F) \quad \forall \mathbb{P} \in \mathcal{M}(\mathbb{R}^N), \mathbf{\Lambda} \in \Omega \\ & (\mathbf{p}, \mathbf{q}) \in \mathcal{P} \times \mathcal{Q}, \theta \in \mathbb{R}_+, \end{array} \quad (\text{CRS})$$

which can be equivalently reformulated as the following robust optimization problem:

$$\begin{array}{ll} \min & \theta \\ \text{s.t.} & S^{-1} \cdot \sum_{s \in [S]} (\mathbf{g}^\top(\ell(\mathbf{\Lambda}, \mathbf{u}, \mathbf{p}) + \epsilon_s - \mathbf{q})^+ - \mathbf{p}^\top(\ell(\mathbf{\Lambda}, \mathbf{u}, \mathbf{p}) + \epsilon_s) - \kappa_\tau^\star \cdot \|\epsilon_s - \hat{\epsilon}_{\mathbf{\Lambda}, s}\|) \\ & - \kappa_\tau^\star \|\mathbf{\Lambda} - \hat{\mathbf{\Lambda}}\|_F \cdot \theta \leq -\mathbf{c}^\top \mathbf{q} - \underline{\tau} \quad \forall \epsilon_s \in \mathbb{R}^N, \mathbf{\Lambda} \in \Omega, s \in [S] \\ & (\mathbf{p}, \mathbf{q}) \in \mathcal{P} \times \mathcal{Q}, \theta \in \mathbb{R}_+, \end{array} \quad (34)$$

Using similar techniques as in approaching **CR0** via **CR0** (with our proposed approximation

strategy based on the enhanced affine recourse adaptation), problem (34) can be approximated by

$$\begin{aligned}
\min \quad & \theta \\
\text{s.t.} \quad & S^{-1} \cdot \sum_{s \in [S]} (\mathbf{p}^\top (\ell(\mathbf{\Lambda}, \mathbf{u}, \mathbf{p}) + \boldsymbol{\epsilon}_s) - \mathbf{g}^\top \mathbf{y}_s(\mathbf{\Lambda}_s, \boldsymbol{\epsilon}_s, \phi_s, \psi_s) \\
& + \kappa_\tau^\star \cdot \phi_s + \mathbf{D}_s \bullet \mathbf{\Lambda}_s + \kappa_\tau^\star \psi_s \cdot \theta) \geq \mathbf{c}^\top \mathbf{q} + \underline{\tau}, \quad \forall (\mathbf{\Lambda}_s, \boldsymbol{\epsilon}_s, \phi_s, \psi_s) \in \mathcal{E}_s, s \in [S] \\
& \mathbf{y}_s(\mathbf{\Lambda}_s, \boldsymbol{\epsilon}_s, \phi_s, \psi_s) \geq \mathbf{A}_s \mathbf{u} + \mathbf{B}_s \mathbf{p} + \boldsymbol{\epsilon}_s - \mathbf{q}, \quad \forall (\mathbf{\Lambda}_s, \boldsymbol{\epsilon}_s, \phi_s, \psi_s) \in \mathcal{E}_s, s \in [S] \\
& \mathbf{y}_s(\mathbf{\Lambda}_s, \boldsymbol{\epsilon}_s, \phi_s, \psi_s) \geq \mathbf{0}, \quad \forall (\mathbf{\Lambda}_s, \boldsymbol{\epsilon}_s, \phi_s, \psi_s) \in \mathcal{E}_s, s \in [S] \\
& \sum_{s \in [S]} \mathbf{D}_s = \mathbf{0} \\
& (\mathbf{p}, \mathbf{q}) \in \mathcal{P} \times \mathcal{Q}, \theta \in \mathbb{R}_+,
\end{aligned} \tag{35}$$

where

$$\mathcal{E}_s = \{(\mathbf{\Lambda}_s, \boldsymbol{\epsilon}_s, \eta_s, \psi_s) \in \Omega \times \mathbb{R}^N \times \mathbb{R}_+ \times \mathbb{R}_+ : \eta_s \geq \|\boldsymbol{\epsilon}_s - \hat{\boldsymbol{\epsilon}}_{\mathbf{\Lambda}_s, s}\|, \psi_s \geq \|\mathbf{\Lambda}_s - \hat{\mathbf{\Lambda}}\|_F\}.$$

Approximation (35) can be further reformulated as a semidefinite program (see, *e.g.*, Theorem 3).

