# Distributionally Robust Inventory Management with Advance Purchase Contracts

Yilin Xue[1], Yongzhen Li[2], Napat Rujeerapaiboon[1]
[1] *Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore*
[2] *Department of Management Science and Engineering, School of Economics and Management, Southeast University, China*
yilin.xue@u.nus.edu, liyongzhen@seu.edu.cn, napat.rujeerapaiboon@nus.edu.sg

Motivated by the worldwide Covid-19 vaccine procurement, we study an inventory problem with an advance purchase contract which requires all ordering decisions to be committed at once. In reality, not only the demand is uncertain, but its distribution can also be ambiguous. Hence, we assume that only the mean and the variance are known and aim at minimizing the worst-case expected cost. We first show that our inventory model reduces to a robust conic optimization problem with a finite yet exponentially-sized uncertainty set. To gain tractability and err on the safe side, we propose two conservative approximations. Then to measure their approximation quality, we develop a progressive approximation based on a scenario reduction technique. All of the approximate models are expressed as standard polynomially-sized conic programs, which scale gracefully and allow us to incorporate additional distributional knowledge via a cone replacement. We quantify the benefit of committing to advance purchases, and we show that all approximations are close to being exact. Besides, we analytically derive the worst-case demand distribution and numerically use it to show that our robust policy is more resilient to the misspecification of the demand distribution than the state-of-the-art non-robust policies.

*Keywords*: Inventory management; distributionally robust optimization; conic programming.

## 1. Introduction

During the past three years, the global population has been rampaged by the spread of Covid-19, and advance purchase contracts have repeatedly been signed by governments of different countries; see *e.g.* Jalelah (2021) and Andres (2022). In such contracts, the governments acting as inventory managers submit a single advance purchase order to the pharmaceutical companies and specify the amount of vaccines needed at different times, to acquire sufficient protective vaccines. This episode has prompted us to revisit the advance purchase agreement in the inventory management problem.

Advance purchase contracts have a unique standing and have been scrutinized extensively in the literature. Suppliers who are on the receiving end of an advance purchase contract can devise an informed production plan that is less likely disrupted by fluctuating order volumes, and they are incentivized to innovate (Price et al. 2020). For these reasons, Özer and Wei (2006) and Tang et al. (2004) argued that the suppliers may assent to offering discounts to attract advance purchases. In return, inventory managers who are willing to commit early can benefit from the reduced purchasing cost and the strengthened relationship with the suppliers (Basciftci et al. 2021). Besides, because of the limited purchasing opportunities, advance purchase represents a norm when dealing with seasonal goods and products with short life cycles, such as comestibles, electronic gadgets and apparels; see *e.g.* Mamani et al. (2017), Tang et al. (2004).

When demands are independent and follow a known distribution, Scarf (1960) derived an inventory policy, famously known as the base-stock policy. Despite its simplicity, the optimal base-stock policy may not be easy to characterize, and in reality it is impossible to obtain the true demand distribution. Bertsimas and Thiele (2006), Mamani et al. (2017), Postek et al. (2018) then identified a robustly optimal advance purchase contract in various settings. Other papers that are similarly robust while allowing the ordering amounts to be decided adaptively include Ben-Tal et al. (2005), See and Sim (2010), Postek et al. (2018), Bertsimas et al. (2019) etc.

In this paper, we propose and analyze a novel inventory model that yields a robustly optimal advance purchase order. Our model specifically aims at minimizing the worst-case expected total cost (consisting of purchasing, holding and backlogging costs) over all possible demand distributions that are consistent with a marginal mean and variance. Relying on the duality theory, we first formulate our distributionally robust inventory problem as a finite convex optimization problem, however, with an exponential number of constraints. To gain tractability robustly, we propose *'L-conservative'* and *'Q-conservative'* approximations. The former is derived by interpreting the problem as an artificial two-stage robust optimization problem and then restricting each second-stage decision to follow a linear decision rule (Ben-Tal et al. 2004), whereas the latter is obtained via constraint partitioning and a series of interchanges between summation and maximization. In addition, we come up with a scenario reduction technique (Hadjiyiannis et al. 2011) to develop a progressive approximation of the model.

For succinctness, we summarize the main contributions of the paper below.

- To our knowledge, we are the first to consider the distributionally robust inventory problem with advance purchase contracts under the mean-variance ambiguity set of the demand distribution. We reduce the problem to a robust second-order cone program with an exponentially-sized uncertainty set which is characterized by the unit holding and backlogging costs.
- To gain tractability, we propose `L`- and `Q`- conservative approximations as well as a progressive approximation, all of which are presented compactly as scalable second-order cone programs. By means of a cone replacement, these conic representations can accommodate additional distributional information of the demands, such as non-negativity and pairwise uncorrelatedness.
- We exploit the progressive approximation to quantify the optimality gaps of our `L`- and `Q`-conservative solutions. Numerically, these gaps appear insignificant, and thus our inventory problem can be solved almost exactly. Besides, we verify that our conservative approximations are tighter than the state-of-the-art alternative inspired by Postek et al. (2018) and prove that our progressive approximation is tight when the problem consists of two periods.
- We show how our distributionally robust inventory policy is more resilient than the stochastic policy even when the demand distribution is only slightly contaminated. Additionally, an adaptive version of our policy (which is obtained by re-solving the problem repeatedly in a shrinking horizon fashion) is almost on par with the optimal base-stock policy even when the demand distribution is known, is correct, and is serially independent.
- Finally, to close the loop we utilize a dataset from a fashion retailer on Amazon to examine a discount level for an advance purchase agreement that simultaneously enables the inventory manager to pay less and the supplier to earn more in comparison to the base-stock policy.

Our techniques could have a far-reaching impact on other operational problems of similar nature, such as an appointment scheduling problem; see Mak et al. (2015) and Padmanabhan et al. (2021). We refer our readers to Appendix A for a technical comparison between these papers and ours.

### 1.1. Literature review

A single-period inventory problem is referred to as a newsvendor problem (Shapiro et al. 2014). Typically, the solution to the risk-neutral newsvendor problem could be characterized by the inverse of the cumulative distribution function of the unknown demand, which however is difficult to accurately empirically estimate. Scarf (1958) addressed this issue by seeking to find an optimal decision that performs best in view of the worst-case distribution amongst all those that share the same mean and variance. Ben-Tal and Hochman (1976) and Das et al. (2021) then extended this seminal work by exploiting other statistical information. Gallego and Moon (1993) showed that this ambiguity set is not overly conservative in the sense that, when the demand is normal, the stochastically optimal and the robustly optimal solutions attain similar expected costs. For completeness, we note that there are other newsvendor models which directly incorporate the historical observations of the demand and/or other influential features and simultaneously account for the demand distribution ambiguity, such as Chen and Xie (2021), Lee et al. (2021) and Fu et al. (2021).

