Submitted to
manuscript (Please, provide the manuscript number!)

A Lifting-and-Splitting Framework for Risk-Averse Distributionally Robust Multi-Item Newsvendor Problems

Yuzhou Liu
Lee Kong Chian School of Business, Singapore Management University, yuzhou.liu.2023@pbs.smu.edu.sg

Sarah Yini Gao
Lee Kong Chian School of Business, Singapore Management University, yngao@smu.edu.sg

Lianmin Zhang
School of Management and Economics, The Chinese University of Hong Kong, Shenzhen, zhanglianmin@cuhk.edu.cn

Zhichao Zheng
Lee Kong Chian School of Business, Singapore Management University, danielzheng@smu.edu.sg

Risk-averse distributionally robust multi-item newsvendor problems provide a fundamental model for inventory decisions under demand uncertainty, limited distributional information, and downside-risk concerns. We study this problem under mean–covariance demand ambiguity, where the decision maker maximizes the worst-case conditional value-at-risk of profit. While cross-item demand correlations are important for portfolio-level inventory decisions, they are difficult to estimate reliably in high dimensions. Moreover, preserving the full covariance information leads to an NP-hard formulation that is computationally challenging for large assortments. We develop a lifting-and-splitting framework that combines exact reformulation, scalable approximation, and performance guarantees. First, we derive an exact semidefinite programming reformulation based on the constrained Boolean quadric polytope, replacing exponentially many positive semidefinite constraints in the classical dual formulation with exponentially many linear constraints. Second, we propose a correlation-splitting approximation that partitions items into clusters, preserves within-cluster covariance information, and relaxes cross-cluster dependence. We establish exactness conditions and derive an explicit cluster-wise upper bound on the optimality gap. We further show that, in the two-cluster case, the approximation becomes increasingly accurate when one cluster has uniformly small demand variability, motivating a simple variance-based splitting rule. Numerical experiments demonstrate improved exact-solution tractability, strong approximation quality, and scalability to large instances.

Key words: distributionally robust optimization; multi-item newsvendor; conditional value-at-risk; semidefinite program; Boolean quadric polytope

1. Introduction

Multi-item inventory decisions arise in many modern retail and service settings. Online retailers and platform-based sellers, for example, often manage large product assortments across multiple sales channels. These firms must determine stocking quantities before demand is realized, while facing limited demand information, dependence across item demands, and exposure to adverse outcomes such as excess inventory, stockouts, markdowns, and working-capital pressure. These concerns are especially salient for small and medium-sized businesses, which may offer many products online but have limited capacity to absorb large losses. In such settings, decision makers are not only interested in maximizing expected profit; they also seek to control downside risk.

The multi-item newsvendor problem provides a canonical framework for studying such inventory decisions under stochastic demand. Overordering leads to excess inventory and potential losses, whereas underordering results in unmet demand and missed revenue opportunities. Balancing these two sources of risk is therefore central to inventory management. However, risk-neutral models focus on expected performance and may not adequately capture decision makers' concerns about adverse outcomes. This limitation has motivated a growing literature on risk-averse newsvendor problems, which explicitly incorporate downside-risk considerations into inventory decisions (Schweitzer and Cachon 2000, Choi and Ruszczyński 2011, Yang et al. 2021).

In addition to downside risk, incomplete distributional information poses a central challenge. When the demand distribution is difficult to estimate reliably, decisions optimized under a misspecified model may perform poorly under the true distribution. Distributionally robust optimization (DRO) addresses this by optimizing against the worst-case distribution within an ambiguity set, i.e., a set of distributions sharing the same partial information (Scarf 1957, Gallego and Moon 1993). DRO has been widely applied in inventory and supply chain management (Perakis and Roels 2008, Gao et al. 2019, Cui et al. 2023, Zhang et al. 2026), mechanism design and pricing (Carrasco et al. 2018, Chen et al. 2022, Chen et al. 2024), and machine learning (Sinha et al. 2018, Kuhn et al. 2019).

We study a risk-averse multi-item newsvendor problem under incomplete demand distribution information. Specifically, we consider a mean–covariance DRO framework, in which only the mean and covariance of random demands are known, and adopt conditional value-at-risk (CVaR) as the risk measure. The mean–covariance ambiguity set is attractive in sparse-data regimes because it requires only modest distributional information while preserving a clear economic interpretation. CVaR is well suited for decision makers who are concerned about substantial downside losses, as it captures expected performance in adverse tail scenarios. It is also a coherent risk measure, which preserves convexity and facilitates both theoretical analysis and computation (Artzner et al. 1999).

The main computational challenge in this setting arises from the interaction between tail-risk control and demand dependence. Correlations across item demands are important for capturing portfolio-level inventory risk, especially when products are substitutes, complements, or exposed to common demand shocks. However, explicitly preserving such dependence in a mean–covariance

DRO model with a CVaR objective leads to difficult optimization problems. Using conic duality and a full expansion of the piecewise-linear loss function, Hanasusanto et al. (2015) show that a closely related risk-averse DRO multi-item newsvendor problem admits an exact semidefinite programming (SDP) reformulation with the number of linear matrix inequalities that grows exponentially in the number of products. They further establish NP-hardness even when the ambiguity set contains only one mixture component, a special case closely related to the mean–covariance setting considered here. Solving such SDPs is computationally demanding, particularly as the number of items grows. To address this intractability, existing approaches primarily rely on decision-rule approximations, which are scalable but often lack theoretical performance guarantees (Hanasusanto et al. 2015, Wang and Delage 2024). In particular, little is known about the suboptimality of even the simplest linear decision rules (Iancu et al. 2013). This creates a gap between risk-aware, correlation-aware modeling and practically reliable solution methods.

In this paper, we address this tractability-reliability gap by developing a unified lifting-and-splitting framework for the mean–covariance DRO multi-item newsvendor problem with a CVaR objective. The lifting step provides an exact reformulation that replaces exponentially many positive semidefinite (PSD) constraints with exponentially many linear constraints, thereby substantially improving computational tractability. The splitting step then yields scalable approximations by retaining salient within-cluster correlations, such as those among substitutes, complements, or items exposed to common shocks, while dropping less reliably estimated cross-cluster dependence. This structure allows us to derive explicit performance guarantees and to develop a simple variance-based rule for constructing effective item partitions.

Our first contribution is an exact lifting reformulation based on the constrained Boolean quadric polytope (CBQP). This reformulation transforms the original problem into an equivalent SDP with exponentially many linear constraints, rather than exponentially many PSD constraints as in Hanasusanto et al. (2015). This shift substantially improves computational tractability and enables more efficient solution through constraint-generation methods. Our approach is related to Natarajan and Teo (2017), which connects the risk-neutral problem to the Boolean quadric polytope (BQP). However, that approach does not directly extend to the CVaR-based setting because the risk measure introduces additional nonlinearity.

Our second contribution is a correlation-splitting approximation for large-scale instances. Building on the lifting formulation, we partition items into clusters and selectively drop cross-cluster correlations. This approximation preserves within-cluster dependence while simplifying the global correlation structure, leading to a polynomial-time solvable relaxation. The resulting approximation is easy to implement, making it suitable for high-dimensional multi-item inventory problems.

Our third contribution is to establish theoretical performance guarantees for the proposed approximation. We identify structural conditions under which the correlation-splitting approximation is exact, despite the removal of cross-cluster correlations. For general settings, we derive an explicit upper bound on the optimality gap. The bound admits a cluster-wise decomposition and depends on both covariance-related and mean-related terms, thereby quantifying the cost of simplifying the dependence structure.

Our fourth contribution is to show that variance heterogeneity across clusters plays a central role in the quality of the correlation-splitting approximation. In the two-cluster case, we prove that the approximation becomes increasingly accurate when one cluster consists of items with sufficiently small demand variability. This result provides structural insight into when omitting cross-cluster correlations is relatively harmless. Motivated by this insight, we propose a simple variance-based splitting rule that prioritizes high-variance items when forming clusters. The rule avoids estimating high-dimensional correlation structures and is well-suited for applications in sparse-data settings.

Finally, we conduct extensive numerical experiments to evaluate the proposed framework. The results show that the CBQP-based lifting reformulation improves the computation of exact solutions, and the correlation-splitting approximation achieves strong scalability and high approximation quality. The experiments also demonstrate the effectiveness of the variance-based splitting rule across a range of problem instances.

Taken together, our results provide a tractable and performance-guaranteed framework for correlation-aware, risk-averse multi-item inventory decisions under moment ambiguity. More broadly, the proposed lifting-and-splitting approach offers a general methodology for managing high-dimensional dependence structures in DRO: lifting enables exact reformulation, splitting yields scalable approximation, and the variance-based rule provides an interpretable guideline for constructing reliable partitions.

Structure of the paper. The remainder of the paper is organized as follows. Section 2 reviews the relevant literature. Section 3 introduces the mean–covariance DRO multi-item newsvendor model and develops a CBQP-based lifting reformulation. Section 4 presents the correlation-splitting approximation, establishes conditions for exactness, and derives a theoretical upper bound on the optimality gap under general problem structures. Section 5 shows that both the optimality gap and its theoretical upper bound vanish at a linear rate in the relevant cluster-level standard deviation, and uses this insight to motivate a variance-based splitting rule for practical item partitioning. Section 6 reports numerical experiments on computational tractability, approximation quality, and the effectiveness of the proposed splitting rule. Finally, Section 7 concludes the paper. All technical proofs are provided in Appendix EC.1.

2. Literature Review

This section reviews three streams of literature related to our work: distributionally robust newsvendor models, tractable reformulations and approximations for multi-item newsvendor problems, and lifting-based approaches for exploiting dependence information.

The distributionally robust newsvendor problem has been extensively studied over the past several decades. The seminal work of Scarf (1957) assumes that only the mean and variance of demand are known and derives a closed-form optimal inventory level. Gallego and Moon (1993) extend this framework to the multi-item setting. Yue et al. (2006) and Perakis and Roels (2008) study regret-based criteria for ordering decisions under distributional ambiguity. Subsequent studies incorporate additional distributional information to mitigate the conservatism inherent in moment-based ambiguity sets. Natarajan et al. (2018) incorporate asymmetry information through mean–variance–semivariance ambiguity sets and demonstrate its benefits under asymmetric demand. Chen and Xie (2021) examine joint demand–yield uncertainty using a type-∞ Wasserstein ambiguity set. Das et al. (2021) show that Scarf's order quantity remains optimal for heavy-tailed demand distributions when only the first and α-th moments are known. A related line of work incorporates risk-averse decision making into distributionally robust newsvendor models. Natarajan et al. (2011a) provide a closed-form solution for a DRO newsvendor model in which risk is measured by CVaR. Han et al. (2014) consider a mean–standard deviation objective and derive an explicit solution that deviates from Scarf's classical order quantity, showing that the optimal inventory level may increase or decrease with the degree of risk aversion depending on the profitability parameters.

Compared with the single-item case, the distributionally robust multi-item newsvendor problem is more practically relevant but also significantly more challenging. Hanasusanto et al. (2015) develop a risk-averse DRO multi-item newsvendor model for multimodal demand distributions, where each mode is characterized by its probability mass, conditional mean, conditional covariance, and support information. They formulate a worst-case mean-risk objective involving CVaR, establish NP-hardness of the resulting problem, derive an exact SDP reformulation with exponentially many linear matrix inequalities, and propose tractable conservative approximations based on quadratic decision rules and partial expansion. Ardestani-Jaafari and Delage (2016) study robust optimization of sums of piecewise-linear functions and apply their framework to robust and distributionally robust multi-item newsvendor problems with budgeted uncertainty sets, mean information, and bounds on first-order partial moments; they derive tractable linear programming (LP) and SDP approximations and identify conditions under which the approximations are exact. Wang and Delage (2024) study a DRO multi-item newsvendor problem under a general event-wise ambiguity set, formulate a conservative approximation based on event-wise affine decision rules, and develop a column generation scheme to accelerate the resulting large-scale linear program. Long et al.

(2024) exploit supermodularity in two-stage DRO problems with support, mean, and mean absolute deviation information, deriving explicit worst-case distributions and applying the framework to several operational problems, including the multi-item newsvendor problem. More generally, Xu and Hanasusanto (2025) develop copositive reformulations and piecewise linear/quadratic decision-rule approximations for multi-stage robust optimization, providing tighter semidefinite approximations than existing decision-rule schemes.

A parallel research stream leverages lifting techniques to obtain tighter reformulations and relaxations. Natarajan et al. (2011b) develop a completely positive programming approach for mixed 0–1 linear programs with mean–variance ambiguity in nonnegative objective coefficients. Natarajan and Teo (2017) develop semidefinite relaxations for the multi-item DRO newsvendor problem with mean–covariance information, showing a connection to the BQP and yielding tighter bounds than Scarf-type bounds under dependence. Padmanabhan et al. (2021) further extend this lifting reformulation when the covariance matrix is block diagonal. These lifting-based approaches provide powerful tools for exploiting dependence information, but they primarily address risk-neutral objectives. Our work bridges this lifting literature with the risk-averse DRO modeling perspective of Hanasusanto et al. (2015): we extend the lifting paradigm to a CVaR-based risk-averse setting by developing an exact CBQP-based reformulation and a performance-guaranteed correlation-splitting approximation for the mean–covariance DRO multi-item newsvendor problem.

From a tractability perspective, the above literature highlights several complementary ways to address the computational difficulty of multi-item DRO newsvendor problems. Decision-rule methods improve scalability by restricting the functional form of recourse or auxiliary decisions, and they have been developed in affine, event-wise affine, quadratic, and piecewise variants (Hanasusanto et al. 2015, Ardestani-Jaafari and Delage 2016, Wang and Delage 2024, Xu and Hanasusanto 2025). Structure-exploiting approaches instead identify special ambiguity sets or objective properties, such as budgeted uncertainty or supermodularity, under which exact reformulations or optimality of certain decision rules can be obtained (Ardestani-Jaafari and Delage 2016, Long et al. 2024). While these methods significantly improve solvability, theoretical guarantees on optimality gaps remain limited in general. Even for linear decision rules, the extent of suboptimality is not well understood (Iancu et al. 2013). Cheramin et al. (2022) propose an alternative principal component analysis (PCA)-based partitioning method and establish performance bounds, although the resulting formulation still involves exponentially many PSD constraints. Jiang et al. (2026) further integrate dimensionality reduction with subsequent optimization, enabling automatic selection of the reduction scheme. Our approach differs from these methods by selectively discarding cross-cluster correlation information rather than approximating the decision rules or reducing dimensionality before optimization. This correlation-splitting strategy yields a scalable relaxation while allowing us to derive performance guarantees and structural insights into when the approximation is tight.

3. Model Formulation and CBQP-Based Lifting

This section develops the model formulation and the CBQP-based lifting reformulation. We first introduce the distributionally robust CVaR newsvendor model under mean–covariance ambiguity. We then review the standard conic-duality reformulation, which leads to an SDP with exponentially many PSD constraints. Finally, we develop an exact lifting reformulation based on CBQP, which replaces the exponentially many PSD constraints with exponentially many linear constraints. This reformulation provides the foundation for the correlation-splitting approximation developed in the subsequent sections.

Notation. Throughout the paper, lowercase letters denote scalars, bold lowercase letters denote vectors, and bold uppercase letters denote matrices. Random variables are represented with a tilde, such as $\tilde{d}$ for random demand. We use $\mathcal{M}(\Xi)$ to denote the set of all probability distributions supported on Ξ. For scalars a and b, let $a \wedge b$ and $a \vee b$ denote their minimum and maximum, respectively. For any scalar a, let $a^+ = \max\{a, 0\}$ and $a^- = \min\{a, 0\}$ denote its positive and negative parts, respectively; these operators are applied componentwise to vectors and matrices. For an integer n, define $[n] = \{1, 2, \ldots, n\}$. The sets of n-dimensional real vectors, nonnegative real vectors, nonpositive real vectors, and $n \times n$ real matrices are denoted by $\mathbb{R}^n$, $\mathbb{R}^n_+$, $\mathbb{R}^n_-$, and $\mathbb{R}^{n\times n}$, respectively. The vector $\boldsymbol{e}_n \in \mathbb{R}^n$ denotes the n-dimensional vector of ones. The symbol $\mathbf{0}$ denotes a zero vector or zero matrix, with its dimension inferred from context when no subscript is specified. For $\boldsymbol{a} \in \mathbb{R}^n$, $\operatorname{diag}(\boldsymbol{a}) \in \mathbb{R}^{n\times n}$ denotes the diagonal matrix with entries $(\operatorname{diag}(\boldsymbol{a}))_{ii} = a_i$ for $i \in [n]$ and zeros elsewhere. For matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times m}$, $\langle \boldsymbol{A}, \boldsymbol{B}\rangle = \sum_{i,j} A_{ij} B_{ij}$ denotes the Frobenius inner product, and $\boldsymbol{A} \circ \boldsymbol{B} = (A_{ij}B_{ij})_{i,j}$ denotes the Hadamard product. For a matrix $\boldsymbol{A}$, $\|\boldsymbol{A}\|_F = \sqrt{\langle \boldsymbol{A}, \boldsymbol{A}\rangle}$ denotes its Frobenius norm, $\boldsymbol{A}^\top$ its transpose, and $\boldsymbol{A}^{-1}$ its inverse, when the inverse exists.

3.1. Risk-Averse DRO Newsvendor Model

We consider a risk-averse multi-item newsvendor who sells n products. Let $\boldsymbol{p} \in \mathbb{R}^n_+$ and $\boldsymbol{c} \in \mathbb{R}^n_+$ denote the vectors of selling prices and unit ordering costs, respectively, with $\boldsymbol{p} > \boldsymbol{c} > \mathbf{0}$ componentwise. Let $\tilde{\boldsymbol{d}} \in \mathbb{R}^n$ denote the random demand vector with distribution F, and let $\boldsymbol{y} \in \mathbb{R}^n$ denote the order quantity vector. The corresponding random profit is $\pi(\boldsymbol{y}, \tilde{\boldsymbol{d}}) = \boldsymbol{p}^\top(\boldsymbol{y} \wedge \tilde{\boldsymbol{d}}) - \boldsymbol{c}^\top \boldsymbol{y}$, where the minimum operator is applied componentwise. The newsvendor evaluates profit using lower-tail CVaR. For a random profit π and a tail probability level $\eta \in (0, 1]$, define the corresponding value-at-risk (VaR) as

$$\mathrm{VaR}_\eta(\pi) := \inf\{r \in \mathbb{R} \mid \mathbb{P}(\pi \le r) \ge \eta\}.$$

For continuous distributions, $\mathrm{CVaR}_\eta(\pi)$ can be interpreted as the expected profit conditional on π falling in its lower η-tail. Thus, a smaller value of η focuses on more adverse profit outcomes and corresponds to a more risk-averse newsvendor (Chen et al. 2009). We use the following equivalent representation of CVaR (Rockafellar and Uryasev 2000):

$$\mathrm{CVaR}_\eta(\pi) = \max_{v\in\mathbb{R}} \left\{ v + \frac{1}{\eta}\mathbb{E}\left[\min(\pi - v, 0)\right] \right\}.$$

If the demand distribution F is known, the risk-averse multi-item newsvendor problem can be written as

$$\max_{\boldsymbol{y}\in\mathbb{R}^n} \max_{v\in\mathbb{R}} \left\{ v + \frac{1}{\eta}\mathbb{E}_F\left[\min\left(\pi(\boldsymbol{y},\tilde{\boldsymbol{d}}) - v, 0\right)\right] \right\}. \tag{1}$$

Problem (1) does not admit a closed-form solution in general. As noted by Hanasusanto et al. (2016), even evaluating $\mathbb{E}_F[\min(\pi(\boldsymbol{y},\tilde{\boldsymbol{d}}) - v, 0)]$ for a fixed pair $(\boldsymbol{y}, v)$ can be difficult.

In practice, the newsvendor often has access to only limited distributional information about demand. We therefore adopt a distributionally robust perspective and model the set of candidate demand distributions by an ambiguity set $\mathscr{F}$. The decision maker then hedges against the most adverse distribution in $\mathscr{F}$. In this paper, we focus on a mean–covariance ambiguity set. Specifically, the demand mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma} \succ 0$ are assumed to be known, and

$$\mathscr{F} := \left\{ F \in \mathcal{M}(\mathbb{R}^n) \;\middle|\; \mathbb{E}_F[\tilde{\boldsymbol{d}}] = \boldsymbol{\mu},\ \mathbb{E}_F[\tilde{\boldsymbol{d}}\tilde{\boldsymbol{d}}^\top] = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\top \right\}.$$

For a demand distribution F, we write $\mathrm{CVaR}_{\eta,F}(\pi(\boldsymbol{y},\tilde{\boldsymbol{d}}))$ to make explicit that the expectation in the CVaR representation is evaluated under F. Within this framework, the distributionally robust risk-averse multi-item newsvendor problem is formulated as

$$\Pi = \max_{\boldsymbol{y}\in\mathbb{R}^n} \min_{F\in\mathscr{F}} \mathrm{CVaR}_{\eta,F}\left(\pi(\boldsymbol{y},\tilde{\boldsymbol{d}})\right) = \max_{\boldsymbol{y}\in\mathbb{R}^n} \min_{F\in\mathscr{F}} \max_{v\in\mathbb{R}} \left\{ v + \frac{1}{\eta}\mathbb{E}_F\left[\min\left(\pi(\boldsymbol{y},\tilde{\boldsymbol{d}}) - v, 0\right)\right] \right\}. \tag{P}$$

3.2. Conic-Duality SDP Reformulation

We first review a standard SDP reformulation obtained through conic duality. This formulation serves as the baseline against which we develop the CBQP-based lifting reformulation in the next subsection. Using the CVaR representation and the stochastic minimax theorem of Shapiro and Kleywegt (2002), and introducing the auxiliary variable $\bar{v} = v - (\boldsymbol{p}-\boldsymbol{c})^\top\boldsymbol{y}$, Problem (P) can be equivalently written as

$$\begin{aligned} \Pi &= \max_{\boldsymbol{y}\in\mathbb{R}^n} \max_{v\in\mathbb{R}} \left\{ v + \frac{1}{\eta}\min_{F\in\mathscr{F}} \mathbb{E}_F\left[\min\left(\boldsymbol{p}^\top(\boldsymbol{y}\wedge\tilde{\boldsymbol{d}}) - \boldsymbol{c}^\top\boldsymbol{y} - v, 0\right)\right] \right\} \\ &= \max_{\boldsymbol{y}\in\mathbb{R}^n} \max_{\bar{v}\in\mathbb{R}} \left\{ \bar{v} + (\boldsymbol{p}-\boldsymbol{c})^\top\boldsymbol{y} + \frac{1}{\eta}\min_{F\in\mathscr{F}} \mathbb{E}_F\left[\min\left(\boldsymbol{p}^\top(\tilde{\boldsymbol{d}}-\boldsymbol{y})^- - \bar{v}, 0\right)\right] \right\}. \end{aligned} \tag{2}$$

The second equality follows from $\boldsymbol{p}^\top(\boldsymbol{y}\wedge\tilde{\boldsymbol{d}}) - \boldsymbol{c}^\top\boldsymbol{y} = (\boldsymbol{p}-\boldsymbol{c})^\top\boldsymbol{y} + \boldsymbol{p}^\top(\tilde{\boldsymbol{d}}-\boldsymbol{y})^-$.

For any fixed $(\boldsymbol{y},\bar{v})$, the inner minimization over F can be written as the following generalized moment problem:

$$
\begin{aligned}
\min_{F} \quad & \int_{\mathbb{R}^n} \min\left(\boldsymbol{p}^\top(\boldsymbol{d}-\boldsymbol{y})^- - \bar{v}, 0\right)\,\mathrm{d}F(\boldsymbol{d}) \\
\text{s.t.} \quad & \int_{\mathbb{R}^n} \boldsymbol{d}\,\mathrm{d}F(\boldsymbol{d}) = \boldsymbol{\mu} \\
& \int_{\mathbb{R}^n} \boldsymbol{d}\boldsymbol{d}^\top\,\mathrm{d}F(\boldsymbol{d}) = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\top \\
& \int_{\mathbb{R}^n} 1\,\mathrm{d}F(\boldsymbol{d}) = 1 \\
& \mathrm{d}F(\boldsymbol{d}) \geq 0,\ \forall \boldsymbol{d}\in\mathbb{R}^n
\end{aligned}
\tag{3}
$$

Since $\boldsymbol{\Sigma}\succ 0$, the moment constraints lie in the interior of the moment cone. By strong duality (Isii 1962), Problem (3) is equivalent to its dual:

$$
\begin{aligned}
\max_{s_1,\boldsymbol{s}_2,\boldsymbol{s}_3} \quad & s_1 + \boldsymbol{s}_2^\top\boldsymbol{\mu} + \left\langle \boldsymbol{s}_3, \boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^\top\right\rangle \\
\text{s.t.} \quad & s_1 + \boldsymbol{s}_2^\top\boldsymbol{d} + \left\langle \boldsymbol{s}_3, \boldsymbol{d}\boldsymbol{d}^\top\right\rangle \leq \min\left(\boldsymbol{p}^\top(\boldsymbol{d}-\boldsymbol{y})^- - \bar{v}, 0\right),\ \forall \boldsymbol{d}\in\mathbb{R}^n
\end{aligned}
\tag{4}
$$

By expanding the piecewise-linear function in (4) and applying the technical reformulations in Appendix EC.1.1, Problem (P) can be equivalently reformulated as the following SDP. This conic-duality reformulation is standard in the DRO literature (Delage and Ye 2010, Han et al. 2014, Hanasusanto et al. 2015, Natarajan et al. 2018).

Proposition 1. *Problem* (P) *is equivalent to*

$$
\begin{aligned}
\max_{\boldsymbol{y},\bar{v},s_1,\boldsymbol{s}_2,\boldsymbol{s}_3} \quad & \bar{v} + (\boldsymbol{p}-\boldsymbol{c})^\top\boldsymbol{y} + \frac{1}{\eta}\left(s_1 + \boldsymbol{s}_2^\top\boldsymbol{\mu} + \left\langle \boldsymbol{s}_3, \boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^\top\right\rangle\right) \\
\text{s.t.} \quad & \begin{pmatrix} -s_1 - (\boldsymbol{p}\circ\boldsymbol{x})^\top\boldsymbol{y} - \bar{v} & \frac{1}{2}(-\boldsymbol{s}_2+\boldsymbol{p}\circ\boldsymbol{x})^\top \\ \frac{1}{2}(-\boldsymbol{s}_2+\boldsymbol{p}\circ\boldsymbol{x}) & -\boldsymbol{s}_3 \end{pmatrix} \succeq 0,\ \forall \boldsymbol{x}\in\{0,1\}^n \\
& \begin{pmatrix} -s_1 & -\frac{1}{2}\boldsymbol{s}_2^\top \\ -\frac{1}{2}\boldsymbol{s}_2 & -\boldsymbol{s}_3 \end{pmatrix} \succeq 0
\end{aligned}
\tag{D}
$$

Proposition 1 shows that Problem (P) admits an SDP reformulation with exponentially many PSD constraints. Because PSD constraints are computationally demanding and their number grows exponentially with the number of items, this formulation is tractable only for relatively small instances. This computational limitation is consistent with the NP-hardness results for closely related risk-averse DRO multi-item newsvendor problems in Hanasusanto et al. (2015).

In the next subsection, we introduce a CBQP-based lifting reformulation. The reformulation preserves exactness while replacing the exponentially many PSD constraints in (D) with exponentially many linear constraints. This shift is crucial for the scalable approximation methods developed later, including the proposed correlation-splitting approximation in Section 4.

3.3. CBQP-Based Lifting Reformulation

We now develop an exact lifting reformulation for the distributionally robust CVaR newsvendor problem. The key idea is to lift the binary representation of the piecewise-linear term into a moment space characterized by CBQP. This reformulation preserves exactness while replacing the exponentially many PSD constraints in formulation (D) with exponentially many linear constraints.

We start from the inner minimization problem in (2):

$$\min_{F\in\mathscr{F}} \mathbb{E}_F\left[\min\left(\boldsymbol{p}^\top(\tilde{\boldsymbol{d}}-\boldsymbol{y})^- - \bar{v}, 0\right)\right]. \tag{5}$$

This problem admits the following equivalent reformulation:

$$\min_{F\in\mathscr{F}} \mathbb{E}_F\left[\min_{x\in\{0,1\}} \left(\boldsymbol{p}^\top(\tilde{\boldsymbol{d}}-\boldsymbol{y})^- - \bar{v}\right) x\right] \tag{6a}$$

$$= \min_{F\in\mathscr{F}} \mathbb{E}_F\left[\min_{\substack{x\in\{0,1\},\\ \boldsymbol{z}\in\{0,1\}^n}} \left(\boldsymbol{p}\circ(\tilde{\boldsymbol{d}}-\boldsymbol{y})\right)^\top \boldsymbol{z}x - \bar{v}x\right]. \tag{6b}$$

Introducing $\boldsymbol{t}=\boldsymbol{z}x$ and enforcing $\boldsymbol{t}\le x\boldsymbol{e}_n$, (6b) can be written as

$$\min_{F\in\mathscr{F}} \mathbb{E}_F\left[\min_{\substack{x\in\{0,1\},\ \boldsymbol{t}\in\{0,1\}^n\\ \boldsymbol{t}\le x\boldsymbol{e}_n}} \left(\boldsymbol{p}\circ(\tilde{\boldsymbol{d}}-\boldsymbol{y})\right)^\top \boldsymbol{t} - \bar{v}x\right]. \tag{7}$$

Indeed, when $x=0$, the constraint $\boldsymbol{t}\le x\boldsymbol{e}_n$ forces $\boldsymbol{t}=\boldsymbol{0}$ and the objective value is zero. When $x=1$, the constraint $\boldsymbol{t}\le\boldsymbol{e}_n$ is redundant for binary $\boldsymbol{t}$, so (6b) and (7) have the same feasible region and objective value.

For a given demand realization $\boldsymbol{d}\in\mathbb{R}^n$, let $\boldsymbol{t}_*(\boldsymbol{d})$ and $x_*(\boldsymbol{d})$ denote minimizers in (7). Since $\tilde{\boldsymbol{d}}$ is random, $\boldsymbol{t}_*(\tilde{\boldsymbol{d}})$ and $x_*(\tilde{\boldsymbol{d}})$ are binary random variables. Thus, (7) can be written as

$$\min_{F\in\mathscr{F}} \mathbb{E}_F\left[\left(\boldsymbol{p}\circ(\tilde{\boldsymbol{d}}-\boldsymbol{y})\right)^\top \boldsymbol{t}_*(\tilde{\boldsymbol{d}}) - \bar{v}x_*(\tilde{\boldsymbol{d}})\right]. \tag{8}$$

The expression in (8) involves products between demand and binary random variables. To obtain a tractable lifted representation, define

$$\begin{aligned} x &:= \mathbb{E}_F\left[x_*(\tilde{\boldsymbol{d}})\right], & \boldsymbol{t} &:= \mathbb{E}_F\left[\boldsymbol{t}_*(\tilde{\boldsymbol{d}})\right], \\ \boldsymbol{Q} &:= \mathbb{E}_F\left[\boldsymbol{t}_*(\tilde{\boldsymbol{d}})\tilde{\boldsymbol{d}}^\top\right], & \boldsymbol{T} &:= \mathbb{E}_F\left[\boldsymbol{t}_*(\tilde{\boldsymbol{d}})\boldsymbol{t}_*(\tilde{\boldsymbol{d}})^\top\right]. \end{aligned} \tag{9}$$

With these variables, the objective in (8) becomes

$$\mathbb{E}_F\left[\left(\boldsymbol{p}\circ(\tilde{\boldsymbol{d}}-\boldsymbol{y})\right)^\top \boldsymbol{t}_*(\tilde{\boldsymbol{d}}) - \bar{v}x_*(\tilde{\boldsymbol{d}})\right] = \langle \mathrm{diag}(\boldsymbol{p}), \boldsymbol{Q}\rangle - (\boldsymbol{p}\circ\boldsymbol{y})^\top\boldsymbol{t} - \bar{v}x. \tag{10}$$

The lifted variables must satisfy two sets of constraints. First, for each realization of $\tilde{\boldsymbol{d}}$, consider the rank-one matrix

$$\begin{pmatrix} \tilde{\boldsymbol{d}} \\ 1 \\ \boldsymbol{t}_*(\tilde{\boldsymbol{d}}) \end{pmatrix} \begin{pmatrix} \tilde{\boldsymbol{d}} \\ 1 \\ \boldsymbol{t}_*(\tilde{\boldsymbol{d}}) \end{pmatrix}^\top.$$

This matrix is PSD by construction, and therefore its expectation is also PSD. Using the definitions of the lifted variables, this expectation can be written as

$$\begin{pmatrix} \boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^\top & \boldsymbol{\mu} & \boldsymbol{Q}^\top \\ \boldsymbol{\mu}^\top & 1 & \boldsymbol{t}^\top \\ \boldsymbol{Q} & \boldsymbol{t} & \boldsymbol{T} \end{pmatrix} = \mathbb{E}_F\left[\begin{pmatrix} \tilde{\boldsymbol{d}} \\ 1 \\ \boldsymbol{t}_*(\tilde{\boldsymbol{d}}) \end{pmatrix}\begin{pmatrix} \tilde{\boldsymbol{d}} \\ 1 \\ \boldsymbol{t}_*(\tilde{\boldsymbol{d}}) \end{pmatrix}^\top\right].$$

Hence, the lifted variables must satisfy the PSD constraint:

$$\begin{pmatrix} \boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^\top & \boldsymbol{\mu} & \boldsymbol{Q}^\top \\ \boldsymbol{\mu}^\top & 1 & \boldsymbol{t}^\top \\ \boldsymbol{Q} & \boldsymbol{t} & \boldsymbol{T} \end{pmatrix} \succeq 0. \tag{11}$$

Second, the binary variables must encode the binary relation that $t_{*,i}(\tilde{\boldsymbol{d}})=1$ is allowed only when $x_*(\tilde{\boldsymbol{d}})=1$, for each item $i\in[n]$. This relation is captured by the constraint $\boldsymbol{t}_*(\tilde{\boldsymbol{d}})\le x_*(\tilde{\boldsymbol{d}})\boldsymbol{e}_n$, and is represented in the lifted formulation through CBQP:

$$\begin{pmatrix} x & x & \boldsymbol{t}^\top \\ x & 1 & \boldsymbol{t}^\top \\ \boldsymbol{t} & \boldsymbol{t} & \boldsymbol{T} \end{pmatrix} = \mathbb{E}_F\left[\begin{pmatrix} x_*(\tilde{\boldsymbol{d}}) \\ 1 \\ \boldsymbol{t}_*(\tilde{\boldsymbol{d}}) \end{pmatrix}\begin{pmatrix} x_*(\tilde{\boldsymbol{d}}) \\ 1 \\ \boldsymbol{t}_*(\tilde{\boldsymbol{d}}) \end{pmatrix}^\top\right] \in \mathrm{CBQP}(n). \tag{12}$$

Here,

$$\mathrm{CBQP}(n) := \mathrm{conv}\left\{\begin{pmatrix} w_0 \\ 1 \\ \boldsymbol{w} \end{pmatrix}\begin{pmatrix} w_0 \\ 1 \\ \boldsymbol{w} \end{pmatrix}^\top \,\middle|\, (w_0,\boldsymbol{w})\in\Upsilon\right\},$$

where

$$\Upsilon := \left\{(w_0,\boldsymbol{w})\in\{0,1\}^{n+1} \mid \boldsymbol{w}\le w_0\boldsymbol{e}_n\right\}.$$

Combining (10), (11), and (12), we obtain the following lifted formulation. Its equivalence to the original inner minimization problem is stated in Proposition 2.