Appendix F The Lost-Sales Model

When the emergency cost $\mathbf{g}$ is replaced by the price $\mathbf{p}$, our multiproduct inventory and pricing problem becomes the lost-sales model, which is equivalent to the multiproduct newsvendor problem with pricing decisions. In this case, the profit function is expressed as:

$$\pi_{\text{LS}}(\boldsymbol{\epsilon}_s, \mathbf{\Lambda}, \mathbf{u}, \mathbf{p}, \mathbf{q}) = \mathbf{p}^\top (\ell(\mathbf{\Lambda}, \mathbf{u}, \mathbf{p}) + \boldsymbol{\epsilon}_s) - \mathbf{p}^\top (\ell(\mathbf{\Lambda}, \mathbf{u}, \mathbf{p}) + \boldsymbol{\epsilon}_s - \mathbf{q})^+ - \mathbf{c}^\top \mathbf{q}.$$

Using the same approximation and reformulation techniques as for the **CRO** model, we can approach the contextual robust lost-sales model

$$\max_{(\mathbf{p}, \mathbf{q}) \in \mathcal{P} \times \mathcal{Q}} \min_{\mathbf{\Lambda} \in \mathcal{W}(\gamma), \mathbb{P} \in \mathcal{F}(\mathbf{\Lambda}, \theta)} \mathbb{E}_{\mathbb{P}} [\pi_{\text{LS}}(\tilde{\boldsymbol{\epsilon}}, \mathbf{\Lambda}, \mathbf{u}, \mathbf{p}, \mathbf{q})]$$

via the following lower bound:

$$\begin{aligned}
& \max \quad S^{-1} \cdot \sum_{s \in [S]} \left(\hat{\mathbf{v}}_s^\top \boldsymbol{\alpha}_s - \mathbf{p}^\top \mathbf{y}_s^0 + (\hat{\mathbf{A}} \hat{\mathbf{U}}^A + \hat{\mathbf{B}} \hat{\mathbf{U}}^B) \bullet \mathbf{R}_s - \gamma \sqrt{S} \eta_s \right) - \mathbf{c}^\top \mathbf{q} + \theta \kappa \\
& \text{s.t.} \quad \boldsymbol{\alpha}_s = \mathbf{p} - \mathbf{Y}_s^\epsilon \mathbf{p} \quad \forall s \in [S] \\
& \quad \mathbf{p}(\mathbf{y}_s^A)^\top - \mathbf{p} \mathbf{u}^\top + \boldsymbol{\alpha}_s \hat{\mathbf{u}}_s^\top + \mathbf{R}_s (\hat{\mathbf{U}}^A)^\top - \mathbf{D}_s^A = \mathbf{0} \quad \forall s \in [S] \\
& \quad \begin{pmatrix} \mathbf{p}(\mathbf{y}_s^B)^\top + \boldsymbol{\alpha}_s \hat{\mathbf{p}}_s^\top + \mathbf{R}_s (\hat{\mathbf{U}}^B)^\top - \mathbf{D}_s^B & \mathbf{p} \\ \mathbf{p}^\top & 1 \end{pmatrix} \succeq \mathbf{0} \quad \forall s \in [S] \\
& \quad \|\boldsymbol{\alpha}_s\|_* \leq -\kappa - \mathbf{p}^\top \mathbf{y}_s^1, \quad \|\mathbf{R}_s\|_F \leq \eta_s \quad \forall s \in [S] \\
& \quad y_{s,n}^0 + (\boldsymbol{\alpha}_{s,n}^1)^\top \hat{\mathbf{v}}_s - \eta_{s,n}^1 \gamma \sqrt{S} + (\hat{\mathbf{A}} \hat{\mathbf{U}}^A + \hat{\mathbf{B}} \hat{\mathbf{U}}^B) \bullet \mathbf{R}_{s,n}^1 + q_n \geq 0 \quad \forall n \in [N], s \in [S] \\
& \quad \boldsymbol{\alpha}_{s,n}^1 = \mathbf{Y}_s^\epsilon e_n - e_n, \quad \boldsymbol{\alpha}_{s,n}^2 = \mathbf{Y}_s^\epsilon e_n \quad \forall n \in [N], s \in [S] \\
& \quad \boldsymbol{\alpha}_{s,n}^1 \hat{\mathbf{u}}_s^\top + \mathbf{R}_{s,n}^1 (\hat{\mathbf{U}}^A)^\top = e_n (\mathbf{y}_s^A - \mathbf{u})^\top \quad \forall n \in [N], s \in [S] \\
& \quad \boldsymbol{\alpha}_{s,n}^1 \hat{\mathbf{p}}_s^\top + \mathbf{R}_{s,n}^1 (\hat{\mathbf{U}}^B)^\top - \mathbf{Q}_{s,n}^1 = e_n (\mathbf{y}_s^B - \mathbf{p})^\top \quad \forall n \in [N], s \in [S] \\
& \quad \|\boldsymbol{\alpha}_{s,n}^1\|_* \leq y_{s,n}^1, \quad \|\mathbf{R}_{s,n}^1\|_F \leq \eta_{s,n}^1 \quad \forall n \in [N], s \in [S] \quad (36) \\
& \quad y_{s,n}^0 + (\boldsymbol{\alpha}_{s,n}^2)^\top \hat{\mathbf{v}}_s - \eta_{s,n}^2 \gamma \sqrt{S} + \mathbf{R}_{s,n}^2 \bullet (\hat{\mathbf{A}} \hat{\mathbf{U}}^A + \hat{\mathbf{B}} \hat{\mathbf{U}}^B) \geq 0 \quad \forall n \in [N], s \in [S] \\
& \quad \forall n \in [N], s \in [S] \\
& \quad \boldsymbol{\alpha}_{s,n}^2 \hat{\mathbf{u}}_s^\top + \mathbf{R}_{s,n}^2 (\hat{\mathbf{U}}^A)^\top = e_n (\mathbf{y}_s^A)^\top \quad \forall n \in [N], s \in [S] \\
& \quad \boldsymbol{\alpha}_{s,n}^2 \hat{\mathbf{p}}_s^\top + \mathbf{R}_{s,n}^2 (\hat{\mathbf{U}}^B)^\top - \mathbf{Q}_{s,n}^2 = e_n (\mathbf{y}_s^B)^\top \quad \forall n \in [N], s \in [S] \\
& \quad \|\boldsymbol{\alpha}_{s,n}^2\|_* \leq y_{s,n}^1, \quad \|\mathbf{R}_{s,n}^2\|_F \leq \eta_{s,n}^2 \quad \forall n \in [N], s \in [S] \\
& \quad \sum_{s \in [S]} \mathbf{D}_s = \mathbf{0}, \quad (\mathbf{p}, \mathbf{q}) \in \mathcal{P} \times \mathcal{Q}, \quad \kappa \leq 0 \\
& \quad \mathbf{y}_s^0 \in \mathbb{R}^N, \quad \mathbf{y}_s^A \in \mathbb{R}^J, \quad \mathbf{y}_s^B \in \mathbb{R}^N, \quad \mathbf{Y}_s^\epsilon \in \mathbb{R}^{N \times N}, \quad \mathbf{y}_s^1 \in \mathbb{R}^N \quad \forall s \in [S] \\
& \quad \mathbf{D}_s \in \mathbb{R}^{N \times (J+N)}, \quad \mathbf{R}_s \in \mathbb{R}^{N \times S}, \quad \boldsymbol{\alpha}_s \in \mathbb{R}^N, \quad \eta_s \in \mathbb{R}_+ \quad \forall s \in [S] \\
& \quad \mathbf{Q}_{s,n}^1 \in \mathbb{S}_+^N, \quad \mathbf{R}_{s,n}^1 \in \mathbb{R}^{N \times S}, \quad \boldsymbol{\alpha}_{s,n}^1 \in \mathbb{R}^N, \quad \eta_{s,n}^1 \in \mathbb{R}_+ \quad \forall n \in [N], s \in [S] \\
& \quad \mathbf{Q}_{s,n}^2 \in \mathbb{S}_+^N, \quad \mathbf{R}_{s,n}^2 \in \mathbb{R}^{N \times S}, \quad \boldsymbol{\alpha}_{s,n}^2 \in \mathbb{R}^N, \quad \eta_{s,n}^2 \in \mathbb{R}_+ \quad \forall n \in [N], s \in [S].
\end{aligned}$$