For a multi-period problem, a risk-neutral inventory manager also aims at minimizing the total expected cost. When the demands are serially independent and follow a known distribution, the purchasing costs consist of a fixed and a linearly variable part, and the holding and the backlogging costs are linear, Scarf (1960) formulated this problem as a dynamic program and showed that there exists an optimal base-stock policy which is characterized by a collection of reordering points and order-up-to levels. In the absence of the fixed ordering cost, it can be further imposed without loss that these two are the same, and thus the optimal policy simplifies (Bertsekas 1995). Still, even the determination of this simplified base-stock policy is provably difficult (Halman et al. 2009), and only a few asymptotic variants admit a reasonably efficient exact solution approach (Veinott and Wagner 1965, Federgruen and Zipkin 1984, Zheng and Federgruen 1991). For the rest, we may need a simulation-based method (Fu 1994) or an interpolation technique (Halman et al. 2009).

In reality, the demand distribution is hardly available but is required for the computation of the optimal policy. Incorporating inaccurate information (*e.g.* using an empirical distribution in lieu of the true demand distribution) in a dynamic program systematically induces *error propagation*, which is neither easy to quantify nor to eliminate (Zhang et al. 2021). Retaining the adaptability of the ordering quantities in later periods, several authors have turned towards robust (*e.g.* Ben-Tal et al. 2005) and distributionally robust (*e.g.* See and Sim 2010, Bertsimas et al. 2019) optimization. Tractability is ensured when they solved their respective variant of the inventory problem using a decision rule approximation. Though, Lu and Sturt (2022) showed that the majority of decision rule coefficients may unexpectedly vanish. Solyalı et al. (2016) relied on a different formulation which is inspired by the facility layout and knapsack problem to improve the computational efficiency even when there is a fixed ordering cost. In a similar vein, Chen et al. (2022) proposed a cycle-based base-stock policy to lessen the financial impact of frequent ordering on the worst-case cost.

Differently from the above, the inventory manager as well as the supplier may have a preference for an advance purchase contract. When no other distributional information besides the support of the uncertain demand is available, Bertsimas and Thiele (2006) studied a multi-stage robust inventory problem with an advance purchase contract and identified a robustly optimal policy that can be interpreted as a base-stock policy, and subsequently, Mamani et al. (2017) determined a closed-form solution of the same problem when the uncertainty set satisfies a certain symmetrical criterion. Note that the robust inventory model studied in these papers is overly conservative in the way it computes the worst-case cost. As a purely robust model, it is also ignorant of almost all distributional information. To address this, Postek et al. (2018) considered a distributionally robust inventory approach that accounts for the mean and the mean absolute deviation of the demands.

In this paper, we develop a novel distributionally robust inventory management model to identify an advance purchase agreement that minimizes the worst-case expected total cost. Comparing to the above, we aim for our model to be less conservative in two ways. First, similarly to Postek et al. (2018) we utilize distributionally robust optimization, but we differently adopt the mean-variance ambiguity set, which has been extensively studied in the literature, see *e.g.* Delage and Ye (2010), Zymler et al. (2013). Second, our model either accurately captures the worst-case expected cost or approximates it with a quantifiable a posteriori bound.

The remainder of the paper is structured as follows. Section 2 details our distributionally robust inventory problem and its equivalent reformulations, one of which is used to derive the worst-case distribution to be exploited in a stress test. Our tractable conservative and progressive approximations are derived in Section 3, and the conic extensions for some restricted ambiguity sets are given in Section 4. Finally, experiments in Section 5 are used to measure the quality of the proposed approximations and to compare our approach with the state-of-the-art benchmarks, and Section 6 reports an additional case study demonstrating the benefit of advance purchase contracts. Section 7 finally concludes the paper. Note that some proofs and discussions are relegated to appendices.

**Notation:** We use boldface lowercase letters (*e.g.* $\boldsymbol{v}$) and uppercase letters (*e.g.* $\mathbf{M}$) to represent vectors and matrices, respectively. Specifically, $\mathbf{1}$ ($\mathbf{0}$) denotes a vector or a matrix of all ones (zeros), and $\mathbf{1}_i$ denotes the $i^{\text{th}}$ canonical basis vector. The set of all real numbers is denoted by $\mathbb{R}$ and its subset of non-negative and strictly positive numbers are denoted by $\mathbb{R}_+$ and $\mathbb{R}_{++}$, respectively. $\mathbb{S}^n$ represents a set of symmetric matrices in $\mathbb{R}^{n\times n}$. For any $\mathbf{A},\mathbf{B}\in\mathbb{S}^n$, $\mathbf{A}\succeq\mathbf{B}$ indicates that $\mathbf{A}-\mathbf{B}$ is positive semidefinite. If $\mathcal{K}$ is a cone, then $\mathcal{K}^\star$ denotes its dual. For any vector $\boldsymbol{v}\in\mathbb{R}^n$, $\text{diag}(\boldsymbol{v})$ is a diagonal matrix in $\mathbb{S}^n$ that has elements of $\boldsymbol{v}$ sitting on its main diagonal, and $\#(\boldsymbol{v},m)$ counts the number of occurrences of $m$ in $\boldsymbol{v}$, and $\boldsymbol{v}^m$, $1\leqslant m\leqslant n$, is a subvector of $\boldsymbol{v}$ containing its first $m$ elements, *i.e.*, $\boldsymbol{v}^m=(v_1,\ldots,v_m)^\top$. Throughout, a division by zero is allowed and $a/0$ is defined to be $=+\infty$ if $a>0$; $=0$ if $a=0$; $=-\infty$ otherwise.

## 2. Mean-variance inventory model and its finite representations

We consider an uncapacitated inventory system that stores a single product for sale, and we denote by $T$ the planning horizon as well as by $c$, $h$ and $b$ the unit ordering, holding and backlogging costs, respectively. We assume that these cost parameters are positive and that they are time-invariant, although these are assumptions that could readily be lifted. Supposing that the inventory manager is risk-neutral, that the demand distribution is perfectly known, and that any unfulfilled order could be indefinitely backlogged, facing the uncertain demands $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_T)^\top$, the inventory manager would be interested in minimizing the total expected cost, *i.e.*, solving

$$\begin{aligned} \text{minimize} \quad & \mathbb{E}_{\mathbb{P}}\left[\sum_{t=1}^{T}\Big(cx_t + \max\left\{hy_t(\boldsymbol{\xi}^t), -by_t(\boldsymbol{\xi}^t)\right\}\Big)\right] \\ \text{subject to} \quad & \boldsymbol{x} \in \mathcal{X},\ y_t : \mathbb{R}^t \mapsto \mathbb{R} \quad \forall t \in \{1, \ldots, T\} \\ & y_t(\boldsymbol{\xi}^t) = y_{t-1}(\boldsymbol{\xi}^{t-1}) + x_t - \xi_t \quad \mathbb{P}\text{-a.s.} \quad \forall t \in \{1, \ldots, T\}, \end{aligned} \tag{1}$$

where $\boldsymbol{x} = (x_1, \ldots, x_T)^\top$ and $\boldsymbol{y} = (y_1, \ldots, y_T)^\top$ collect all decision variables representing the order quantities and the end-of-period inventory levels, respectively, and $y_0 \in \mathbb{R}$ denotes the initial inventory level, which is given. For the ease of exposition, we assume that all orders are instantaneous, *i.e.*, the lead times are zero. Besides, we follow Bertsimas and Thiele (2006) and Mamani et al. (2017) in imposing that $\boldsymbol{x}$ is chosen here-and-now and the feasible set $\mathcal{X} \subseteq \mathbb{R}^T_+$ is uncertainty-free.