PROPOSITION 2. *Define*

$$\Pi_B = \max_{\boldsymbol{y},\bar{v}} \quad \bar{v} + (\boldsymbol{p}-\boldsymbol{c})^\top\boldsymbol{y} + \frac{1}{\eta}\Phi_B(\boldsymbol{y},\bar{v}), \tag{B}$$

where

$$\begin{aligned} \Phi_B(\boldsymbol{y},\bar{v}) := \min_{x,\boldsymbol{t},\boldsymbol{Q},\boldsymbol{T}} \quad & \langle \mathrm{diag}(\boldsymbol{p}),\boldsymbol{Q}\rangle - (\boldsymbol{p}\circ\boldsymbol{y})^\top\boldsymbol{t} - \bar{v}x \\ \text{s.t.} \quad & \begin{pmatrix} \boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^\top & \boldsymbol{\mu} & \boldsymbol{Q}^\top \\ \boldsymbol{\mu}^\top & 1 & \boldsymbol{t}^\top \\ \boldsymbol{Q} & \boldsymbol{t} & \boldsymbol{T} \end{pmatrix} \succeq 0 \\ & \begin{pmatrix} x & x & \boldsymbol{t}^\top \\ x & 1 & \boldsymbol{t}^\top \\ \boldsymbol{t} & \boldsymbol{t} & \boldsymbol{T} \end{pmatrix} \in \mathrm{CBQP}(n) \end{aligned} \tag{B-I}$$

Then, for any given $(\boldsymbol{y},\bar{v})$*, Problem* (5) *is equivalent to Problem* (B-I)*. Moreover,* $\Pi=\Pi_B$*, i.e., Problem* (P) *is equivalent to Problem* (B).

Proposition 2 extends the lifting approach of Natarajan and Teo (2017), which addresses the risk-neutral problem through BQP. The CVaR objective introduces an additional binary indicator for the lower-tail truncation, and the constraint $\boldsymbol{t} \le x\boldsymbol{e}_n$ links this indicator with the item-level binary variables. This additional structure motivates the use of CBQP and allows us to extend the lifting paradigm to the risk-averse setting. Compared with (D), the lifted formulation replaces exponentially many PSD constraints with a CBQP constraint. The CBQP constraint can be represented by exponentially many linear inequalities, as stated in the following lemma.

Lemma 1. *The CBQP constraint in Problem* (B-I) *is equivalent to the following linear system:*

$$\begin{cases} \boldsymbol{t} = \sum_{\boldsymbol{w}\in\{0,1\}^n} \eta_{\boldsymbol{w}} \boldsymbol{w} \\ \boldsymbol{T} = \sum_{\boldsymbol{w}\in\{0,1\}^n} \eta_{\boldsymbol{w}} \boldsymbol{w}\boldsymbol{w}^\top \\ \sum_{\boldsymbol{w}\in\{0,1\}^n} \eta_{\boldsymbol{w}} = x \\ x \le 1 \\ \eta_{\boldsymbol{w}} \ge 0, \ \forall \boldsymbol{w} \in \{0,1\}^n \end{cases}$$

Using this representation, we dualize the inner minimization in (B-I) and obtain the following joint maximization problem.

Theorem 1. *Problem* (P) *is equivalent to the following SDP:*

$$\begin{aligned} \max_{\substack{\boldsymbol{y},\bar{v},\alpha,\tau,\gamma,\\ \boldsymbol{\beta},\boldsymbol{\delta},\boldsymbol{\Gamma},\boldsymbol{V}}} \quad & \bar{v} + (\boldsymbol{p}-\boldsymbol{c})^\top \boldsymbol{y} - \frac{1}{\eta}\left(\alpha - \gamma + \boldsymbol{\mu}^\top\boldsymbol{\beta} + \left\langle \boldsymbol{\Gamma}, \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\top \right\rangle\right) \\ \text{s.t.} \quad & \begin{pmatrix} \boldsymbol{\Gamma} & \frac{1}{2}\boldsymbol{\beta} & \frac{1}{2}\mathrm{diag}(\boldsymbol{p}) \\ \frac{1}{2}\boldsymbol{\beta}^\top & \alpha & \frac{1}{2}(\boldsymbol{\delta} - \boldsymbol{p}\circ\boldsymbol{y})^\top \\ \frac{1}{2}\mathrm{diag}(\boldsymbol{p}) & \frac{1}{2}(\boldsymbol{\delta} - \boldsymbol{p}\circ\boldsymbol{y}) & \boldsymbol{V} \end{pmatrix} \succeq 0 \\ & \gamma + \tau + \bar{v} = 0 \\ & \boldsymbol{\delta}^\top\boldsymbol{w} + \boldsymbol{w}^\top\boldsymbol{V}\boldsymbol{w} - \tau \le 0, \ \forall \boldsymbol{w}\in\{0,1\}^n \\ & \gamma \le 0 \end{aligned} \tag{B-R}$$

The SDP (B-R) replaces the exponentially many PSD constraints in (D) with exponentially many linear constraints. This replacement is the main computational advantage of the CBQP-based lifting reformulation. Although the number of constraints remains exponential, linear constraints are much easier to separate and add iteratively than PSD constraints. Specifically, separating the family of constraints $\boldsymbol{\delta}^\top\boldsymbol{w} + \boldsymbol{w}^\top\boldsymbol{V}\boldsymbol{w} - \tau \le 0, \forall \boldsymbol{w}\in\{0,1\}^n$, in (B-R) reduces to optimizing a binary quadratic function over $\{0,1\}^n$. Although this separation problem is NP-hard, it can be handled in practice using off-the-shelf mixed-integer quadratic programming (MIQP) solvers. This enables a practical constraint-generation scheme: after solving a restricted SDP, one searches for violated linear constraints and adds them iteratively. The constraint-generation algorithm in Appendix EC.3 implements this procedure by adding only violated constraints as needed. By contrast, applying an analogous constraint-generation strategy directly to (D) would require solving mixed-integer SDPs, for which reliable off-the-shelf solvers are currently unavailable.

Despite this computational advantage, the constraint-generation approach still requires solving a sequence of increasingly large SDPs. Moreover, before convergence, the intermediate solutions do not provide an explicit performance guarantee relative to the original problem. In the next section, we therefore introduce a complementary approximation scheme based on correlation splitting. This approximation further enhances tractability and admits explicit performance guarantees.

4. Correlation-Splitting Approximation and Performance Guarantees

Building on the CBQP-based lifting reformulation in Section 3, we now introduce a correlation-splitting approximation to further improve tractability. The approximation enlarges the ambiguity set by retaining only within-cluster covariance information and omitting cross-cluster covariance information. This section addresses three questions. First, how can the full-information model be approximated in a polynomial-time solvable manner? Second, under what conditions is this approximation exact? Third, when exactness does not hold, how large can the resulting approximation error be? To answer these questions, Section 4.1 presents the correlation-splitting formulation and its tractability properties, Section 4.2 establishes exactness conditions, and Section 4.3 derives an explicit upper bound on the optimality gap.

4.1. Correlation-Splitting Formulation and Tractability

To address the computational difficulty of the distributionally robust multi-item CVaR newsvendor problem (P), we introduce a correlation-splitting approximation. The key idea is to partition items into clusters, retain covariance information within each cluster, and omit cross-cluster covariance information. This approximation is motivated by both modeling and statistical considerations. From a modeling perspective, the most salient dependence relationships, such as strong positive correlations among complementary products, strong negative correlations among substitutes, or common demand shocks within a product group, can often be organized at the cluster level. From a statistical perspective, estimating a full high-dimensional covariance matrix can be unreliable when data are limited, whereas within-cluster covariance information may be more stable or more readily justified. Correlation splitting therefore allows the decision maker to specify the dependence information that is most credible and operationally meaningful, while avoiding potentially unreliable cross-cluster covariance estimates. This modeling perspective is related to DRO approaches based on partially specified correlation structures (Ahipasaoglu et al. 2019, Padmanabhan et al. 2021). We show that, under appropriately chosen cluster sizes, the resulting formulation admits a polynomial-time representation.

Let $N_1, \ldots, N_R$ be a partition of the item set $N = \{1, 2, \ldots, n\}$ into R disjoint clusters. For each $r \in [R]$, let $n_r = |N_r|$ denote the number of items in cluster r. For any vector $\boldsymbol{x}$ and matrix $\boldsymbol{X}$, we use $\boldsymbol{x}^r$ and $\boldsymbol{X}^r$ to denote the corresponding subvector and principal submatrix restricted to cluster r. For example, $\boldsymbol{\mu}^r$ and $\boldsymbol{\Sigma}^r$ denote the mean vector and covariance matrix of the items in cluster r, respectively.

Under the correlation-splitting approximation, the ambiguity set is relaxed to

$$\mathscr{F}^c = \left\{ F \in \mathcal{M}(\mathbb{R}^n) \;\middle|\; \mathbb{E}_F[\tilde{\boldsymbol{d}}] = \boldsymbol{\mu},\ \mathbb{E}_F\left[\tilde{\boldsymbol{d}}^r (\tilde{\boldsymbol{d}}^r)^\top\right] = \boldsymbol{\Sigma}^r + \boldsymbol{\mu}^r (\boldsymbol{\mu}^r)^\top,\ \forall r \in [R] \right\},$$

where $\tilde{\boldsymbol{d}}^r \in \mathbb{R}^{n_r}$ denotes the demand subvector associated with cluster r. Compared with $\mathscr{F}$, the ambiguity set $\mathscr{F}^c$ preserves the marginal mean information and the within-cluster second-moment information, but does not impose constraints on cross-cluster second moments. Since $\mathscr{F}^c \supseteq \mathscr{F}$, this relaxation leads to the following conservative approximation of Problem (P):

$$\underline{\Pi} = \max_{\boldsymbol{y} \in \mathbb{R}^n} \min_{F \in \mathscr{F}^c} \mathrm{CVaR}_{\eta, F}\left(\pi(\boldsymbol{y}, \tilde{\boldsymbol{d}})\right). \tag{P-C}$$

For notational convenience, we assume without loss of generality that the items are indexed according to the cluster order. That is, for any $r_1 < r_2$, we have $i < j$ for all $i \in N_{r_1}$ and $j \in N_{r_2}$. Any partition can be brought into this form by a suitable permutation of item indices.

Given a fixed partition, the worst-case CVaR problem under the relaxed ambiguity set admits a lifted reformulation analogous to Proposition 2.

Proposition 3. *Define*

$$\Pi_C = \max_{\boldsymbol{y}, \bar{v}} \quad \bar{v} + \sum_{r=1}^{R} (\boldsymbol{p}^r - \boldsymbol{c}^r)^\top \boldsymbol{y}^r + \frac{1}{\eta} \Phi_C(\boldsymbol{y}, \bar{v}), \tag{C}$$

where

$$\begin{aligned} \Phi_C(\boldsymbol{y}, \bar{v}) := \min_{x, \boldsymbol{t}^r, \boldsymbol{Q}^r, \boldsymbol{T}^r} \quad & \sum_{r=1}^{R} \left(\langle \mathrm{diag}(\boldsymbol{p}^r), \boldsymbol{Q}^r \rangle - (\boldsymbol{p}^r \circ \boldsymbol{y}^r)^\top \boldsymbol{t}^r \right) - \bar{v} x \\ \text{s.t.} \quad & \begin{pmatrix} \boldsymbol{\Sigma}^r + \boldsymbol{\mu}^r (\boldsymbol{\mu}^r)^\top & \boldsymbol{\mu}^r & (\boldsymbol{Q}^r)^\top \\ (\boldsymbol{\mu}^r)^\top & 1 & (\boldsymbol{t}^r)^\top \\ \boldsymbol{Q}^r & \boldsymbol{t}^r & \boldsymbol{T}^r \end{pmatrix} \succeq 0, \ \forall r \in [R] \\ & \begin{pmatrix} x & x & (\boldsymbol{t}^r)^\top \\ x & 1 & (\boldsymbol{t}^r)^\top \\ \boldsymbol{t}^r & \boldsymbol{t}^r & \boldsymbol{T}^r \end{pmatrix} \in \mathrm{CBQP}(n_r), \ \forall r \in [R] \end{aligned} \tag{C-I}$$

Then $\underline{\Pi} = \Pi_C$, i.e., the correlation-splitting problem (P-C) *is equivalent to Problem* (C)*.*

Proposition 3 extends the risk-neutral partial-correlation reformulation of Padmanabhan et al. (2021) to the risk-averse CVaR setting. By construction, $\mathscr{F}^c$ relaxes the original moment constraints and satisfies $\mathscr{F}^c \supseteq \mathscr{F}$. Therefore, for any fixed order quantity, the worst-case CVaR under $\mathscr{F}^c$ is no larger than that under $\mathscr{F}$. Hence, the correlation-splitting problem provides a lower bound on the original problem value. Correlation splitting improves tractability by enlarging the ambiguity set in a structured way, at the cost of a possible approximation gap.

COROLLARY 1. $\Pi_C \le \Pi_B$, *i.e., Problem* (C) *provides a lower bound for Problem* (B).

Similar to Lemma 1, the CBQP constraints in (C-I) can be represented by linear constraints. By dualizing Problem (C-I) and substituting the resulting formulation into Problem (C), we obtain the following SDP reformulation of the correlation-splitting approximation.

THEOREM 2. *Problem* (C) *is equivalent to the following SDP:*

$$
\begin{aligned}
\max_{\substack{\boldsymbol{y},\bar{v},\gamma,\\ \alpha^r,\tau^r,\boldsymbol{\beta}^r,\boldsymbol{\delta}^r,\boldsymbol{\Gamma}^r,\boldsymbol{V}^r}} \quad & \bar{v} + (\boldsymbol{p}-\boldsymbol{c})^\top \boldsymbol{y} + \frac{\gamma}{\eta} - \frac{1}{\eta}\sum_{r=1}^{R}\left(\alpha^r + (\boldsymbol{\mu}^r)^\top \boldsymbol{\beta}^r + \left\langle \boldsymbol{\Gamma}^r, \boldsymbol{\Sigma}^r + \boldsymbol{\mu}^r(\boldsymbol{\mu}^r)^\top \right\rangle\right) \\
\text{s.t.} \quad & \begin{pmatrix} \boldsymbol{\Gamma}^r & \frac{1}{2}\boldsymbol{\beta}^r & \frac{1}{2}\mathrm{diag}(\boldsymbol{p}^r) \\ \frac{1}{2}(\boldsymbol{\beta}^r)^\top & \alpha^r & \frac{1}{2}(\boldsymbol{\delta}^r - \boldsymbol{p}^r \circ \boldsymbol{y}^r)^\top \\ \frac{1}{2}\mathrm{diag}(\boldsymbol{p}^r) & \frac{1}{2}(\boldsymbol{\delta}^r - \boldsymbol{p}^r \circ \boldsymbol{y}^r) & \boldsymbol{V}^r \end{pmatrix} \succeq 0, \ \forall r \in [R] \\
& \gamma + \sum_{r=1}^{R} \tau^r + \bar{v} = 0 \\
& (\boldsymbol{\delta}^r)^\top \boldsymbol{w}^r + (\boldsymbol{w}^r)^\top \boldsymbol{V}^r \boldsymbol{w}^r - \tau^r \le 0, \ \forall \boldsymbol{w}^r \in \{0,1\}^{n_r}, \ \forall r \in [R] \\
& \gamma \le 0
\end{aligned}
\tag{C-R}
$$

Compared with (B-R), which has one PSD constraint of dimension $(2n+1)\times(2n+1)$ and 2^n linear constraints, the correlation-splitting formulation (C-R) has R PSD constraints of dimensions $(2n_r+1)\times(2n_r+1)$ and $\sum_{r\in[R]} 2^{n_r}$ linear constraints. Thus, the computational complexity depends critically on the cluster sizes $\{n_r\}_{r\in[R]}$. In particular, keeping the cluster sizes sufficiently small can lead to substantial computational gains. The following lemma quantifies this gain using the cutting-plane method for SDPs in Jiang et al. (2020).

LEMMA 2. *(a) The complexity of solving* (B-R) *by the cutting-plane method in Jiang et al. (2020) is* $\widetilde{O}(8^n)$.

(b) Fix any $a>1$. *If the items are partitioned into* $R = \lceil n/\lfloor \log_a n \rfloor \rceil$ *clusters such that*

$$n_r = \lfloor \log_a n \rfloor, \ r = 1,\ldots,R-1, \ n_R = n - (R-1)\lfloor \log_a n \rfloor,$$

then the complexity of solving (C-R) *is* $\widetilde{O}(n^{1+3\log_a 2})$.

Lemma 2 shows that, while the full-information formulation has exponential complexity in n, the correlation-splitting formulation becomes polynomial-time solvable under a balanced partition with cluster size on the order of $\log_a n$. In particular, choosing $a=2$ yields complexity $\widetilde{O}(n^4)$. Thus, correlation splitting yields a polynomial-size approximation of the full-information formulation under suitable cluster-size choices while preserving the within-cluster dependence structure.

More generally, the choice of cluster size captures the trade-off between tractability and approximation quality. Larger clusters preserve more correlation information and therefore tend to yield tighter approximations, but at a higher computational cost. Smaller clusters yield more tractable formulations by omitting more cross-cluster covariance information. In the extreme case where $n_r = 1$ for all r, all cross-item covariance information is omitted. The next two subsections characterize this trade-off further by identifying when the approximation is exact and by deriving an upper bound on the optimality gap in general.

4.2. Exactness and Decomposition

We now investigate when the correlation-splitting approximation is exact. Our goal is to characterize conditions under which the lower bound from correlation splitting coincides with the full-information optimal value, that is, when $\Pi = \underline{\Pi}$.

We proceed in two steps. First, we study the structure of the correlation-splitting problem itself. In particular, we show that when each cluster contains a single item, the relaxed problem decomposes exactly into independent single-item subproblems, despite the non-additivity of CVaR. Second, we examine when this decomposed value also coincides with the full-information value of Problem (P).

Recall that the full-information optimal value of Problem (P) can be written as

$$\Pi = \max_{\boldsymbol{y},\bar{v}} \min_{F\in\mathscr{F}} \left\{ \bar{v} + (\boldsymbol{p}-\boldsymbol{c})^\top \boldsymbol{y} + \frac{1}{\eta}\mathbb{E}_F\left[\min\left(\boldsymbol{p}^\top(\tilde{\boldsymbol{d}}-\boldsymbol{y})^- - \bar{v}, 0\right)\right]\right\}.$$

Similarly, the optimal value of the correlation-splitting approximation Problem (P-C) can be written as

$$\begin{aligned}\underline{\Pi} &= \max_{\boldsymbol{y},\bar{v}} \min_{F\in\mathscr{F}^c} \left\{ \bar{v} + (\boldsymbol{p}-\boldsymbol{c})^\top \boldsymbol{y} + \frac{1}{\eta}\mathbb{E}_F\left[\min\left(\boldsymbol{p}^\top(\tilde{\boldsymbol{d}}-\boldsymbol{y})^- - \bar{v}, 0\right)\right]\right\} \\ &= \max_{\boldsymbol{y},\bar{v}^r: r\in[R]} \min_{F\in\mathscr{F}^c} \left\{ \sum_{r=1}^R \bar{v}^r + (\boldsymbol{p}-\boldsymbol{c})^\top \boldsymbol{y} + \frac{1}{\eta}\mathbb{E}_F\left[\min\left(\boldsymbol{p}^\top(\tilde{\boldsymbol{d}}-\boldsymbol{y})^- - \sum_{r=1}^R \bar{v}^r, 0\right)\right]\right\},\end{aligned}$$

where the second equality follows by decomposing the scalar variable $\bar{v}$ into cluster-level terms $\bar{v} = \sum_{r=1}^R \bar{v}^r$.

For each cluster r, let F_r denote the marginal distribution of $\tilde{\boldsymbol{d}}^r$. Define the corresponding cluster-level ambiguity set as

$$\mathscr{F}_r := \left\{ F_r \in \mathcal{M}(\mathbb{R}^{n_r}) \;\middle|\; \mathbb{E}_{F_r}[\tilde{\boldsymbol{d}}^r] = \boldsymbol{\mu}^r,\ \mathbb{E}_{F_r}\left[\tilde{\boldsymbol{d}}^r(\tilde{\boldsymbol{d}}^r)^\top\right] = \boldsymbol{\Sigma}^r + \boldsymbol{\mu}^r(\boldsymbol{\mu}^r)^\top \right\}.$$

The corresponding cluster subproblem is

$$\Pi_r := \max_{\boldsymbol{y}^r,\bar{v}^r} \min_{F_r\in\mathscr{F}_r} \left\{ \bar{v}^r + (\boldsymbol{p}^r-\boldsymbol{c}^r)^\top \boldsymbol{y}^r + \frac{1}{\eta}\mathbb{E}_{F_r}\left[\min\left((\boldsymbol{p}^r)^\top(\tilde{\boldsymbol{d}}^r-\boldsymbol{y}^r)^- - \bar{v}^r, 0\right)\right]\right\}. \tag{13}$$

For any fixed $\{\boldsymbol{y}^r, \bar{v}^r\}_{r\in[R]}$, we have

$$\min\left(\boldsymbol{p}^\top(\tilde{\boldsymbol{d}}-\boldsymbol{y})^- - \sum_{r=1}^R \bar{v}^r, 0\right) \geq \sum_{r=1}^R \min\left((\boldsymbol{p}^r)^\top(\tilde{\boldsymbol{d}}^r-\boldsymbol{y}^r)^- - \bar{v}^r, 0\right).$$

It follows that

$$\begin{aligned}\underline{\Pi} &\geq \max_{\boldsymbol{y},\bar{v}^r:r\in[R]} \min_{F\in\mathscr{F}^c} \left\{\sum_{r=1}^R \left(\bar{v}^r + (\boldsymbol{p}^r-\boldsymbol{c}^r)^\top\boldsymbol{y}^r + \frac{1}{\eta}\mathbb{E}_F\left[\min\left((\boldsymbol{p}^r)^\top(\tilde{\boldsymbol{d}}^r-\boldsymbol{y}^r)^- - \bar{v}^r, 0\right)\right]\right)\right\} \\ &= \sum_{r=1}^R \Pi_r.\end{aligned} \tag{14}$$

where the equality holds because the objective is separable across clusters and each term depends only on the marginal distribution of $\tilde{\boldsymbol{d}}^r$. Thus, the sum of the cluster subproblem values provides a natural benchmark for studying the correlation-splitting approximation.

The inequality in (14) does not imply equality in general. Unlike expectation, CVaR is not additive, because the tail behavior of a sum depends on the dependence structure across its components, not only on their marginal distributions. Even independence is generally insufficient to guarantee separability. The following result shows that, under the correlation-splitting relaxation, exact separability nevertheless holds when each cluster contains a single item.

Proposition 4. *Suppose $R = n$. Then the correlation-splitting approximation decomposes exactly:*

$$\underline{\Pi} = \sum_{i\in[n]} \Pi_i.$$

Proposition 4 shows that, when all clusters are singletons, Problem (P-C) is equivalent to solving n one-dimensional subproblems. The key mechanism is that the relaxation leaves the joint distribution across items unspecified. Hence, the “adversary” may choose a coupling of the marginal worst-case distributions that aligns adverse outcomes across items, so that the lower tail of the aggregate profit matches the sum of the marginal lower tails.

The exact decomposition in Proposition 4 does not, by itself, imply exactness relative to the full-information problem, that is, $\Pi = \underline{\Pi}$. In the full-information model, the covariance constraints restrict the feasible set of joint distributions, so the adversary can no longer choose an arbitrary coupling of the marginal distributions. The following proposition identifies a case in which, despite this restriction, the full-information problem still coincides with the decomposed value.

Proposition 5. *Suppose $R = n$ and all items are mutually uncorrelated. If*

$$\prod_{i\in[n]}\left(1-\frac{p_i-c_i}{p_i}\eta\right) + \eta - 1 \geq 0,$$

then the original problem is also separable:

$$\Pi = \underline{\Pi} = \sum_{i \in [n]} \Pi_i.$$

In particular, when $n = 2$, this condition is equivalent to

$$\eta \geq 1 - \frac{c_1 c_2}{(p_1 - c_1)(p_2 - c_2)}.$$

Proposition 5 highlights why exactness with respect to the full-information problem is nontrivial. Even when demands are mutually uncorrelated, the covariance constraints restrict the admissible couplings of the marginal worst-case distributions and limit the adversary's ability to synchronize adverse outcomes across items. In particular, when the marginal worst-case distributions are two-point distributions, the uncorrelated covariance condition imposes restrictions on their joint distribution. As a result, the aggregate lower tail under the full-information ambiguity set may differ from that under the correlation-splitting relaxation, leading to a positive gap $\Pi - \underline{\Pi}$.

The condition in Proposition 5 can be interpreted through the interplay between risk aversion and marginal profitability. When the risk-aversion level is moderate, in the sense that η is sufficiently large, the lower-tail region is less extreme, and the tail behavior of the aggregate profit under the admissible uncorrelated coupling matches the tail behavior under the worst-case coupling allowed by the correlation-splitting relaxation. In this regime, $\Pi = \underline{\Pi}$. In particular, $\eta = 1$ always satisfies the condition in Proposition 5, implying decomposability in the corresponding risk-neutral case under the uncorrelated setting. The condition is also easier to satisfy when marginal profits are small, i.e., when $p_i - c_i$ is close to zero. In this case, each item's downside contribution is less sensitive to the precise coordination of adverse outcomes across items, making exact decomposition more likely.

The proof of Proposition 5 constructs a joint distribution by coupling the marginal worst-case distributions obtained from the single-item subproblems. This construction relies on a closed-form characterization of the single-item CVaR-based DRO newsvendor problem, including an explicit worst-case two-point distribution. This single-item characterization may also be of independent interest for the literature on risk-averse and ambiguity-averse inventory management.

Proposition 5 identifies regimes in which the correlation-splitting approximation is exact. We next move beyond exactness and develop a general performance guarantee for the approximation.

4.3. General Optimality Gap Bound

We now turn to the general case, where the correlation-splitting approximation Problem (P-C) may introduce a nonzero gap relative to the full-information problem (P). Our goal is to derive an explicit upper bound on the optimality gap $\Pi - \underline{\Pi}$ and to characterize how omitting cross-cluster covariance information affects solution quality.

To facilitate the analysis, we first rewrite both formulations using square-root factorizations of the covariance matrices. This transformation simplifies the resulting SDPs and enables a direct comparison between the full-information and correlation-splitting models. Let $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2}$, where $\boldsymbol{\Sigma}^{1/2}$ denotes the unique positive square root of $\boldsymbol{\Sigma}$. Similarly, for each cluster $r \in [R]$, write $\boldsymbol{\Sigma}^r = (\boldsymbol{\Sigma}^r)^{1/2}(\boldsymbol{\Sigma}^r)^{1/2}$. We use $(\boldsymbol{\Sigma}^{1/2})^r$ to denote the principal block of the full square root $\boldsymbol{\Sigma}^{1/2}$ corresponding to cluster r, which is generally different from $(\boldsymbol{\Sigma}^r)^{1/2}$. Define the normalized demands $\tilde{\boldsymbol{d}}^D$ and $\tilde{\boldsymbol{d}}^D_c$ as

$$\tilde{\boldsymbol{d}}^D := \boldsymbol{\Sigma}^{-1/2}(\tilde{\boldsymbol{d}} - \boldsymbol{\mu}) \text{ and } (\tilde{\boldsymbol{d}}^D_c)^r := (\boldsymbol{\Sigma}^r)^{-1/2}(\tilde{\boldsymbol{d}}^r - \boldsymbol{\mu}^r),\ r \in [R].$$

This normalization yields equivalent factorized formulations of both Problem (B-I) and Problem (C-I), as stated below.

Proposition 6. *For any given* $(\boldsymbol{y}, \bar{v})$*, Problem* (B-I) *is equivalent to*

$$\begin{aligned}
\min_{x,\boldsymbol{t},\boldsymbol{Q},\boldsymbol{T}} \quad & \left\langle \operatorname{diag}(\boldsymbol{p}), \boldsymbol{Q}\boldsymbol{\Sigma}^{1/2} \right\rangle + \left[\boldsymbol{p} \circ (\boldsymbol{\mu} - \boldsymbol{y})\right]^\top \boldsymbol{t} - \bar{v}x \\
\text{s.t.} \quad & \begin{pmatrix} \boldsymbol{I}_n & \boldsymbol{0} & \boldsymbol{Q}^\top \\ \boldsymbol{0}^\top & 1 & \boldsymbol{t}^\top \\ \boldsymbol{Q} & \boldsymbol{t} & \boldsymbol{T} \end{pmatrix} \succeq 0 \\
& \begin{pmatrix} x & x & \boldsymbol{t}^\top \\ x & 1 & \boldsymbol{t}^\top \\ \boldsymbol{t} & \boldsymbol{t} & \boldsymbol{T} \end{pmatrix} \in \operatorname{CBQP}(n)
\end{aligned} \tag{B-F}$$

Moreover, for any given partition into R clusters, Problem (C-I) *is equivalent to*

$$\begin{aligned}
\Phi_{CF}(\boldsymbol{y}, \bar{v}) := \min_{x,\boldsymbol{t}^r,\boldsymbol{Q}^r,\boldsymbol{T}^r} \quad & \sum_{r=1}^{R} \Big(\left\langle \operatorname{diag}(\boldsymbol{p}^r), \boldsymbol{Q}^r(\boldsymbol{\Sigma}^r)^{1/2} \right\rangle + \left[\boldsymbol{p}^r \circ (\boldsymbol{\mu}^r - \boldsymbol{y}^r)\right]^\top \boldsymbol{t}^r \Big) - \bar{v}x \\
\text{s.t.} \quad & \begin{pmatrix} \boldsymbol{I}_{n_r} & \boldsymbol{0} & (\boldsymbol{Q}^r)^\top \\ \boldsymbol{0}^\top & 1 & (\boldsymbol{t}^r)^\top \\ \boldsymbol{Q}^r & \boldsymbol{t}^r & \boldsymbol{T}^r \end{pmatrix} \succeq 0,\ \forall r \in [R] \\
& \begin{pmatrix} x & x & (\boldsymbol{t}^r)^\top \\ x & 1 & (\boldsymbol{t}^r)^\top \\ \boldsymbol{t}^r & \boldsymbol{t}^r & \boldsymbol{T}^r \end{pmatrix} \in \operatorname{CBQP}(n_r),\ \forall r \in [R]
\end{aligned} \tag{C-F}$$

Recall from Corollary 1 that the correlation-splitting problem provides a lower bound on the full-information value. Since $\Pi_B = \Pi$ and $\Pi_C = \underline{\Pi}$, this gap can be written as $\Pi - \underline{\Pi}$. To assess the quality of the relaxation, we derive an upper bound on this gap by using the optimal solution of the factorized relaxation (C-F), evaluated at the optimal decision $(\boldsymbol{y}^*, \bar{v}^*)$ of Problem (C-R).

The following theorem provides an explicit upper bound on the optimality gap between Problem (P) and Problem (P-C). The bound depends on the cluster structure and on the optimal solution of the correlation-splitting relaxation.

THEOREM 3. *Suppose* $(\boldsymbol{y}_*, \bar{v}_*)$ *is an optimal solution to Problem* (C-R). *For Problem* (C-F) *evaluated at* $(\boldsymbol{y}, \bar{v}) = (\boldsymbol{y}_*, \bar{v}_*)$, *let* $\Phi_{CF}(\boldsymbol{y}_*, \bar{v}_*)$ *denote its optimal value, and let* $(\{\boldsymbol{T}_*^r, \boldsymbol{Q}_*^r, \boldsymbol{t}_*^r\}_{r\in[R]}, x_*)$ *be an associated optimal solution. Define* $\Delta_\Sigma^r = \left\langle \mathrm{diag}(\boldsymbol{p}^r), \boldsymbol{Q}_*^r (\boldsymbol{\Sigma}^{1/2})^r \right\rangle$, $\Delta_\mu^r = \left[\boldsymbol{p}^r \circ (\boldsymbol{\mu}^r - \boldsymbol{y}_*^r)\right]^\top \boldsymbol{t}_*^r$, *and* $\Delta^r = \Delta_\Sigma^r + \Delta_\mu^r$ *for all* $r \in [R]$. *Then*

$$0 \le \eta\big(\Pi - \underline{\Pi}\big) \le \min_{r\in[R]} G^r - \Phi_{CF}(\boldsymbol{y}_*, \bar{v}_*) - \bar{v}_* x_*, \tag{15}$$

where $G^r = \min\{\Delta^r, \Delta_\mu^r, 0\}$ *for all* $r \in [R]$.

A key feature of the bound (15) is that the effect of omitting cross-cluster covariance information is captured through cluster-level terms. Each cluster contributes a term G^r, which selects the most favorable value among Δ^r, Δ_μ^r, and zero. The term $\min_{r\in[R]} G^r$ then identifies the cluster that yields the tightest member of this family of bounds. Full derivations are provided in Appendix EC.1.12.

The proof of Theorem 3 is based on constructing a parametric family of paired feasible solutions that connects the full-information model and the correlation-splitting relaxation. Rather than comparing the original inner minimization problems (B-I) and (C-I) directly, Proposition 6 allows us to work with the factorized counterparts (B-F) and (C-F). Starting from the optimal solution $(\{\boldsymbol{T}_*^r, \boldsymbol{Q}_*^r, \boldsymbol{t}_*^r\}_{r\in[R]}, x_*)$ of Problem (C-F), we construct feasible solutions for the full-information model (B-F). The construction uses parameters that interpolate among several feasible solution mappings, enabling a direct comparison of objective values. Optimizing over these parameters yields the upper bound in Theorem 3. In particular, the bound vanishes when $R = 1$, because no cross-cluster covariance is dropped and the relaxation coincides with the full-information model.

The closed-form expression for G^r reveals three possible construction regimes. When $G^r = \Delta^r$, the construction uses $\boldsymbol{T}_*^r$, $\boldsymbol{Q}_*^r$, and $\boldsymbol{t}_*^r$, so both covariance-related and mean-related components contribute to tightening the bound. When $G^r = \Delta_\mu^r$, the construction retains $\boldsymbol{T}_*^r$ and $\boldsymbol{t}_*^r$ but excludes $\boldsymbol{Q}_*^r$, so only the mean-related component is active. When $G^r = 0$, neither component improves the bound, and the construction effectively discards them. Thus, the bound automatically selects the most favorable construction according to the problem data.

Because a smaller value of G^r leads to a tighter upper bound, it is useful to identify when G^r becomes strictly negative. In particular, if $\Delta_\mu^r < 0$, then $G^r < 0$. The term $\Delta_\mu^r = [\boldsymbol{p}^r \circ (\boldsymbol{\mu}^r - \boldsymbol{y}_*^r)]^\top \boldsymbol{t}_*^r$ can be interpreted as a tail-weighted safety-stock effect within cluster r. When $\Delta_\mu^r < 0$, the relaxed solution places weight on states where order quantities exceed mean demand, indicating overage exposure in the tail. In such cases, the corresponding cluster can sharpen the gap bound.

Theorem 3 provides an upper bound on the optimality gap under arbitrary covariance structures. The bound quantifies the worst-case loss from dropping cross-cluster covariance information, but it does not by itself identify specific problem regimes in which the approximation becomes tight. This naturally raises the question of how the optimality gap behaves as the problem structure varies. We next study this question through the lens of variance heterogeneity across clusters.

5. Variance-Based Splitting Rule

Section 4.3 establishes a general upper bound on the optimality gap of the correlation-splitting approximation. We now use this bound to motivate a practical clustering rule for constructing item partitions. In particular, we show that variance heterogeneity across clusters plays a central role in determining the quality of the approximation.

From an implementation perspective, correlation splitting involves two design choices: the cluster sizes and the assignment of items to clusters. The former is primarily driven by tractability considerations and is linked to the polynomial complexity bound in Lemma 2. For a fixed size specification, however, the assignment determines which dependence relationships are retained within clusters and which are left unspecified across clusters. This assignment can therefore have a substantial impact on approximation quality.

While one could form clusters directly from the estimated covariance or correlation matrix, doing so may be unreliable in sparse-data regimes, precisely where moment-based DRO is most relevant. We therefore seek a simpler assignment rule based on marginal variances, which are typically easier to estimate than high-dimensional correlation patterns. The key question is whether such a rule can still preserve good approximation quality. The next result answers this question in the two-cluster case: when one cluster contains only items with sufficiently small demand variability, both the optimality gap and its theoretical upper bound vanish linearly as that variability decreases. This result clarifies when omitting cross-cluster covariance information is relatively harmless and motivates a practical variance-based splitting rule. The following theorem formalizes this result.