Here, the bilinear terms (*e.g.*, $\mathbf{p}^\top \mathbf{y}_s^0$, $\mathbf{Y}_s^\epsilon \mathbf{p}$, $\mathbf{p}(\mathbf{y}_s^A)^\top$, $\mathbf{p}(\mathbf{y}_s^B)^\top$, and $\mathbf{p}^\top \mathbf{y}_s^1$) render problem (36) a challenging-to-solve bilinear optimization problem. It's common to restrict prices to a discrete set (*i.e.*, a price ladder) so that the problem can be reformulated as a mixed-integer conic program with binary variables and big-M constraints. This formulation is solvable when the discrete set of possible prices is of moderate size or the number of products is relatively small. For our numerical experiment in the single-product setting, we discretize the price p into the set $[0, 0.02, 0.04, \dots, 1.5\bar{p}]$, and for each fixed candidate price, we solve problem (36). The price yielding the highest objective value is then selected as the optimal price.

References

- ApS, MOSEK. 2019. *MOSEK Optimizer API for Julia 10.1.21*. URL <https://docs.mosek.com/latest/juliaapi/index.html>.
- den Boer, Arnoud. 2014. Dynamic pricing with multiple products and partially specified demand distribution. *Mathematics of Operations Research* **39**(3) 863–888.
- Ge, Dongdong, Qi Huangfu, Zizhuo Wang, Jian Wu, Yinyu Ye. 2023. Cardinal Optimizer (COPT) user guide. URL <https://guide.coap.online/copt/en-doc>.
- Hsu, Daniel, Sham Kakade, Tong Zhang. 2012. A tail inequality for quadratic forms of subgaussian random vectors. *Electronic Communications in Probability* **17**(52) 1–6.
- Kallenberg, Olav. 1997. *Foundations of modern probability*. Springer.
- Kannan, Rohit, Güzin Bayraksan, James R Luedtke. 2024. Residuals-based distributionally robust optimization with covariate information. *Mathematical Programming* **207**(1-2) 369–425.
- Kuhn, Daniel, Peyman Mohajerin Esfahani, Viet Anh Nguyen, Soroosh Shafieezadeh-Abadeh. 2019. Wasserstein distributionally robust optimization: Theory and applications in machine learning. *Operations research & management science in the age of analytics*. INFORMS, 130–166.
- Qi, Meng, Ying Cao, Zuo-Jun Max Shen. 2022. Distributionally robust conditional quantile prediction with fixed design. *Management Science* **68**(3) 1639–1658.
- Sim, Melvyn, Qinshen Tang, Minglong Zhou, Taozeng Zhu. 2025. The analytics of robust satisficing: Predict, optimize, satisfice, then fortify. *Operations Research* **73**(5) 2708–2728.
- Wainwright, Martin. 2019. *High-dimensional statistics: A non-asymptotic viewpoint*. Cambridge University Press.