In the actual reality, not only $\boldsymbol{\xi}$ is uncertain but its probability distribution $\mathbb{P}$, which needs to be empirically estimated, is also ambiguous and may only be known to belong to $\mathcal{P}$. Therefore, to hedge against such an estimation risk, the inventory manager may robustify Problem (1) and solve

$$\begin{aligned} \text{minimize} \quad & \max_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[\sum_{t=1}^{T}\left(cx_t + \max\left\{h\left(y_0 + \sum_{\tau=1}^{t}(x_\tau - \xi_\tau)\right), -b\left(y_0 + \sum_{\tau=1}^{t}(x_\tau - \xi_\tau)\right)\right\}\right)\right] \\ \text{subject to} \quad & \boldsymbol{x} \in \mathcal{X} \end{aligned} \tag{2}$$

Remark 1. The pure robust inventory model studied in, *e.g.*, Bertsimas and Thiele (2006) and Mamani et al. (2017) assumes that the demands $\boldsymbol{\xi}$ belong to an uncertainty set $\Xi$. For tractability reasons, it captures the worst-case holding and backlogging cost in period $t \in \{1, \ldots, T\}$ through

$$\max_{\boldsymbol{\xi} \in \Xi}\left\{\max\left\{h\left(y_0 + \sum_{\tau=1}^{t}(x_\tau - \xi_\tau)\right), -b\left(y_0 + \sum_{\tau=1}^{t}(x_\tau - \xi_\tau)\right)\right\}\right\}.$$

Here, the nature can adversarially choose different worst-case $\boldsymbol{\xi}$'s for different $t$'s. A more faithful and less conservative model should only allow for a single worst-case $\boldsymbol{\xi}$ to be chosen. We refer our readers to Gorissen and den Hertog (2013) for a detailed discussion on this subtle intricacy, which could be overcome by tighter approximations from Gorissen and den Hertog (2013) and Ardestani-Jaafari and Delage (2016) or by iterative heuristics from Bienstock and Özbay (2008) and Rodrigues et al. (2021). To our knowledge, however, these techniques are not directly applicable to the distributionally robust variant considered in this paper. With neither them nor the above simplification, we could foresee the extra computational burden associated with our inventory model. □

We henceforth express Problem (2) compactly as $\min_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x})$ where the objective function $f$ characterizes the worst-case expected cost incurred from $\boldsymbol{x}$, *i.e.*,

$$f(\boldsymbol{x}) = \max_{\mathbb{P}\in\mathcal{P}} \ \mathbb{E}_{\mathbb{P}}\left[\sum_{t=1}^{T}\left(cx_t + \max\left\{h\left(y_0+\sum_{\tau=1}^{t}(x_\tau-\xi_\tau)\right), -b\left(y_0+\sum_{\tau=1}^{t}(x_\tau-\xi_\tau)\right)\right\}\right)\right]. \quad (3)$$

We note that $f$ is convex but not necessarily easy to evaluate because it involves solving an infinite optimization problem over the probability distribution $\mathbb{P}\in\mathcal{P}$. Following the seminal work of Scarf (1958), we choose to work with the following mean-variance ambiguity set:

$$\mathcal{P} = \left\{\mathbb{P}\in\mathcal{M}_+(\mathbb{R}^T) : \mathbb{P}\left(\boldsymbol{\xi}\in\mathbb{R}^T\right)=1,\ \mathbb{E}_{\mathbb{P}}[\xi_t]=\mu,\ \mathbb{E}_{\mathbb{P}}[\xi_t^2]=\mu^2+\sigma^2 \quad \forall t=1,\ldots,T\right\}, \quad (4)$$

where $\mathcal{M}_+$ denotes the cone of all non-negative measures supported on the input set, because these two summary statistics can be readily estimated empirically. We remark that this basic ambiguity set allows negative demands to capture possible product returns. When returns are disallowed (perhaps because of a hygienic reason or the product's short shelf life), we will consider a restricted ambiguity set in Section 4.

Under this $\mathcal{P}$, we will first show that the worst-case expected total cost $f(\boldsymbol{x})$, with $\boldsymbol{x}$ fixed, can be determined by solving a finite-optimization problem or its dual. Before presenting these results, we introduce a cone $\mathcal{K}_t$, $t\in\{1,\ldots,T\}$, that is instrumental to our subsequent analyses.

$$\mathcal{K}_t = \left\{(\alpha,\boldsymbol{\beta},\boldsymbol{\gamma})\in\mathbb{R}_+\times\mathbb{R}^t\times\mathbb{R}_+^t : 4\alpha \geqslant \sum_{\tau=1}^{t}\frac{\beta_\tau^2}{\gamma_\tau}\right\}.$$

Note that $\mathcal{K}_t$ is proper and is representable as an intersection of multiple second-order cones, *i.e.*,

$$\mathcal{K}_t = \left\{(\alpha,\boldsymbol{\beta},\boldsymbol{\gamma})\in\mathbb{R}_+\times\mathbb{R}^t\times\mathbb{R}_+^t : \exists\boldsymbol{\theta}\in\mathbb{R}_+^t,\ \alpha\geqslant\sum_{\tau=1}^{t}\theta_\tau,\ \|(\beta_\tau,\theta_\tau-\gamma_\tau)^\top\|_2\leqslant\theta_\tau+\gamma_\tau\ \forall\tau=1,\ldots,t\right\}.$$

Moreover, we also introduce a discrete set $\mathcal{E}=\{h,-b\}^T$ whose cardinality is $2^T$. We will use it to capture the $2^T$ underlying linear lower bounds of the total holding and backlogging cost, which appears as a part of the objective function of Problems (1) and (2), *i.e.*,

$$\sum_{t=1}^{T}\max\left\{hy_t(\boldsymbol{\xi}^t), -by_t(\boldsymbol{\xi}^t)\right\} = \max_{\boldsymbol{e}\in\mathcal{E}}\sum_{t=1}^{T}e_t\left(y_0+\sum_{\tau=1}^{t}(x_\tau-\xi_\tau)\right).$$

This exponential complexity arises as a result of that, at the end of each period $t \in \{1, \ldots, T\}$, the inventory manager will have to pay for either the holding cost if the end-of-period inventory level is positive or the backlogging cost otherwise.