THEOREM 4. *Suppose $R=2$ and, for some cluster $\tilde{r}\in\{1,2\}$, the marginal variances satisfy $\Sigma_{ii}\le\sigma^2$ for all $i\in N_{\tilde{r}}$. Then:*

(a) $\Pi-\underline{\Pi}=O(\sigma)$. *That is, the optimality gap vanishes at rate $O(\sigma)$ as $\sigma\to 0$.*

(b) $\min_{r\in[R]} G^r-\Phi_{CF}(\boldsymbol{y}_*,\bar{v}_*)-\bar{v}_*x_*=O(\sigma)$. *Hence, the upper bound in* (15) *also vanishes at rate $O(\sigma)$.*

Theorem 4 shows that the correlation-splitting approximation becomes asymptotically tight when one cluster contains only low-variability items. In this regime, both the optimality gap and the bound in (15) vanish at rate $O(\sigma)$, regardless of the cross-cluster correlation pattern. The intuition is decision-driven: low-variability items have demands concentrated around their means, so their optimal order quantities are close to mean demand and their contribution to the tail loss $\boldsymbol{p}^\top(\tilde{\boldsymbol{d}}-\boldsymbol{y})^-$ is small and stable. Moreover, when the variances in one cluster are small, the associated cross-cluster covariance terms are also small in magnitude. As a result, omitting these cross-cluster covariances has only a limited effect on the worst-case CVaR and the induced ordering decision. Although the result focuses on two clusters, the same intuition suggests that in multi-cluster settings, low-variability clusters should have a limited impact on the approximation quality.

Theorem 4 motivates a practical *variance-based splitting rule* for cluster assignment. After fixing the cluster sizes, we sort items in descending order of marginal variance and assign them to clusters sequentially according to this ordering. This rule gives priority to high-variance items, which are more likely to drive tail risk and therefore benefit more from retaining within-cluster dependence information. It requires only marginal variance estimates and avoids relying on the full covariance matrix for cluster construction.

The detailed implementation of the variance-based splitting rule is outlined in Algorithm 1. Section 6.2 provides numerical evidence that this rule captures a substantial portion of the performance gains of correlation splitting and becomes increasingly effective as the problem size grows.

Algorithm 1 Correlation-Splitting Approximation with Variance-Based Splitting Rule

Input: Number of clusters R, cluster sizes $\{n_r\}_{r\in[R]}$, covariance matrix $\boldsymbol{\Sigma}$, mean vector $\boldsymbol{\mu}$.
Output: Inventory level $\boldsymbol{y}_*$.

1: Sort the marginal variances $\mathrm{diag}(\boldsymbol{\Sigma})$ in descending order, and let $\boldsymbol{B}$ be the corresponding permutation matrix.
2: **for all** $r \in [R]$ **do**
3: Let $\boldsymbol{B}^r$ be the submatrix consisting of rows $\sum_{\bar{r}=1}^{r-1} n_{\bar{r}} + 1$ to $\sum_{\bar{r}=1}^{r} n_{\bar{r}}$ of $\boldsymbol{B}$.
4: Set $\boldsymbol{\Sigma}^r = \boldsymbol{B}^r \boldsymbol{\Sigma} (\boldsymbol{B}^r)^\top$ and $\boldsymbol{\mu}^r = \boldsymbol{B}^r \boldsymbol{\mu}$.
5: Solve Problem (C-R) with $\{\boldsymbol{\Sigma}^r\}_{r\in[R]}$ and $\{\boldsymbol{\mu}^r\}_{r\in[R]}$ to obtain $\bar{\boldsymbol{y}}_*$.
6: **return** $\boldsymbol{y}_* = \boldsymbol{B}^\top \bar{\boldsymbol{y}}_*$.

6. Numerical Experiments

In this section, we conduct numerical experiments to evaluate the proposed lifting-and-splitting framework. Recall that our approach consists of two components: an exact CBQP-based lifting reformulation and a correlation-splitting approximation, together with a variance-based splitting rule. The experiments are designed to assess three main aspects of the framework: (i) whether the CBQP-based lifting reformulation improves the tractability of exact solution methods; (ii) whether the correlation-splitting approximation achieves strong solution quality with substantial computational gains; and (iii) whether the variance-based splitting rule provides an effective and practical clustering strategy.

We first evaluate the computational efficiency of the exact CBQP-based lifting reformulation in Section 6.1. We then study the performance of the correlation-splitting approximation and the variance-based splitting rule in Section 6.2. Section 6.3 compares the proposed method with the quadratic decision rule (QDR) benchmark of Hanasusanto et al. (2015) and shows that correlation splitting achieves substantially better solution quality while being significantly faster. Finally, Section 6.4 demonstrates the scalability of the proposed approach on large-scale instances.

We use the experimental setup in Natarajan and Teo (2017). In the main experiments, we set the tail probability level to $\eta = 0.95$; additional robustness results for other values of η are provided in Appendix EC.4. For each item $i \in [n]$, the selling price is $p_i = 10$, and the unit cost c_i is randomly drawn from $[3, 8]$. The mean μ_i and standard deviation σ_i of demand $\tilde{d}_i$ are randomly generated from $[1, 10]$ and $[0.1\mu_i, \mu_i]$, respectively. To generate the covariance matrix, we first generate a random matrix $\boldsymbol{U} \in \mathbb{R}^{n\times n}$ with entries independently drawn from the uniform distribution on $[0, 1]$. Letting $\boldsymbol{S} = \boldsymbol{U}^\top \boldsymbol{U}$ and $\boldsymbol{s} = 1/\sqrt{\mathrm{diag}(\boldsymbol{S})}$, we construct the correlation matrix as $\boldsymbol{C} = \mathrm{diag}(\boldsymbol{s})\boldsymbol{S}\mathrm{diag}(\boldsymbol{s})$, and set the covariance matrix to $\boldsymbol{\Sigma} = \mathrm{diag}(\boldsymbol{\sigma})\boldsymbol{C}\mathrm{diag}(\boldsymbol{\sigma})$. We vary the number of items n to test problem instances of different sizes. For each value of n, we generate 50 independent instances with randomly sampled parameters. All experiments are implemented in MATLAB R2023b using the YALMIP interface. The optimization problems are solved using MOSEK 10.1.11 and Gurobi 10.0.2 on a 64-bit Windows system with an AMD Ryzen 9 7940H CPU at 4.00 GHz and 16 GB of RAM.

6.1. Tractability of the Exact CBQP-Based Lifting Reformulation

We first evaluate the computational performance of different exact solution approaches. The methods considered are:

- *PD*: The conic-duality SDP formulation (D), which involves exponentially many PSD constraints and follows the standard reformulation approach used in Hanasusanto et al. (2015).
- *PB*: The proposed CBQP-based lifting reformulation (B-R), which replaces the exponentially many PSD constraints with exponentially many linear constraints.
- *PB-CG*: The constraint-generation algorithm in Algorithm 2 of Appendix EC.3, applied to the CBQP-based lifting reformulation.

The first two methods solve their corresponding exact formulations directly, while *PB-CG* solves the CBQP-based formulation by iteratively adding violated linear constraints. Table 1 summarizes the average computation times over 50 independently generated instances for each value of n. The symbol "–" indicates that the average computation time exceeded the 100-second time limit.

The results show that *PB* is substantially more efficient than *PD* for solving the exact problem. For example, when $n = 14$, *PB* reduces the average computation time from 79.71 seconds to 0.72 seconds, a speedup of more than 100 times relative to *PD*. This improvement is consistent with the theoretical development in Section 3, where the CBQP-based lifting reformulation shifts the main exponential complexity from PSD constraints to linear constraints.

The constraint-generation version *PB-CG* further demonstrates the computational advantage of the CBQP-based lifting reformulation. Because *PB* represents the exponential component through

Table 1 Average Computation Times for Exact Solution Methods over 50 Instances

Size (n)	*PD* Time (s)	*PB* Time (s)	*PB-CG* Time (s)
7	0.27	0.02	1.77
8	0.63	0.03	2.47
9	1.43	0.05	3.12
10	3.36	0.08	4.72
11	8.27	0.14	6.28
12	13.92	0.14	8.43
13	29.81	0.46	10.67
14	79.71	0.72	13.95
15	–	1.65	15.10
16	–	4.16	18.18
17	–	8.54	18.20
18	–	25.21	24.03
19	–	63.28	29.29
20	–	–	36.11
21	–	–	58.16
22	–	–	72.59
23	–	–	88.16

linear constraints, it enables a practical constraint-generation strategy: after solving a restricted SDP, violated linear constraints are identified by solving a binary quadratic optimization problem and then added iteratively. As reported in Table 1, this strategy substantially reduces computation time for larger instances and allows exact solution of problems beyond the range of direct solution methods. In particular, when $n \geq 18$, *PB-CG* becomes faster than directly solving *PB*, and it solves instances up to $n = 23$ within the 100-second time limit on average.

6.2. Performance of the Correlation-Splitting Approximation

The results in Section 6.1 show that the CBQP-based lifting reformulation substantially improves the tractability of exact solution methods. Nevertheless, its computational cost still grows quickly with the number of items, making an exact solution impractical for larger instances. We therefore evaluate the correlation-splitting approximation, with a particular focus on the variance-based splitting rule developed in Section 5.

Following the cluster-size construction in Lemma 2, we determine the cluster sizes using the parameter a. Specifically, we set $R = \lceil n/\lfloor \log_a n \rfloor \rceil$, construct $R-1$ clusters of size $\lfloor \log_a n \rfloor$, and place the remaining items in the last cluster. We consider two values, $a = 2$ and $a = 1.5$. A smaller value of a leads to larger clusters, which retain more within-cluster correlation information and typically improve approximation quality, but at the cost of higher computation time. We first compare different splitting rules in Section 6.2.1. We then examine the resulting quality–efficiency trade-off in Section 6.2.2.

6.2.1. Benefit of the Variance-Based Splitting Rule. As discussed in Section 5, once the cluster sizes are fixed, a key remaining design choice is how to assign items to clusters. We compare the variance-based splitting rule with a random splitting benchmark. For the random benchmark, items are randomly assigned to clusters with the prescribed cluster sizes. We denote the correlation-splitting methods using this random splitting rule by *CS-R-2* and *CS-R-1.5* for $a = 2$ and $a = 1.5$, respectively. The variance-based splitting rule, summarized in Algorithm 1, sorts items in descending order of marginal variance and assigns them sequentially to clusters according to the prescribed cluster sizes. This rule is motivated by Theorem 4, which shows that variance heterogeneity across clusters plays an important role in determining the optimality gap. We denote the correlation-splitting methods using the variance-based splitting rule by *CS-V-2* and *CS-V-1.5* for $a = 2$ and $a = 1.5$, respectively.

To quantify the incremental benefit of variance-based clustering, for each problem instance and each value of a, we define the *Variance Rule Improvement* (VRI) as

$$\mathrm{VRI} = \frac{\text{Value of } \textit{CS-V-a} - \mathbb{E}\left[\text{Value of } \textit{CS-R-a}\right]}{\mathbb{E}\left[\text{Value of } \textit{CS-R-a}\right]} \times 100\%,$$

where the expectation is taken over the random splitting rule, conditional on the given problem instance and the prescribed cluster sizes determined by a. In implementation, we estimate this expectation using repeated random partitions for each instance. We start with at least 30 random partitions and construct a 95% confidence interval for the induced VRI. Sampling continues until this confidence interval lies entirely above or below zero, so that the two methods can be statistically distinguished, or until its width is less than 0.1%. The resulting sample mean of the random-splitting objective values is then used to compute VRI. A positive VRI indicates that the variance-based splitting rule yields a higher objective value than random splitting.

Table 2 reports summary statistics of VRI across the 50 independently generated instances for each problem size. The table includes the 10th percentile, mean, and 90th percentile of VRI, as well as the fraction of instances in which the variance-based rule strictly outperforms the random splitting benchmark. The results show that the variance-based splitting rule improves solution quality with high probability across all tested problem sizes and for both choices of a.

We first consider the case $a = 2$. The mean VRI is consistently positive across all problem sizes, indicating that variance-based splitting improves approximation quality on average. Although the 10th percentile VRI is slightly negative for some smaller instances, its magnitude is limited, suggesting that the downside risk of using the variance-based rule is small. As the problem size increases, the distribution of VRI also becomes more concentrated: the gap between the 10th and 90th percentiles narrows, while the mean remains relatively stable. This suggests that the benefit of variance-based splitting becomes more robust as the problem dimension grows. In addition, *CS-V-2* strictly outperforms *CS-R-2* in at least 76% of the instances for all tested problem sizes, and this probability exceeds 90% for larger instances.

For $a = 1.5$, similar patterns are observed. The mean VRI remains positive across all problem sizes, and the probability that *CS-V-1.5* outperforms *CS-R-1.5* is consistently high. The improvements under $a = 1.5$ are generally larger than those under $a = 2$, suggesting that the variance-based rule becomes more effective when larger clusters are used and more within-cluster correlation information is retained. In particular, when $n = 100$, *CS-V-1.5* outperforms *CS-R-1.5* in all generated instances.

Overall, these results show that the variance-based splitting rule provides a simple and effective enhancement to the correlation-splitting framework. It translates the theoretical insight on variance heterogeneity into measurable performance gains, while requiring only marginal variance information for cluster assignment.

Table 2 Improvement from the Variance-Based Splitting Rule

	Size (n)							
	30	40	50	60	70	80	90	100
CS-V-2 *vs.* ***CS-R-2***								
VRI – 10th percentile	-0.18	-0.12	-0.03	-0.14	-0.06	-0.01	0.11	0.07
VRI – mean	0.25	0.20	0.32	0.28	0.28	0.30	0.30	0.32
VRI – 90th percentile	0.73	0.57	0.66	0.66	0.60	0.62	0.55	0.58
Prob(*CS-V-2* > *CS-R-2*)	0.76	0.78	0.88	0.76	0.84	0.90	0.92	0.96
CS-V-1.5 *vs.* ***CS-R-1.5***								
VRI – 10th percentile	-0.23	-0.01	0.05	-0.02	-0.06	0.02	0.13	0.09
VRI – mean	0.37	0.34	0.30	0.32	0.26	0.35	0.36	0.36
VRI – 90th percentile	0.87	0.69	0.56	0.57	0.54	0.64	0.55	0.61
Prob(*CS-V-1.5* > *CS-R-1.5*)	0.82	0.88	0.96	0.90	0.86	0.92	0.98	1.00

6.2.2. Trade-off for Varying Cluster Sizes. We next examine the trade-off between solution quality and computational efficiency induced by different cluster sizes. Specifically, we compare *CS-V-2* and *CS-V-1.5*, which use the same variance-based splitting rule but different cluster-size specifications. Since $a = 1.5$ produces larger clusters than $a = 2$, *CS-V-1.5* retains more within-cluster correlation information, while *CS-V-2* uses smaller clusters and is computationally faster.

Table 3 summarizes the results. Generally, *CS-V-1.5* achieves slightly higher objective values than *CS-V-2*, reflecting the benefit of retaining more correlation information within clusters. The improvement, however, is moderate. For example, the mean objective values differ only by a small percentage across all tested values of n.

The computational difference is more pronounced. *CS-V-2* solves all tested instances with up to 100 items in about 0.1 seconds on average, whereas *CS-V-1.5* requires noticeably longer runtimes as n increases. Thus, increasing the cluster size improves approximation quality, but the incremental gain comes at a non-negligible computational cost. The percentile patterns are also similar across the two methods, suggesting that larger clusters mainly shift the objective level upward rather than changing the variability of performance across instances.

Taken together, these results illustrate the trade-off between solution quality and computational efficiency in choosing cluster sizes. In large-scale settings where speed is critical, smaller clusters such as those used in *CS-V-2* can offer an attractive balance, while larger clusters such as those used in *CS-V-1.5* may be preferable when higher approximation quality is desired and additional computation time is acceptable.

Table 3 Comparison between CS-V-1.5 and CS-V-2

	Size (n)							
	30	40	50	60	70	80	90	100
CS-V-2								
Objective Value – 10th percentile	2152	2962	3689	4241	5123	6257	7039	7618
Objective Value – mean	2851	3953	4772	5530	6358	7532	8142	9077
Objective Value – 90th percentile	3732	4870	5982	6820	7359	8630	9289	10929
Time (s)	0.02	0.03	0.06	0.08	0.09	0.10	0.10	0.12
CS-V-1.5								
Objective Value – 10th percentile	2162	2984	3712	4268	5148	6298	7083	7659
Objective Value – mean	2864	3973	4798	5558	6392	7574	8185	9116
Objective Value – 90th percentile	3742	4883	6014	6843	7386	8666	9347	10903
Time (s)	0.07	0.11	0.25	0.30	0.50	0.67	1.00	1.06

6.3. Benchmark Comparison with QDR

We next benchmark the correlation-splitting method *CS-V-2* against the *QDR* approximation of Hanasusanto et al. (2015). The comparison focuses on both approximation quality and computational efficiency on medium-scale instances.

The *QDR* approach provides a tractable conservative approximation by introducing auxiliary variables and semidefinite constraints. While *QDR* is scalable relative to exact conic-duality reformulations, it does not provide explicit optimality-gap guarantees relative to the original problem. For completeness, the *QDR* formulation used in our experiments is provided in Appendix EC.2.

For a given problem instance, we define the *Correlation-Splitting Improvement* (CSI) as

$$\mathrm{CSI} = \frac{\text{Value of } \textit{CS-V-2} - \text{Value of } \textit{QDR}}{\text{Value of } \textit{QDR}} \times 100\%.$$

This metric measures the relative improvement in objective value achieved by *CS-V-2* over *QDR*. Since both methods provide lower bounds on the original problem value, a positive CSI indicates that *CS-V-2* yields a tighter lower bound than *QDR* for the corresponding instance.

Table 4 reports the CSI statistics across the generated instances, including the 10th percentile, mean, and 90th percentile, together with the probability that *CS-V-2* outperforms *QDR* and the average computation times of both methods. The results show that *CS-V-2* consistently achieves higher objective values than *QDR* across all tested problem sizes. In particular, the improvement probability is 100% for every value of n, and the mean CSI remains between 1.51% and 1.91%.

The computational advantage of *CS-V-2* is also substantial. While both methods are tractable at these problem sizes, the runtime of QDR grows much faster with n. For example, when $n = 140$, *CS-V-2* solves the problem in 0.27 seconds on average, whereas QDR requires 65.66 seconds on average. These results suggest that correlation splitting provides a tighter approximation than the QDR benchmark while also delivering significantly lower computation times, making it attractive for larger-scale applications.

Table 4 Comparison in Solution Quality and Computation Time between CS-V-2 and QDR

	Size (n)											
	30	40	50	60	70	80	90	100	110	120	130	140
CS-V-2* vs. *QDR												
CSI – 10th percentile	0.76	0.82	0.96	1.23	1.08	1.15	1.17	1.33	1.29	1.28	1.27	1.34
CSI – mean	1.62	1.51	1.73	1.83	1.76	1.71	1.67	1.91	1.82	1.80	1.62	1.80
CSI – 90th percentile	2.69	2.34	2.63	2.65	2.45	2.33	2.17	2.64	2.47	2.38	2.01	2.21
Prob(*CS-V-2* > *QDR*)	1	1	1	1	1	1	1	1	1	1	1	1
CS-V-2												
Time (s)	0.04	0.04	0.05	0.05	0.06	0.08	0.10	0.14	0.16	0.19	0.22	0.27
QDR												
Time (s)	0.15	0.24	0.45	1.04	2.12	3.53	6.90	11.26	18.04	28.46	44.65	65.66

6.4. Scalability to Large-Scale Instances

We further evaluate the scalability of the proposed approach by testing *CS-V-2* on large-scale instances. This experiment examines whether the computational advantages observed for medium-scale instances in Table 4 persist when the number of items increases substantially. Table 5 reports the average computation times for problem sizes ranging from $n = 200$ to 4000. The results show that *CS-V-2* remains computationally tractable even for instances with thousands of items.

To characterize the empirical growth rate of the computation time, we fit a polynomial relationship of the form $T(n) \approx Cn^p$, where $T(n)$ denotes runtime and $C > 0$ and $p > 0$ are constants.

Taking logarithms gives $\log T(n) = \log C + p \log n$, so the exponent p can be estimated from the slope of a log–log regression of computation time on problem size.

Figure 1 presents the log–log plot for *CS-V-2*, together with a least-squares fit. The approximately linear pattern on the log–log scale suggests polynomial growth in computation time, with an empirical rate close to quadratic. Taken together, Table 5 and Figure 1 demonstrate the scalability of the correlation-splitting approximation and its ability to handle large-scale instances efficiently.

Table 5 **Average Computation Times for Large-Scale Instances Using CS-V-2**

	Size (n)						
	200	400	700	1000	2000	3000	4000
CS-V-2							
Time (s)	0.23	1.13	1.88	4.30	20.37	59.83	77.67

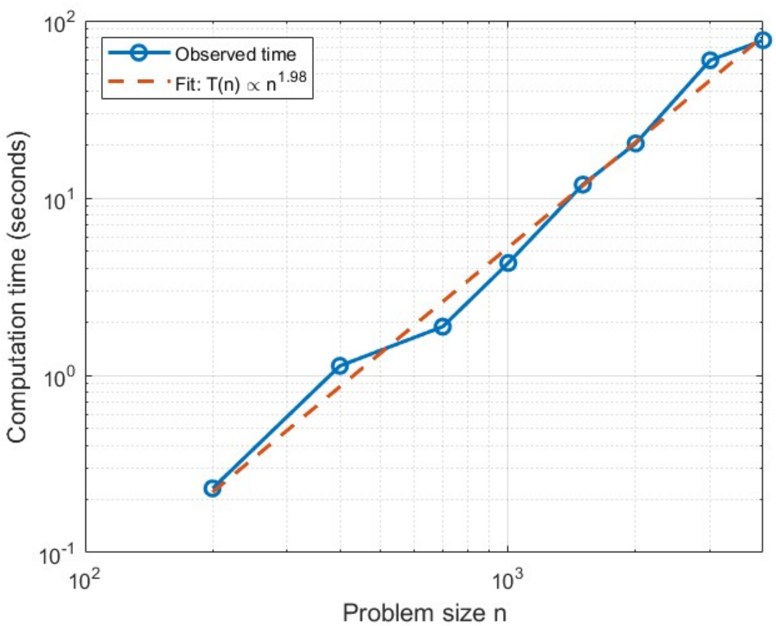

Figure 1 **Log–log plot of computation time versus problem size for CS-V-2. The dashed line represents the least-squares fit on the log–log scale.**

7. Conclusions

This paper studies a mean–covariance distributionally robust multi-item newsvendor problem with a CVaR objective. To address the computational difficulty of the resulting NP-hard problem, we develop a lifting-and-splitting framework that combines exact reformulation, scalable approximation, and performance guarantees.

The first component of the framework is an exact CBQP-based lifting reformulation. This reformulation converts the exponentially many PSD constraints in the standard conic-duality SDP into exponentially many linear constraints, thereby substantially improving the tractability of exact solution methods. The second component is a correlation-splitting approximation, which retains within-cluster covariance information while omitting cross-cluster covariance information. This approximation yields a polynomial-time solvable lower bound under suitable cluster-size choices. We further establish exactness conditions and derive an explicit upper bound on the optimality gap, providing a theoretical basis for evaluating the quality of the approximation.

The analysis also identifies variance heterogeneity as an important driver of approximation quality. In the two-cluster case, when one cluster contains only low-variability items, both the optimality gap and its theoretical upper bound vanish linearly as the variability of that cluster decreases. This insight motivates a simple variance-based splitting rule that relies only on marginal variance information. Numerical experiments show that the CBQP-based lifting reformulation improves exact solution times, the correlation-splitting approximation achieves strong solution quality with substantial computational gains, and the variance-based splitting rule provides an effective clustering strategy for large-scale instances.

Overall, the proposed framework offers a tractable and performance-guaranteed approach for risk-averse, correlation-aware inventory decisions under moment ambiguity. More broadly, it suggests that high-dimensional dependence in DRO can be managed by combining exact lifting reformulations with structured approximations that retain the most relevant dependence information.

Several directions for future research remain. First, the CBQP-based lifting approach may be extended to other practically relevant settings, such as models with nonnegative demand support, which are closely related to copositive programming (Natarajan et al. 2011b), or to other classes of distributionally robust inventory and pricing problems. Second, while this paper highlights the role of variance heterogeneity in correlation splitting, other structural features may also affect approximation quality. Identifying such features could lead to more refined clustering strategies and improved approximation methods.

References

Ahipasaoglu SD, Natarajan K, Shi D (2019) Distributionally robust project crashing with partial or no correlation information. *Networks* 74(1):79–106.

Ardestani-Jaafari A, Delage E (2016) Robust optimization of sums of piecewise linear functions with application to inventory problems. *Operations Research* 64(2):474–494.

Artzner P, Delbaen F, Eber JM, Heath D (1999) Coherent measures of risk. *Mathematical Finance* 9(3):203–228.

Carrasco V, Luz VF, Kos N, Messner M, Monteiro P, Moreira H (2018) Optimal selling mechanisms under moment conditions. *Journal of Economic Theory* 177:245–279.

Chen H, Hu M, Perakis G (2022) Distribution-free pricing. *Manufacturing & Service Operations Management* 24(4):1939–1958.

Chen Y, Xu M, Zhang ZG (2009) A risk-averse newsvendor model under the CVaR criterion. *Operations Research* 57(4):1040–1044.

Chen Z, Hu Z, Wang R (2024) Screening with limited information: A dual perspective. *Operations Research* 72(4):1487–1504.

Chen Z, Xie W (2021) Regret in the newsvendor model with demand and yield randomness. *Production and Operations Management* 30(11):4176–4197.

Cheramin M, Cheng J, Jiang R, Pan K (2022) Computationally efficient approximations for distributionally robust optimization under moment and Wasserstein ambiguity. *INFORMS Journal on Computing* 34(3):1768–1794.

Choi S, Ruszczyński A (2011) A multi-product risk-averse newsvendor with exponential utility function. *European Journal of Operational Research* 214(1):78–84.

Cui Z, Long DZ, Qi J, Zhang L (2023) The inventory routing problem under uncertainty. *Operations Research* 71(1):378–395.

Das B, Dhara A, Natarajan K (2021) On the heavy-tail behavior of the distributionally robust newsvendor. *Operations Research* 69(4):1077–1099.

Delage E, Ye Y (2010) Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research* 58(3):595–612.

Dhaene J, Denuit M, Goovaerts MJ, Kaas R, Vyncke D (2002) The concept of comonotonicity in actuarial science and finance: theory. *Insurance: Mathematics and Economics* 31(1):3–33.

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

Gao SY, Simchi-Levi D, Teo CP, Yan Z (2019) Disruption risk mitigation in supply chains: The risk exposure index revisited. *Operations Research* 67(3):831–852.

Han Q, Du D, Zuluaga LF (2014) A risk- and ambiguity-averse extension of the max-min newsvendor order formula. *Operations Research* 62(3):535–542.

Hanasusanto GA, Kuhn D, Wallace SW, Zymler S (2015) Distributionally robust multi-item newsvendor problems with multimodal demand distributions. *Mathematical Programming* 152(1):1–32.

Hanasusanto GA, Kuhn D, Wiesemann W (2016) A comment on "computational complexity of stochastic programming problems". *Mathematical Programming* 159(1):557–569.

Iancu DA, Sharma M, Sviridenko M (2013) Supermodularity and affine policies in dynamic robust optimization. *Operations Research* 61(4):941–956.

Isii K (1962) On sharpness of tchebycheff-type inequalities. *Annals of the Institute of Statistical Mathematics* 14(1):185–197.

Jiang H, Lee YT, Song Z, Wong SCW (2020) An improved cutting plane method for convex optimization, convex-concave games, and its applications. *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, 944–953.

Jiang S, Cheng J, Pan K, Shen ZJM (2026) Optimized dimensionality reduction for moment-based distributionally robust optimization. *Operations Research* 74(3):1648–1669.

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

Long DZ, Qi J, Zhang A (2024) Supermodularity in two-stage distributionally robust optimization. *Management Science* 70(3):1394–1409.

Natarajan K, Sim M, Teo CP (2011a) Beyond risk: Ambiguity in supply chains. *The Handbook of Integrated Risk Management in Global Supply Chains* 103–124.

Natarajan K, Sim M, Uichanco J (2018) Asymmetry and ambiguity in newsvendor models. *Management Science* 64(7):3146–3167.

Natarajan K, Teo CP (2017) On reduced semidefinite programs for second order moment bounds with applications. *Mathematical Programming* 161(1-2):487–518.

Natarajan K, Teo CP, Zheng Z (2011b) Mixed 0-1 linear programs under objective uncertainty: A completely positive representation. *Operations Research* 59(3):713–728.

Padmanabhan D, Natarajan K, Murthy K (2021) Exploiting partial correlations in distributionally robust optimization. *Mathematical Programming* 186(1):209–255.

Perakis G, Roels G (2008) Regret in the newsvendor model with partial information. *Operations Research* 56(1):188–203.

Rockafellar RT, Uryasev S (2000) Optimization of conditional value-at-risk. *Journal of Risk* 2:21–42.

Scarf HE (1957) *A min-max solution of an inventory problem* (Rand Corporation Santa Monica).

Schweitzer ME, Cachon GP (2000) Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Science* 46(3):404–420.

Shapiro A, Kleywegt A (2002) Minimax analysis of stochastic problems. *Optimization Methods and Software* 17(3):523–542.

Sinha A, Namkoong H, Duchi J (2018) Certifying some distributional robustness with principled adversarial training. *International Conference on Learning Representations.*

Wang S, Delage E (2024) A column generation scheme for distributionally robust multi-item newsvendor problems. *INFORMS Journal on Computing* 36(3):849–867.

Wihler TP (2009) On the Hölder continuity of matrix functions for normal matrices. *Journal of Inequalities in Pure and Applied Mathematics* 10(10).

Xu G, Hanasusanto GA (2025) Improved decision rule approximations for multistage robust optimization via copositive programming. *Operations Research* 73(2):842–861.

Yang C, Hu Z, Zhou SX (2021) Multilocation newsvendor problem: Centralization and inventory pooling. *Management Science* 67(1):185–200.

Yue J, Chen B, Wang MC (2006) Expected value of distribution information for the newsvendor problem. *Operations Research* 54(6):1128–1136.

Zhang X, Tang Q, Chen Z, Chen L (2026) Managing inventory and pricing with contextual robust optimization, available at SSRN 6768258.

EC.1. Technical Proofs

EC.1.1. Proof of Proposition 1

The constraint in (4) can be decomposed into

$$s_1 + \boldsymbol{s}_2^\top \boldsymbol{d} + \left\langle \boldsymbol{s}_3, \boldsymbol{d}\boldsymbol{d}^\top \right\rangle \le 0,\ \forall \boldsymbol{d} \in \mathbb{R}^n,\ \text{and}$$

$$s_1 + \boldsymbol{s}_2^\top \boldsymbol{d} + \left\langle \boldsymbol{s}_3, \boldsymbol{d}\boldsymbol{d}^\top \right\rangle \le \boldsymbol{p}^\top (\boldsymbol{d} - \boldsymbol{y})^- - \bar{v},\ \forall \boldsymbol{d} \in \mathbb{R}^n. \tag{EC.1}$$

In the following, we show that these two semi-infinite quadratic constraints are equivalent to positive semidefinite (PSD) constraints. We prove the equivalence for the second constraint; the first follows from the same argument by replacing the right-hand side constant with zero.

Since for any realization $\boldsymbol{d}$, $\boldsymbol{p}^\top (\boldsymbol{d} - \boldsymbol{y})^- - \bar{v} = \min_{\boldsymbol{x} \in \{0,1\}^n} \sum_{i=1}^n p_i (d_i - y_i) x_i - \bar{v}$, (EC.1) is equivalent to

$$s_1 + \boldsymbol{s}_2^\top \boldsymbol{d} + \left\langle \boldsymbol{s}_3, \boldsymbol{d}\boldsymbol{d}^\top \right\rangle \le \sum_{i=1}^n p_i (d_i - y_i) x_i - \bar{v},\ \forall \boldsymbol{d} \in \mathbb{R}^n,\ \forall \boldsymbol{x} \in \{0,1\}^n.$$

Equivalently,

$$\begin{pmatrix} 1 & \boldsymbol{d}^\top \end{pmatrix} \begin{pmatrix} s_1 + (\boldsymbol{p} \circ \boldsymbol{x})^\top \boldsymbol{y} + \bar{v} & \frac{1}{2}(\boldsymbol{s}_2 - \boldsymbol{p} \circ \boldsymbol{x})^\top \\ \frac{1}{2}(\boldsymbol{s}_2 - \boldsymbol{p} \circ \boldsymbol{x}) & \boldsymbol{s}_3 \end{pmatrix} \begin{pmatrix} 1 \\ \boldsymbol{d} \end{pmatrix} \le 0,\ \forall \boldsymbol{d} \in \mathbb{R}^n,\ \forall \boldsymbol{x} \in \{0,1\}^n. \tag{EC.2}$$

We now show that

$$\text{(EC.2)} \iff \begin{pmatrix} -s_1 - (\boldsymbol{p} \circ \boldsymbol{x})^\top \boldsymbol{y} - \bar{v} & \frac{1}{2}(-\boldsymbol{s}_2 + \boldsymbol{p} \circ \boldsymbol{x})^\top \\ \frac{1}{2}(-\boldsymbol{s}_2 + \boldsymbol{p} \circ \boldsymbol{x}) & -\boldsymbol{s}_3 \end{pmatrix} \succeq 0,\ \forall \boldsymbol{x} \in \{0,1\}^n. \tag{EC.3}$$

The direction "$\Leftarrow$" follows directly from the definition of a PSD matrix. For the direction "$\Rightarrow$", fix any $\boldsymbol{x} \in \{0,1\}^n$ and let $\boldsymbol{M}$ denote the matrix in (EC.3). We want to prove $\boldsymbol{M} \succeq 0$, which is equivalent to $(\lambda_0, \boldsymbol{\lambda})^\top \boldsymbol{M} (\lambda_0, \boldsymbol{\lambda}) \ge 0$ for all $(\lambda_0, \boldsymbol{\lambda}) \in \mathbb{R}^{n+1}$. To this end, we consider two cases: $\lambda_0 \neq 0$ and $\lambda_0 = 0$.

If $\lambda_0 \neq 0$, by letting $\boldsymbol{d} = \boldsymbol{\lambda}/\lambda_0$, (EC.2) implies $(\lambda_0, \boldsymbol{\lambda})^\top \boldsymbol{M} (\lambda_0, \boldsymbol{\lambda}) = \lambda_0^2 (1, \boldsymbol{d})^\top \boldsymbol{M} (1, \boldsymbol{d}) \ge 0$.

It remains to consider $\lambda_0 = 0$. Suppose, for contradiction, that there exists $\boldsymbol{\lambda} \neq \boldsymbol{0}$ such that

$$(0, \boldsymbol{\lambda})^\top \boldsymbol{M} (0, \boldsymbol{\lambda}) = \boldsymbol{\lambda}^\top (-\boldsymbol{s}_3) \boldsymbol{\lambda} < 0. \tag{EC.4}$$

Define $\alpha := \boldsymbol{\lambda}^\top (-\boldsymbol{s}_3) \boldsymbol{\lambda} < 0$, $\beta := (-\boldsymbol{s}_2 + \boldsymbol{p} \circ \boldsymbol{x})^\top \boldsymbol{\lambda}$, and $\gamma := -s_1 - (\boldsymbol{p} \circ \boldsymbol{x})^\top \boldsymbol{y} - \bar{v}$. Then, for any $\epsilon > 0$, $(\epsilon, \boldsymbol{\lambda})^\top \boldsymbol{M} (\epsilon, \boldsymbol{\lambda}) = \epsilon^2 \gamma + \epsilon \beta + \alpha$. Since $\alpha < 0$, we can choose $\epsilon > 0$ sufficiently small such that $\epsilon^2 |\gamma| + \epsilon |\beta| < -\alpha/2$, which implies $(\epsilon, \boldsymbol{\lambda})^\top \boldsymbol{M} (\epsilon, \boldsymbol{\lambda}) \le \epsilon^2 |\gamma| + \epsilon |\beta| + \alpha < \alpha/2 < 0$.