THEOREM 1. *We have that*

$$
\begin{aligned}
f(\boldsymbol{x}) = \text{minimize} \quad & c\mathbf{1}^\top \boldsymbol{x} + \alpha + \mu \mathbf{1}^\top \boldsymbol{\beta} + (\mu^2 + \sigma^2)\mathbf{1}^\top \boldsymbol{\gamma} \\
\text{subject to} \quad & \alpha \in \mathbb{R},\ \boldsymbol{\beta} \in \mathbb{R}^T,\ \boldsymbol{\gamma} \in \mathbb{R}^T \\
& (\alpha(\boldsymbol{x}, \boldsymbol{e}), \boldsymbol{\beta}(\boldsymbol{e}), \boldsymbol{\gamma}) \succeq_{\mathcal{K}_T} \mathbf{0} \quad \forall \boldsymbol{e} \in \mathcal{E},
\end{aligned}
\tag{5}
$$

*where* $\alpha(\boldsymbol{x}, \boldsymbol{e}) = \alpha - y_0 \mathbf{1}^\top \boldsymbol{e} - \sum_{t=1}^T x_t \sum_{\tau=t}^T e_\tau$, $\beta_t(\boldsymbol{e}) = \beta_t + \sum_{\tau=t}^T e_\tau$, *for all* $\boldsymbol{x} \in \mathbb{R}^T_+$.

We can interpret Problem (5) as a robust conic program with an auxiliary uncertain vector $\boldsymbol{e} \in \mathcal{E}$. Note that we can straightforwardly extend Theorem 1 to account for a fixed ordering cost at the expense of adding relevant step functions to the objective of Problem (5).

### 2.1. Worst-case demand distribution

This section determines a probability distribution from the ambiguity set $\mathcal{P}$ that solves the maximization problem on the right-hand side of (3), *i.e.,* that attains the worst-case expected cost, for any $\boldsymbol{x} \in \mathcal{X}$. Such a distribution is often used in a stress test (Bertsimas et al. 2010), and it can also help validate or invalidate the independence assumption typically imposed (*e.g.* Scarf 1960).

The worst-case distribution could be constructed from the dual of Problem (5) derived below.

THEOREM 2. *For all* $\boldsymbol{x} \in \mathbb{R}^T_+$, *we have that*

$$
\begin{aligned}
f(\boldsymbol{x}) = \text{maximize} \quad & c\mathbf{1}^\top \boldsymbol{x} + \sum_{\substack{\boldsymbol{e} \in \mathcal{E}: \\ \overline{\alpha}(\boldsymbol{e}) \neq 0}} \overline{\alpha}(\boldsymbol{e}) \sum_{t=1}^T e_t \left( y_0 + \sum_{\tau=1}^t \left( x_\tau - \frac{\overline{\beta}_\tau(\boldsymbol{e})}{\overline{\alpha}(\boldsymbol{e})} \right) \right) \\
\text{subject to} \quad & \overline{\alpha} : \mathcal{E} \mapsto \mathbb{R},\ \overline{\boldsymbol{\beta}} : \mathcal{E} \mapsto \mathbb{R}^T,\ \overline{\boldsymbol{\gamma}} : \mathcal{E} \mapsto \mathbb{R}^T \\
& \left( \overline{\alpha}(\boldsymbol{e}), \overline{\boldsymbol{\beta}}(\boldsymbol{e}), \overline{\boldsymbol{\gamma}}(\boldsymbol{e}) \right) \succeq_{\mathcal{K}_T^\star} \mathbf{0} \quad \forall \boldsymbol{e} \in \mathcal{E} \\
& \sum_{\boldsymbol{e} \in \mathcal{E}} \overline{\alpha}(\boldsymbol{e}) = 1 \\
& \sum_{\boldsymbol{e} \in \mathcal{E}} \overline{\boldsymbol{\beta}}(\boldsymbol{e}) = \mu \mathbf{1} \\
& \sum_{\boldsymbol{e} \in \mathcal{E}} \overline{\boldsymbol{\gamma}}(\boldsymbol{e}) = (\mu^2 + \sigma^2)\mathbf{1}.
\end{aligned}
\tag{6}
$$

With $\mathcal{K}_T^\star$ derived in Proposition 1 in Appendix C, we solve Problem (6) to determine an optimal solution $(\overline{\alpha}^\star, \overline{\boldsymbol{\beta}}^\star, \overline{\boldsymbol{\gamma}}^\star)$ that satisfies $\overline{\boldsymbol{\beta}}^\star(\boldsymbol{e}) = \overline{\boldsymbol{\gamma}}^\star(\boldsymbol{e}) = \boldsymbol{0},\ \forall \boldsymbol{e} \in \mathcal{E} : \overline{\alpha}^\star(\boldsymbol{e}) = 0$. Note that this extra condition can be imposed without any loss of optimality (see Lemma 1 in Appendix C). Then, for any $\boldsymbol{e} \in \mathcal{E}$ such that $\overline{\alpha}^\star(\boldsymbol{e}) \neq 0$, we construct a probability distribution $\mathbb{P}^{\boldsymbol{e},\varepsilon} \in \mathcal{M}_+(\mathbb{R}^T)$, $\varepsilon \in (0,1)$, via

$$\mathbb{P}^{\boldsymbol{e},\varepsilon}\left(\boldsymbol{\xi} = \underline{\boldsymbol{\xi}}^{\boldsymbol{e},\varepsilon}\right) = 1 - \varepsilon \quad \text{and} \quad \mathbb{P}^{\boldsymbol{e},\varepsilon}\left(\boldsymbol{\xi} = \overline{\boldsymbol{\xi}}^{\boldsymbol{e},\varepsilon}\right) = \varepsilon, \tag{7}$$

where the atoms $\underline{\boldsymbol{\xi}}^{\boldsymbol{e},\varepsilon} \in \mathbb{R}^T$ and $\overline{\boldsymbol{\xi}}^{\boldsymbol{e},\varepsilon} \in \mathbb{R}^T$ are chosen as

$$\begin{aligned}
\underline{\xi}_t^{\boldsymbol{e},\varepsilon} &= \frac{\overline{\beta}_t^\star(\boldsymbol{e})}{\overline{\alpha}^\star(\boldsymbol{e})} - \sqrt{\frac{\varepsilon}{1-\varepsilon}} \frac{\sqrt{\overline{\alpha}^\star(\boldsymbol{e})\overline{\gamma}_t^\star(\boldsymbol{e}) - (\overline{\beta}_t^\star(\boldsymbol{e}))^2}}{\overline{\alpha}^\star(\boldsymbol{e})} & \forall t = 1, \ldots, T, \\
\overline{\xi}_t^{\boldsymbol{e},\varepsilon} &= \frac{\overline{\beta}_t^\star(\boldsymbol{e})}{\overline{\alpha}^\star(\boldsymbol{e})} + \sqrt{\frac{1-\varepsilon}{\varepsilon}} \frac{\sqrt{\overline{\alpha}^\star(\boldsymbol{e})\overline{\gamma}_t^\star(\boldsymbol{e}) - (\overline{\beta}_t^\star(\boldsymbol{e}))^2}}{\overline{\alpha}^\star(\boldsymbol{e})} & \forall t = 1, \ldots, T.
\end{aligned}$$

Note that $\sqrt{\overline{\alpha}^\star(\boldsymbol{e})\overline{\gamma}_t^\star(\boldsymbol{e}) - (\overline{\beta}_t^\star(\boldsymbol{e}))^2}$ is well-defined because the term inside is non-negative, which is due to the characterization of $\mathcal{K}_T^\star$ from Proposition 1 provided in Appendix C. Next, we construct a mixture distribution

$$\mathbb{P}^\varepsilon = \sum_{\boldsymbol{e} \in \mathcal{E} : \overline{\alpha}^\star(\boldsymbol{e}) > 0} \overline{\alpha}^\star(\boldsymbol{e}) \mathbb{P}^{\boldsymbol{e},\varepsilon} \tag{8}$$

and assert that this demand distribution attains the worst-case inventory cost in (3) as $\varepsilon \downarrow 0$.

Theorem 3. *The demand distribution $\mathbb{P}^\varepsilon$, $\varepsilon \in (0,1)$, has the following properties:*

- $\mathbb{P}^\varepsilon \in \mathcal{P}$ *for all* $\varepsilon \in (0,1)$,
- $\lim_{\varepsilon \downarrow 0} \mathbb{E}_{\mathbb{P}^\varepsilon}\left[\sum_{t=1}^T \max\left\{h\left(y_0 + \sum_{\tau=1}^t (x_\tau - \xi_\tau)\right), -b\left(y_0 + \sum_{\tau=1}^t (x_\tau - \xi_\tau)\right)\right\}\right] = f(\boldsymbol{x}) - c\mathbf{1}^\top \boldsymbol{x}$.

Under the constructed worst-case distribution, the demands appear to be highly dependent across time periods, and we verify this claim in Section 5. An ambiguity-averse inventory manager should hence be heedful of the stochastic *in*dependence, which is typically assumed.

## 3. Efficiently solvable approximation

Theorems 1 and 2 ensure that we can evaluate $f(\boldsymbol{x})$ by solving a finite convex optimization problem. Nevertheless, the difficulty arises when $T$ gets large as the uncertainty set $\mathcal{E}$ is exponentially-sized. In this section, we provide two distinct ways of approximating $f(\boldsymbol{x})$ from above. One is a result of interpreting Problem (5) as an artificial two-stage robust optimization problem and restricting the second-stage decisions using affine adaptation, whereas the other is due to constraint categorization and a series of interchanges between maximization and summation. To measure their respective optimality gap, we also propose a progressive approximation based on scenario reduction.

### 3.1. Efficiently solvable conservative approximation

Our first approximation, $\overline{f}_{\mathtt{L}}(\boldsymbol{x})$, leverages Theorem 1 to express $f(\boldsymbol{x})$ as the optimal objective value of a two-stage robust optimization problem and then apply a linear decision rule approximation to the second-stage decisions to derive the following upper bound on $f(\boldsymbol{x})$.

$$\begin{aligned}
\overline{f}_{\mathtt{L}}(\boldsymbol{x}) \;=\; \text{minimize} \quad & c\mathbf{1}^\top \boldsymbol{x} + \alpha + \mu \mathbf{1}^\top \boldsymbol{\beta} + (\mu^2+\sigma^2)\mathbf{1}^\top \boldsymbol{\gamma} \\
\text{subject to} \quad & \alpha \in \mathbb{R},\ \boldsymbol{\beta}\in\mathbb{R}^T,\ \boldsymbol{\gamma}\in\mathbb{R}^T,\ \boldsymbol{\kappa}\in\mathbb{R}_+^{2T+1},\ (\pi_t,\boldsymbol{\pi}_t^u,\boldsymbol{\pi}_t^v)\in\mathbb{R}\times\mathbb{R}^T\times\mathbb{R}^T \ \ \forall t=1,\ldots,T \\
& (\pi_t-\kappa_t,\boldsymbol{\pi}_t^u,\boldsymbol{\pi}_t^v) \succeq_{\mathcal{K}_T} \mathbf{0} \quad \forall t=1,\ldots,T \\
& \left(\pi_t + y_0 + \sum_{\tau=1}^t x_\tau - \kappa_{T+t}, \boldsymbol{\pi}_t^u - (1,\ldots,1,0,\ldots,0)^\top, \boldsymbol{\pi}_t^v\right) \succeq_{\mathcal{K}_T} \mathbf{0} \quad \forall t=1,\ldots,T \\
& \Bigg(\alpha - hTy_0 - h\sum_{t=1}^T x_t(T-t+1) - (b+h)\sum_{t=1}^T \pi_t - \kappa_{2T+1}, \\
& \qquad \boldsymbol{\beta} - (b+h)\sum_{t=1}^T \boldsymbol{\pi}_t^u + h(T,T-1,\ldots,1)^\top, \boldsymbol{\gamma} - (b+h)\sum_{t=1}^T \boldsymbol{\pi}_t^v\Bigg) \succeq_{\mathcal{K}_T} \mathbf{0},
\end{aligned}$$

where the vector $(1,\ldots,1,0,\ldots,0)^\top$ in the third line of constraints has $t$ components equal to one and $T-t$ components equal to zero.

Theorem 4 (**L-conservative approximation**). *We have that $f(\boldsymbol{x}) \leqslant \overline{f}_{\mathtt{L}}(\boldsymbol{x})$ for all $\boldsymbol{x}\in\mathbb{R}_+^T$.*

*Proof.* We first re-express the robust constraints of Problem (5), which characterizes $f(\boldsymbol{x})$, as:

$$\min_{\boldsymbol{e}\in\mathcal{E}} \max_{\omega\in\mathbb{R}} \{\omega : (\alpha(\boldsymbol{x},\boldsymbol{e}),\boldsymbol{\beta}(\boldsymbol{e}),\boldsymbol{\gamma}) \succeq_{\mathcal{K}_T} (\omega,\mathbf{0},\mathbf{0})\} \geqslant 0.$$