This contradicts the nonnegativity already established for $(\lambda_0, \boldsymbol{\lambda})^\top \boldsymbol{M} (\lambda_0, \boldsymbol{\lambda})$ for all $\lambda_0 \neq 0$. Therefore, (EC.4) cannot occur, and $(\lambda_0, \boldsymbol{\lambda})^\top \boldsymbol{M} (\lambda_0, \boldsymbol{\lambda}) \ge 0$ also holds for $\lambda_0 = 0$ and all $\boldsymbol{\lambda} \in \mathbb{R}^n$.

Combining the two cases yields $\boldsymbol{M} \succeq 0$. Hence, (EC.2) is equivalent to (EC.3). Applying the same argument to the first constraint gives $\begin{pmatrix} -s_1 & -\frac{1}{2}\boldsymbol{s}_2^\top \\ -\frac{1}{2}\boldsymbol{s}_2 & -\boldsymbol{s}_3 \end{pmatrix} \succeq 0$. Combining this PSD constraint with (EC.3) for all $\boldsymbol{x} \in \{0,1\}^n$ gives the semidefinite programming (SDP) reformulation in Proposition 1.

EC.1.2. Proof of Proposition 2

Fix any $(\boldsymbol{y},\bar{v})$. For brevity, let Φ_B denote the optimal value $\Phi_B(\boldsymbol{y},\bar{v})$ of Problem (B-I). It is sufficient to show the equivalence between Problem (B-I) and the inner minimization problem

$$\Phi_n = \min_{F\in\mathscr{F}} \mathbb{E}_F\left[\min\left(\boldsymbol{p}^\top(\tilde{\boldsymbol{d}}-\boldsymbol{y})^- - \bar{v},0\right)\right] = \min_{F\in\mathscr{F}} \mathbb{E}_F\left[\left(\boldsymbol{p}\circ(\tilde{\boldsymbol{d}}-\boldsymbol{y})\right)^\top \boldsymbol{t}_*(\tilde{\boldsymbol{d}}) - \bar{v}x_*(\tilde{\boldsymbol{d}})\right], \qquad \text{(EC.5)}$$

where the second equality follows from (8). Because the pointwise minimization in (7) is over a finite set, the minimizers $x_*(\tilde{\boldsymbol{d}})$ and $\boldsymbol{t}_*(\tilde{\boldsymbol{d}})$ can be selected measurably.

Step 1: To show $\Phi_n \geq \Phi_B$.

Consider any random vector $\tilde{\boldsymbol{d}}$ with distribution $\Lambda_{\tilde{\boldsymbol{d}}} \in \mathscr{F}$, i.e., $\mathbb{E}_{\Lambda_{\tilde{\boldsymbol{d}}}}[\tilde{\boldsymbol{d}}] = \boldsymbol{\mu}$ and $\mathbb{E}_{\Lambda_{\tilde{\boldsymbol{d}}}}[\tilde{\boldsymbol{d}}\tilde{\boldsymbol{d}}^\top] = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\top$. By introducing the decision variables defined in (9) with respect to distribution $\Lambda_{\tilde{\boldsymbol{d}}}$, the following expectation matrix for $\tilde{\boldsymbol{d}}$ and $\boldsymbol{t}_*(\tilde{\boldsymbol{d}})$ is PSD by construction:

$$\mathbb{E}_{\Lambda_{\tilde{\boldsymbol{d}}}}\left[\begin{pmatrix}\tilde{\boldsymbol{d}}\\ 1\\ \boldsymbol{t}_*(\tilde{\boldsymbol{d}})\end{pmatrix}\begin{pmatrix}\tilde{\boldsymbol{d}}\\ 1\\ \boldsymbol{t}_*(\tilde{\boldsymbol{d}})\end{pmatrix}^\top\right] = \mathbb{E}_{\Lambda_{\tilde{\boldsymbol{d}}}}\left[\begin{pmatrix}\tilde{\boldsymbol{d}}\tilde{\boldsymbol{d}}^\top & \tilde{\boldsymbol{d}} & \tilde{\boldsymbol{d}}\boldsymbol{t}_*(\tilde{\boldsymbol{d}})^\top\\ \tilde{\boldsymbol{d}}^\top & 1 & \boldsymbol{t}_*(\tilde{\boldsymbol{d}})^\top\\ \boldsymbol{t}_*(\tilde{\boldsymbol{d}})\tilde{\boldsymbol{d}}^\top & \boldsymbol{t}_*(\tilde{\boldsymbol{d}}) & \boldsymbol{t}_*(\tilde{\boldsymbol{d}})\boldsymbol{t}_*(\tilde{\boldsymbol{d}})^\top\end{pmatrix}\right] \succeq 0.$$

Hence,

$$\begin{pmatrix}\boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^\top & \boldsymbol{\mu} & \boldsymbol{Q}^\top\\ \boldsymbol{\mu}^\top & 1 & \boldsymbol{t}^\top\\ \boldsymbol{Q} & \boldsymbol{t} & \boldsymbol{T}\end{pmatrix} = \mathbb{E}_{\Lambda_{\tilde{\boldsymbol{d}}}}\left[\begin{pmatrix}\tilde{\boldsymbol{d}}\\ 1\\ \boldsymbol{t}_*(\tilde{\boldsymbol{d}})\end{pmatrix}\begin{pmatrix}\tilde{\boldsymbol{d}}\\ 1\\ \boldsymbol{t}_*(\tilde{\boldsymbol{d}})\end{pmatrix}^\top\right] \succeq 0.$$

On the other hand, the feasibility of the binary random variables $(x_*(\tilde{\boldsymbol{d}}),\boldsymbol{t}_*(\tilde{\boldsymbol{d}}))$ implies that

$$\begin{pmatrix}x_*(\tilde{\boldsymbol{d}})\\ 1\\ \boldsymbol{t}_*(\tilde{\boldsymbol{d}})\end{pmatrix}\begin{pmatrix}x_*(\tilde{\boldsymbol{d}})\\ 1\\ \boldsymbol{t}_*(\tilde{\boldsymbol{d}})\end{pmatrix}^\top \in \left\{\begin{pmatrix}w_0\\ 1\\ \boldsymbol{w}\end{pmatrix}\begin{pmatrix}w_0\\ 1\\ \boldsymbol{w}\end{pmatrix}^\top \,\middle|\, (w_0,\boldsymbol{w})\in\Upsilon\right\}.$$

Taking expectations gives a convex combination of such rank-one matrices. Therefore, we have the following.

$$\begin{pmatrix}x & x & \boldsymbol{t}^\top\\ x & 1 & \boldsymbol{t}^\top\\ \boldsymbol{t} & \boldsymbol{t} & \boldsymbol{T}\end{pmatrix} = \mathbb{E}_{\Lambda_{\tilde{\boldsymbol{d}}}}\left[\begin{pmatrix}x_*(\tilde{\boldsymbol{d}})\\ 1\\ \boldsymbol{t}_*(\tilde{\boldsymbol{d}})\end{pmatrix}\begin{pmatrix}x_*(\tilde{\boldsymbol{d}})\\ 1\\ \boldsymbol{t}_*(\tilde{\boldsymbol{d}})\end{pmatrix}^\top\right] \in \mathrm{CBQP}(n).$$

Lastly, from the decision variables defined in (9), it follows that

$$\mathbb{E}_{\Lambda_{\tilde{\boldsymbol{d}}}}\left[\left(\boldsymbol{p}\circ(\tilde{\boldsymbol{d}}-\boldsymbol{y})\right)^\top \boldsymbol{t}_*(\tilde{\boldsymbol{d}}) - \bar{v}x_*(\tilde{\boldsymbol{d}})\right] = \langle \mathrm{diag}(\boldsymbol{p}),\boldsymbol{Q}\rangle - (\boldsymbol{p}\circ\boldsymbol{y})^\top\boldsymbol{t} - \bar{v}x.$$

Thus, any feasible distribution of Problem (EC.5) induces a feasible solution of Problem (B-I) with the same objective value for the given $(\boldsymbol{y},\bar{v})$. Since both problems are minimization problems, we conclude that $\Phi_n \geq \Phi_B$.

Step 2: To show $\Phi_n \leq \Phi_B$.

Consider an optimal solution $(x_*, \boldsymbol{t}_*, \boldsymbol{Q}_*, \boldsymbol{T}_*)$ to Problem (B-I) for the given $(\boldsymbol{y}, \bar{v})$. Then

$$\begin{pmatrix} \boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^\top & \boldsymbol{\mu} & \boldsymbol{Q}_*^\top \\ \boldsymbol{\mu}^\top & 1 & \boldsymbol{t}_*^\top \\ \boldsymbol{Q}_* & \boldsymbol{t}_* & \boldsymbol{T}_* \end{pmatrix} \succeq 0, \text{ and } \begin{pmatrix} x_* & x_* & \boldsymbol{t}_*^\top \\ x_* & 1 & \boldsymbol{t}_*^\top \\ \boldsymbol{t}_* & \boldsymbol{t}_* & \boldsymbol{T}_* \end{pmatrix} \in \mathrm{CBQP}(n).$$

We will construct a feasible distribution for Problem (EC.5) whose objective value is no larger than the objective value of this lifted solution. Throughout this step, we continue to focus on the inner minimization problem. By the definition of the constrained Boolean quadric polytope (CBQP), the set CBQP(n) is a polytope. By the symmetry of the matrices and Carathéodory's theorem, there exists a subset $\Omega = \{(w_0^1, \boldsymbol{w}^1), \ldots, (w_0^K, \boldsymbol{w}^K)\} \subseteq \Upsilon$ containing K points, with $K \le (n^2+3n)/2+2$, such that

$$\begin{pmatrix} x_* & x_* & \boldsymbol{t}_*^\top \\ x_* & 1 & \boldsymbol{t}_*^\top \\ \boldsymbol{t}_* & \boldsymbol{t}_* & \boldsymbol{T}_* \end{pmatrix} = \sum_{k\in[K]} \lambda_k \begin{pmatrix} w_0^k \\ 1 \\ \boldsymbol{w}^k \end{pmatrix} \begin{pmatrix} w_0^k \\ 1 \\ \boldsymbol{w}^k \end{pmatrix}^\top,$$

where $\lambda_k > 0$ and $\sum_{k\in[K]} \lambda_k = 1$. In particular, considering only $\boldsymbol{t}_*$ and $\boldsymbol{T}_*$ gives

$$\begin{pmatrix} 1 & \boldsymbol{t}_*^\top \\ \boldsymbol{t}_* & \boldsymbol{T}_* \end{pmatrix} = \sum_{k\in[K]} \lambda_k \begin{pmatrix} 1 \\ \boldsymbol{w}^k \end{pmatrix} \begin{pmatrix} 1 \\ \boldsymbol{w}^k \end{pmatrix}^\top. \tag{EC.6}$$

We next use the following decomposition lemma.

Lemma EC.1. *(Natarajan and Teo 2017, Theorem 1) Assume that the following two conditions hold:*

(a) The matrix $\boldsymbol{M}$ is a $(n_1+n_2)\times(n_1+n_2)$ PSD block matrix of the form $\boldsymbol{M} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B}^\top \\ \boldsymbol{B} & \boldsymbol{C} \end{pmatrix} \succeq 0$, where $\boldsymbol{A} \in \mathcal{S}^{n_1}$, $\boldsymbol{C} \in \mathcal{S}^{n_2}$, and $\mathcal{S}^n$ denotes the set of symmetric matrices of size $n \times n$.

(b) The matrix $\boldsymbol{C}$ lies in $\mathcal{C}(\mathcal{K})$ with an explicit factorization $\boldsymbol{C} = \boldsymbol{V}\boldsymbol{V}^\top$, where each column vector of $\boldsymbol{V}$ lies in the closed convex cone $\mathcal{K}$, and

$$\mathcal{C}(\mathcal{K}) = \left\{ \boldsymbol{X} \in \mathcal{S}^n : \exists \boldsymbol{v}_1, \ldots, \boldsymbol{v}_p \in \mathcal{K} \text{ such that } \boldsymbol{X} = \sum_{k\in[p]} \boldsymbol{v}_k \boldsymbol{v}_k^\top \right\}.$$

Then $\boldsymbol{M}$ lies in the generalized completely positive cone $\mathcal{C}(\mathbb{R}^{n_1} \times \mathcal{K})$ with factorization

$$\boldsymbol{M} = \begin{pmatrix} \boldsymbol{B}^\top (\boldsymbol{V}^\dagger)^\top \\ \boldsymbol{V} \end{pmatrix} \begin{pmatrix} \boldsymbol{B}^\top (\boldsymbol{V}^\dagger)^\top \\ \boldsymbol{V} \end{pmatrix}^\top + \begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{0} \end{pmatrix}^\top,$$

where $\boldsymbol{U}$ is defined such that $\boldsymbol{A} - \boldsymbol{B}^\top \boldsymbol{C}^\dagger \boldsymbol{B} = \boldsymbol{U}\boldsymbol{U}^\top \succeq 0$, and $\boldsymbol{C}^\dagger$ is the Moore–Penrose pseudoinverse of $\boldsymbol{C}$.

Define $\boldsymbol{A} = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\top$, $\boldsymbol{B} = \begin{pmatrix} \boldsymbol{\mu}^\top \\ \boldsymbol{Q}_* \end{pmatrix}$, and $\boldsymbol{C} = \begin{pmatrix} 1 & \boldsymbol{t}_*^\top \\ \boldsymbol{t}_* & \boldsymbol{T}_* \end{pmatrix}$, and $\boldsymbol{V} = \begin{pmatrix} \sqrt{\lambda_1} & \cdots & \sqrt{\lambda_K} \\ \sqrt{\lambda_1}\boldsymbol{w}^1 & \cdots & \sqrt{\lambda_K}\boldsymbol{w}^K \end{pmatrix}$. By (EC.6), we have $\boldsymbol{C} = \boldsymbol{V}\boldsymbol{V}^\top$. For each $k \in [K]$, let $\boldsymbol{d}^k \in \mathbb{R}^n$ be the kth column of $\boldsymbol{B}^\top (\boldsymbol{V}^\dagger)^\top / \sqrt{\lambda_k}$. Applying Lemma EC.1, we can decompose the matrix as

$$\begin{pmatrix} \boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^\top & \boldsymbol{\mu} & \boldsymbol{Q}_*^\top \\ \boldsymbol{\mu}^\top & 1 & \boldsymbol{t}_*^\top \\ \boldsymbol{Q}_* & \boldsymbol{t}_* & \boldsymbol{T}_* \end{pmatrix} = \sum_{k\in[K]} \lambda_k \begin{pmatrix} \boldsymbol{d}^k \\ 1 \\ \boldsymbol{w}^k \end{pmatrix} \begin{pmatrix} \boldsymbol{d}^k \\ 1 \\ \boldsymbol{w}^k \end{pmatrix}^\top + \begin{pmatrix} \boldsymbol{\Delta} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0}^\top & 0 & \boldsymbol{0}^\top \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}, \tag{EC.7}$$

where $\boldsymbol{\Delta} := \boldsymbol{A} - \boldsymbol{B}^\top \boldsymbol{C}^\dagger \boldsymbol{B} \succeq 0$.

We now construct a probability distribution θ_* for the random demand $\tilde{\boldsymbol{d}}$. For each $k \in [K]$, let $\tilde{\boldsymbol{\xi}}_k \sim N(\boldsymbol{d}^k, \boldsymbol{\Delta})$ be a multivariate normal random vector with mean $\boldsymbol{d}^k$ and covariance matrix $\boldsymbol{\Delta}$. Since $\boldsymbol{\Delta} \succeq 0$, this distribution is well defined, and its second moment matrix is $\boldsymbol{d}^k(\boldsymbol{d}^k)^\top + \boldsymbol{\Delta}$.

Let θ_* be the mixture distribution that selects component k with probability (w.p.) λ_k and, conditional on selecting component k, draws $\tilde{\boldsymbol{d}}$ according to the distribution of $\tilde{\boldsymbol{\xi}}_k$. The mixture distribution is well defined because $\lambda_k > 0$ for all $k \in [K]$ and $\sum_{k\in[K]} \lambda_k = 1$.

We can verify that θ_* is feasible for Problem (EC.5), i.e., $\theta_* \in \mathscr{F}$:

$$\mathbb{E}_{\theta_*}\left[\begin{pmatrix}\tilde{\boldsymbol{d}}\\1\end{pmatrix}\begin{pmatrix}\tilde{\boldsymbol{d}}\\1\end{pmatrix}^\top\right] = \sum_{k\in[K]} \lambda_k \begin{pmatrix}\boldsymbol{d}^k(\boldsymbol{d}^k)^\top + \boldsymbol{\Delta} & \boldsymbol{d}^k\\ (\boldsymbol{d}^k)^\top & 1\end{pmatrix} = \begin{pmatrix}\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\top & \boldsymbol{\mu}\\ \boldsymbol{\mu}^\top & 1\end{pmatrix}.$$

On the other hand,

$$\begin{aligned}
&\min_{F\in\mathscr{F}} \mathbb{E}_F\left[\min\left(\boldsymbol{p}^\top(\tilde{\boldsymbol{d}}-\boldsymbol{y})^- - \bar{v}, 0\right)\right]\\
\text{(By (7))} &= \min_{F\in\mathscr{F}} \mathbb{E}_F\left[\min_{(x,\boldsymbol{t})\in\Upsilon}\left(\boldsymbol{p}\circ(\tilde{\boldsymbol{d}}-\boldsymbol{y})\right)^\top \boldsymbol{t} - \bar{v}x\right]\\
\text{(Since } \theta_*\in\mathscr{F}) &\le \mathbb{E}_{\theta_*}\left[\min_{(x,\boldsymbol{t})\in\Upsilon}(\boldsymbol{p}\circ\tilde{\boldsymbol{d}})^\top\boldsymbol{t} - (\boldsymbol{p}\circ\boldsymbol{y})^\top\boldsymbol{t} - \bar{v}x\right]\\
\text{(Since every point in } \Omega \text{ belongs to } \Upsilon) &\le \mathbb{E}_{\theta_*}\left[\min_{(x,\boldsymbol{t})\in\Omega}(\boldsymbol{p}\circ\tilde{\boldsymbol{d}})^\top\boldsymbol{t} - (\boldsymbol{p}\circ\boldsymbol{y})^\top\boldsymbol{t} - \bar{v}x\right]\\
\text{(By the definition of } \theta_*) &= \sum_{k\in[K]}\lambda_k\mathbb{E}\left[\min_{(x,\boldsymbol{t})\in\Omega}(\boldsymbol{p}\circ\tilde{\boldsymbol{\xi}}_k)^\top\boldsymbol{t} - (\boldsymbol{p}\circ\boldsymbol{y})^\top\boldsymbol{t} - \bar{v}x\right]\\
\text{(Setting } (x,\boldsymbol{t}) = (w_0^k,\boldsymbol{w}^k)) &\le \sum_{k\in[K]}\lambda_k\mathbb{E}\left[(\boldsymbol{p}\circ\tilde{\boldsymbol{\xi}}_k)^\top\boldsymbol{w}^k - (\boldsymbol{p}\circ\boldsymbol{y})^\top\boldsymbol{w}^k - \bar{v}w_0^k\right]\\
\text{(By the definition of } \tilde{\boldsymbol{\xi}}_k) &= \sum_{k\in[K]}\lambda_k\left[(\boldsymbol{p}\circ\boldsymbol{d}^k)^\top\boldsymbol{w}^k - (\boldsymbol{p}\circ\boldsymbol{y})^\top\boldsymbol{w}^k - \bar{v}w_0^k\right]\\
\text{(By (EC.7))} &= \langle \operatorname{diag}(\boldsymbol{p}), \boldsymbol{Q}_*\rangle - (\boldsymbol{p}\circ\boldsymbol{y})^\top\boldsymbol{t}_* - \bar{v}x_*.
\end{aligned}$$

The last expression is the objective value of Problem (B-I) at the optimal solution $(x_*, \boldsymbol{t}_*, \boldsymbol{Q}_*, \boldsymbol{T}_*)$ for the given $(\boldsymbol{y}, \bar{v})$. Hence, the constructed distribution yields an objective value no larger than Φ_B, and therefore $\Phi_n \le \Phi_B$.

Combining Steps 1 and 2, we obtain $\Phi_n = \Phi_B$ for any given $(\boldsymbol{y}, \bar{v})$. Therefore, Problem (B-I) is equivalent to the inner minimization problem (EC.5), and consequently Proposition 2 follows.

EC.1.3. Proof of Lemma 1

By the definition of CBQP(n), the CBQP constraint in Problem (B-I) is equivalent to requiring the existence of nonnegative weights $\{\lambda_{(w_0,\boldsymbol{w})}\}_{(w_0,\boldsymbol{w})\in\Upsilon}$ such that

$$\begin{cases} x = \sum_{(w_0,\boldsymbol{w})\in\Upsilon} \lambda_{(w_0,\boldsymbol{w})} w_0 \\ \boldsymbol{t} = \sum_{(w_0,\boldsymbol{w})\in\Upsilon} \lambda_{(w_0,\boldsymbol{w})} \boldsymbol{w} \\ \boldsymbol{T} = \sum_{(w_0,\boldsymbol{w})\in\Upsilon} \lambda_{(w_0,\boldsymbol{w})} \boldsymbol{w}\boldsymbol{w}^\top \\ \sum_{(w_0,\boldsymbol{w})\in\Upsilon} \lambda_{(w_0,\boldsymbol{w})} = 1 \\ \lambda_{(w_0,\boldsymbol{w})} \geq 0, \ \forall (w_0,\boldsymbol{w}) \in \Upsilon \end{cases}$$

With a slight abuse of notation, we use $\lambda_{(w_0,\boldsymbol{w})}$ to denote the weight associated with the point $(w_0,\boldsymbol{w}) \in \Upsilon$. Since $\boldsymbol{w} \leq w_0 \boldsymbol{e}_n$ and $w_0 \in \{0,1\}$, when $w_0 = 0$ we must have $\boldsymbol{w} = \boldsymbol{0}$.

Only the weights $\lambda_{(w_0,\boldsymbol{w})}$ with $w_0 = 1$ affect the values of x, $\boldsymbol{t}$, and $\boldsymbol{T}$. Define the reduced weights $\eta_{\boldsymbol{w}} = \lambda_{(1,\boldsymbol{w})}$ for all $\boldsymbol{w} \in \{0,1\}^n$. Then $x = \sum_{(1,\boldsymbol{w})\in\Upsilon} \lambda_{(1,\boldsymbol{w})} = \sum_{\boldsymbol{w}\in\{0,1\}^n} \eta_{\boldsymbol{w}}$, and the normalization constraint becomes $\sum_{\boldsymbol{w}\in\{0,1\}^n} \eta_{\boldsymbol{w}} + \lambda_{(0,\boldsymbol{0})} = 1$. Hence $x \leq 1$. Moreover, $\boldsymbol{t} = \sum_{\boldsymbol{w}\in\{0,1\}^n} \eta_{\boldsymbol{w}} \boldsymbol{w}$ and $\boldsymbol{T} = \sum_{\boldsymbol{w}\in\{0,1\}^n} \eta_{\boldsymbol{w}} \boldsymbol{w}\boldsymbol{w}^\top$. Thus, the CBQP constraint implies the linear system

$$\begin{cases} \boldsymbol{t} = \sum_{\boldsymbol{w}\in\{0,1\}^n} \eta_{\boldsymbol{w}} \boldsymbol{w} \\ \boldsymbol{T} = \sum_{\boldsymbol{w}\in\{0,1\}^n} \eta_{\boldsymbol{w}} \boldsymbol{w}\boldsymbol{w}^\top \\ \sum_{\boldsymbol{w}\in\{0,1\}^n} \eta_{\boldsymbol{w}} = x \\ x \leq 1 \\ \eta_{\boldsymbol{w}} \geq 0, \ \forall \boldsymbol{w} \in \{0,1\}^n \end{cases}$$

Conversely, suppose there exist weights $\{\eta_{\boldsymbol{w}}\}_{\boldsymbol{w}\in\{0,1\}^n}$ satisfying the above linear system. Define $\lambda_{(1,\boldsymbol{w})} = \eta_{\boldsymbol{w}}$ for all $\boldsymbol{w} \in \{0,1\}^n$, and $\lambda_{(0,\boldsymbol{0})} = 1 - x$. Because $\eta_{\boldsymbol{w}} \geq 0$ and $x \leq 1$, all these weights are nonnegative. Moreover, $\sum_{\boldsymbol{w}\in\{0,1\}^n} \lambda_{(1,\boldsymbol{w})} + \lambda_{(0,\boldsymbol{0})} = x + (1-x) = 1$. Substituting these weights into the convex-hull representation of CBQP(n) yields the CBQP constraint in Problem (B-I). Therefore, the CBQP constraint is equivalent to the stated linear system. This proves the lemma.

EC.1.4. Proof of Theorem 1

By substituting the linear representation of CBQP(n) from Lemma 1 into Problem (B-I) and taking the conic dual of the resulting SDP, we obtain the following joint maximization problem:

$$\begin{aligned} \max_{\substack{\boldsymbol{y},\bar{v},\alpha,\tau,\gamma,\boldsymbol{\beta},\boldsymbol{\kappa},\boldsymbol{\delta},\\ \boldsymbol{\Gamma},\boldsymbol{K},\boldsymbol{V}}} \quad & \bar{v} + (\boldsymbol{p}-\boldsymbol{c})^\top \boldsymbol{y} - \frac{1}{\eta}\left(\alpha - \gamma + \boldsymbol{\mu}^\top\boldsymbol{\beta} + \left\langle \boldsymbol{\Gamma}, \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\top \right\rangle\right) \\ \text{s.t.} \quad & \begin{pmatrix} \boldsymbol{\Gamma} & \frac{1}{2}\boldsymbol{\beta} & \frac{1}{2}\operatorname{diag}(\boldsymbol{p}) \\ \frac{1}{2}\boldsymbol{\beta}^\top & \alpha & \frac{1}{2}\boldsymbol{\kappa}^\top \\ \frac{1}{2}\operatorname{diag}(\boldsymbol{p}) & \frac{1}{2}\boldsymbol{\kappa} & \boldsymbol{K} \end{pmatrix} \succeq 0 \\ & \gamma + \tau = -\bar{v} \\ & \boldsymbol{\kappa} = \boldsymbol{\delta} - \boldsymbol{p} \circ \boldsymbol{y} \\ & \boldsymbol{K} = \boldsymbol{V} \\ & \tau \geq \boldsymbol{\delta}^\top \boldsymbol{w} + \boldsymbol{w}^\top \boldsymbol{V} \boldsymbol{w}, \ \forall \boldsymbol{w} \in \{0,1\}^n \\ & \gamma \leq 0 \end{aligned}$$

Eliminating the auxiliary variables $\boldsymbol{\kappa}$ and $\boldsymbol{K}$ gives the SDP in (B-R). Moreover, under the assumption $\boldsymbol{\Sigma} \succ 0$, one can verify that Slater's condition holds for both the primal SDP and its dual. Hence, there is no duality gap. By Proposition 2, Problem (B-R) is equivalent to Problem (P).

EC.1.5. Proof of Proposition 3

The inner minimization problem for the correlation-splitting approximation Problem (P-C) is

$$\begin{aligned}\overline{\Phi} &= \min_{F\in\mathscr{F}^c} \mathbb{E}_F\left[\min\left(\boldsymbol{p}^\top(\tilde{\boldsymbol{d}}-\boldsymbol{y})^- - \bar{v}, 0\right)\right] \\ &= \min_{\substack{\mathbb{E}_F[\tilde{\boldsymbol{d}}]=\boldsymbol{\mu},\\ \mathbb{E}_F[\tilde{\boldsymbol{d}}^r(\tilde{\boldsymbol{d}}^r)^\top]=\boldsymbol{\Sigma}^r+\boldsymbol{\mu}^r(\boldsymbol{\mu}^r)^\top,\ \forall r\in[R]}} \mathbb{E}_F\left[\min\left(\boldsymbol{p}^\top(\tilde{\boldsymbol{d}}-\boldsymbol{y})^- - \bar{v}, 0\right)\right] \\ &= \min_{\overline{\boldsymbol{\Sigma}}:\overline{\boldsymbol{\Sigma}}^r=\boldsymbol{\Sigma}^r,\ \forall r\in[R]} \ \min_{\substack{\mathbb{E}_F[\tilde{\boldsymbol{d}}]=\boldsymbol{\mu},\\ \mathbb{E}_F[\tilde{\boldsymbol{d}}\tilde{\boldsymbol{d}}^\top]=\overline{\boldsymbol{\Sigma}}+\boldsymbol{\mu}\boldsymbol{\mu}^\top}} \mathbb{E}_F\left[\min\left(\boldsymbol{p}^\top(\tilde{\boldsymbol{d}}-\boldsymbol{y})^- - \bar{v}, 0\right)\right]. \end{aligned} \tag{EC.8}$$

By Proposition 2, for any feasible $\overline{\boldsymbol{\Sigma}}$ such that $\overline{\boldsymbol{\Sigma}}^r = \boldsymbol{\Sigma}^r$ for all $r\in[R]$, the inner minimization problem is equivalent to the following SDP:

$$\begin{aligned} \min_{x,\boldsymbol{t},\boldsymbol{Q},\boldsymbol{T}} \quad & \langle \mathrm{diag}(\boldsymbol{p}), \boldsymbol{Q}\rangle - (\boldsymbol{p}\circ\boldsymbol{y})^\top \boldsymbol{t} - \bar{v}x \\ \text{s.t.} \quad & \begin{pmatrix} \overline{\boldsymbol{\Sigma}}+\boldsymbol{\mu}\boldsymbol{\mu}^\top & \boldsymbol{\mu} & \boldsymbol{Q}^\top \\ \boldsymbol{\mu}^\top & 1 & \boldsymbol{t}^\top \\ \boldsymbol{Q} & \boldsymbol{t} & \boldsymbol{T} \end{pmatrix} \succeq 0 \\ & \begin{pmatrix} x & x & \boldsymbol{t}^\top \\ x & 1 & \boldsymbol{t}^\top \\ \boldsymbol{t} & \boldsymbol{t} & \boldsymbol{T} \end{pmatrix} \in \mathrm{CBQP}(n) \end{aligned}$$

Therefore, (EC.8) is equivalent to

$$\begin{aligned} \overline{\Phi}_C := \min_{x,\boldsymbol{t},\boldsymbol{Q},\boldsymbol{T},\overline{\boldsymbol{\Sigma}}} \quad & \langle \mathrm{diag}(\boldsymbol{p}), \boldsymbol{Q}\rangle - (\boldsymbol{p}\circ\boldsymbol{y})^\top \boldsymbol{t} - \bar{v}x \\ \text{s.t.} \quad & \begin{pmatrix} \overline{\boldsymbol{\Sigma}}+\boldsymbol{\mu}\boldsymbol{\mu}^\top & \boldsymbol{\mu} & \boldsymbol{Q}^\top \\ \boldsymbol{\mu}^\top & 1 & \boldsymbol{t}^\top \\ \boldsymbol{Q} & \boldsymbol{t} & \boldsymbol{T} \end{pmatrix} \succeq 0 \\ & \overline{\boldsymbol{\Sigma}}^r = \boldsymbol{\Sigma}^r,\ \forall r\in[R] \\ & \begin{pmatrix} x & x & \boldsymbol{t}^\top \\ x & 1 & \boldsymbol{t}^\top \\ \boldsymbol{t} & \boldsymbol{t} & \boldsymbol{T} \end{pmatrix} \in \mathrm{CBQP}(n) \end{aligned} \tag{EC.9}$$

For any given $\boldsymbol{y}$ and $\bar{v}$, let Φ_C denote the optimal value of Problem (C-I). It is sufficient to prove that $\Phi_C = \overline{\Phi}_C$.

Step 1: To show $\overline{\Phi}_C \geq \Phi_C$.

For any feasible solution $(\bar{x}, \bar{\boldsymbol{t}}, \bar{\boldsymbol{Q}}, \bar{\boldsymbol{T}}, \overline{\boldsymbol{\Sigma}})$ of (EC.9), set $x=\bar{x}$, $\boldsymbol{t}^r = \bar{\boldsymbol{t}}^r$, $\boldsymbol{Q}^r = \bar{\boldsymbol{Q}}^r$, and $\boldsymbol{T}^r = \bar{\boldsymbol{T}}^r$ for all $r\in[R]$, where $\bar{\boldsymbol{t}}^r$ is the subvector of $\bar{\boldsymbol{t}}$ restricted to cluster r, and $\bar{\boldsymbol{Q}}^r$ and $\bar{\boldsymbol{T}}^r$ are the corresponding cluster submatrices of $\bar{\boldsymbol{Q}}$ and $\bar{\boldsymbol{T}}$, respectively. Then, for each $r\in[R]$,

$$\begin{pmatrix} \boldsymbol{\Sigma}^r+\boldsymbol{\mu}^r(\boldsymbol{\mu}^r)^\top & \boldsymbol{\mu}^r & (\boldsymbol{Q}^r)^\top \\ (\boldsymbol{\mu}^r)^\top & 1 & (\boldsymbol{t}^r)^\top \\ \boldsymbol{Q}^r & \boldsymbol{t}^r & \boldsymbol{T}^r \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}^r+\boldsymbol{\mu}^r(\boldsymbol{\mu}^r)^\top & \boldsymbol{\mu}^r & (\bar{\boldsymbol{Q}}^r)^\top \\ (\boldsymbol{\mu}^r)^\top & 1 & (\bar{\boldsymbol{t}}^r)^\top \\ \bar{\boldsymbol{Q}}^r & \bar{\boldsymbol{t}}^r & \bar{\boldsymbol{T}}^r \end{pmatrix}.$$

The matrix on the right-hand side is PSD because it is a principal submatrix of a PSD matrix.

From the CBQP(n) constraint in (EC.9), the corresponding cluster-restricted moment matrix also belongs to CBQP(n_r). Indeed, any binary vector $\boldsymbol{w} \in \{0,1\}^n$ satisfying $\boldsymbol{w} \le w_0 \boldsymbol{e}_n$ induces a restricted vector $\boldsymbol{w}^r \in 0, 1^{n_r}$ satisfying $\boldsymbol{w}^r \le w_0 \boldsymbol{e}_{n_r}$, and the convex-hull weights are preserved under this projection. Hence, for each $r \in [R]$, $\begin{pmatrix} x & x & (\boldsymbol{t}^r)^\top \\ x & 1 & (\boldsymbol{t}^r)^\top \\ \boldsymbol{t}^r & \boldsymbol{t}^r & \boldsymbol{T}^r \end{pmatrix} \in \text{CBQP}(n_r)$. Thus, $(x, \{\boldsymbol{t}^r, \boldsymbol{Q}^r, \boldsymbol{T}^r\}_{r\in[R]})$ is feasible for Problem (C-I). Moreover,

$$\begin{aligned}\langle \text{diag}(\boldsymbol{p}), \boldsymbol{Q}\rangle - (\boldsymbol{p} \circ \boldsymbol{y})^\top \boldsymbol{t} - \bar{v}x &= \sum_{r=1}^{R} \left(\langle \text{diag}(\boldsymbol{p}^r), \boldsymbol{Q}^r \rangle - (\boldsymbol{p}^r \circ \boldsymbol{y}^r)^\top \boldsymbol{t}^r \right) - \bar{v}x \\ &= \sum_{r=1}^{R} \left(\left\langle \text{diag}(\boldsymbol{p}^r), \bar{\boldsymbol{Q}}^r \right\rangle - (\boldsymbol{p}^r \circ \boldsymbol{y}^r)^\top \bar{\boldsymbol{t}}^r \right) - \bar{v}\bar{x}.\end{aligned}$$

Therefore, for any feasible solution of (EC.9), there exists a feasible solution of Problem (C-I) with the same objective value. Since both problems are minimization problems, we obtain $\overline{\Phi}_C \ge \Phi_C$.

Step 2: To show $\overline{\Phi}_C \le \Phi_C$.

Suppose $(x_*, \{\boldsymbol{t}^r_*, \boldsymbol{Q}^r_*, \boldsymbol{T}^r_*\}_{r\in[R]})$ is an optimal solution to Problem (C-I). We construct a feasible solution for (EC.9) with the same objective value.