Then, by dualizing the maximization problem on the left-hand side of this inequality, we attain:

$$\begin{aligned}
& \min_{\boldsymbol{e}\in\mathcal{E}} \min_{\boldsymbol{u},\boldsymbol{v}\in\mathbb{R}^T} \left\{\alpha(\boldsymbol{x},\boldsymbol{e}) + \boldsymbol{u}^\top\boldsymbol{\beta}(\boldsymbol{e}) + \boldsymbol{\gamma}^\top\boldsymbol{v} : (1,\boldsymbol{u},\boldsymbol{v}) \succeq_{\mathcal{K}_T^\star} \mathbf{0}\right\} \geqslant 0 \\
\iff \quad & \min_{\boldsymbol{u},\boldsymbol{v}\in\mathbb{R}^T} \left\{\boldsymbol{\gamma}^\top\boldsymbol{v} + \min_{\boldsymbol{e}\in\mathcal{E}}\left\{\alpha(\boldsymbol{x},\boldsymbol{e}) + \boldsymbol{u}^\top\boldsymbol{\beta}(\boldsymbol{e})\right\} : (1,\boldsymbol{u},\boldsymbol{v}) \succeq_{\mathcal{K}_T^\star} \mathbf{0}\right\} \geqslant 0 \qquad (9) \\
\iff \quad & \min_{\boldsymbol{u},\boldsymbol{v}\in\mathbb{R}^T} \left\{\boldsymbol{\gamma}^\top\boldsymbol{v} + \min_{\boldsymbol{e}\in\mathrm{conv}(\mathcal{E})}\left\{\alpha(\boldsymbol{x},\boldsymbol{e}) + \boldsymbol{u}^\top\boldsymbol{\beta}(\boldsymbol{e})\right\} : (1,\boldsymbol{u},\boldsymbol{v}) \succeq_{\mathcal{K}_T^\star} \mathbf{0}\right\} \geqslant 0,
\end{aligned}$$

where the second equivalence is because $\alpha(\boldsymbol{x},\boldsymbol{e})$ and $\boldsymbol{\beta}(\boldsymbol{e})$ are linear in $\boldsymbol{e}$ and because any solvable linear program has a vertex solution. Next, we expand the inner minimization problem over $\boldsymbol{e}$ as

$$\begin{array}{ll} \text{minimize} & \left[\alpha - y_0\mathbf{1}^\top\boldsymbol{e} - \sum_{t=1}^T x_t \sum_{\tau=t}^T e_\tau\right] + \left[\boldsymbol{\beta}^\top\boldsymbol{u} + \sum_{t=1}^T u_t \sum_{\tau=t}^T e_\tau\right] \\ \text{subject to} & \boldsymbol{e} \geqslant -b\mathbf{1}, \quad \boldsymbol{e} \leqslant h\mathbf{1} \end{array}$$

thanks to the definition of $\alpha(\boldsymbol{x},\boldsymbol{e})$, $\boldsymbol{\beta}(\boldsymbol{e})$ and $\mathcal{E}$, which admits a dual:

$$\begin{array}{ll} \text{maximize} & \alpha + \boldsymbol{\beta}^\top\boldsymbol{u} - b\mathbf{1}^\top\boldsymbol{\phi} - h\mathbf{1}^\top\boldsymbol{\psi} \\ \text{subject to} & \boldsymbol{\phi} \in \mathbb{R}^T_+, \ \boldsymbol{\psi} \in \mathbb{R}^T_+ \\ & \phi_t - \psi_t = -y_0 + \sum_{\tau=1}^t (u_\tau - x_\tau) \quad \forall t = 1,\ldots,T. \end{array}$$

Replacing the inner minimization problem in (9) with its dual, we have that, for all $\boldsymbol{x} \in \mathbb{R}^T_+$,

$$\begin{array}{rll} f(\boldsymbol{x}) = & \text{minimize} & c\mathbf{1}^\top\boldsymbol{x} + \alpha + \mu\mathbf{1}^\top\boldsymbol{\beta} + (\mu^2+\sigma^2)\mathbf{1}^\top\boldsymbol{\gamma} \\ & \text{subject to} & \alpha \in \mathbb{R}, \ \boldsymbol{\beta} \in \mathbb{R}^T, \ \boldsymbol{\gamma} \in \mathbb{R}^T, \ \boldsymbol{\phi} : \mathbb{R}^{2T} \mapsto \mathbb{R}^T_+, \ \boldsymbol{\psi} : \mathbb{R}^{2T} \mapsto \mathbb{R}^T_+ \\ & & \alpha + \boldsymbol{\beta}^\top\boldsymbol{u} + \boldsymbol{\gamma}^\top\boldsymbol{v} \geqslant b\mathbf{1}^\top\boldsymbol{\phi}(\boldsymbol{u},\boldsymbol{v}) + h\mathbf{1}^\top\boldsymbol{\psi}(\boldsymbol{u},\boldsymbol{v}) \\ & & \qquad \forall(\boldsymbol{u},\boldsymbol{v}) : (1,\boldsymbol{u},\boldsymbol{v}) \in \mathcal{K}^\star_T \\ & & y_0 + \phi_t(\boldsymbol{u},\boldsymbol{v}) - \psi_t(\boldsymbol{u},\boldsymbol{v}) = \sum_{\tau=1}^t (u_\tau - x_\tau) \quad \forall t = 1,\ldots,T \\ & & \qquad \forall(\boldsymbol{u},\boldsymbol{v}) : (1,\boldsymbol{u},\boldsymbol{v}) \in \mathcal{K}^\star_T. \end{array}$$

Our next step entails simplifying this new exact characterization of $f(\boldsymbol{x})$ by replacing the adaptive decisions $\psi_t(\boldsymbol{u},\boldsymbol{v})$ by $y_0 + \phi_t(\boldsymbol{u},\boldsymbol{v}) + \sum_{\tau=1}^t (x_\tau - u_\tau)$, $1 \leqslant t \leqslant T$. We can then obtain a conservative approximation by restricting each $\phi_t(\boldsymbol{u},\boldsymbol{v})$ to an affine form: $\phi_t(\boldsymbol{u},\boldsymbol{v}) = \pi_t + \boldsymbol{u}^\top\boldsymbol{\pi}^u_t + \boldsymbol{v}^\top\boldsymbol{\pi}^v_t$ for some linear decision rule coefficients $\pi_t \in \mathbb{R}$, $\boldsymbol{\pi}^u_t \in \mathbb{R}^T$, $\boldsymbol{\pi}^v_t \in \mathbb{R}^T$. As a result, we have the following upper bound of $f(\boldsymbol{x})$ that involves only here-and-now decisions.

$$\begin{array}{ll} \text{minimize} & c\mathbf{1}^\top\boldsymbol{x} + \alpha + \mu\mathbf{1}^\top\boldsymbol{\beta} + (\mu^2+\sigma^2)\mathbf{1}^\top\boldsymbol{\gamma} \\ \text{subject to} & \alpha \in \mathbb{R}, \ \boldsymbol{\beta} \in \mathbb{R}^T, \ \boldsymbol{\gamma} \in \mathbb{R}^T, \ (\pi_t, \boldsymbol{\pi}^u_t, \boldsymbol{\pi}^v_t) \in \mathbb{R} \times \mathbb{R}^T \times \mathbb{R}^T \ \forall t = 1,\ldots,T \\ & \pi_t + \boldsymbol{u}^\top\boldsymbol{\pi}^u_t + \boldsymbol{v}^\top\boldsymbol{\pi}^v_t \geqslant 0 \quad \forall(\boldsymbol{u},\boldsymbol{v}) : (1,\boldsymbol{u},\boldsymbol{v}) \in \mathcal{K}^\star_T, \ t = 1,\ldots,T \end{array}$$