Define $\Upsilon^r := \{(w_0, \boldsymbol{w}^r) \in \{0,1\}^{n_r+1} \mid \boldsymbol{w}^r \le w_0 \boldsymbol{e}_{n_r}\}$. For each $r \in [R]$, by the definition of CBQP(n_r), there exist weights $\{\lambda^r_{(w_0,\boldsymbol{w}^r)}\}_{(w_0,\boldsymbol{w}^r)\in\Upsilon^r}$ such that $\lambda^r_{(w_0,\boldsymbol{w}^r)} \ge 0$ for all $(w_0, \boldsymbol{w}^r) \in \Upsilon^r$, $\sum_{(w_0,\boldsymbol{w}^r)\in\Upsilon^r} \lambda^r_{(w_0,\boldsymbol{w}^r)} = 1$, and

$$\begin{pmatrix} x_* & x_* & (\boldsymbol{t}^r_*)^\top \\ x_* & 1 & (\boldsymbol{t}^r_*)^\top \\ \boldsymbol{t}^r_* & \boldsymbol{t}^r_* & \boldsymbol{T}^r_* \end{pmatrix} = \sum_{(w_0,\boldsymbol{w}^r)\in\Upsilon^r} \lambda^r_{(w_0,\boldsymbol{w}^r)} \begin{pmatrix} w_0 \\ 1 \\ \boldsymbol{w}^r \end{pmatrix} \begin{pmatrix} w_0 \\ 1 \\ \boldsymbol{w}^r \end{pmatrix}^\top. \tag{EC.10}$$

We first construct $\boldsymbol{Y} = \begin{pmatrix} x & x & \boldsymbol{t}^\top \\ x & 1 & \boldsymbol{t}^\top \\ \boldsymbol{t} & \boldsymbol{t} & \boldsymbol{T} \end{pmatrix}$ such that $\boldsymbol{Y} \in \text{CBQP}(n)$ and $\boldsymbol{t}^r = \boldsymbol{t}^r_*$, $\boldsymbol{T}^r = \boldsymbol{T}^r_*$ for all $r \in [R]$.

Case 1: $x_* = 0$**.** For each $r \in [R]$, by the definition of Υ^r, $\begin{pmatrix} x_* & x_* & (\boldsymbol{t}^r_*)^\top \\ x_* & 1 & (\boldsymbol{t}^r_*)^\top \\ \boldsymbol{t}^r_* & \boldsymbol{t}^r_* & \boldsymbol{T}^r_* \end{pmatrix} = \begin{pmatrix} 0 & 0 & \boldsymbol{0}^\top \\ 0 & 1 & \boldsymbol{0}^\top \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}$. Let $\boldsymbol{Y} = \begin{pmatrix} 0 & 0 & \boldsymbol{0}^\top \\ 0 & 1 & \boldsymbol{0}^\top \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}$. It is straightforward that $\boldsymbol{Y} \in \text{CBQP}(n)$. Furthermore, $\boldsymbol{t}^r = \boldsymbol{t}^r_* = \boldsymbol{0}$ and $\boldsymbol{T}^r = \boldsymbol{T}^r_* = \boldsymbol{0}$ for all $r \in [R]$.

Case 2: $x_* > 0$**.** From the definition of Υ^r, we have $\lambda^r_{(0,\boldsymbol{0}_{n_r})} = 1 - x_*$ and $\sum_{\boldsymbol{w}^r\in\{0,1\}^{n_r}} \lambda^r_{(1,\boldsymbol{w}^r)} = x_*$ for all $r \in [R]$. For each $r \in [R]$ and $\boldsymbol{w}^r \in \{0,1\}^{n_r}$, define $\eta^r_{\boldsymbol{w}^r} = \lambda^r_{(1,\boldsymbol{w}^r)}/x_*$. It follows that

$$\sum_{\boldsymbol{w}^r\in\{0,1\}^{n_r}} \eta^r_{\boldsymbol{w}^r} = 1, \ \forall r \in [R]. \tag{EC.11}$$

We further define the global weights by $\lambda_{(0,\boldsymbol{0}_n)} = 1 - x_*$ and $\lambda_{(1,\boldsymbol{w})} = x_* \prod_{r\in[R]} \eta^r_{\boldsymbol{w}^r}$ for all $\boldsymbol{w} \in \{0,1\}^n$. Let $x := x_*$, $\boldsymbol{t} := \sum_{\boldsymbol{w}\in\{0,1\}^n} \lambda_{(1,\boldsymbol{w})} \boldsymbol{w}$, and $\boldsymbol{T} := \sum_{\boldsymbol{w}\in\{0,1\}^n} \lambda_{(1,\boldsymbol{w})} \boldsymbol{w}\boldsymbol{w}^\top$. We first show that $\boldsymbol{Y} \in$

CBQP(n). To this end, it suffices to verify that $\sum_{\boldsymbol{w}\in\{0,1\}^n}\lambda_{(1,\boldsymbol{w})}=x$ and $\sum_{\boldsymbol{w}\in\{0,1\}^n}\lambda_{(1,\boldsymbol{w})}+\lambda_{(0,\mathbf{0}_n)}=1$. The first equality follows from

$$\sum_{\boldsymbol{w}\in\{0,1\}^n}\lambda_{(1,\boldsymbol{w})}=\sum_{\boldsymbol{w}\in\{0,1\}^n}x_*\prod_{r\in[R]}\eta^r_{\boldsymbol{w}^r}=x\sum_{\boldsymbol{w}^1\in\{0,1\}^{n_1}}\cdots\sum_{\boldsymbol{w}^R\in\{0,1\}^{n_R}}\eta^1_{\boldsymbol{w}^1}\cdots\eta^R_{\boldsymbol{w}^R}=x\prod_{r\in[R]}\left(\sum_{\boldsymbol{w}^r\in\{0,1\}^{n_r}}\eta^r_{\boldsymbol{w}^r}\right)=x,$$

where the last equality follows from (EC.11). Hence, $\sum_{\boldsymbol{w}\in\{0,1\}^n}\lambda_{(1,\boldsymbol{w})}+\lambda_{(0,\mathbf{0}_n)}=1$ follows directly from $\lambda_{(0,\mathbf{0}_n)}=1-x$.

We next show that, for each $r\in[R]$, $\boldsymbol{t}^r=\boldsymbol{t}^r_*$ and $\boldsymbol{T}^r=\boldsymbol{T}^r_*$. Since $\boldsymbol{t}=\sum_{\boldsymbol{w}\in\{0,1\}^n}\lambda_{(1,\boldsymbol{w})}\boldsymbol{w}$, we have

$$\boldsymbol{t}^r=\sum_{\boldsymbol{w}\in\{0,1\}^n}\lambda_{(1,\boldsymbol{w})}\boldsymbol{w}^r=x\prod_{\bar r\in[R]\setminus\{r\}}\left(\sum_{\boldsymbol{w}^{\bar r}\in\{0,1\}^{n_{\bar r}}}\eta^{\bar r}_{\boldsymbol{w}^{\bar r}}\right)\sum_{\boldsymbol{w}^r\in\{0,1\}^{n_r}}\eta^r_{\boldsymbol{w}^r}\boldsymbol{w}^r=\sum_{\boldsymbol{w}^r\in\{0,1\}^{n_r}}\lambda^r_{(1,\boldsymbol{w}^r)}\boldsymbol{w}^r=\boldsymbol{t}^r_*,$$

where the last equality follows from (EC.10). Similarly

$$\boldsymbol{T}^r=\sum_{\boldsymbol{w}\in\{0,1\}^n}\lambda_{(1,\boldsymbol{w})}\boldsymbol{w}^r(\boldsymbol{w}^r)^\top=x\prod_{\bar r\in[R]\setminus\{r\}}\left(\sum_{\boldsymbol{w}^{\bar r}\in\{0,1\}^{n_{\bar r}}}\eta^{\bar r}_{\boldsymbol{w}^{\bar r}}\right)\sum_{\boldsymbol{w}^r\in\{0,1\}^{n_r}}\eta^r_{\boldsymbol{w}^r}\boldsymbol{w}^r(\boldsymbol{w}^r)^\top=\sum_{\boldsymbol{w}^r\in\{0,1\}^{n_r}}\lambda^r_{(1,\boldsymbol{w}^r)}\boldsymbol{w}^r(\boldsymbol{w}^r)^\top=\boldsymbol{T}^r_*,$$

again by (EC.10).

So far, we have constructed $\boldsymbol{Y}$ such that $\boldsymbol{Y}\in$ CBQP(n) and $\boldsymbol{t}^r=\boldsymbol{t}^r_*$, $\boldsymbol{T}^r=\boldsymbol{T}^r_*$ for all $r\in[R]$. It remains to construct $(\overline{\boldsymbol{\Sigma}},\boldsymbol{Q})$ for (EC.9) such that

$$\begin{pmatrix}\overline{\boldsymbol{\Sigma}}+\boldsymbol{\mu}\boldsymbol{\mu}^\top & \boldsymbol{\mu} & \boldsymbol{Q}^\top\\ \boldsymbol{\mu}^\top & 1 & \boldsymbol{t}^\top\\ \boldsymbol{Q} & \boldsymbol{t} & \boldsymbol{T}\end{pmatrix}\succeq 0,\ \overline{\boldsymbol{\Sigma}}^r=\boldsymbol{\Sigma}^r,\ \boldsymbol{Q}^r=\boldsymbol{Q}^r_*,\ \forall r\in[R]. \tag{EC.12}$$

This is guaranteed by the following lemma.

Lemma EC.2. *(Padmanabhan et al. 2021, Lemma 3) There exists matrices $\overline{\boldsymbol{\Sigma}}$ and $\boldsymbol{Q}$ such that* (EC.12) *holds.*

Hence, we can construct a feasible solution $(x,\boldsymbol{t},\boldsymbol{Q},\boldsymbol{T},\overline{\boldsymbol{\Sigma}})$ for (EC.9). It follows that

$$\begin{aligned}\overline{\Phi}_C\le\langle\mathrm{diag}(\boldsymbol{p}),\boldsymbol{Q}\rangle-(\boldsymbol{p}\circ\boldsymbol{y})^\top\boldsymbol{t}-\bar v x&=\sum_{r=1}^R\left(\langle\mathrm{diag}(\boldsymbol{p}^r),\boldsymbol{Q}^r\rangle-(\boldsymbol{p}^r\circ\boldsymbol{y}^r)^\top\boldsymbol{t}^r\right)-\bar v x\\&=\sum_{r=1}^R\left(\langle\mathrm{diag}(\boldsymbol{p}^r),\boldsymbol{Q}^r_*\rangle-(\boldsymbol{p}^r\circ\boldsymbol{y}^r)^\top\boldsymbol{t}^r_*\right)-\bar v x_*=\Phi_C.\end{aligned}$$

Combining Steps 1 and 2 yields $\overline{\Phi}_C=\Phi_C$, which completes the proof of Proposition 3.

EC.1.6. Proof of Theorem 2

The proof follows the same dualization argument as in the proof of Theorem 1. Specifically, we substitute the linear representation of each CBQP(n_r) constraint from Lemma 1 into Problem (C-I) and take the dual of the resulting SDP. Eliminating auxiliary variables yields the formulation in (C-R). Since $\boldsymbol{\Sigma}\succ 0$, each principal submatrix $\boldsymbol{\Sigma}^r$ also satisfies $\boldsymbol{\Sigma}^r\succ 0$. Hence, Slater's condition holds and there is no duality gap. Therefore, Problem (C) is equivalent to Problem (C-R).

EC.1.7. Proof of Lemma 2

We study the computational complexity using the cutting-plane bound in Jiang et al. (2020).

For part (a), the formulation (B-R) contains one PSD constraint of size $(2n+1)\times(2n+1)$ and 2^n linear constraints. Applying the cutting-plane complexity bound $\widetilde{O}\big(m(md^2+m^2+d^\theta)\big)$, where d is the PSD matrix dimension, m is the number of linear constraints, and θ is the exponent of matrix multiplication, and substituting $d=2n+1$ and $m=2^n$, we obtain $\widetilde{O}\left(2^n\left(2^n(2n+1)^2+2^{2n}+(2n+1)^\theta\right)\right)$. Equivalently, this is $\widetilde{O}\left(2^{2n}(2n+1)^2+2^{3n}+2^n(2n+1)^\theta\right)$. Since the term 2^{3n} dominates asymptotically, the complexity is $\widetilde{O}(2^{3n})=\widetilde{O}(8^n)$.

For part (b), the correlation-splitting formulation (C-R) has one PSD block of dimension $2n_r+1$ and 2^{n_r} linear constraints for each cluster $r\in[R]$. Applying the same cutting-plane bound cluster by cluster and summing over $r\in[R]$, we obtain $\widetilde{O}\left(\sum_{r=1}^R 2^{n_r}\left(2^{n_r}(2n_r+1)^2+2^{2n_r}+(2n_r+1)^\theta\right)\right)$. Equivalently, this is $\widetilde{O}\left(\sum_{r=1}^R\left(2^{2n_r}(2n_r+1)^2+2^{3n_r}+2^{n_r}(2n_r+1)^\theta\right)\right)$.

Fix any $a>1$ and set $R=\lceil n/\lfloor\log_a n\rfloor\rceil$. Assume n is large enough so that $\lfloor\log_a n\rfloor\geq 1$. Let $n_r=\lfloor\log_a n\rfloor$ for $r=1,\ldots,R-1$, and let the last cluster contain the remaining items, so that $n_R\leq\lfloor\log_a n\rfloor$. Then, for all $r\in[R]$, $2^{n_r}\leq 2^{\log_a n}=n^{\log_a 2}$ and $2n_r+1\leq 2\log_a n+1$. Hence, $2^{2n_r}(2n_r+1)^2\leq n^{2\log_a 2}(2\log_a n+1)^2$, $2^{3n_r}\leq n^{3\log_a 2}$, and $2^{n_r}(2n_r+1)^\theta\leq n^{\log_a 2}(2\log_a n+1)^\theta$. Summing over all clusters gives $\widetilde{O}\left(R\left(n^{2\log_a 2}(2\log_a n+1)^2+n^{3\log_a 2}+n^{\log_a 2}(2\log_a n+1)^\theta\right)\right)$. Since $R\leq n/\lfloor\log_a n\rfloor+1$, the complexity is bounded by $\widetilde{O}\left((n/\lfloor\log_a n\rfloor+1)\left(n^{2\log_a 2}(2\log_a n+1)^2+n^{3\log_a 2}+n^{\log_a 2}(2\log_a n+1)^\theta\right)\right)$. The three components scale, up to logarithmic factors, as $n^{1+2\log_a 2}$, $n^{1+3\log_a 2}$, and $n^{1+\log_a 2}$. Therefore, the dominant polynomial term is $n^{1+3\log_a 2}$. Since logarithmic factors are absorbed by the $\widetilde{O}$ notation, the overall complexity is $\widetilde{O}\left(n^{1+3\log_a 2}\right)$.

EC.1.8. Proof of Proposition 4

Under $R=n$, the ambiguity set of the correlation-splitting problem is

$$\mathscr{F}^c=\left\{F\in\mathcal{M}(\mathbb{R}^n)\;\middle|\;\mathbb{E}_F[\tilde{d}_i]=\mu_i,\ \mathbb{E}_F[\tilde{d}_i^2]=\sigma_i^2+\mu_i^2,\ \forall i\in[n]\right\}.$$

For each $i\in[n]$, let $\mathscr{F}(\mu_i,\sigma_i^2)=\{F_i\in\mathcal{M}(\mathbb{R})\mid\mathbb{E}_{F_i}[\tilde{d}_i]=\mu_i,\ \mathbb{E}_{F_i}[\tilde{d}_i^2]=\sigma_i^2+\mu_i^2\}$. Thus, any joint distribution whose ith marginal distribution belongs to $\mathscr{F}(\mu_i,\sigma_i^2)$ for every $i\in[n]$ is feasible for $\mathscr{F}^c$.

For any fixed $\boldsymbol{y}$, define $\varphi^c(\boldsymbol{y}):=\min_{F\in\mathscr{F}^c}\mathrm{CVaR}_{\eta,F}(\pi(\boldsymbol{y},\tilde{\boldsymbol{d}}))$. For each item i, define the single-item value function $\varphi_i(y_i):=\min_{F_i\in\mathscr{F}(\mu_i,\sigma_i^2)}\mathrm{CVaR}_{\eta,F_i}(\pi_i(y_i,\tilde{d}_i))$, where $\pi_i(y_i,\tilde{d}_i):=p_i(y_i\wedge\tilde{d}_i)-c_iy_i=(p_i-c_i)y_i+p_i(\tilde{d}_i-y_i)^-$. Then the single-item optimal value is $\Pi^i=\max_{y_i}\varphi_i(y_i)$. We first show that, for every fixed $\boldsymbol{y}$,

$$\varphi^c(\boldsymbol{y})=\sum_{i\in[n]}\varphi_i(y_i).\tag{EC.13}$$

We begin with the lower bound (i.e., the "$\geq$" direction). Fix any $F \in \mathscr{F}^c$, and let F_i be its ith marginal distribution. For arbitrary $v_i \in \mathbb{R}$, $i \in [n]$, the conditional value-at-risk (CVaR) representation gives

$$\begin{aligned}
\mathrm{CVaR}_{\eta,F}\left(\pi(\boldsymbol{y},\tilde{\boldsymbol{d}})\right) &= \max_{v\in\mathbb{R}}\left\{v + \frac{1}{\eta}\mathbb{E}_F\left[\min\left(\pi(\boldsymbol{y},\tilde{\boldsymbol{d}}) - v, 0\right)\right]\right\} \\
&= \max_{v\in\mathbb{R}}\left\{v + \frac{1}{\eta}\mathbb{E}_F\left[\min\left(\sum_{i\in[n]}\pi_i(y_i,\tilde{d}_i) - v, 0\right)\right]\right\} \\
&\geq \sum_{i\in[n]} v_i + \frac{1}{\eta}\mathbb{E}_F\left[\min\left(\sum_{i\in[n]}\left(\pi_i(y_i,\tilde{d}_i) - v_i\right), 0\right)\right] \\
&\geq \sum_{i\in[n]}\left\{v_i + \frac{1}{\eta}\mathbb{E}_{F_i}\left[\min\left(\pi_i(y_i,\tilde{d}_i) - v_i, 0\right)\right]\right\}.
\end{aligned}$$

The first inequality evaluates the maximization over v at $v = \sum_{i\in[n]} v_i$. The second inequality follows from $\min\left(\sum_{i\in[n]} a_i, 0\right) \geq \sum_{i\in[n]} \min(a_i, 0)$ for all $a_i \in \mathbb{R}$. Taking the maximum over v_i for each $i \in [n]$ yields $\mathrm{CVaR}_{\eta,F}(\pi(\boldsymbol{y},\tilde{\boldsymbol{d}})) \geq \sum_{i\in[n]} \mathrm{CVaR}_{\eta,F_i}(\pi_i(y_i,\tilde{d}_i)) \geq \sum_{i\in[n]} \varphi_i(y_i)$. Since this holds for every $F \in \mathscr{F}^c$, we obtain $\varphi^c(\boldsymbol{y}) \geq \sum_{i\in[n]} \varphi_i(y_i)$.

We next prove the reverse inequality. Fix any $\epsilon > 0$. For each $i \in [n]$, choose an ϵ/n-optimal distribution $F_i^\epsilon \in \mathscr{F}(\mu_i, \sigma_i^2)$ such that $\mathrm{CVaR}_{\eta,F_i^\epsilon}(\pi_i(y_i,\tilde{d}_i)) \leq \varphi_i(y_i) + \epsilon/n$. Let Q_i^ϵ be the generalized quantile function of F_i^ϵ, and let U be uniformly distributed on $[0,1]$. Define $\tilde{d}_i^\epsilon = Q_i^\epsilon(U)$ for $i \in [n]$. The joint distribution F^ϵ of $\tilde{\boldsymbol{d}}^\epsilon = (\tilde{d}_1^\epsilon, \ldots, \tilde{d}_n^\epsilon)$ preserves the marginal distribution F_i^ϵ for each $i \in [n]$, and hence $F^\epsilon \in \mathscr{F}^c$. Moreover, $\pi_i(y_i, d_i)$ is nondecreasing in d_i. Therefore, the random variables $\pi_i(y_i, \tilde{d}_i^\epsilon)$, $i \in [n]$, are nondecreasing functions of the same uniform random variable U. Hence they are comonotonic. Using the quantile representation of lower-tail CVaR and the additivity of quantiles under comonotonicity (Dhaene et al. 2002, Theorem 6), we obtain

$$\mathrm{CVaR}_{\eta,F^\epsilon}\left(\pi(\boldsymbol{y},\tilde{\boldsymbol{d}}^\epsilon)\right) = \mathrm{CVaR}_{\eta,F^\epsilon}\left(\sum_{i\in[n]}\pi_i(y_i,\tilde{d}_i^\epsilon)\right) = \sum_{i\in[n]}\mathrm{CVaR}_{\eta,F_i^\epsilon}\left(\pi_i(y_i,\tilde{d}_i)\right) \leq \sum_{i\in[n]}\varphi_i(y_i) + \epsilon.$$

Since $F^\epsilon \in \mathscr{F}^c$, it follows that $\varphi^c(\boldsymbol{y}) = \min_{F\in\mathscr{F}^c} \mathrm{CVaR}_{\eta,F}(\pi(\boldsymbol{y},\tilde{\boldsymbol{d}})) \leq \sum_{i\in[n]} \varphi_i(y_i) + \epsilon$. Letting $\epsilon \downarrow 0$ gives $\varphi^c(\boldsymbol{y}) \leq \sum_{i\in[n]} \varphi_i(y_i)$. Combining the two inequalities proves (EC.13).

Finally, maximizing both sides of (EC.13) over $\boldsymbol{y}$ gives $\underline{\Pi} = \max_{\boldsymbol{y}} \varphi^c(\boldsymbol{y}) = \max_{\boldsymbol{y}} \sum_{i\in[n]} \varphi_i(y_i) = \sum_{i\in[n]} \max_{y_i} \varphi_i(y_i) = \sum_{i\in[n]} \Pi_i$. This completes the proof.

EC.1.9. Proof of Proposition 5.

By Proposition 4, when $R = n$, we have $\underline{\Pi} = \sum_{i\in[n]} \Pi^i$. Moreover, since the full-information ambiguity set $\mathscr{F}$ is a subset of the correlation-splitting ambiguity set $\mathscr{F}^c$, it follows that $\Pi \geq \underline{\Pi}$. It remains to show that $\Pi \leq \sum_{i\in[n]} \Pi^i$ under the stated condition of Proposition 5.

We first introduce the following single-item result, which will be used in the proof. The proof is given in Section EC.1.10.

Lemma EC.3. *Consider Problem* (P) *under the single-item setting, that is, considering demand* $\tilde{d}$ *with mean* μ *and variance* σ^2. *Then* $\Pi = (p-c)\left(\mu - \sigma\sqrt{(1-\eta\beta)/(\eta\beta)}\right)$, *where* $\beta = (p-c)/p$ *is the critical ratio. The corresponding optimal inventory level is* $y_* = \mu + \frac{\sigma}{2}\frac{2\eta\beta-1}{\sqrt{\eta\beta(1-\eta\beta)}}$, *and the worst-case distribution is*

$$\tilde{d}_* = \begin{cases} \mu - \sigma\sqrt{\dfrac{1-\eta\beta}{\eta\beta}}, \text{ w.p. } \eta\beta, \\ \mu + \sigma\sqrt{\dfrac{\eta\beta}{1-\eta\beta}}, \text{ w.p. } 1-\eta\beta. \end{cases}$$

Define $\beta_i = (p_i - c_i)/p_i$ and $q_i = \eta\beta_i$ for $i \in [n]$. By Lemma EC.3, for each $i \in [n]$, $\Pi^i = (p_i - c_i)\left(\mu_i - \sigma_i\sqrt{(1-q_i)/q_i}\right)$, and the optimal inventory level is $y_{i,*} = \mu_i + \frac{\sigma_i}{2}\left(\sqrt{q_i/(1-q_i)} - \sqrt{(1-q_i)/q_i}\right)$. The corresponding worst-case marginal distribution is

$$\tilde{d}_{i,*} = \begin{cases} d_i^L := \mu_i - \sigma_i\sqrt{\dfrac{1-q_i}{q_i}}, \text{ w.p. } q_i, \\ d_i^H := \mu_i + \sigma_i\sqrt{\dfrac{q_i}{1-q_i}}, \text{ w.p. } 1-q_i. \end{cases}$$

Note that $y_{i,*} = (d_i^L + d_i^H)/2$ and $d_i^L < y_{i,*} < d_i^H$.

Let F_* be the independent product distribution of the marginal distributions $\{F_{i,*}\}_{i\in[n]}$. Since all items are mutually uncorrelated, the covariance matrix $\boldsymbol{\Sigma}$ is diagonal with entries σ_i^2. Therefore, F_* preserves the given means, variances, and zero cross-covariances, and hence F_* is feasible for the full-information ambiguity set $\mathscr{F}$.

For each item, define the single-item profit contribution $\pi_i(y_i, \tilde{d}_i) := p_i(y_i \wedge \tilde{d}_i) - c_i y_i = (p_i - c_i)y_i + p_i(\tilde{d}_i - y_i)^-$. Then $\pi(\boldsymbol{y}, \tilde{\boldsymbol{d}}) = \sum_{i\in[n]} \pi_i(y_i, \tilde{d}_i)$. Define $G(\boldsymbol{y}) := \mathrm{CVaR}_{\eta,F_*}(\pi(\boldsymbol{y}, \tilde{\boldsymbol{d}}))$. Since $F_* \in \mathscr{F}$, for any $\boldsymbol{y}$, $\min_{F\in\mathscr{F}} \mathrm{CVaR}_{\eta,F}(\pi(\boldsymbol{y}, \tilde{\boldsymbol{d}})) \le G(\boldsymbol{y})$. Thus, to show $\Pi \le \sum_{i\in[n]} \Pi^i$, it suffices to show that $\max_{\boldsymbol{y}} G(\boldsymbol{y}) = \sum_{i\in[n]} \Pi^i$.

We first evaluate G at $\boldsymbol{y}_* = (y_{1,*}, \ldots, y_{n,*})$. Let $X_i^L := \pi_i(y_{i,*}, d_i^L)$ and $X_i^H := \pi_i(y_{i,*}, d_i^H)$, and define $S^H := \sum_{i\in[n]} X_i^H$. Since $d_i^L < y_{i,*} < d_i^H$, we have $X_i^L < X_i^H$ for all $i \in [n]$. Thus, S^H is the maximum possible aggregate profit under F_*. Under the independent product distribution F_*, the probability of the all-high-demand state is $P_H = \prod_{i\in[n]}(1-q_i) = \prod_{i\in[n]}(1-\eta\beta_i)$. The condition in Proposition 5 is equivalent to $P_H \ge 1-\eta$. Therefore, $\mathbb{P}_{F_*}(\pi(\boldsymbol{y}_*, \tilde{\boldsymbol{d}}) < S^H) = 1 - P_H \le \eta$ and $\mathbb{P}_{F_*}(\pi(\boldsymbol{y}_*, \tilde{\boldsymbol{d}}) \le S^H) = 1 \ge \eta$. Hence, by the quantile characterization of the optimal CVaR threshold, S^H is an optimal threshold for $\mathrm{CVaR}_{\eta,F_*}(\pi(\boldsymbol{y}_*, \tilde{\boldsymbol{d}}))$. Since S^H is the maximum possible aggregate profit, we have $\min(\pi(\boldsymbol{y}_*, \tilde{\boldsymbol{d}}) - S^H, 0) = \pi(\boldsymbol{y}_*, \tilde{\boldsymbol{d}}) - S^H$. It follows that

$$G(\boldsymbol{y}_*) = S^H + \frac{1}{\eta}\mathbb{E}_{F_*}\left[\pi(\boldsymbol{y}_*, \tilde{\boldsymbol{d}}) - S^H\right] = \sum_{i\in[n]}\left(X_i^H + \frac{1}{\eta}\mathbb{E}_{F_{i,*}}\left[\pi_i(y_{i,*}, \tilde{d}_i) - X_i^H\right]\right) = \sum_{i\in[n]}\left(X_i^H + \frac{q_i}{\eta}\left(X_i^L - X_i^H\right)\right).$$

For the single-item problem, X_i^H is an optimal CVaR threshold under $F_{i,*}$, and hence $\Pi^i = X_i^H + \frac{q_i}{\eta}(X_i^L - X_i^H)$. Therefore, $G(\boldsymbol{y}_*) = \sum_{i\in[n]} \Pi^i$.

It remains to show that $\boldsymbol{y}_*$ maximizes $G(\boldsymbol{y})$. For any fixed distribution, $\pi(\boldsymbol{y},\tilde{\boldsymbol{d}})$ is concave in $\boldsymbol{y}$, and lower-tail CVaR preserves concavity in the profit random variable. Therefore, $G(\boldsymbol{y})$ is concave in $\boldsymbol{y}$. It is thus sufficient to show that $\mathbf{0}$ is a supergradient of G at $\boldsymbol{y}_*$.

At $\boldsymbol{y}_*$, the derivative of $\pi_i(y_i,d_i)$ with respect to y_i is

$$\frac{\partial\pi_i(y_i,d_i)}{\partial y_i}=\begin{cases}-c_i, & d_i=d_i^L,\\ p_i-c_i, & d_i=d_i^H.\end{cases}$$

Using the standard supergradient representation of CVaR for a finite distribution, assign weight $1/\eta$ to all states satisfying $\pi(\boldsymbol{y}_*,\tilde{\boldsymbol{d}})<S^H$, and assign the remaining total weight $(\eta-\mathbb{P}_{F_*}(\pi(\boldsymbol{y}_*,\tilde{\boldsymbol{d}})<S^H))/\eta=(\eta-1+P_H)/\eta$ to the all-high-demand state. This is valid because $P_H\geq 1-\eta$. For any event A, let $\mathbb{I}\{A\}$ denote the indicator of A, i.e., $\mathbb{I}\{A\}=1$ if A occurs and $\mathbb{I}\{A\}=0$ otherwise. The ith component of the resulting supergradient is

$$\begin{aligned}&\frac{1}{\eta}\mathbb{E}_{F_*}\left[\frac{\partial\pi_i(y_{i,*},\tilde{d}_i)}{\partial y_i}\mathbb{I}\left\{\pi(\boldsymbol{y}_*,\tilde{\boldsymbol{d}})<S^H\right\}\right]+\frac{\eta-1+P_H}{\eta}(p_i-c_i)\\ =&\frac{1}{\eta}\left[\mathbb{E}_{F_*}\left[\frac{\partial\pi_i(y_{i,*},\tilde{d}_i)}{\partial y_i}\right]-P_H(p_i-c_i)+(\eta-1+P_H)(p_i-c_i)\right]\\ =&\frac{1}{\eta}\left[q_i(-c_i)+(1-q_i)(p_i-c_i)+(\eta-1)(p_i-c_i)\right]\\ =&\frac{1}{\eta}\left[(p_i-c_i)-q_ip_i+(\eta-1)(p_i-c_i)\right]\\ =&\frac{1}{\eta}\left[\eta(p_i-c_i)-q_ip_i\right]=0,\end{aligned}$$

where the last equality follows from $q_i=\eta(p_i-c_i)/p_i$. Thus, $\mathbf{0}$ is a supergradient of the concave function G at $\boldsymbol{y}_*$. Hence, $\boldsymbol{y}_*$ is a global maximizer of G, and $\max_{\boldsymbol{y}}G(\boldsymbol{y})=G(\boldsymbol{y}_*)=\sum_{i\in[n]}\Pi^i$. Consequently, $\Pi=\max_{\boldsymbol{y}}\min_{F\in\mathscr{F}}\mathrm{CVaR}_{\eta,F}(\pi(\boldsymbol{y},\tilde{\boldsymbol{d}}))\leq\max_{\boldsymbol{y}}G(\boldsymbol{y})=\sum_{i\in[n]}\Pi^i$. Together with $\Pi\geq\underline{\Pi}=\sum_{i\in[n]}\Pi^i$, we obtain $\Pi=\underline{\Pi}=\sum_{i\in[n]}\Pi^i$.

Finally, when $n=2$, the condition becomes $(1-\eta\beta_1)(1-\eta\beta_2)+\eta-1\geq 0$. Expanding and simplifying gives $\eta\beta_1\beta_2-\beta_1-\beta_2+1\geq 0$, which is equivalent to $\eta\geq(\beta_1+\beta_2-1)/(\beta_1\beta_2)=1-(1-\beta_1)(1-\beta_2)/(\beta_1\beta_2)=1-c_1c_2/((p_1-c_1)(p_2-c_2))$. This completes the proof.

EC.1.10. Proof of Lemma EC.3

Recall that $\Pi=\max_y\min_{F\in\mathscr{F}}\mathrm{CVaR}_{\eta,F}(\pi(y,\tilde{d}))$. For any given $y\in\mathbb{R}$, by the stochastic minimax theorem (Shapiro and Kleywegt 2002), the inner minimization problem can be written as

$$\min_{F\in\mathscr{F}}\mathrm{CVaR}_{\eta,F}(\pi(y,\tilde{d}))=\max_{v\in\mathbb{R}}\left\{v+\frac{1}{\eta}\min_{F\in\mathscr{F}}\mathbb{E}_F\left[\min(\pi(y,\tilde{d})-v,0)\right]\right\}. \tag{EC.14}$$

Step 1: Worst-case CVaR for a given y.

We first calculate the worst-case CVaR value in (EC.14) for any given y. For any given (y,v), the inner problem is the moment problem

$$\begin{aligned}
\min_{F\in\mathscr{F}} \quad & \int_{\mathbb{R}} \min(p(y\wedge d)-cy-v,0)\,\mathrm{d}F(d) \\
\text{s.t.} \quad & \int_{\mathbb{R}} d\,\mathrm{d}F(d)=\mu \\
& \int_{\mathbb{R}} d^2\,\mathrm{d}F(d)=\mu^2+\sigma^2 \\
& \int_{\mathbb{R}} 1\,\mathrm{d}F(d)=1 \\
& \mathrm{d}F(d)\ge 0,\ \forall d\in\mathbb{R}
\end{aligned}$$

Its dual is

$$\begin{aligned}
\max_{s_1,s_2,s_3} \quad & \mu s_1+(\mu^2+\sigma^2)s_2+s_3 \\
\text{s.t.} \quad & s_1 d+s_2 d^2+s_3\le \min(p(y\wedge d)-cy-v,0),\ \forall d\in\mathbb{R}
\end{aligned} \tag{EC.15}$$

We analyze two cases depending on v. In each case, we construct feasible solutions to the primal and dual problems that attain the same objective value.