$$
\begin{aligned}
&\pi_t + y_0 + \sum_{\tau=1}^{t} x_\tau + \boldsymbol{u}^\top \boldsymbol{\pi}_t^u - \sum_{\tau=1}^{t} u_\tau + \boldsymbol{v}^\top \boldsymbol{\pi}_t^v \geqslant 0 \quad \forall (\boldsymbol{u},\boldsymbol{v}) : (1,\boldsymbol{u},\boldsymbol{v}) \in \mathcal{K}_T^\star,\ t = 1,\ldots,T \\
&\alpha - hTy_0 - h\sum_{t=1}^{T} x_t(T-t+1) - (b+h)\sum_{t=1}^{T}\pi_t + \boldsymbol{u}^\top\left(\boldsymbol{\beta} - (b+h)\sum_{t=1}^{T}\boldsymbol{\pi}_t^u\right) \\
&\qquad + h\sum_{t=1}^{T} u_t(T-t+1) + \boldsymbol{v}^\top\left(\boldsymbol{\gamma} - (b+h)\sum_{t=1}^{T}\boldsymbol{\pi}_t^v\right) \geqslant 0 \quad \forall(\boldsymbol{u},\boldsymbol{v}) : (1,\boldsymbol{u},\boldsymbol{v}) \in \mathcal{K}_T^\star.
\end{aligned}
$$

Leveraging the primal-dual pair involving cone $\mathcal{K}_T$ familiar from a derivation leading to (9), we can derive the robust counterpart of the first constraint as follows.

$$
\begin{aligned}
\min_{\boldsymbol{u},\boldsymbol{v}}\left\{\pi_t + \boldsymbol{u}^\top\boldsymbol{\pi}_t^u + \boldsymbol{v}^\top\boldsymbol{\pi}_t^v :\ (1,\boldsymbol{u},\boldsymbol{v}) \in \mathcal{K}_T^\star\right\} \geqslant 0 \iff & \max_{\kappa_t}\left\{\kappa_t :\ (\pi_t - \kappa_t, \boldsymbol{\pi}_t^u, \boldsymbol{\pi}_t^v) \in \mathcal{K}_T\right\} \geqslant 0 \\
\iff & (\pi_t - \kappa_t, \boldsymbol{\pi}_t^u, \boldsymbol{\pi}_t^v) \succeq_{\mathcal{K}_T} \boldsymbol{0} \quad \exists \kappa_t \in \mathbb{R}_+.
\end{aligned}
$$

Applying the same routine for the remaining constraints, which are similarly linear in the conically constrained uncertain parameters $\boldsymbol{u}$ and $\boldsymbol{v}$, completes the proof. □

We emphasize that the linear decision rule approximation adopted here is different than those used in the earlier papers, such as Ben-Tal et al. (2005), See and Sim (2010), and Bertsimas et al. (2019), because unlike theirs our main operational decisions $\boldsymbol{x} \in \mathcal{X}$ are to be chosen here and now. We only apply the affine restriction to the artificial wait-and-see decisions $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ in the proof of Theorem 4, which we introduce as a means to capture the worst-case expected total cost $f(\boldsymbol{x})$. Note that the L-conservative approximation $\overline{f}_{\mathtt{L}}(\boldsymbol{x})$ involves only the primal cone $\mathcal{K}_T$ and that we only use the dual cone $\mathcal{K}_T^\star$ to construct the worst-case demand distribution; see Section 2.1.

Next, we derive an alternative conservative approximation of $f(\boldsymbol{x})$ which is obtained by constraint categorization and a series of interchanges between maximization and summation:

$$
\begin{aligned}
\overline{f}_{\mathtt{Q}}(\boldsymbol{x}) = \text{minimize} \quad & c\mathbf{1}^\top\boldsymbol{x} + \alpha + \mu\mathbf{1}^\top\boldsymbol{\beta} + (\mu^2+\sigma^2)\mathbf{1}^\top\boldsymbol{\gamma} \\
\text{subject to} \quad & \alpha \in \mathbb{R},\ \boldsymbol{\beta} \in \mathbb{R}^T,\ \boldsymbol{\gamma} \in \mathbb{R}^T,\ \boldsymbol{\lambda}^0,\ldots,\boldsymbol{\lambda}^T \in \mathbb{R}^T \\
& \alpha + by_0(T-k) - hy_0k \geqslant \mathbf{1}^\top\boldsymbol{\lambda}^k \quad \forall k = 0,\ldots,T \\
& \left(\lambda_t^k - x_t\textstyle\sum_{\tau=t}^{T}\check{e}_\tau^k, \beta_t(\check{\boldsymbol{e}}^k), \gamma_t\right) \succeq_{\mathcal{K}_1} \boldsymbol{0} \quad \forall k = 0,\ldots,T\ \forall t = 1,\ldots,T \\
& \left(\lambda_t^k - x_t\textstyle\sum_{\tau=t}^{T}\hat{e}_\tau^k, \beta_t(\hat{\boldsymbol{e}}^k), \gamma_t\right) \succeq_{\mathcal{K}_1} \boldsymbol{0} \quad \forall k = 0,\ldots,T\ \forall t = 1,\ldots,T,
\end{aligned} \tag{10}
$$

where the auxiliary vectors $\widehat{\boldsymbol{e}}^k \in \mathbb{R}^T$ and $\breve{\boldsymbol{e}}^k \in \mathbb{R}^T$ are defined through

$$\widehat{\boldsymbol{e}}^k = (-b, \ldots, -b, \underbrace{+h, \ldots, +h}_{k \text{ components}})^\top \quad \text{and} \quad \breve{\boldsymbol{e}}^k = (\underbrace{+h, \ldots, +h}_{k \text{ components}}, -b, \ldots, -b)^\top.$$

The correctness of this upper bound is validated in the next theorem.

Theorem 5 **(Q-conservative approximation)**. *We have that* $f(\boldsymbol{x}) \leqslant \overline{f}_{\mathtt{Q}}(\boldsymbol{x})$ *for all* $\boldsymbol{x} \in \mathbb{R}^T_+$.