Case 1: $v\ge(p-c)y$**.** In this case, the dual problem (EC.15) becomes

$$\begin{aligned}
\max_{s_1,s_2,s_3} \quad & \mu s_1+(\mu^2+\sigma^2)s_2+s_3 \\
\text{s.t.} \quad & s_1 d+s_2 d^2+s_3\le pd-cy-v,\ \forall d\le y, \\
& s_1 d+s_2 d^2+s_3\le py-cy-v,\ \forall d\ge y.
\end{aligned}$$

For the primal problem, consider the feasible two-point distribution $x_1=y-\Delta$, $x_2=y+\Delta$, and $\Delta:=\sqrt{(\mu-y)^2+\sigma^2}$, with probabilities $\rho_1=\sigma^2/((\mu-y+\Delta)^2+\sigma^2)$ and $\rho_2=(\mu-y+\Delta)^2/((\mu-y+\Delta)^2+\sigma^2)$. For the dual problem, the solution $s_1=p(y+\Delta)/(2\Delta)$, $s_2=-p/(4\Delta)$, and $s_3=(p-c)y-v-p(y+\Delta)^2/(4\Delta)$ is feasible and attains the same objective value as the primal two-point distribution above. Hence, by strong duality,

$$\min_{F\in\mathscr{F}}\mathbb{E}_F\left[\min(\pi(y,\tilde{d})-v,0)\right]=(p-c)y-v+p\left[\frac{\mu-y}{2}-\frac{1}{2}\sqrt{(\mu-y)^2+\sigma^2}\right].$$

Substituting this expression into (EC.14) gives

$$\max_{v\ge(p-c)y}\left\{\left(1-\frac{1}{\eta}\right)v+\frac{p-c}{\eta}y+\frac{p}{\eta}\left(\frac{\mu-y}{2}-\frac{1}{2}\sqrt{(\mu-y)^2+\sigma^2}\right)\right\}.$$

Since $1-1/\eta\le 0$, the maximizer is $v_*=(p-c)y$, yielding

$$h(y)=(p-c)y+\frac{p}{\eta}\left(\frac{\mu-y}{2}-\frac{1}{2}\sqrt{(\mu-y)^2+\sigma^2}\right).$$

The corresponding worst-case distribution is

$$\tilde{d}_*=\begin{cases} y-\Delta, & \text{w.p. } \dfrac{\sigma^2}{(\mu-y+\Delta)^2+\sigma^2}, \\[2ex] y+\Delta, & \text{w.p. } \dfrac{(\mu-y+\Delta)^2}{(\mu-y+\Delta)^2+\sigma^2}. \end{cases} \tag{EC.16}$$

Case 2: $v \le (p-c)y$. Let $y_0 := (cy+v)/p \le y$. Then $v \le (p-c)y$ is equivalent to $y_0 \le y$. The dual problem (EC.15) becomes

$$\begin{aligned}\max_{s_1,s_2,s_3} \quad & \mu s_1 + (\mu^2+\sigma^2)s_2 + s_3 \\ \text{s.t.} \quad & s_1 d + s_2 d^2 + s_3 \le pd - py_0, \ \forall d \le y_0, \\ & s_1 d + s_2 d^2 + s_3 \le 0, \ \forall d \ge y_0.\end{aligned}$$

Consider the primal solution $x_1 = y_0 - \Delta$, $x_2 = y_0 + \Delta$, and $\Delta := \sqrt{(\mu - y_0)^2 + \sigma^2}$, with probabilities $\rho_1 = \sigma^2/((\mu - y_0 + \Delta)^2 + \sigma^2)$ and $\rho_2 = (\mu - y_0 + \Delta)^2/((\mu - y_0 + \Delta)^2 + \sigma^2)$, and the dual solution $s_1 = p(y_0 + \Delta)/(2\Delta)$, $s_2 = -p/(4\Delta)$, and $s_3 = -p(y_0+\Delta)^2/(4\Delta)$. The primal and dual solutions constructed above are feasible and attain the same objective value. Hence, by strong duality, repeating the argument in Case 1, the problem in (EC.14) becomes $\max_{y_0 \le y} g(y_0)$, where

$$g(y_0) = py_0 - cy + \frac{p}{\eta}\left(\frac{\mu - y_0}{2} - \frac{1}{2}\sqrt{(\mu - y_0)^2 + \sigma^2}\right).$$

The function $g(y_0)$ is concave and attains its maximum at $\bar{y}_0 = \mu + (2\eta - 1)\sigma/(2\sqrt{\eta(1-\eta)})$, with value $g(\bar{y}_0) = p\left(\mu - \sigma\sqrt{(1-\eta)/\eta}\right) - cy$. At $y_0 = \bar{y}_0$, the corresponding worst-case distribution is

$$\tilde{d}_* = \begin{cases} \mu - \sigma\sqrt{\dfrac{1-\eta}{\eta}}, & \text{w.p. } \eta, \\ \mu + \sigma\sqrt{\dfrac{\eta}{1-\eta}}, & \text{w.p. } 1-\eta. \end{cases}$$

Therefore,

$$\max_{y_0 \le y} g(y_0) = \begin{cases} g(y) = (p-c)y + \dfrac{p}{\eta}\left(\dfrac{\mu - y}{2} - \dfrac{1}{2}\sqrt{(\mu - y)^2 + \sigma^2}\right), & \text{if } y \le \bar{y}_0, \\ g(\bar{y}_0) = p\left(\mu - \sigma\sqrt{\dfrac{1-\eta}{\eta}}\right) - cy, & \text{if } y > \bar{y}_0. \end{cases}$$

Combining Case 1 and Case 2 yields

$$\begin{aligned}\min_{F \in \mathscr{F}} \mathrm{CVaR}_{\eta,F}(\pi(y,\tilde{d})) &= \max\{h(y), g(y), g(\bar{y}_0)\} = \max\{g(y), g(\bar{y}_0)\} \\ &= \begin{cases} (p-c)y + \dfrac{p}{\eta}\left(\dfrac{\mu - y}{2} - \dfrac{1}{2}\sqrt{(\mu - y)^2 + \sigma^2}\right), & \text{if } y \le \bar{y}_0, \\ p\left(\mu - \sigma\sqrt{\dfrac{1-\eta}{\eta}}\right) - cy, & \text{if } y > \bar{y}_0. \end{cases}\end{aligned} \tag{EC.17}$$

Step 2: Optimization over y.

Let $L(y)$ denote the above function. On the first piece $y \le \bar{y}_0$, $L(y)$ is differentiable and concave. Setting $L'(y) = 0$ gives $y_* = \mu + \frac{\sigma}{2}\frac{2\eta\beta - 1}{\sqrt{\eta\beta(1-\eta\beta)}}$, where $\beta = (p-c)/p$. The function $(2t-1)/\sqrt{t(1-t)}$ is increasing on $(0,1)$. Since $\eta\beta \le \eta$ and $\beta \in (0,1)$, we have

$$y_* = \mu + \frac{\sigma}{2}\frac{2\eta\beta - 1}{\sqrt{\eta\beta(1-\eta\beta)}} \le \mu + \frac{(2\eta - 1)\sigma}{2\sqrt{\eta(1-\eta)}} = \bar{y}_0.$$

Thus, the stationary point lies in the first region. On the second piece, $L(y) = p(\mu - \sigma\sqrt{(1-\eta)/\eta}) - cy$ is decreasing in y. Hence, this piece is maximized at the boundary $y = \bar{y}_0$, and it cannot dominate the concave maximum attained at y_*. Therefore, y_* is the global maximizer of $L(y)$, and the corresponding optimal threshold is $v_* = (p-c)y_*$. Substituting y_* into $L(y)$ and (EC.16) gives

$$\Pi = \min_F \mathrm{CVaR}_{\eta,F}(\pi(y_*, \tilde{d})) = (p-c)\left(\mu - \sigma\sqrt{\frac{1-\eta\beta}{\eta\beta}}\right),$$

with the worst-case distribution

$$\tilde{d}_* = \begin{cases} \mu - \sigma\sqrt{\dfrac{1-\eta\beta}{\eta\beta}}, & \text{w.p. } \eta\beta, \\ \mu + \sigma\sqrt{\dfrac{\eta\beta}{1-\eta\beta}}, & \text{w.p. } 1-\eta\beta. \end{cases}$$

This completes the proof.

EC.1.11. Proof of Proposition 6

We first focus on Problem (B-I). Define $\tilde{\boldsymbol{d}}^D = \boldsymbol{\Sigma}^{-1/2}(\tilde{\boldsymbol{d}} - \boldsymbol{\mu})$. Since $\boldsymbol{\Sigma} \succ 0$, this transformation is well defined and one-to-one. Moreover, $\mathbb{E}[\tilde{\boldsymbol{d}}^D] = \boldsymbol{0}$ and $\mathbb{E}[\tilde{\boldsymbol{d}}^D(\tilde{\boldsymbol{d}}^D)^\top] = \boldsymbol{\Sigma}^{-1/2}\mathbb{E}[(\tilde{\boldsymbol{d}} - \boldsymbol{\mu})(\tilde{\boldsymbol{d}} - \boldsymbol{\mu})^\top]\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{I}_n$. By Proposition 2, for any given $(\boldsymbol{y}, \bar{v})$, Problem (B-I) is equivalent to the inner minimization problem $\min_{F\in\mathscr{F}} \mathbb{E}_F\left[\min\left(\boldsymbol{p}^\top(\tilde{\boldsymbol{d}} - \boldsymbol{y})^- - \bar{v}, 0\right)\right]$. Under the transformation from $\tilde{\boldsymbol{d}}$ to $\tilde{\boldsymbol{d}}^D$, this problem can be written as

$$\min_{F\in\mathscr{F}^D} \mathbb{E}_F\left[\min\left(\boldsymbol{p}^\top(\boldsymbol{\Sigma}^{1/2}\tilde{\boldsymbol{d}}^D + \boldsymbol{\mu} - \boldsymbol{y})^- - \bar{v}, 0\right)\right], \tag{EC.18}$$

where $\mathscr{F}^D = \left\{F \in \mathcal{M}(\mathbb{R}^n) \;\middle|\; \mathbb{E}_F[\tilde{\boldsymbol{d}}^D] = \boldsymbol{0},\ \mathbb{E}_F[\tilde{\boldsymbol{d}}^D(\tilde{\boldsymbol{d}}^D)^\top] = \boldsymbol{I}_n\right\}$.

Using the same lifting argument as in the proof of Proposition 2 in Section EC.1.2, Problem (EC.18) admits the following equivalent formulation:

$$\begin{aligned} \min_{x,\boldsymbol{t},\boldsymbol{Q},\boldsymbol{T}} \quad & \left\langle \mathrm{diag}(\boldsymbol{p}), \boldsymbol{Q}\boldsymbol{\Sigma}^{1/2}\right\rangle + [\boldsymbol{p}\circ(\boldsymbol{\mu} - \boldsymbol{y})]^\top \boldsymbol{t} - \bar{v}x \\ \text{s.t.} \quad & \begin{pmatrix} \boldsymbol{I}_n & \boldsymbol{0} & \boldsymbol{Q}^\top \\ \boldsymbol{0}^\top & 1 & \boldsymbol{t}^\top \\ \boldsymbol{Q} & \boldsymbol{t} & \boldsymbol{T} \end{pmatrix} \succeq 0 \\ & \begin{pmatrix} x & x & \boldsymbol{t}^\top \\ x & 1 & \boldsymbol{t}^\top \\ \boldsymbol{t} & \boldsymbol{t} & \boldsymbol{T} \end{pmatrix} \in \mathrm{CBQP}(n) \end{aligned}$$

This is exactly Problem (B-F).

The proof for the second part is analogous. For each cluster $r \in [R]$, define $(\tilde{\boldsymbol{d}}_c^D)^r = (\boldsymbol{\Sigma}^r)^{-1/2}(\tilde{\boldsymbol{d}}^r - \boldsymbol{\mu}^r)$. Since $\boldsymbol{\Sigma}^r$ is a principal submatrix of $\boldsymbol{\Sigma} \succ 0$, we have $\boldsymbol{\Sigma}^r \succ 0$, and hence this transformation is well defined. It gives $\mathbb{E}[(\tilde{\boldsymbol{d}}_c^D)^r] = \boldsymbol{0}, \mathbb{E}[(\tilde{\boldsymbol{d}}_c^D)^r((\tilde{\boldsymbol{d}}_c^D)^r)^\top] = \boldsymbol{I}_{n_r}, \forall r \in [R]$. Applying the same lifting argument cluster by cluster to Problem (C-I) yields Problem (C-F). This proves the proposition.

EC.1.12. Proof of Theorem 3

We begin by establishing a decomposition result for the block matrix structure arising in our formulation. This result reduces the positive semidefiniteness analysis of a high-dimensional matrix to a set of lower-dimensional block conditions. We then use the optimal solution of Problem (C-F) to construct a family of feasible solutions to Problem (B-F), thereby generating paired solutions that connect the correlation-splitting model and the full-information model. Building on this construction, we derive a family of optimality-gap bounds and then minimize this family to obtain the bound in Theorem 3. The following lemma provides the block decomposition result. Its proof is given in Section EC.1.13.

LEMMA EC.4. *Consider the following PSD matrix of dimension* $(2n+1)\times(2n+1)$*:*

$$
\boldsymbol{A} = \begin{array}{c}
\left(\begin{array}{c|ccc|ccc}
s & \boldsymbol{u}_1^\top & \cdots & \boldsymbol{u}_R^\top & \boldsymbol{v}_1^\top & \cdots & \boldsymbol{v}_R^\top \\ \hline
\boldsymbol{u}_1 & \boldsymbol{U}_1 & & \boldsymbol{0} & \boldsymbol{W}_1^\top & & \boldsymbol{0} \\
\vdots & & \ddots & & & \ddots & \\
\boldsymbol{u}_R & \boldsymbol{0} & & \boldsymbol{U}_R & \boldsymbol{0} & & \boldsymbol{W}_R^\top \\ \hline
\boldsymbol{v}_1 & \boldsymbol{W}_1 & & \boldsymbol{0} & \boldsymbol{V}_1 & & \boldsymbol{0} \\
\vdots & & \ddots & & & \ddots & \\
\boldsymbol{v}_R & \boldsymbol{0} & & \boldsymbol{W}_R & \boldsymbol{0} & & \boldsymbol{V}_R
\end{array}\right)
\begin{array}{l}
\left.\vphantom{\begin{array}{c}a\\a\\a\end{array}}\right\}n \\
\left.\vphantom{\begin{array}{c}a\\a\\a\end{array}}\right\}n
\end{array} \\
\qquad\underbrace{\hphantom{\boldsymbol{U}_1\ \cdots\ \boldsymbol{U}_R}}_{n}\ \underbrace{\hphantom{\boldsymbol{V}_1\ \cdots\ \boldsymbol{V}_R}}_{n}
\end{array},
$$

where $s\in\mathbb{R}$*,* $\boldsymbol{u}_r,\boldsymbol{v}_r\in\mathbb{R}^{n_r}$*, and* $\boldsymbol{U}_r,\boldsymbol{V}_r\in\mathbb{R}^{n_r\times n_r}$ *for all* $r\in[R]$*, with* $\sum_{r=1}^R n_r = n$*. Then* $\boldsymbol{A}\succeq 0$ *if and only if there exist scalars* s_r*,* $r\in[R]$*, such that* $\sum_{r=1}^R s_r = s$ *and*

$$
\boldsymbol{A}_r = \begin{pmatrix} s_r & \boldsymbol{u}_r^\top & \boldsymbol{v}_r^\top \\ \boldsymbol{u}_r & \boldsymbol{U}_r & \boldsymbol{W}_r^\top \\ \boldsymbol{v}_r & \boldsymbol{W}_r & \boldsymbol{V}_r \end{pmatrix} \succeq 0,\ \forall r\in[R].
$$

We next characterize a family of upper bounds on the optimality gap. The proof is given in Section EC.1.14.

LEMMA EC.5. *Suppose* $(\boldsymbol{y}_*,\bar{v}_*)$ *is an optimal solution to Problem* (C-R)*. For Problem* (C-F) *evaluated at* $(\boldsymbol{y},\bar{v})=(\boldsymbol{y}_*,\bar{v}_*)$*, let* $(\{\boldsymbol{T}_*^r,\boldsymbol{Q}_*^r,\boldsymbol{t}_*^r\}_{r\in[R]},x_*)$ *be an associated optimal solution. Then*

$$
0\le \eta(\Pi_B-\Pi_C)\le \inf_{(a_r,b_r,c_r)_{r\in[R]}\in\Omega}\sum_{r\in[R]} c_r g_r(a_r,b_r) - \sum_{r\in[R]}\left(\Delta_{\Sigma,c}^r+\Delta_\mu^r\right), \tag{EC.19}
$$

where

$$
\Omega = \left\{ (a_r,b_r,c_r)_{r\in[R]} \;\middle|\; a_r,b_r,c_r\ge 0,\ 0\le a_r\le 1\le b_r,\ a_rb_r\le 1,\ \sum_{r\in[R]} c_r = 1 \right\},
$$

$$
g_r(a_r,b_r) = a_r\Delta_\Sigma^r + a_rb_r\Delta_\mu^r, \Delta_\Sigma^r = \left\langle \operatorname{diag}(\boldsymbol{p}^r), \boldsymbol{Q}_*^r(\boldsymbol{\Sigma}^{1/2})^r\right\rangle,
$$

$$
\Delta_{\Sigma,c}^r = \left\langle \operatorname{diag}(\boldsymbol{p}^r), \boldsymbol{Q}_*^r(\boldsymbol{\Sigma}^r)^{1/2}\right\rangle, \textit{and } \Delta_\mu^r = \left[\boldsymbol{p}^r\circ(\boldsymbol{\mu}^r-\boldsymbol{y}_*^r)\right]^\top \boldsymbol{t}_*^r.
$$

Lemma EC.5 expresses the gap as the infimum of a convex combination of the functions $\{g_r(a_r,b_r)\}_{r\in[R]}$ minus the term $\sum_{r\in[R]}(\Delta^r_{\Sigma,c}+\Delta^r_\mu)$. The expression in (EC.19) depends on the auxiliary variables (a_r,b_r,c_r). To obtain the tightest bound within this family, we first minimize $g_r(a_r,b_r)$ over a_r and b_r for each cluster:

$$G^r = \inf_{\substack{0\le a_r\le 1\le b_r\\ a_rb_r\le 1}} a_r\Delta^r_\Sigma + a_rb_r\Delta^r_\mu. \tag{EC.20}$$

Although (EC.20) is a quadratic program, its optimal value admits a closed-form expression, as stated in the following lemma. The proof is given in Section EC.1.15.

Lemma EC.6. *The following properties hold for any $r\in[R]$:*

(a) If $\Delta^r_\Sigma\le 0$ or $\Delta^r_\mu\ge 0$, then $G^r=\min\{\Delta^r_\Sigma+\Delta^r_\mu,0\}$. Moreover, $\Delta^r_\Sigma+\Delta^r_\mu$ is achieved at $(a_r,b_r)=(1,1)$; 0 is achieved at $a_r=0$ and any $b_r\ge 1$;

(b) If $\Delta^r_\Sigma>0$ and $\Delta^r_\mu<0$, then $G^r=\Delta^r_\mu$. Moreover, Δ^r_μ is approached when $b_r=1/a_r$ and $a_r\downarrow 0$.

Lemma EC.6 simplifies the gap (EC.19) by establishing the infimum of function $g_r(a_r,b_r)$. After eliminating a_r and b_r, the only remaining variable is c_r, subject to $c_r\ge 0$ and $\sum_{r\in[R]}c_r=1$. Naturally, the minimum of (EC.19) is attained when all weight is assigned to the cluster r_* corresponding to the smallest g_{r_*} value.

Lemma EC.6 implies that $G^r=\min\{\Delta^r_\Sigma+\Delta^r_\mu,\Delta^r_\mu,0\}$ for all $r\in[R]$. After minimizing over a_r and b_r, the only remaining variables in (EC.19) are the weights c_r, which satisfy $c_r\ge 0$ and $\sum_{r\in[R]}c_r=1$. Hence, the best bound in this family is obtained by assigning all weight to a cluster with the smallest value of G^r. In particular, $\inf_{(a_r,b_r,c_r)_{r\in[R]}\in\Omega}\sum_{r\in[R]}c_rg_r(a_r,b_r)=\min_{r\in[R]}G^r$.

We now complete the proof of Theorem 3. Combining Lemmas EC.5 and EC.6, we obtain

$$\begin{aligned}0\le\eta(\Pi_B-\Pi_C)&\le\inf_{(a_r,b_r,c_r)_{r\in[R]}\in\Omega}\sum_{r\in[R]}c_rg_r(a_r,b_r)-\sum_{r\in[R]}(\Delta^r_{\Sigma,c}+\Delta^r_\mu)\\&=\min_{\substack{c_r\ge 0\\ \sum_{r\in[R]}c_r=1}}\sum_{r\in[R]}c_rG^r-\sum_{r\in[R]}(\Delta^r_{\Sigma,c}+\Delta^r_\mu).\end{aligned}$$

By the definition of Φ_{CF}, $\Phi_{CF}(\boldsymbol{y}_*,\bar v_*)=\sum_{r\in[R]}(\Delta^r_{\Sigma,c}+\Delta^r_\mu)-\bar v_*x_*$. Moreover, because the minimization over c is over the probability simplex and the objective is linear,

$$\min_{\substack{c_r\ge 0\\ \sum_{r\in[R]}c_r=1}}\sum_{r\in[R]}c_rG^r=\min_{r\in[R]}G^r.$$

Therefore, $0\le\eta(\Pi_B-\Pi_C)\le\min_{r\in[R]}G^r-\Phi_{CF}(\boldsymbol{y}_*,\bar v_*)-\bar v_*x_*$.

Finally, by Lemma EC.6, $G^r=\min\{\Delta^r_\Sigma+\Delta^r_\mu,\Delta^r_\mu,0\}$. Defining $\Delta^r:=\Delta^r_\Sigma+\Delta^r_\mu=\left\langle\operatorname{diag}(\boldsymbol{p}^r),\boldsymbol{Q}^r_*(\boldsymbol{\Sigma}^{1/2})^r\right\rangle+\left[\boldsymbol{p}^r\circ(\boldsymbol{\mu}^r-\boldsymbol{y}^r_*)\right]^\top\boldsymbol{t}^r_*$, we obtain $G^r=\min\{\Delta^r,\Delta^r_\mu,0\}$. Since $\Pi_B=\Pi$ and $\Pi_C=\underline{\Pi}$, the desired bound follows.

EC.1.13. Proof of Lemma EC.4

Let $\widehat{\boldsymbol{A}}$ be the matrix obtained from $\boldsymbol{A}$ by simultaneously permuting its rows and columns so that the variables are ordered cluster by cluster:

$$\widehat{\boldsymbol{A}} = \begin{pmatrix} s & \boldsymbol{u}_1^\top & \boldsymbol{v}_1^\top & \cdots & \boldsymbol{u}_R^\top & \boldsymbol{v}_R^\top \\ \boldsymbol{u}_1 & \boldsymbol{U}_1 & \boldsymbol{W}_1^\top & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{v}_1 & \boldsymbol{W}_1 & \boldsymbol{V}_1 & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{u}_R & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{U}_R & \boldsymbol{W}_R^\top \\ \boldsymbol{v}_R & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{W}_R & \boldsymbol{V}_R \end{pmatrix}.$$

Because simultaneous row and column permutations preserve positive semidefiniteness, $\boldsymbol{A} \succeq 0$ if and only if $\widehat{\boldsymbol{A}} \succeq 0$.

By definition, $\widehat{\boldsymbol{A}} \succeq 0$ is equivalent to $\boldsymbol{x}^\top \widehat{\boldsymbol{A}} \boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^{2n+1}$, or equivalently, $\inf_{\boldsymbol{x}\in\mathbb{R}^{2n+1}} \boldsymbol{x}^\top \widehat{\boldsymbol{A}} \boldsymbol{x} \geq 0$. Using the same argument as in (EC.3), it is enough to consider vectors of the form $\boldsymbol{x} = (1, \boldsymbol{\lambda}_1, \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\lambda}_R, \boldsymbol{\gamma}_R)$, where $\boldsymbol{\lambda}_r, \boldsymbol{\gamma}_r \in \mathbb{R}^{n_r}$ for all $r \in [R]$. Hence, $\widehat{\boldsymbol{A}} \succeq 0$ is equivalent to

$$\inf_{\{\boldsymbol{\lambda}_r, \boldsymbol{\gamma}_r \in \mathbb{R}^{n_r}\}_{r\in[R]}} \left\{ s + \sum_{r=1}^{R} \left(\boldsymbol{\lambda}_r^\top \boldsymbol{U}_r \boldsymbol{\lambda}_r + \boldsymbol{\gamma}_r^\top \boldsymbol{V}_r \boldsymbol{\gamma}_r + 2\boldsymbol{u}_r^\top \boldsymbol{\lambda}_r + 2\boldsymbol{v}_r^\top \boldsymbol{\gamma}_r + 2\boldsymbol{\lambda}_r^\top \boldsymbol{W}_r \boldsymbol{\gamma}_r \right) \right\} \geq 0.$$

Since the minimization variables are separable across clusters, this condition is equivalent to

$$s + \sum_{r=1}^{R} \inf_{\boldsymbol{\lambda}_r, \boldsymbol{\gamma}_r \in \mathbb{R}^{n_r}} \left(\boldsymbol{\lambda}_r^\top \boldsymbol{U}_r \boldsymbol{\lambda}_r + \boldsymbol{\gamma}_r^\top \boldsymbol{V}_r \boldsymbol{\gamma}_r + 2\boldsymbol{u}_r^\top \boldsymbol{\lambda}_r + 2\boldsymbol{v}_r^\top \boldsymbol{\gamma}_r + 2\boldsymbol{\lambda}_r^\top \boldsymbol{W}_r \boldsymbol{\gamma}_r \right) \geq 0. \tag{EC.21}$$

For each $r \in [R]$, define $\alpha_r = \inf_{\boldsymbol{\lambda}_r, \boldsymbol{\gamma}_r \in \mathbb{R}^{n_r}} \left(\boldsymbol{\lambda}_r^\top \boldsymbol{U}_r \boldsymbol{\lambda}_r + \boldsymbol{\gamma}_r^\top \boldsymbol{V}_r \boldsymbol{\gamma}_r + 2\boldsymbol{u}_r^\top \boldsymbol{\lambda}_r + 2\boldsymbol{v}_r^\top \boldsymbol{\gamma}_r + 2\boldsymbol{\lambda}_r^\top \boldsymbol{W}_r \boldsymbol{\gamma}_r \right)$. Then (EC.21) becomes $s + \sum_{r\in[R]} \alpha_r \geq 0$.

We next show that this condition holds if and only if there exist scalars $\{s_r\}_{r\in[R]}$ such that $\sum_{r\in[R]} s_r = s$ and $s_r + \alpha_r \geq 0$ for all $r \in [R]$. To prove this claim, first suppose that $s + \sum_{r\in[R]} \alpha_r \geq 0$. Define $s_1 = -\alpha_1 + s + \sum_{r\in[R]} \alpha_r$ and $s_r = -\alpha_r$, $r = 2, \ldots, R$. Then $\sum_{r\in[R]} s_r = s$ and $s_r + \alpha_r \geq 0$ for all $r \in [R]$. Conversely, if such $\{s_r\}_{r\in[R]}$ exist, then $s + \sum_{r\in[R]} \alpha_r = \sum_{r\in[R]} (s_r + \alpha_r) \geq 0$. This proves the claim.

Finally, for each $r \in [R]$, the condition $s_r + \alpha_r \geq 0$ is equivalent to

$$\begin{aligned} & s_r + \inf_{\boldsymbol{\lambda}_r, \boldsymbol{\gamma}_r \in \mathbb{R}^{n_r}} \left(\boldsymbol{\lambda}_r^\top \boldsymbol{U}_r \boldsymbol{\lambda}_r + \boldsymbol{\gamma}_r^\top \boldsymbol{V}_r \boldsymbol{\gamma}_r + 2\boldsymbol{u}_r^\top \boldsymbol{\lambda}_r + 2\boldsymbol{v}_r^\top \boldsymbol{\gamma}_r + 2\boldsymbol{\lambda}_r^\top \boldsymbol{W}_r \boldsymbol{\gamma}_r \right) \geq 0 \\ \iff \quad & s_r + \boldsymbol{\lambda}_r^\top \boldsymbol{U}_r \boldsymbol{\lambda}_r + \boldsymbol{\gamma}_r^\top \boldsymbol{V}_r \boldsymbol{\gamma}_r + 2\boldsymbol{u}_r^\top \boldsymbol{\lambda}_r + 2\boldsymbol{v}_r^\top \boldsymbol{\gamma}_r + 2\boldsymbol{\lambda}_r^\top \boldsymbol{W}_r \boldsymbol{\gamma}_r \geq 0, \ \forall \boldsymbol{\lambda}_r, \boldsymbol{\gamma}_r \in \mathbb{R}^{n_r} \\ \iff \quad & \begin{pmatrix} 1 \\ \boldsymbol{\lambda}_r \\ \boldsymbol{\gamma}_r \end{pmatrix}^\top \begin{pmatrix} s_r & \boldsymbol{u}_r^\top & \boldsymbol{v}_r^\top \\ \boldsymbol{u}_r & \boldsymbol{U}_r & \boldsymbol{W}_r^\top \\ \boldsymbol{v}_r & \boldsymbol{W}_r & \boldsymbol{V}_r \end{pmatrix} \begin{pmatrix} 1 \\ \boldsymbol{\lambda}_r \\ \boldsymbol{\gamma}_r \end{pmatrix} \geq 0, \ \forall \boldsymbol{\lambda}_r, \boldsymbol{\gamma}_r \in \mathbb{R}^{n_r} \\ \iff \quad & \boldsymbol{A}_r \succeq 0, \end{aligned} \tag{EC.22}$$

where the last equivalence follows from the same argument as in (EC.3). Therefore, $\boldsymbol{A} \succeq 0$ if and only if there exist scalars $\{s_r\}_{r\in[R]}$ such that $\sum_{r\in[R]} s_r = s$ and $\boldsymbol{A}_r \succeq 0$ for all $r \in [R]$. This completes the proof.

EC.1.14. Proof of Lemma EC.5

Suppose $(\boldsymbol{y}_*, \bar{v}_*)$ is an optimal solution to Problem (C-R). For Problem (C-F) evaluated at $(\boldsymbol{y}, \bar{v}) = (\boldsymbol{y}_*, \bar{v}_*)$, let $(\{\boldsymbol{T}_*^r, \boldsymbol{Q}_*^r, \boldsymbol{t}_*^r\}_{r\in[R]}, x_*)$ be an associated optimal solution. We construct a feasible solution $(\boldsymbol{X}_0, \boldsymbol{Y}_0)$ to the full-information factorized model (B-F) and compare its objective value with that of Problem (C-F). This yields the bound in (EC.19).

For each $r \in [R]$, feasibility of Problem (C-F) gives

$$\boldsymbol{X}_*^r := \begin{pmatrix} 1 & \boldsymbol{0}^\top & (\boldsymbol{t}_*^r)^\top \\ \boldsymbol{0} & \boldsymbol{I}_{n_r} & (\boldsymbol{Q}_*^r)^\top \\ \boldsymbol{t}_*^r & \boldsymbol{Q}_*^r & \boldsymbol{T}_*^r \end{pmatrix} \succeq 0, \; \boldsymbol{Y}_*^r := \begin{pmatrix} x_* & x_* & (\boldsymbol{t}_*^r)^\top \\ x_* & 1 & (\boldsymbol{t}_*^r)^\top \\ \boldsymbol{t}_*^r & \boldsymbol{t}_*^r & \boldsymbol{T}_*^r \end{pmatrix} \in \mathrm{CBQP}(n_r).$$

Define the joint parameter set

$$\Omega := \left\{ (a_r, b_r, c_r)_{r\in[R]} \;\middle|\; a_r, b_r, c_r \ge 0,\; 0 \le a_r \le 1 \le b_r,\; a_r b_r \le 1,\; \sum_{r\in[R]} c_r = 1 \right\}.$$

For any $(a_r, b_r, c_r)_{r\in[R]} \in \Omega$, define the following block matrices, with blocks ordered by cluster:

$$\boldsymbol{X}_0 = \left(\begin{array}{c|ccc|ccc} 1 & \boldsymbol{0}^\top & \cdots & \boldsymbol{0}^\top & a_1b_1c_1(\boldsymbol{t}_*^1)^\top & \cdots & a_Rb_Rc_R(\boldsymbol{t}_*^R)^\top \\ \hline \boldsymbol{0} & & & & a_1c_1(\boldsymbol{Q}_*^1)^\top & & \boldsymbol{0} \\ \vdots & & \boldsymbol{I}_n & & & \ddots & \\ \boldsymbol{0} & & & & \boldsymbol{0} & & a_Rc_R(\boldsymbol{Q}_*^R)^\top \\ \hline a_1b_1c_1\boldsymbol{t}_*^1 & a_1c_1\boldsymbol{Q}_*^1 & & \boldsymbol{0} & a_1b_1c_1\boldsymbol{T}_*^1 & & \boldsymbol{0} \\ \vdots & & \ddots & & & \ddots & \\ a_Rb_Rc_R\boldsymbol{t}_*^R & \boldsymbol{0} & & a_Rc_R\boldsymbol{Q}_*^R & \boldsymbol{0} & & a_Rb_Rc_R\boldsymbol{T}_*^R \end{array}\right), \text{and}$$

$$\boldsymbol{Y}_0 = \begin{pmatrix} x_* & x_* & a_1b_1c_1(\boldsymbol{t}_*^1)^\top & \cdots & a_Rb_Rc_R(\boldsymbol{t}_*^R)^\top \\ x_* & 1 & a_1b_1c_1(\boldsymbol{t}_*^1)^\top & \cdots & a_Rb_Rc_R(\boldsymbol{t}_*^R)^\top \\ a_1b_1c_1\boldsymbol{t}_*^1 & a_1b_1c_1\boldsymbol{t}_*^1 & a_1b_1c_1\boldsymbol{T}_*^1 & & \boldsymbol{0} \\ \vdots & \vdots & & \ddots & \\ a_Rb_Rc_R\boldsymbol{t}_*^R & a_Rb_Rc_R\boldsymbol{t}_*^R & \boldsymbol{0} & & a_Rb_Rc_R\boldsymbol{T}_*^R \end{pmatrix}.$$

The parameters c_r allocate the leading scalar in the block decomposition of Lemma EC.4, while a_r and b_r provide flexibility in scaling the contributions from $\boldsymbol{Q}_*^r$, $\boldsymbol{t}_*^r$, and $\boldsymbol{T}_*^r$. We next verify that $(\boldsymbol{X}_0, \boldsymbol{Y}_0)$ is feasible for Problem (B-F).

Step 1: Verifying the PSD constraint in Problem (B-F)**.**

By Lemma EC.4, it suffices to show that, for each $r \in [R]$,

$$\boldsymbol{X}_0^r := \begin{pmatrix} c_r & \boldsymbol{0}^\top & a_rb_rc_r(\boldsymbol{t}_*^r)^\top \\ \boldsymbol{0} & \boldsymbol{I}_{n_r} & a_rc_r(\boldsymbol{Q}_*^r)^\top \\ a_rb_rc_r\boldsymbol{t}_*^r & a_rc_r\boldsymbol{Q}_*^r & a_rb_rc_r\boldsymbol{T}_*^r \end{pmatrix} \succeq 0.$$

If $c_r = 0$, then $\boldsymbol{X}_0^r = \begin{pmatrix} 0 & \boldsymbol{0}^\top & \boldsymbol{0}^\top \\ \boldsymbol{0} & \boldsymbol{I}_{n_r} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \succeq 0$. If $a_r = 0$, then $\boldsymbol{X}_0^r = \begin{pmatrix} c_r & \boldsymbol{0}^\top & \boldsymbol{0}^\top \\ \boldsymbol{0} & \boldsymbol{I}_{n_r} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \succeq 0$. It remains to consider the case $c_r > 0$ and $a_r > 0$. Let $\epsilon_r := \sqrt{a_r/b_r}$. Since $0 < a_r \le 1 \le b_r$ and $a_r b_r \le 1$, we have $0 < \epsilon_r \le 1$ and $a_r/\epsilon_r \le b_r$. We decompose $\boldsymbol{X}_0^r = c_r(\boldsymbol{X}_1^r + \boldsymbol{X}_2^r + \boldsymbol{X}_3^r)$, where

$$\boldsymbol{X}_1^r = \begin{pmatrix} \epsilon_r & \mathbf{0}^\top & a_r(\boldsymbol{t}_*^r)^\top \\ \mathbf{0} & \epsilon_r \boldsymbol{I}_{n_r} & a_r(\boldsymbol{Q}_*^r)^\top \\ a_r \boldsymbol{t}_*^r & a_r \boldsymbol{Q}_*^r & \frac{a_r^2}{\epsilon_r}\boldsymbol{T}_*^r \end{pmatrix}, \ \boldsymbol{X}_2^r = \begin{pmatrix} 1-\epsilon_r & \mathbf{0}^\top & a_r(b_r-1)(\boldsymbol{t}_*^r)^\top \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ a_r(b_r-1)\boldsymbol{t}_*^r & \mathbf{0} & a_r\left(b_r - \frac{a_r}{\epsilon_r}\right)\boldsymbol{T}_*^r \end{pmatrix}, \ \boldsymbol{X}_3^r = \begin{pmatrix} 0 & \mathbf{0}^\top & \mathbf{0}^\top \\ \mathbf{0} & \left(\frac{1}{c_r} - \epsilon_r\right)\boldsymbol{I}_{n_r} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

We show that each of the three matrices is PSD.

First, $\boldsymbol{X}_1^r \succeq 0$ follows from $\boldsymbol{X}_*^r \succeq 0$. Indeed, $\boldsymbol{X}_1^r$ can be obtained by the congruence transformation $\boldsymbol{X}_1^r = \boldsymbol{D}_r \boldsymbol{X}_*^r \boldsymbol{D}_r$, where

$$\boldsymbol{D}_r = \begin{pmatrix} \sqrt{\epsilon_r} & \mathbf{0}^\top & \mathbf{0}^\top \\ \mathbf{0} & \sqrt{\epsilon_r}\boldsymbol{I}_{n_r} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dfrac{a_r}{\sqrt{\epsilon_r}}\boldsymbol{I}_{n_r} \end{pmatrix}.$$

Second, we verify $\boldsymbol{X}_2^r \succeq 0$. Since $0 < \epsilon_r \le 1$ and $a_r/\epsilon_r \le b_r$, the diagonal coefficients $1-\epsilon_r$ and $a_r(b_r - a_r/\epsilon_r)$ are nonnegative. If $\epsilon_r = 1$, then $a_r = b_r = 1$ and $\boldsymbol{X}_2^r = \mathbf{0}$. Otherwise, $1 - \epsilon_r > 0$. From $\boldsymbol{X}_*^r \succeq 0$, taking the Schur complement of the leading identity block gives $\boldsymbol{T}_*^r - \boldsymbol{t}_*^r(\boldsymbol{t}_*^r)^\top - \boldsymbol{Q}_*^r(\boldsymbol{Q}_*^r)^\top \succeq 0$, and hence $\boldsymbol{T}_*^r \succeq \boldsymbol{t}_*^r(\boldsymbol{t}_*^r)^\top$. Therefore, it is sufficient to verify $(1-\epsilon_r)a_r(b_r - a_r/\epsilon_r) - a_r^2(b_r-1)^2 \ge 0$, or equivalently, $(1-\epsilon_r)(b_r - a_r/\epsilon_r) - a_r(b_r-1)^2 \ge 0$. Substituting $\epsilon_r = \sqrt{a_r/b_r}$ gives

$$\begin{aligned}(1-\epsilon_r)\left(b_r - \frac{a_r}{\epsilon_r}\right) - a_r(b_r-1)^2 =& (\sqrt{b_r} - \sqrt{a_r})^2 - a_r(b_r-1)^2 \\ =& (\sqrt{b_r} - \sqrt{a_r} - \sqrt{a_r}(b_r-1))(\sqrt{b_r} - \sqrt{a_r} + \sqrt{a_r}(b_r-1)) \\ =& \sqrt{b_r}(1 - \sqrt{a_r b_r})(\sqrt{b_r} - \sqrt{a_r} + \sqrt{a_r}(b_r-1)) \ge 0,\end{aligned}$$

where the last inequality follows from $a_r b_r \le 1$, $b_r \ge 1$, and $b_r \ge a_r$. Hence, $\boldsymbol{X}_2^r \succeq 0$.

Third, $\boldsymbol{X}_3^r \succeq 0$ holds because $0 < c_r \le 1$ and $0 < \epsilon_r \le 1$, which imply $1/c_r - \epsilon_r \ge 0$. Therefore, $\boldsymbol{X}_0^r \succeq 0$ for all $r \in [R]$, and Lemma EC.4 implies $\boldsymbol{X}_0 \succeq 0$.

Step 2: Verifying the CBQP constraint in Problem (B-F)**.**

It remains to show that $\boldsymbol{Y}_0 \in \mathrm{CBQP}(n)$. Since $\boldsymbol{Y}_*^r \in \mathrm{CBQP}(n_r)$, embedding this matrix into the full n-dimensional space by placing its nonzero $\boldsymbol{t}_*^r$ and $\boldsymbol{T}_*^r$ components in cluster r and zeros elsewhere gives a matrix $\boldsymbol{M}_r \in \mathrm{CBQP}(n)$:

$$\boldsymbol{M}_r = \begin{pmatrix} x_* & x_* & \mathbf{0}^\top & (\boldsymbol{t}_*^r)^\top & \mathbf{0}^\top \\ x_* & 1 & \mathbf{0}^\top & (\boldsymbol{t}_*^r)^\top & \mathbf{0}^\top \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{t}_*^r & \boldsymbol{t}_*^r & \mathbf{0} & \boldsymbol{T}_*^r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix},$$

where the displayed block structure places cluster r in its corresponding coordinates. Also, $\boldsymbol{M}_0 := \begin{pmatrix} x_* & x_* & \mathbf{0}^\top \\ x_* & 1 & \mathbf{0}^\top \\ \mathbf{0} & \mathbf{0} & \mathbf{0}_{n\times n} \end{pmatrix} \in \mathrm{CBQP}(n)$. Because $\mathrm{CBQP}(n)$ is convex and $0 \le a_r b_r \le 1$, for each $r \in [R]$, $\boldsymbol{Y}_r :=$

$a_rb_r\boldsymbol{M}_r + (1-a_rb_r)\boldsymbol{M}_0 \in \text{CBQP}(n)$. Finally, since $c_r \geq 0$ and $\sum_{r\in[R]} c_r = 1$, $\boldsymbol{Y}_0 = \sum_{r\in[R]} c_r\boldsymbol{Y}_r \in \text{CBQP}(n)$. Thus, $(\boldsymbol{X}_0, \boldsymbol{Y}_0)$ is feasible for Problem (B-F).

Step 3: Objective comparison.

Plugging $(\boldsymbol{X}_0, \boldsymbol{Y}_0)$ into Problem (B-F) gives $\bar{\Pi}_B = \bar{v}_* + (\boldsymbol{p}-\boldsymbol{c})^\top \boldsymbol{y}_* + \frac{1}{\eta}\Phi_B$, where

$$\Phi_B = \left\langle \text{diag}(\boldsymbol{p}), \begin{pmatrix} a_1c_1\boldsymbol{Q}_*^1 & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & a_Rc_R\boldsymbol{Q}_*^R \end{pmatrix} \boldsymbol{\Sigma}^{1/2} \right\rangle + \sum_{r\in[R]} \left[\boldsymbol{p}^r \circ (\boldsymbol{\mu}^r - \boldsymbol{y}_*^r)\right]^\top (a_rb_rc_r\boldsymbol{t}_*^r) - \bar{v}_*x_*.$$

Since $(\boldsymbol{\Sigma}^{1/2})^r$ is the rth diagonal block of $\boldsymbol{\Sigma}^{1/2}$, Φ_B can be rewritten as $\Phi_B = \sum_{r\in[R]} a_rc_r \left\langle \text{diag}(\boldsymbol{p}^r), \boldsymbol{Q}_*^r(\boldsymbol{\Sigma}^{1/2})^r \right\rangle + \sum_{r\in[R]} a_rb_rc_r \left[\boldsymbol{p}^r \circ (\boldsymbol{\mu}^r - \boldsymbol{y}_*^r)\right]^\top \boldsymbol{t}_*^r - \bar{v}_*x_*$. Recall that $\Pi_C = \bar{v}_* + (\boldsymbol{p}-\boldsymbol{c})^\top \boldsymbol{y}_* + \frac{1}{\eta}\Phi_C$, where $\Phi_C = \sum_{r\in[R]} \left(\left\langle \text{diag}(\boldsymbol{p}^r), \boldsymbol{Q}_*^r(\boldsymbol{\Sigma}^r)^{1/2} \right\rangle + \left[\boldsymbol{p}^r \circ (\boldsymbol{\mu}^r - \boldsymbol{y}_*^r)\right]^\top \boldsymbol{t}_*^r \right) - \bar{v}_*x_*$.

Define $\Delta_\Sigma^r = \left\langle \text{diag}(\boldsymbol{p}^r), \boldsymbol{Q}_*^r(\boldsymbol{\Sigma}^{1/2})^r \right\rangle$, $\Delta_{\Sigma,c}^r = \left\langle \text{diag}(\boldsymbol{p}^r), \boldsymbol{Q}_*^r(\boldsymbol{\Sigma}^r)^{1/2} \right\rangle$, and $\Delta_\mu^r = \left[\boldsymbol{p}^r \circ (\boldsymbol{\mu}^r - \boldsymbol{y}_*^r)\right]^\top \boldsymbol{t}_*^r$. Then

$$\begin{aligned}\eta(\bar{\Pi}_B - \Pi_C) = \sum_{r\in[R]} \left(a_rc_r\Delta_\Sigma^r - \Delta_{\Sigma,c}^r + (a_rb_rc_r - 1)\Delta_\mu^r\right) &= \sum_{r\in[R]} \left(a_rc_r\Delta_\Sigma^r + a_rb_rc_r\Delta_\mu^r\right) - \sum_{r\in[R]} (\Delta_{\Sigma,c}^r + \Delta_\mu^r) \\ &= \sum_{r\in[R]} c_rg_r(a_r,b_r) - \sum_{r\in[R]} (\Delta_{\Sigma,c}^r + \Delta_\mu^r),\end{aligned}$$

where $g_r(a_r,b_r) = a_r\Delta_\Sigma^r + a_rb_r\Delta_\mu^r$.

Step 4: Deriving the bound.

Since $(\boldsymbol{X}_0, \boldsymbol{Y}_0)$ is feasible for Problem (B-F), which is a minimization problem, its objective value provides an upper bound on the optimal value of Problem (B-F) evaluated at $(\boldsymbol{y}_*, \bar{v}_*)$. Therefore, $\Pi_B \leq \bar{\Pi}_B$. Moreover, the correlation-splitting model provides a lower bound for the full-information model, so $\Pi_B \geq \Pi_C$. Combining these inequalities with the objective comparison above gives

$$0 \leq \eta(\Pi_B - \Pi_C) \leq \eta(\bar{\Pi}_B - \Pi_C) \quad = \sum_{r\in[R]} c_rg_r(a_r,b_r) - \sum_{r\in[R]} (\Delta_{\Sigma,c}^r + \Delta_\mu^r).$$

Since this holds for any $(a_r,b_r,c_r)_{r\in[R]} \in \Omega$, taking the infimum over Ω yields

$$0 \leq \eta(\Pi_B - \Pi_C) \leq \inf_{(a_r,b_r,c_r)_{r\in[R]}\in\Omega} \sum_{r\in[R]} c_rg_r(a_r,b_r) - \sum_{r\in[R]} (\Delta_{\Sigma,c}^r + \Delta_\mu^r).$$

This proves the lemma.

EC.1.15. Proof of Lemma EC.6

Recall that $g_r(a_r,b_r) = a_r\Delta_\Sigma^r + a_rb_r\Delta_\mu^r$, $0 \leq a_r \leq 1 \leq b_r$, and $a_rb_r \leq 1$.

Proof of part (a).

The condition in part (a) can be divided into two cases: (i) $\Delta_\mu^r \geq 0$ and (ii) $\Delta_\mu^r < 0$ with $\Delta_\Sigma^r \leq 0$.

First, suppose that $\Delta_\mu^r \geq 0$. Since $b_r \geq 1$, we have $g_r(a_r,b_r) \geq a_r(\Delta_\Sigma^r + \Delta_\mu^r)$. If $\Delta_\Sigma^r + \Delta_\mu^r \leq 0$, then $a_r(\Delta_\Sigma^r + \Delta_\mu^r) \geq \Delta_\Sigma^r + \Delta_\mu^r$, because $a_r \leq 1$. The lower bound is achieved at $(a_r,b_r) = (1,1)$. If

$\Delta_\Sigma^r + \Delta_\mu^r > 0$, then $g_r(a_r, b_r) \geq 0$, and 0 is achieved at $a_r = 0$ with any $b_r \geq 1$. Hence, when $\Delta_\mu^r \geq 0$, $\inf g_r(a_r, b_r) = \min\{\Delta_\Sigma^r + \Delta_\mu^r, 0\}$.

Second, suppose that $\Delta_\mu^r < 0$ and $\Delta_\Sigma^r \leq 0$. Since $a_r b_r \leq 1$ and $a_r \leq 1$, we have $a_r \Delta_\Sigma^r \geq \Delta_\Sigma^r$ and $a_r b_r \Delta_\mu^r \geq \Delta_\mu^r$. Therefore, $g_r(a_r, b_r) \geq \Delta_\Sigma^r + \Delta_\mu^r$. This lower bound is achieved at $(a_r, b_r) = (1, 1)$. Thus, whenever $\Delta_\mu^r \geq 0$ or $\Delta_\Sigma^r \leq 0$, $G^r = \inf_{\substack{0 \leq a_r \leq 1 \leq b_r \\ a_r b_r \leq 1}} g_r(a_r, b_r) = \min\{\Delta_\Sigma^r + \Delta_\mu^r, 0\}$.

Proof of part (b).

Now suppose that $\Delta_\Sigma^r > 0$ and $\Delta_\mu^r < 0$. Since $\Delta_\mu^r < 0$ and $a_r b_r \leq 1$, we have $g_r(a_r, b_r) \geq a_r \Delta_\Sigma^r + \Delta_\mu^r$. Because $\Delta_\Sigma^r > 0$ and $a_r \geq 0$, the right-hand side is bounded below by Δ_μ^r. This lower bound is not attained, but it is approached by taking $b_r = 1/a_r$ and letting $a_r \downarrow 0$. Therefore, $G^r = \inf_{\substack{0 \leq a_r \leq 1 \leq b_r \\ a_r b_r \leq 1}} g_r(a_r, b_r) = \Delta_\mu^r$. This completes the proof.

EC.1.16. Proof of Theorem 4

Without loss of generality, suppose $\tilde{r} = 2$ and $\sigma_i \leq \sigma$ for all $i \in N_2$. In the proof of this theorem and in the subsequent related proofs, we use superscripts (1) and (2) to denote block operators associated with the first and second clusters, respectively. Accordingly, we write $G^{(r)}$, $\Delta^{(r)}$, and $\Delta_c^{(r)}$ for the corresponding quantities G^r, Δ^r, and Δ_c^r, respectively, under this two-cluster block notation. Throughout the proof, write $\boldsymbol{d} = (\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)})$ and $\boldsymbol{y} = (\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)})$, and define $\phi(\boldsymbol{d}; \boldsymbol{y}, \bar{v}) := \min(\boldsymbol{p}^\top (\boldsymbol{d} - \boldsymbol{y})^- - \bar{v}, 0)$. For each $r = 1, 2$, let $\mathscr{F}_r = \left\{ F_r \in \mathcal{M}(\mathbb{R}^{n_r}) \;\middle|\; \mathbb{E}_{F_r}[\tilde{\boldsymbol{d}}^{(r)}] = \boldsymbol{\mu}^{(r)},\ \mathbb{E}_{F_r}[\tilde{\boldsymbol{d}}^{(r)} (\tilde{\boldsymbol{d}}^{(r)})^\top] = \boldsymbol{\Sigma}^{(r)} + \boldsymbol{\mu}^{(r)} (\boldsymbol{\mu}^{(r)})^\top \right\}$. We first establish a bound on the difference between the expectation of a Lipschitz function and its value evaluated at the mean. The proof of Lemma EC.7 is provided in Section EC.1.17.

LEMMA EC.7. *For any random vector $\tilde{\boldsymbol{x}} \in \mathbb{R}^m$ with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, and any measurable function $f : \mathbb{R}^m \to \mathbb{R}$ that is L-Lipschitz with respect to $\|\cdot\|_1$, it holds that $|\mathbb{E}[f(\tilde{\boldsymbol{x}})] - f(\boldsymbol{\mu})| \leq L\sqrt{m \operatorname{tr}(\boldsymbol{\Sigma})}$.*

Proof of part (a).

We use Lemma EC.7 to construct an upper bound on Π and a lower bound on $\underline{\Pi}$, from which an upper bound on $\Pi - \underline{\Pi}$ follows.

Step 1: A uniform perturbation bound in cluster 2.

Fix $(\boldsymbol{y}, \bar{v})$ and any feasible joint distribution $F \in \mathscr{F}$ for $\tilde{\boldsymbol{d}} = (\tilde{\boldsymbol{d}}^{(1)}, \tilde{\boldsymbol{d}}^{(2)})$. For each fixed realization $\boldsymbol{d}^{(1)}$, define a function of $\boldsymbol{z} \in \mathbb{R}^{n_2}$ by $g(\boldsymbol{z}; \boldsymbol{d}^{(1)}, \boldsymbol{y}, \bar{v}) := \min\big((\boldsymbol{p}^{(1)})^\top (\boldsymbol{d}^{(1)} - \boldsymbol{y}^{(1)})^- + (\boldsymbol{p}^{(2)})^\top (\boldsymbol{z} - \boldsymbol{y}^{(2)})^- - \bar{v}, 0\big)$. Here the semicolon indicates that g is viewed as a function of $\boldsymbol{z}$, while $\boldsymbol{d}^{(1)}$, $\boldsymbol{y}$, and $\bar{v}$ are fixed parameters. It can be verified that this function is $\|\boldsymbol{p}^{(2)}\|_\infty$-Lipschitz with respect to $\|\cdot\|_1$. Hence, for every realization $(\tilde{\boldsymbol{d}}^{(1)}, \tilde{\boldsymbol{d}}^{(2)})$, $g(\tilde{\boldsymbol{d}}^{(2)}; \tilde{\boldsymbol{d}}^{(1)}, \boldsymbol{y}, \bar{v}) \leq g(\boldsymbol{\mu}^{(2)}; \tilde{\boldsymbol{d}}^{(1)}, \boldsymbol{y}, \bar{v}) + \|\boldsymbol{p}^{(2)}\|_\infty \|\tilde{\boldsymbol{d}}^{(2)} - \boldsymbol{\mu}^{(2)}\|_1$.

Taking expectation with respect to the full joint distribution F gives $\mathbb{E}_F[\phi(\tilde{\boldsymbol{d}};\boldsymbol{y},\bar{v})] = \mathbb{E}_F[g(\tilde{\boldsymbol{d}}^{(2)};\tilde{\boldsymbol{d}}^{(1)},\boldsymbol{y},\bar{v})] \le \mathbb{E}_F[g(\boldsymbol{\mu}^{(2)};\tilde{\boldsymbol{d}}^{(1)},\boldsymbol{y},\bar{v})] + \|\boldsymbol{p}^{(2)}\|_\infty \mathbb{E}_F[\|\tilde{\boldsymbol{d}}^{(2)} - \boldsymbol{\mu}^{(2)}\|_1]$. Applying Lemma EC.7 to the marginal distribution of $\tilde{\boldsymbol{d}}^{(2)}$, with $f(\boldsymbol{z}) = \|\boldsymbol{z} - \boldsymbol{\mu}^{(2)}\|_1$, gives $\mathbb{E}_F[\|\tilde{\boldsymbol{d}}^{(2)} - \boldsymbol{\mu}^{(2)}\|_1] \le \sqrt{n_2 \operatorname{tr}(\boldsymbol{\Sigma}^{(2)})}$. Therefore, $\mathbb{E}_F[\phi(\tilde{\boldsymbol{d}};\boldsymbol{y},\bar{v})] \le \mathbb{E}_F[\phi((\tilde{\boldsymbol{d}}^{(1)},\boldsymbol{\mu}^{(2)});\boldsymbol{y},\bar{v})] + \|\boldsymbol{p}^{(2)}\|_\infty \sqrt{n_2 \operatorname{tr}(\boldsymbol{\Sigma}^{(2)})} \le \mathbb{E}_F[\phi((\tilde{\boldsymbol{d}}^{(1)},\boldsymbol{\mu}^{(2)});\boldsymbol{y},\bar{v})] + \|\boldsymbol{p}^{(2)}\|_\infty n_2\sigma$, where the last inequality uses $\operatorname{tr}(\boldsymbol{\Sigma}^{(2)}) \le n_2\sigma^2$.

Step 2: Upper bound on Π.

Taking the infimum over feasible joint distributions on both sides yields

$$\min_{F\in\mathscr{F}} \mathbb{E}_F[\phi(\tilde{\boldsymbol{d}};\boldsymbol{y},\bar{v})] \le \min_{F_1\in\mathscr{F}_1} \mathbb{E}_{F_1}\left[\phi((\tilde{\boldsymbol{d}}^{(1)},\boldsymbol{\mu}^{(2)});\boldsymbol{y},\bar{v})\right] + \|\boldsymbol{p}^{(2)}\|_\infty n_2\sigma.$$

The minimization on the right-hand side is over $\mathscr{F}_1$ because any $F_1 \in \mathscr{F}_1$ can be extended to a feasible joint distribution in $\mathscr{F}$ with the prescribed full covariance matrix. For example, one can construct $\tilde{\boldsymbol{d}}^{(2)}$ as an affine function of $\tilde{\boldsymbol{d}}^{(1)}$ plus independent noise using the Schur complement of $\boldsymbol{\Sigma}^{(1)}$ in $\boldsymbol{\Sigma}$. Substituting the above inequality into the definition of Π gives

$$\begin{aligned}\Pi &\le \max_{\boldsymbol{y},\bar{v}}\left\{\bar{v} + (\boldsymbol{p}-\boldsymbol{c})^\top\boldsymbol{y} + \frac{1}{\eta}\left(\min_{F_1\in\mathscr{F}_1}\mathbb{E}_{F_1}\left[\phi((\tilde{\boldsymbol{d}}^{(1)},\boldsymbol{\mu}^{(2)});\boldsymbol{y},\bar{v})\right] + \|\boldsymbol{p}^{(2)}\|_\infty n_2\sigma\right)\right\}\\ &= \max_{\boldsymbol{y},\bar{v}}\left\{\bar{v} + (\boldsymbol{p}-\boldsymbol{c})^\top\boldsymbol{y} + \frac{1}{\eta}\left(\min_{F_1\in\mathscr{F}_1}\mathbb{E}_{F_1}\left[\min\left((\boldsymbol{p}^{(1)})^\top(\tilde{\boldsymbol{d}}^{(1)}-\boldsymbol{y}^{(1)})^- + (\boldsymbol{p}^{(2)})^\top(\boldsymbol{\mu}^{(2)}-\boldsymbol{y}^{(2)})^- - \bar{v}, 0\right)\right] + \|\boldsymbol{p}^{(2)}\|_\infty n_2\sigma\right)\right\}.\end{aligned}$$

Define $\hat{v} := \bar{v} - (\boldsymbol{p}^{(2)})^\top(\boldsymbol{\mu}^{(2)} - \boldsymbol{y}^{(2)})^-$. Then the preceding inequality can be written as

$$\begin{aligned}\Pi \le &\max_{\boldsymbol{y}^{(1)},\hat{v}}\left\{\hat{v} + (\boldsymbol{p}^{(1)}-\boldsymbol{c}^{(1)})^\top\boldsymbol{y}^{(1)} + \frac{1}{\eta}\min_{F_1\in\mathscr{F}_1}\mathbb{E}_{F_1}\left[\min\left((\boldsymbol{p}^{(1)})^\top(\tilde{\boldsymbol{d}}^{(1)}-\boldsymbol{y}^{(1)})^- - \hat{v}, 0\right)\right]\right\}\\ &+ \max_{\boldsymbol{y}^{(2)}}\left\{(\boldsymbol{p}^{(2)}-\boldsymbol{c}^{(2)})^\top\boldsymbol{y}^{(2)} + (\boldsymbol{p}^{(2)})^\top(\boldsymbol{\mu}^{(2)}-\boldsymbol{y}^{(2)})^-\right\} + \frac{\|\boldsymbol{p}^{(2)}\|_\infty n_2}{\eta}\sigma.\end{aligned}$$

The first maximization is exactly $\Pi^{(1)}$ by definition. For the second maximization, one can verify that the maximizer is $\boldsymbol{y}^{(2)}_* = \boldsymbol{\mu}^{(2)}$, and the maximum value is $(\boldsymbol{p}^{(2)} - \boldsymbol{c}^{(2)})^\top\boldsymbol{\mu}^{(2)}$. Hence,

$$\Pi \le \Pi^{(1)} + (\boldsymbol{p}^{(2)} - \boldsymbol{c}^{(2)})^\top\boldsymbol{\mu}^{(2)} + \frac{n_2\|\boldsymbol{p}^{(2)}\|_\infty}{\eta}\sigma. \tag{EC.23}$$

Step 3: Lower bound on $\underline{\Pi}$.

From (14), we have $\underline{\Pi} \ge \Pi^{(1)} + \Pi^{(2)}$. We now bound $\Pi^{(2)}$ from below. Indeed,

$$\Pi^{(2)} = \max_{\boldsymbol{y}^{(2)},\bar{v}}\left\{\bar{v} + (\boldsymbol{p}^{(2)}-\boldsymbol{c}^{(2)})^\top\boldsymbol{y}^{(2)} + \frac{1}{\eta}\min_{F_2\in\mathscr{F}_2}\mathbb{E}_{F_2}\left[\min\left((\boldsymbol{p}^{(2)})^\top(\tilde{\boldsymbol{d}}^{(2)}-\boldsymbol{y}^{(2)})^- - \bar{v}, 0\right)\right]\right\}.$$

Applying Lemma EC.7 to the function $\boldsymbol{z} \mapsto \min\left((\boldsymbol{p}^{(2)})^\top(\boldsymbol{z} - \boldsymbol{y}^{(2)})^- - \bar{v}, 0\right)$, which can be verified to be $\|\boldsymbol{p}^{(2)}\|_\infty$-Lipschitz with respect to $\|\cdot\|_1$, gives

$$\Pi^{(2)} \ge \max_{\boldsymbol{y}^{(2)},\bar{v}}\left\{\bar{v} + (\boldsymbol{p}^{(2)}-\boldsymbol{c}^{(2)})^\top\boldsymbol{y}^{(2)} + \frac{1}{\eta}\min\left((\boldsymbol{p}^{(2)})^\top(\boldsymbol{\mu}^{(2)}-\boldsymbol{y}^{(2)})^- - \bar{v}, 0\right)\right\} - \frac{\|\boldsymbol{p}^{(2)}\|_\infty n_2}{\eta}\sigma.$$

For any given $\boldsymbol{y}^{(2)}$, the optimal value of $\bar{v}$ is $\bar{v}_* = (\boldsymbol{p}^{(2)})^\top(\boldsymbol{\mu}^{(2)} - \boldsymbol{y}^{(2)})^-$. Therefore,

$$\Pi^{(2)} \geq \max_{\boldsymbol{y}^{(2)}} \left\{ (\boldsymbol{p}^{(2)})^\top(\boldsymbol{\mu}^{(2)} - \boldsymbol{y}^{(2)})^- + (\boldsymbol{p}^{(2)} - \boldsymbol{c}^{(2)})^\top \boldsymbol{y}^{(2)} \right\} - \frac{\|\boldsymbol{p}^{(2)}\|_\infty n_2}{\eta}\sigma = (\boldsymbol{p}^{(2)} - \boldsymbol{c}^{(2)})^\top \boldsymbol{\mu}^{(2)} - \frac{\|\boldsymbol{p}^{(2)}\|_\infty n_2}{\eta}\sigma,$$

where the last equality follows because the maximizer is $\boldsymbol{y}_*^{(2)} = \boldsymbol{\mu}^{(2)}$. Hence,

$$\underline{\Pi} \geq \Pi^{(1)} + (\boldsymbol{p}^{(2)} - \boldsymbol{c}^{(2)})^\top \boldsymbol{\mu}^{(2)} - \frac{n_2\|\boldsymbol{p}^{(2)}\|_\infty}{\eta}\sigma. \tag{EC.24}$$

Step 4: Upper bound on $\Pi - \underline{\Pi}$.

Combining (EC.23) and (EC.24), we obtain

$$\Pi - \underline{\Pi} \leq \Pi^{(1)} + (\boldsymbol{p}^{(2)} - \boldsymbol{c}^{(2)})^\top \boldsymbol{\mu}^{(2)} + \frac{n_2\|\boldsymbol{p}^{(2)}\|_\infty}{\eta}\sigma - \left(\Pi^{(1)} + (\boldsymbol{p}^{(2)} - \boldsymbol{c}^{(2)})^\top \boldsymbol{\mu}^{(2)} - \frac{n_2\|\boldsymbol{p}^{(2)}\|_\infty}{\eta}\sigma \right) = \frac{2n_2\|\boldsymbol{p}^{(2)}\|_\infty}{\eta}\sigma,$$

which proves part (a).

Proof of part (b).

Let $(\boldsymbol{y}_*, \bar{v}_*)$ be an optimal solution to Problem (C-R), and let $(\{\boldsymbol{T}_*^{(r)}, \boldsymbol{Q}_*^{(r)}, \boldsymbol{t}_*^{(r)}\}_{r\in[R]}, x_*)$ be a corresponding optimal solution of Problem (C-F). Recall the definitions $\Delta^{(r)} = \Delta_\Sigma^{(r)} + \Delta_\mu^{(r)}$ and $\Delta_c^{(r)} = \Delta_{\Sigma,c}^{(r)} + \Delta_\mu^{(r)}$, $r \in [R]$, where $\Delta_\Sigma^{(r)} = \left\langle \operatorname{diag}(\boldsymbol{p}^{(r)}), \boldsymbol{Q}_*^{(r)}(\boldsymbol{\Sigma}^{1/2})^{(r)} \right\rangle$, $\Delta_{\Sigma,c}^{(r)} = \left\langle \operatorname{diag}(\boldsymbol{p}^{(r)}), \boldsymbol{Q}_*^{(r)}(\boldsymbol{\Sigma}^{(r)})^{1/2} \right\rangle$, and $\Delta_\mu^{(r)} = \left[\boldsymbol{p}^{(r)} \circ (\boldsymbol{\mu}^{(r)} - \boldsymbol{y}_*^{(r)}) \right]^\top \boldsymbol{t}_*^{(r)}$. By Theorem 3, and since $\Phi_{CF}(\boldsymbol{y}_*, \bar{v}_*) = \sum_{r\in[R]}(\Delta_{\Sigma,c}^{(r)} + \Delta_\mu^{(r)}) - \bar{v}_* x_*$, we have

$$\min_{r\in[R]} G^{(r)} - \Phi_{CF}(\boldsymbol{y}_*, \bar{v}_*) - \bar{v}_* x_* \leq \min_{r'\in[R]} \Delta^{(r')} - \sum_{r\in[R]} \Delta_c^{(r)}. \tag{EC.25}$$

To prove part (b), we first bound the difference between $(\boldsymbol{\Sigma}^{1/2})^{(r)}$ and $(\boldsymbol{\Sigma}^{(r)})^{1/2}$ under the Frobenius norm. Building on this norm bound, we derive the orders of $\Delta^{(1)} - \sum_{r\in[R]} \Delta_c^{(r)}$ and $\Delta^{(2)} - \sum_{r\in[R]} \Delta_c^{(r)}$. Finally, combining these bounds yields $\min_{r\in[R]} G^{(r)} - \Phi_{CF}(\boldsymbol{y}_*, \bar{v}_*) - \bar{v}_* x_* = O(\sigma)$. We now proceed step by step.

Step 1: Relating $(\boldsymbol{\Sigma}^{1/2})^{(r)}$ and $(\boldsymbol{\Sigma}^{(r)})^{1/2}$.

Write $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}^{(1)} & \boldsymbol{C}^\top \\ \boldsymbol{C} & \boldsymbol{\Sigma}^{(2)} \end{pmatrix}$, $\boldsymbol{\Sigma}^{1/2} = \begin{pmatrix} (\boldsymbol{\Sigma}^{1/2})^{(1)} & \boldsymbol{D}^\top \\ \boldsymbol{D} & (\boldsymbol{\Sigma}^{1/2})^{(2)} \end{pmatrix}$. From $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2}$, the $(1,1)$ block equality gives $\boldsymbol{\Sigma}^{(1)} = (\boldsymbol{\Sigma}^{1/2})^{(1)}(\boldsymbol{\Sigma}^{1/2})^{(1)} + \boldsymbol{D}^\top\boldsymbol{D}$. Hence, $\boldsymbol{\Sigma}^{(1)} - \boldsymbol{D}^\top\boldsymbol{D} \succeq 0$ and $(\boldsymbol{\Sigma}^{1/2})^{(1)} = (\boldsymbol{\Sigma}^{(1)} - \boldsymbol{D}^\top\boldsymbol{D})^{1/2}$. By the 1/2-Hölder continuity of the PSD matrix square-root map (Wihler 2009, Eq. (3.2)), $\left\|(\boldsymbol{\Sigma}^{(1)} - \boldsymbol{D}^\top\boldsymbol{D})^{1/2} - (\boldsymbol{\Sigma}^{(1)})^{1/2}\right\|_F \leq n_1^{1/4}\|\boldsymbol{D}^\top\boldsymbol{D}\|_F^{1/2}$. Using $\|\boldsymbol{D}^\top\boldsymbol{D}\|_F \leq \|\boldsymbol{D}^\top\|_F\|\boldsymbol{D}\|_F = \|\boldsymbol{D}\|_F^2$, we obtain $\left\|(\boldsymbol{\Sigma}^{1/2})^{(1)} - (\boldsymbol{\Sigma}^{(1)})^{1/2}\right\|_F \leq n_1^{1/4}\|\boldsymbol{D}\|_F$. Finally, the $(2,2)$ block equality gives $\boldsymbol{\Sigma}^{(2)} = \boldsymbol{D}\boldsymbol{D}^\top + (\boldsymbol{\Sigma}^{1/2})^{(2)}(\boldsymbol{\Sigma}^{1/2})^{(2)} \succeq \boldsymbol{D}\boldsymbol{D}^\top$. Therefore, $\|\boldsymbol{D}\|_F^2 = \operatorname{tr}(\boldsymbol{D}^\top\boldsymbol{D}) = \operatorname{tr}(\boldsymbol{D}\boldsymbol{D}^\top) \leq \operatorname{tr}(\boldsymbol{\Sigma}^{(2)})$. Since $\operatorname{tr}(\boldsymbol{\Sigma}^{(2)}) \leq n_2\sigma^2$, we have

$$\left\|(\boldsymbol{\Sigma}^{1/2})^{(1)} - (\boldsymbol{\Sigma}^{(1)})^{1/2}\right\|_F \leq n_1^{1/4} n_2^{1/2}\sigma = O(\sigma). \tag{EC.26}$$

With similar analysis, one can show that

$$\left\|(\boldsymbol{\Sigma}^{1/2})^{(2)} - (\boldsymbol{\Sigma}^{(2)})^{1/2}\right\|_F \leq n_2^{1/4}\|\boldsymbol{D}\|_F \leq n_2^{3/4}\sigma = O(\sigma). \tag{EC.27}$$

Step 2: Bounding $\Delta^{(1)} - \sum_r \Delta_c^{(r)}$.

Note that

$$\begin{aligned}\Delta^{(1)} - \sum_{r\in[R]} \Delta_c^{(r)} &= \left\langle \operatorname{diag}(\boldsymbol{p}^{(1)})\boldsymbol{Q}_*^{(1)}, (\boldsymbol{\Sigma}^{1/2})^{(1)} - (\boldsymbol{\Sigma}^{(1)})^{1/2} \right\rangle - \Delta_c^{(2)} \\ &\leq \left\|\operatorname{diag}(\boldsymbol{p}^{(1)})\boldsymbol{Q}_*^{(1)}\right\|_F \left\|(\boldsymbol{\Sigma}^{1/2})^{(1)} - (\boldsymbol{\Sigma}^{(1)})^{1/2}\right\|_F - \Delta_c^{(2)}.\end{aligned} \tag{EC.28}$$

We next bound $\|\operatorname{diag}(\boldsymbol{p}^{(1)})\boldsymbol{Q}_*^{(1)}\|_F$ using feasibility of the SDP. From the PSD constraint in (C-F), the Schur complement implies $\boldsymbol{T}_*^{(1)} \succeq \boldsymbol{Q}_*^{(1)}(\boldsymbol{Q}_*^{(1)})^\top$. Taking traces yields $\|\boldsymbol{Q}_*^{(1)}\|_F^2 = \operatorname{tr}\left(\boldsymbol{Q}_*^{(1)}(\boldsymbol{Q}_*^{(1)})^\top\right) \leq \operatorname{tr}(\boldsymbol{T}_*^{(1)})$. Moreover, the CBQP($n_1$) constraint implies $\operatorname{tr}(\boldsymbol{T}_*^{(1)}) \leq n_1$. Therefore, $\|\boldsymbol{Q}_*^{(1)}\|_F \leq \sqrt{n_1}$. Finally, since $\|\operatorname{diag}(\boldsymbol{p}^{(1)})\|_F = \|\boldsymbol{p}^{(1)}\|_2$, we have

$$\left\|\operatorname{diag}(\boldsymbol{p}^{(1)})\boldsymbol{Q}_*^{(1)}\right\|_F \leq \|\operatorname{diag}(\boldsymbol{p}^{(1)})\|_F \|\boldsymbol{Q}_*^{(1)}\|_F \leq \|\boldsymbol{p}^{(1)}\|_2\sqrt{n_1}. \tag{EC.29}$$

Combining (EC.26), (EC.28), and (EC.29) gives

$$\Delta^{(1)} - \sum_{r\in[R]} \Delta_c^{(r)} \leq \|\boldsymbol{p}^{(1)}\|_2 n_1^{3/4} n_2^{1/2}\sigma - \Delta_c^{(2)}. \tag{EC.30}$$

It remains to bound $|\Delta_c^{(2)}|$. The following lemma states that $|\Delta_c^{(2)}| = O(\sigma)$; its proof is given in Section EC.1.18.

Lemma EC.8. *It holds that $|\Delta_c^{(2)}| = O(\sigma)$.*

Combining Lemma EC.8 with (EC.30) gives $\Delta^{(1)} - \sum_{r\in[R]} \Delta_c^{(r)} = O(\sigma)$.

Step 3: Bounding $\Delta^{(2)} - \sum_r \Delta_c^{(r)}$.

Similarly to Step 2, $\Delta^{(2)} - \sum_{r\in[R]} \Delta_c^{(r)} = \left\langle \operatorname{diag}(\boldsymbol{p}^{(2)})\boldsymbol{Q}_*^{(2)}, (\boldsymbol{\Sigma}^{1/2})^{(2)} - (\boldsymbol{\Sigma}^{(2)})^{1/2} \right\rangle - \Delta_c^{(1)} \leq \left\|\operatorname{diag}(\boldsymbol{p}^{(2)})\boldsymbol{Q}_*^{(2)}\right\|_F \left\|(\boldsymbol{\Sigma}^{1/2})^{(2)} - (\boldsymbol{\Sigma}^{(2)})^{1/2}\right\|_F - \Delta_c^{(1)}$. As above, $\|\operatorname{diag}(\boldsymbol{p}^{(2)})\boldsymbol{Q}_*^{(2)}\|_F \leq \|\boldsymbol{p}^{(2)}\|_2\sqrt{n_2}$, and by (EC.27), $\|(\boldsymbol{\Sigma}^{1/2})^{(2)} - (\boldsymbol{\Sigma}^{(2)})^{1/2}\|_F = O(\sigma)$. Therefore, $\left\langle \operatorname{diag}(\boldsymbol{p}^{(2)})\boldsymbol{Q}_*^{(2)}, (\boldsymbol{\Sigma}^{1/2})^{(2)} - (\boldsymbol{\Sigma}^{(2)})^{1/2} \right\rangle = O(\sigma)$, and hence $\Delta^{(2)} - \sum_{r\in[R]} \Delta_c^{(r)} = -\Delta_c^{(1)} + O(\sigma)$.

Step 4: Bounding $\min_{r\in[R]} G^{(r)} - \Phi_{CF}(\boldsymbol{y}_*, \bar{v}_*) - \bar{v}_* x_*$.

By (EC.25), Step 2, and Step 3, $\min_{r\in[R]} G^{(r)} - \Phi_{CF}(\boldsymbol{y}_*, \bar{v}_*) - \bar{v}_* x_* \leq \min\{O(\sigma), -\Delta_c^{(1)} + O(\sigma)\}$.

Next, we show that $\Delta_c^{(r)} \leq 0$ for each $r \in [R]$. To this end, let F_* denote the worst-case distribution of $\tilde{\boldsymbol{d}}_c^D$ for the following problem:

$$\min_{F\in\mathscr{F}^{D,c}} \mathbb{E}_F\left[\min_{\substack{x\in\{0,1\},\ \boldsymbol{t}^{(r)}\in\{0,1\}^{n_r} \\ \boldsymbol{t}\leq x\boldsymbol{e}_n}} \sum_{r\in[R]} \left(\boldsymbol{p}^{(r)} \circ \left((\boldsymbol{\Sigma}^{(r)})^{1/2}(\tilde{\boldsymbol{d}}_c^D)^{(r)} + \boldsymbol{\mu}^{(r)} - \boldsymbol{y}_*^{(r)}\right)\right)^\top \boldsymbol{t}^{(r)} - \bar{v}x\right], \tag{EC.31}$$

where $\mathscr{F}^{D,c} = \left\{F \in \mathcal{M}(\mathbb{R}^n) \;\middle|\; \mathbb{E}_F[\tilde{\boldsymbol{d}}_c^D] = \boldsymbol{0},\ \mathbb{E}_F[(\tilde{\boldsymbol{d}}_c^D)^{(r)}((\tilde{\boldsymbol{d}}_c^D)^{(r)})^\top] = \boldsymbol{I}_{n_r},\ \forall r \in [R]\right\}$. For each $r \in [R]$, let $\boldsymbol{t}_*^{(r)}(\tilde{\boldsymbol{d}}_c^D)$ be an associated inner minimizer of $\boldsymbol{t}^{(r)}$ under F_*.

Using the same derivation as in (5)–(8), together with the transformation $(\tilde{\boldsymbol{d}}_c^D)^{(r)} = (\boldsymbol{\Sigma}^r)^{-1/2}(\tilde{\boldsymbol{d}}^r - \boldsymbol{\mu}^r)$, $\forall r \in [R]$, in the proof of Proposition 6 in Section EC.1.11, one can verify that (C-F) is equivalent to (EC.31). Moreover,

$$\begin{aligned}\Delta_c^{(r)} &= \left\langle \operatorname{diag}(\boldsymbol{p}^{(r)}), \boldsymbol{Q}_*^{(r)}(\boldsymbol{\Sigma}^{(r)})^{1/2}\right\rangle + \left[\boldsymbol{p}^{(r)} \circ (\boldsymbol{\mu}^{(r)} - \boldsymbol{y}_*^{(r)})\right]^\top \boldsymbol{t}_*^{(r)} \\ &= \mathbb{E}_{F_*}\left[\left(\boldsymbol{p}^{(r)} \circ \left((\boldsymbol{\Sigma}^{(r)})^{1/2}(\tilde{\boldsymbol{d}}_c^D)^{(r)} + \boldsymbol{\mu}^{(r)} - \boldsymbol{y}_*^{(r)}\right)\right)^\top \boldsymbol{t}_*^{(r)}(\tilde{\boldsymbol{d}}_c^D)\right] \le 0.\end{aligned}$$

The inequality follows from the inner minimization in (EC.31): because $\boldsymbol{t}^{(r)}$ is binary and constrained by $\boldsymbol{t} \le x\boldsymbol{e}_n$, the minimizer selects only components that weakly reduce the objective value, while choosing $x = 0$ and $\boldsymbol{t} = \boldsymbol{0}$ is always feasible.

Therefore, $\Delta_c^{(1)} \le 0$. Combining this with the previous bound gives $\min\{O(\sigma), -\Delta_c^{(1)} + O(\sigma)\} = O(\sigma)$. Thus, $\min_{r\in[R]} G^{(r)} - \Phi_{CF}(\boldsymbol{y}_*, \bar{v}_*) - \bar{v}_* x_* = O(\sigma)$, which proves part (b).

EC.1.17. Proof of Lemma EC.7

Since f is L-Lipschitz with respect to $\|\cdot\|_1$, we have $|\mathbb{E}[f(\tilde{\boldsymbol{x}})] - f(\boldsymbol{\mu})| \le \mathbb{E}\left[|f(\tilde{\boldsymbol{x}}) - f(\boldsymbol{\mu})|\right] \le L\,\mathbb{E}\left[\|\tilde{\boldsymbol{x}} - \boldsymbol{\mu}\|_1\right]$. By the Cauchy–Schwarz inequality, $\mathbb{E}\left[\|\tilde{\boldsymbol{x}} - \boldsymbol{\mu}\|_1\right] \le \sqrt{\mathbb{E}\left[\|\tilde{\boldsymbol{x}} - \boldsymbol{\mu}\|_1^2\right]}$. Moreover, using $\|\boldsymbol{z}\|_1 \le \sqrt{m}\|\boldsymbol{z}\|_2$ for all $\boldsymbol{z} \in \mathbb{R}^m$, we obtain $\mathbb{E}\left[\|\tilde{\boldsymbol{x}} - \boldsymbol{\mu}\|_1^2\right] \le m\,\mathbb{E}\left[\|\tilde{\boldsymbol{x}} - \boldsymbol{\mu}\|_2^2\right] = m\operatorname{tr}(\boldsymbol{\Sigma})$, where the last equality follows from the definition of covariance. Combining the above inequalities yields $|\mathbb{E}[f(\tilde{\boldsymbol{x}})] - f(\boldsymbol{\mu})| \le L\sqrt{m\operatorname{tr}(\boldsymbol{\Sigma})}$. This proves Lemma EC.7.

EC.1.18. Proof of Lemma EC.8

Recall that $\Delta_c^{(2)} = \left\langle \operatorname{diag}(\boldsymbol{p}^{(2)}), \boldsymbol{Q}_*^{(2)}(\boldsymbol{\Sigma}^{(2)})^{1/2}\right\rangle + \left[\boldsymbol{p}^{(2)} \circ (\boldsymbol{\mu}^{(2)} - \boldsymbol{y}_*^{(2)})\right]^\top \boldsymbol{t}_*^{(2)}$. We bound the two terms separately.

Step 1: Bounding the first term.

By the same Schur complement argument used in the proof of Theorem 4, feasibility of the SDP implies $\|\boldsymbol{Q}_*^{(2)}\|_F \le \sqrt{\operatorname{tr}(\boldsymbol{T}_*^{(2)})} \le \sqrt{n_2}$. Hence,

$$\begin{aligned}\left|\left\langle \operatorname{diag}(\boldsymbol{p}^{(2)}), \boldsymbol{Q}_*^{(2)}(\boldsymbol{\Sigma}^{(2)})^{1/2}\right\rangle\right| &= \left|\left\langle \operatorname{diag}(\boldsymbol{p}^{(2)})\boldsymbol{Q}_*^{(2)}, (\boldsymbol{\Sigma}^{(2)})^{1/2}\right\rangle\right| \le \left\|\operatorname{diag}(\boldsymbol{p}^{(2)})\boldsymbol{Q}_*^{(2)}\right\|_F \left\|(\boldsymbol{\Sigma}^{(2)})^{1/2}\right\|_F \\ &\le \|\boldsymbol{p}^{(2)}\|_2 \|\boldsymbol{Q}_*^{(2)}\|_F \sqrt{\operatorname{tr}(\boldsymbol{\Sigma}^{(2)})} \le \|\boldsymbol{p}^{(2)}\|_2 n_2 \sigma. \end{aligned} \tag{EC.32}$$

Step 2: Bounding $\|\boldsymbol{\mu}^{(2)} - \boldsymbol{y}_*^{(2)}\|_1$.

Let $\varphi^c(\boldsymbol{y}) := \min_{F\in\mathscr{F}^c} \operatorname{CVaR}_{\eta,F}(\pi(\boldsymbol{y}, \tilde{\boldsymbol{d}}))$ denote the correlation-splitting value function for a fixed order quantity $\boldsymbol{y}$, where $\mathscr{F}^c$ is the two-cluster ambiguity set. Since $\boldsymbol{y}_*$ is an optimal solution to Problem (C-R), it maximizes $\varphi^c(\boldsymbol{y})$. We compare φ^c with a surrogate problem in which the demand of cluster 2 is fixed at its mean. For $r = 1, 2$, define $\pi^{(r)}(\boldsymbol{y}^{(r)}, \boldsymbol{d}^{(r)}) := (\boldsymbol{p}^{(r)} - \boldsymbol{c}^{(r)})^\top \boldsymbol{y}^{(r)} + (\boldsymbol{p}^{(r)})^\top(\boldsymbol{d}^{(r)} - \boldsymbol{y}^{(r)})^-$. Define $\varphi^{(1)}(\boldsymbol{y}^{(1)}) := \min_{F_1\in\mathscr{F}_1} \operatorname{CVaR}_{\eta,F_1}(\pi^{(1)}(\boldsymbol{y}^{(1)}, \tilde{\boldsymbol{d}}^{(1)}))$ and $D^{(2)}(\boldsymbol{y}^{(2)}) := \pi^{(2)}(\boldsymbol{y}^{(2)}, \boldsymbol{\mu}^{(2)}) = (\boldsymbol{p}^{(2)} - \boldsymbol{c}^{(2)})^\top \boldsymbol{y}^{(2)} + (\boldsymbol{p}^{(2)})^\top(\boldsymbol{\mu}^{(2)} - \boldsymbol{y}^{(2)})^-$. The surrogate value function is $\varphi_0(\boldsymbol{y}) := \varphi^{(1)}(\boldsymbol{y}^{(1)}) + D^{(2)}(\boldsymbol{y}^{(2)})$. This equality follows from translation equivariance of CVaR, because $D^{(2)}(\boldsymbol{y}^{(2)})$ is deterministic.

We first establish a uniform deviation bound between φ^c and φ_0. Fix any $\boldsymbol{y}$ and any feasible distribution $F \in \mathscr{F}^c$. Let $F^{(1)}$ denote the marginal distribution of F on the first cluster, and define $Z_F := \pi^{(1)}(\boldsymbol{y}^{(1)}, \tilde{\boldsymbol{d}}^{(1)}) + \pi^{(2)}(\boldsymbol{y}^{(2)}, \tilde{\boldsymbol{d}}^{(2)})$ and $Z_F^0 := \pi^{(1)}(\boldsymbol{y}^{(1)}, \tilde{\boldsymbol{d}}^{(1)}) + \pi^{(2)}(\boldsymbol{y}^{(2)}, \boldsymbol{\mu}^{(2)})$. For any two random variables Z and Z' defined on the same probability space, the CVaR representation implies $|\mathrm{CVaR}_\eta(Z) - \mathrm{CVaR}_\eta(Z')| \le \mathbb{E}[|Z - Z'|]/\eta$. Moreover, $|Z_F - Z_F^0| \le \|\boldsymbol{p}^{(2)}\|_\infty \|\tilde{\boldsymbol{d}}^{(2)} - \boldsymbol{\mu}^{(2)}\|_1$. Applying Lemma EC.7 to the marginal distribution of $\tilde{\boldsymbol{d}}^{(2)}$, with $f(\boldsymbol{z}) = \|\boldsymbol{z} - \boldsymbol{\mu}^{(2)}\|_1$, gives $\mathbb{E}_F[\|\tilde{\boldsymbol{d}}^{(2)} - \boldsymbol{\mu}^{(2)}\|_1] \le \sqrt{n_2 \,\mathrm{tr}(\boldsymbol{\Sigma}^{(2)})} \le n_2\sigma$. Therefore,

$$\left|\mathrm{CVaR}_{\eta,F}(Z_F) - \mathrm{CVaR}_{\eta,F^{(1)}}(Z_F^0)\right| \le \frac{\|\boldsymbol{p}^{(2)}\|_\infty n_2}{\eta}\sigma. \tag{EC.33}$$

We now translate this bound to the value functions. Inequality (EC.33) implies two one-sided bounds. First, for any feasible $F \in \mathscr{F}^c$, $\mathrm{CVaR}_{\eta,F}(Z_F) \ge \mathrm{CVaR}_{\eta,F^{(1)}}(Z_F^0) - \|\boldsymbol{p}^{(2)}\|_\infty n_2\sigma/\eta$. Since $F^{(1)} \in \mathscr{F}_1$, we further have $\mathrm{CVaR}_{\eta,F^{(1)}}(Z_F^0) \ge \varphi_0(\boldsymbol{y})$. Therefore, $\mathrm{CVaR}_{\eta,F}(Z_F) \ge \varphi_0(\boldsymbol{y}) - \|\boldsymbol{p}^{(2)}\|_\infty n_2\sigma/\eta$. Taking the minimum over $F \in \mathscr{F}^c$ gives $\varphi^c(\boldsymbol{y}) \ge \varphi_0(\boldsymbol{y}) - \|\boldsymbol{p}^{(2)}\|_\infty n_2\sigma/\eta$.

Second, for any $F_1 \in \mathscr{F}_1$, we can extend F_1 to a feasible distribution in $\mathscr{F}^c$ by combining it with any feasible second-cluster marginal distribution with mean $\boldsymbol{\mu}^{(2)}$ and covariance $\boldsymbol{\Sigma}^{(2)}$. Applying the other side of (EC.33) to this extension gives $\mathrm{CVaR}_{\eta,F}(Z_F) \le \mathrm{CVaR}_{\eta,F_1}(Z_F^0) + \|\boldsymbol{p}^{(2)}\|_\infty n_2\sigma/\eta$, where F denotes the resulting feasible joint distribution whose first-cluster marginal is F_1. Since this holds for every $F_1 \in \mathscr{F}_1$, we obtain $\varphi^c(\boldsymbol{y}) \le \varphi_0(\boldsymbol{y}) + \|\boldsymbol{p}^{(2)}\|_\infty n_2\sigma/\eta$. Thus,

$$|\varphi^c(\boldsymbol{y}) - \varphi_0(\boldsymbol{y})| \le \frac{\|\boldsymbol{p}^{(2)}\|_\infty n_2}{\eta}\sigma, \ \forall \boldsymbol{y} \in \mathbb{R}^n. \tag{EC.34}$$

Let $\widehat{\boldsymbol{y}} := (\boldsymbol{y}_*^{(1)}, \boldsymbol{\mu}^{(2)})$. Since $\boldsymbol{y}_*$ maximizes φ^c, we have $\varphi^c(\boldsymbol{y}_*) \ge \varphi^c(\widehat{\boldsymbol{y}})$. Combining this optimality with (EC.34), we obtain $\varphi_0(\boldsymbol{y}_*) \ge \varphi^c(\boldsymbol{y}_*) - \|\boldsymbol{p}^{(2)}\|_\infty n_2\sigma/\eta \ge \varphi^c(\widehat{\boldsymbol{y}}) - \|\boldsymbol{p}^{(2)}\|_\infty n_2\sigma/\eta \ge \varphi_0(\widehat{\boldsymbol{y}}) - 2\|\boldsymbol{p}^{(2)}\|_\infty n_2\sigma/\eta$. Rearranging gives

$$\varphi_0(\widehat{\boldsymbol{y}}) - \varphi_0(\boldsymbol{y}_*) \le \frac{2\|\boldsymbol{p}^{(2)}\|_\infty n_2}{\eta}\sigma. \tag{EC.35}$$

Since $\widehat{\boldsymbol{y}}$ and $\boldsymbol{y}_*$ have the same first-cluster component, and since $\varphi_0(\boldsymbol{y}) = \varphi^{(1)}(\boldsymbol{y}^{(1)}) + D^{(2)}(\boldsymbol{y}^{(2)})$, inequality (EC.35) implies $D^{(2)}(\boldsymbol{\mu}^{(2)}) - D^{(2)}(\boldsymbol{y}_*^{(2)}) \le 2\|\boldsymbol{p}^{(2)}\|_\infty n_2\sigma/\eta$.

We next use the sharpness of the deterministic second-cluster profit around $\boldsymbol{\mu}^{(2)}$. For each $i \in N_2$, define $f_i(y_i) = (p_i - c_i)y_i + p_i(\mu_i - y_i)^-$. Then $D^{(2)}(\boldsymbol{y}^{(2)}) = \sum_{i\in N_2} f_i(y_i)$. If $y_i \le \mu_i$, then $f_i(\mu_i) - f_i(y_i) = (p_i - c_i)(\mu_i - y_i)$. If $y_i \ge \mu_i$, then $f_i(\mu_i) - f_i(y_i) = c_i(y_i - \mu_i)$. Therefore, for all y_i, $f_i(\mu_i) - f_i(y_i) \ge \min\{p_i - c_i, c_i\}|y_i - \mu_i|$. Let $\kappa := \min_{i\in N_2}\min\{p_i - c_i, c_i\}$. Since $\boldsymbol{p} > \boldsymbol{c} > \boldsymbol{0}$, we have $\kappa > 0$. Summing the preceding inequality over all items in the second cluster gives $D^{(2)}(\boldsymbol{\mu}^{(2)}) - D^{(2)}(\boldsymbol{y}_*^{(2)}) \ge \kappa\|\boldsymbol{y}_*^{(2)} - \boldsymbol{\mu}^{(2)}\|_1$. Combining the last two inequalities yields

$$\|\boldsymbol{y}_*^{(2)} - \boldsymbol{\mu}^{(2)}\|_1 \le \frac{2\|\boldsymbol{p}^{(2)}\|_\infty n_2}{\kappa\eta}\sigma. \tag{EC.36}$$

Step 3: Putting the bounds together.

Using (EC.32), (EC.36), and the fact that $\mathbf{0} \le \boldsymbol{t}_*^{(2)} \le \boldsymbol{e}_{n_2}$, we obtain

$$\begin{aligned}|\Delta_c^{(2)}| &\le \left|\left\langle \mathrm{diag}(\boldsymbol{p}^{(2)}), \boldsymbol{Q}_*^{(2)}(\boldsymbol{\Sigma}^{(2)})^{1/2}\right\rangle\right| + \left|\left[\boldsymbol{p}^{(2)} \circ (\boldsymbol{\mu}^{(2)} - \boldsymbol{y}_*^{(2)})\right]^\top \boldsymbol{t}_*^{(2)}\right| \\ &\le \|\boldsymbol{p}^{(2)}\|_2 n_2 \sigma + \|\boldsymbol{p}^{(2)} \circ \boldsymbol{t}_*^{(2)}\|_\infty \|\boldsymbol{\mu}^{(2)} - \boldsymbol{y}_*^{(2)}\|_1 \\ &\le \|\boldsymbol{p}^{(2)}\|_2 n_2 \sigma + \|\boldsymbol{p}^{(2)}\|_\infty \frac{2\|\boldsymbol{p}^{(2)}\|_\infty n_2}{\kappa\eta}\sigma = \left(\|\boldsymbol{p}^{(2)}\|_2 n_2 + \frac{2n_2\|\boldsymbol{p}^{(2)}\|_\infty^2}{\kappa\eta}\right)\sigma = O(\sigma).\end{aligned}$$

This proves Lemma EC.8.

EC.2. Formulation of the Quadratic Decision Rule (QDR) Benchmark

In deriving a tractable benchmark for Problem (P), the QDR method avoids directly imposing the exponentially many PSD constraints. Instead, it constructs a quadratic lower approximation of $(\boldsymbol{d} - \boldsymbol{y})^-$ by introducing auxiliary variables $\boldsymbol{q}, \boldsymbol{r}, \boldsymbol{z}$ and imposing

$$\begin{cases} \boldsymbol{d} \circ \boldsymbol{q} \circ \boldsymbol{d} + \boldsymbol{r} \circ \boldsymbol{d} + \boldsymbol{z} \le \boldsymbol{d} - \boldsymbol{y}, \ \forall \boldsymbol{d} \in \mathbb{R}^n, \\ \boldsymbol{d} \circ \boldsymbol{q} \circ \boldsymbol{d} + \boldsymbol{r} \circ \boldsymbol{d} + \boldsymbol{z} \le \mathbf{0}, \qquad \forall \boldsymbol{d} \in \mathbb{R}^n. \end{cases}$$

These constraints ensure that $\boldsymbol{d} \circ \boldsymbol{q} \circ \boldsymbol{d} + \boldsymbol{r} \circ \boldsymbol{d} + \boldsymbol{z} \le (\boldsymbol{d} - \boldsymbol{y})^-$ for all $\boldsymbol{d} \in \mathbb{R}^n$. Therefore, replacing the constraint $s_1 + \boldsymbol{s}_2^\top \boldsymbol{d} + \left\langle \boldsymbol{s}_3, \boldsymbol{d}\boldsymbol{d}^\top \right\rangle \le \boldsymbol{p}^\top (\boldsymbol{d} - \boldsymbol{y})^- - \bar{v}$ with $s_1 + \boldsymbol{s}_2^\top \boldsymbol{d} + \left\langle \boldsymbol{s}_3, \boldsymbol{d}\boldsymbol{d}^\top \right\rangle \le \boldsymbol{p}^\top (\boldsymbol{d} \circ \boldsymbol{q} \circ \boldsymbol{d} + \boldsymbol{r} \circ \boldsymbol{d} + \boldsymbol{z}) - \bar{v}$ yields the following QDR lower bound:

$$\begin{aligned} \max_{\substack{\boldsymbol{y}, \bar{v}, s_1, \boldsymbol{s}_2, \boldsymbol{s}_3 \\ \boldsymbol{q}, \boldsymbol{r}, \boldsymbol{z}}} \quad & \bar{v} + (\boldsymbol{p} - \boldsymbol{c})^\top \boldsymbol{y} + \frac{1}{\eta}\left(s_1 + \boldsymbol{s}_2^\top \boldsymbol{\mu} + \left\langle \boldsymbol{s}_3, \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\top \right\rangle\right) \\ \text{s.t.} \quad & \begin{pmatrix} -s_1 + \boldsymbol{p}^\top \boldsymbol{z} - \bar{v} & \frac{1}{2}(-\boldsymbol{s}_2 + \boldsymbol{p} \circ \boldsymbol{r})^\top \\ \frac{1}{2}(-\boldsymbol{s}_2 + \boldsymbol{p} \circ \boldsymbol{r}) & -\boldsymbol{s}_3 + \mathrm{diag}(\boldsymbol{p} \circ \boldsymbol{q}) \end{pmatrix} \succeq 0 \\ & \begin{pmatrix} -s_1 & -\frac{1}{2}\boldsymbol{s}_2^\top \\ -\frac{1}{2}\boldsymbol{s}_2 & -\boldsymbol{s}_3 \end{pmatrix} \succeq 0 \\ & \begin{pmatrix} -z_i - y_i & -\frac{1}{2}(r_i - 1) \\ -\frac{1}{2}(r_i - 1) & -q_i \end{pmatrix} \succeq 0, \ \forall i \in [n] \\ & \begin{pmatrix} -z_i & -\frac{1}{2}r_i \\ -\frac{1}{2}r_i & -q_i \end{pmatrix} \succeq 0, \ \forall i \in [n] \end{aligned}$$

EC.3. Constraint Generation Algorithm for the CBQP-Based Reformulation

As shown in (B-R), the number of linear constraints $\boldsymbol{\delta}^\top \boldsymbol{w} + \boldsymbol{w}^\top \boldsymbol{V} \boldsymbol{w} - \tau \le 0$, $\forall \boldsymbol{w} \in \{0,1\}^n$, grows exponentially with n, which creates substantial computational challenges. However, at optimality, typically only a small subset of these constraints is active. We therefore use a constraint-generation (CG) algorithm that iteratively solves a restricted master problem and adds violated constraints when they are identified. We begin with an initial subset $\Theta_0 \subseteq \{0,1\}^n$ and solve the following master problem:

$$
\begin{aligned}
\max_{\substack{\boldsymbol{y},\bar{v},\alpha,\tau,\gamma,\boldsymbol{\beta},\boldsymbol{\delta},\\ \boldsymbol{\Gamma},\boldsymbol{V}}} \quad & \bar{v} + (\boldsymbol{p}-\boldsymbol{c})^\top \boldsymbol{y} - \frac{1}{\eta}\left(\alpha - \gamma + \boldsymbol{\mu}^\top\boldsymbol{\beta} + \left\langle \boldsymbol{\Gamma}, \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\top \right\rangle\right) \\
\text{s.t.} \quad & \begin{pmatrix} \boldsymbol{\Gamma} & \frac{1}{2}\boldsymbol{\beta} & \frac{1}{2}\operatorname{diag}(\boldsymbol{p}) \\ \frac{1}{2}\boldsymbol{\beta}^\top & \alpha & \frac{1}{2}(\boldsymbol{\delta} - \boldsymbol{p}\circ\boldsymbol{y})^\top \\ \frac{1}{2}\operatorname{diag}(\boldsymbol{p}) & \frac{1}{2}(\boldsymbol{\delta} - \boldsymbol{p}\circ\boldsymbol{y}) & \boldsymbol{V} \end{pmatrix} \succeq 0 \\
& \gamma + \tau + \bar{v} = 0 \\
& \boldsymbol{\delta}^\top \boldsymbol{w} + \boldsymbol{w}^\top \boldsymbol{V} \boldsymbol{w} - \tau \le 0, \ \forall \boldsymbol{w} \in \Theta_0 \\
& \gamma \le 0
\end{aligned}
\tag{EC.37}
$$

After solving (EC.37), we check whether the resulting solution is feasible for (B-R). If it is feasible, then it is also optimal for (B-R). Otherwise, violated constraints are identified by solving the separation subproblem

$$
l_* = \max_{\boldsymbol{w}\in\{0,1\}^n} \boldsymbol{\delta}_*^\top \boldsymbol{w} + \boldsymbol{w}^\top \boldsymbol{V}_* \boldsymbol{w} - \tau_*, \tag{EC.38}
$$

where $(\tau_*, \boldsymbol{\delta}_*, \boldsymbol{V}_*)$ is obtained from the current master problem. If $l_* \le 0$, then the current solution satisfies all constraints in (B-R). If $l_* > 0$, then the subproblem solution $\boldsymbol{w}_*$ satisfies $\boldsymbol{\delta}_*^\top \boldsymbol{w}_* + \boldsymbol{w}_*^\top \boldsymbol{V}_* \boldsymbol{w}_* - \tau_* > 0$. The corresponding violated constraint $\boldsymbol{\delta}^\top \boldsymbol{w}_* + \boldsymbol{w}_*^\top \boldsymbol{V} \boldsymbol{w}_* - \tau \le 0$ is then added to the master problem by updating $\Theta_{j+1} = \Theta_j \cup \{\boldsymbol{w}_*\}$ for the current iteration j. In practice, multiple violated constraints can be identified and added in each iteration to accelerate convergence. This CG approach avoids enumerating all exponentially many constraints and introduces them only when needed. The procedure is summarized in Algorithm 2.

Algorithm 2 Constraint Generation Algorithm for the CBQP-Based Reformulation

Input: Initial constraint set Θ_0, separation tolerance ϵ.
Output: Inventory level $\boldsymbol{y}_*$.

1: Set iteration counter $j \leftarrow 0$.
2: **loop**
3:    Solve the master problem (EC.37) with constraint set Θ_j, and let $(\boldsymbol{y}_j, \tau_j, \boldsymbol{\delta}_j, \boldsymbol{V}_j)$ denote the relevant components of an optimal solution.
4:    Solve the subproblem (EC.38) with $(\tau_j, \boldsymbol{\delta}_j, \boldsymbol{V}_j)$, obtaining the maximum violation l_j and an optimizer $\boldsymbol{w}_j$.
5:    **if** $l_j \le \epsilon$ **then**
6:       **return** $\boldsymbol{y}_* = \boldsymbol{y}_j$.
7:    **else**
8:       Update $\Theta_{j+1} \leftarrow \Theta_j \cup \{\boldsymbol{w}_j\}$ and set $j \leftarrow j+1$.

Table EC.1 Improvement (in Percentages) from the Variance-Based Splitting Rule for $\eta = 0.99$ and $\eta = 0.7$

	Size (n)							
	30	40	50	60	70	80	90	100
$\eta = 0.99$								
CS-V-2 *vs.* ***CS-R-2***								
VRI – 10th percentile	-0.10	-0.11	-0.02	-0.13	-0.05	0.01	0.11	0.07
VRI – mean	0.25	0.20	0.31	0.27	0.27	0.30	0.28	0.31
VRI – 90th percentile	0.71	0.56	0.63	0.63	0.58	0.60	0.53	0.55
Prob(*CS-V-2* > *CS-R-2*)	0.78	0.78	0.88	0.76	0.84	0.90	0.92	0.96
$\eta = 0.7$								
CS-V-2 *vs.* ***CS-R-2***								
VRI – 10th percentile	-0.46	-0.17	-0.11	-0.31	-0.17	-0.06	0.14	-0.01
VRI – mean	0.52	0.32	0.56	0.49	0.48	0.48	0.48	0.55
VRI – 90th percentile	1.54	1.00	1.23	1.23	1.13	1.01	0.96	1.09
Prob(*CS-V-2* > *CS-R-2*)	0.74	0.78	0.82	0.78	0.82	0.84	0.92	0.90

EC.4. Additional Numerical Results

In the main numerical experiments, we set the risk-aversion parameter to $\eta = 0.95$. To assess the robustness of our numerical findings, we conduct additional experiments with $\eta = 0.99$ and $\eta = 0.7$. We focus on the key comparisons, including the benefit of the variance-based splitting rule and the comparison with the *QDR* benchmark.

Table EC.1 reports the variance rule improvement (VRI) statistics for $\eta = 0.99$ and $\eta = 0.7$. The results are consistent with those in Section 6.2.1: the variance-based splitting rule continues to improve upon random partitioning, with positive mean VRI and a high probability of improvement. Comparing the two values of η, we observe that the magnitude of VRI tends to increase when η decreases from 0.99 to 0.7, while the dispersion also becomes more pronounced. This pattern is consistent with the intuition that, when the objective places greater emphasis on more adverse tail outcomes, the choice of clustering rule becomes more consequential.

Table EC.2 presents the comparison between *CS-V-2* and *QDR* under $\eta = 0.99$ and $\eta = 0.7$. Consistent with Section 6.3, *CS-V-2* continues to outperform *QDR* across all instances, with positive CSI observed throughout, and remains substantially faster than *QDR*. Although the magnitude of improvement decreases as η increases, the improvement of correlation splitting becomes more concentrated and the superiority remains evident, indicating that the advantage over decision-rule-based approximations is robust to the choice of η.

Overall, the additional experiments show that the main numerical conclusions in Section 6 are robust to different choices of η. The variance-based splitting rule continues to improve upon random splitting, and the proposed *CS-V-2* method consistently dominates *QDR* in both solution quality and computational efficiency. The results also suggest that, as the objective becomes more sensitive to adverse tail outcomes, the value of variance-based clustering may become more pronounced.

Table EC.2 Comparison in Solution Quality and Computation Time between CS-V-2 and QDR for $\eta = 0.99$ and $\eta = 0.7$

	Size (n)											
	30	40	50	60	70	80	90	100	110	120	130	140
$\eta = 0.99$												
CS-V-2 *vs.* ***QDR***												
CSI – 10th percentile	0.80	0.83	0.99	1.25	1.09	1.18	1.20	1.36	1.32	1.32	1.30	1.37
CSI – mean	1.61	1.50	1.72	1.82	1.75	1.71	1.67	1.91	1.82	1.81	1.64	1.81
CSI – 90th percentile	2.57	2.30	2.56	2.58	2.43	2.30	2.14	2.59	2.45	2.37	2.02	2.21
Prob(*CS-V-2* > *QDR*)	1	1	1	1	1	1	1	1	1	1	1	1
CS-V-2												
Time (s)	0.02	0.03	0.04	0.05	0.06	0.08	0.10	0.13	0.16	0.19	0.22	0.26
QDR												
Time (s)	0.10	0.22	0.47	1.11	2.31	3.62	6.89	10.98	18.01	28.46	44.07	65.42
$\eta = 0.7$												
CS-V-2 *vs.* ***QDR***												
CSI – 10th percentile	0.70	0.57	0.71	0.98	0.71	0.93	0.86	1.05	0.94	0.95	0.85	0.94
CSI – mean	3.24	1.86	2.11	2.78	2.03	1.77	1.71	2.02	1.79	1.66	1.31	1.65
CSI – 90th percentile	4.42	3.14	3.68	3.77	3.54	2.80	2.59	3.48	2.72	2.59	1.93	2.27
Prob(*CS-V-2* > *QDR*)	1	1	1	1	1	1	1	1	1	1	1	1
CS-V-2												
Time (s)	0.02	0.03	0.04	0.05	0.06	0.08	0.10	0.13	0.15	0.19	0.21	0.25
QDR												
Time (s)	0.09	0.21	0.47	1.07	2.10	3.23	6.22	9.85	16.12	25.62	38.38	59.06