*Proof.* It is sufficient to show that, for any $(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}^0, \ldots, \boldsymbol{\lambda}^T)$ that is feasible in Problem (10), $(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma})$ is feasible in Problem (5). To achieve this, we first observe that, for each $k \in \{0, \ldots, T\}$ and $t \in \{1, \ldots, T\}$, the two $\mathcal{K}_1$-inequalities of Problem (10) imply that

$$4\lambda_t^k \;\geqslant\; \max\left\{ \frac{\beta_t(\breve{\boldsymbol{e}}^k)^2}{\gamma_t} + 4x_t \sum_{\tau=t}^T \breve{e}_\tau^k,\; \frac{\beta_t(\widehat{\boldsymbol{e}}^k)^2}{\gamma_t} + 4x_t \sum_{\tau=t}^T \widehat{e}_\tau^k \right\}.$$

Next, fixing the index $k \in \{0, \ldots, T\}$, summing up the above inequality over $t \in \{1, \ldots, T\}$ and finally using the remaining (linear) constraint of Problem (10) results in

$$\begin{aligned} 4\alpha + 4by_0(T-k) - 4hy_0k \;&\geqslant\; \sum_{t=1}^T \max\left\{ \frac{\beta_t(\breve{\boldsymbol{e}}^k)^2}{\gamma_t} + 4x_t \sum_{\tau=t}^T \breve{e}_\tau^k,\; \frac{\beta_t(\widehat{\boldsymbol{e}}^k)^2}{\gamma_t} + 4x_t \sum_{\tau=t}^T \widehat{e}_\tau^k \right\} \\ &= \sum_{t=1}^T \max_{\boldsymbol{e} \in \mathcal{E}} \left\{ \frac{\beta_t(\boldsymbol{e})^2}{\gamma_t} + 4x_t \sum_{\tau=t}^T e_\tau : \#(\boldsymbol{e}, h) = k \right\} \\ &\geqslant \max_{\boldsymbol{e} \in \mathcal{E}} \left\{ \sum_{t=1}^T \left( \frac{\beta_t(\boldsymbol{e})^2}{\gamma_t} + 4x_t \sum_{\tau=t}^T e_\tau \right) : \#(\boldsymbol{e}, h) = k \right\}, \end{aligned}$$

where the equality holds because the quadratic expression $\frac{\beta_t(\boldsymbol{e})^2}{\gamma_t} + 4x_t \sum_{\tau=t}^T e_\tau$ is convex in $\sum_{\tau=t}^T e_\tau$ and hence its maximum value over the discrete feasible set $\{\boldsymbol{e} \in \mathcal{E} :\ \#(\boldsymbol{e}, h) = k\}$ is attained when $\boldsymbol{e} = \widehat{\boldsymbol{e}}^k$ or $\boldsymbol{e} = \breve{\boldsymbol{e}}^k$, and the last inequality follows as a result of the interchange between summation and maximization. A slight rearrangement of terms further yields

$$\min_{\boldsymbol{e} \in \mathcal{E}} \left\{ 4\alpha - 4y_0 \mathbf{1}^\top \boldsymbol{e} - 4 \sum_{t=1}^T x_t \sum_{\tau=t}^T e_\tau - \sum_{t=1}^T \frac{\beta_t(\boldsymbol{e})^2}{\gamma_t} : \#(\boldsymbol{e}, h) = k \right\} \geqslant 0 \quad \forall k = 0, \ldots, T$$

$$\iff \quad (\alpha(\boldsymbol{x}, \boldsymbol{e}), \boldsymbol{\beta}(\boldsymbol{e}), \boldsymbol{\gamma}) \succeq_{\mathcal{K}_T} \mathbf{0} \quad \forall \boldsymbol{e} \in \mathcal{E} : \#(\boldsymbol{e}, h) = k \quad \forall k = 0, \ldots, T$$

$$\iff \quad (\alpha(\boldsymbol{x}, \boldsymbol{e}), \boldsymbol{\beta}(\boldsymbol{e}), \boldsymbol{\gamma}) \succeq_{\mathcal{K}_T} \mathbf{0} \quad \forall \boldsymbol{e} \in \mathcal{E}$$

and correspondingly the feasibility of $(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma})$ in Problem in (5). The proof is now completed. □

REMARK 2. Another way to upper bound $f(\boldsymbol{x})$ is to use a larger ambiguity set, such as

$$\widehat{\mathcal{P}} = \left\{ \mathbb{P} \in \mathcal{M}_+(\mathbb{R}^T) : \mathbb{P}\left(\boldsymbol{\xi} \in \mathbb{R}^T\right) = 1,\ \mathbb{E}_{\mathbb{P}}\left[\sum_{\tau=1}^{t} \xi_\tau\right] = t\mu,\ \mathbb{E}_{\mathbb{P}}\left[\left|\sum_{\tau=1}^{t} \xi_\tau - t\mu\right|\right] \leqslant t\sigma\ \forall t = 1, \ldots, T \right\} \supset \mathcal{P}$$

which describes the means and the mean absolute deviations of the cumulative demands. Inspired by Postek et al. (2018), we could then derive the following analytical bound for $f(\boldsymbol{x})$:

$$\overline{f}_{\texttt{MAD}}(\boldsymbol{x}) = c\mathbf{1}^\top \boldsymbol{x} + \sum_{t=1}^{T} \max\left\{ h\left(y_0 + \sum_{\tau=1}^{t} x_t - t\mu\right), -b\left(y_0 + \sum_{\tau=1}^{t} x_t - t\mu\right)\right\} + \frac{\sigma T(T+1)(h+b)}{4}.$$

Though, we numerically confirm that this bound is significantly inferior to both $\overline{f}_{\texttt{L}}(\boldsymbol{x})$ and $\overline{f}_{\texttt{Q}}(\boldsymbol{x})$. Besides, regardless of $\sigma$, an $\boldsymbol{x} \in \mathcal{X}$ that minimizes $\overline{f}_{\texttt{MAD}}$ is evidently the same as the solution of a deterministic optimization problem which uses $\mu$ in lieu of each $\xi_t$, and thus the nominal solution is the most robust in view of the enlarged ambiguity set. The same observation does not hold true for the original mean-variance ambiguity set $\mathcal{P}$. □

### 3.2. Efficiently solvable progressive approximation

Previously, we have proposed two conservative approximations for $f(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^T_+$, with however no way of quantifying how accurate they are. We now complement our earlier results with a progressive approximation $\underline{f}(\boldsymbol{x})$ which itself is another optimization problem whose size grows linearly with $T$:

$$\begin{aligned} \underline{f}(\boldsymbol{x}) = \text{minimize} \quad & c\mathbf{1}^\top \boldsymbol{x} + \alpha + \mu\mathbf{1}^\top \boldsymbol{\beta} + (\mu^2 + \sigma^2)\mathbf{1}^\top \boldsymbol{\gamma} \\ \text{subject to} \quad & \alpha \in \mathbb{R},\ \boldsymbol{\beta} \in \mathbb{R}^T,\ \boldsymbol{\gamma} \in \mathbb{R}^T \\ & \left(\alpha(\boldsymbol{x}, \check{\boldsymbol{e}}^k), \boldsymbol{\beta}(\check{\boldsymbol{e}}^k), \boldsymbol{\gamma}\right) \succeq_{\mathcal{K}_T} \mathbf{0} \quad \forall k = 0, \ldots, T, \end{aligned}$$

where, for any $\boldsymbol{x} \in \mathbb{R}^T_+$ and $\boldsymbol{e} \in \mathcal{E}$, $\alpha(\boldsymbol{x}, \boldsymbol{e})$ and $\boldsymbol{\beta}(\boldsymbol{e})$ are defined as in Theorem 1.

THEOREM 6 **(Progressive approximation)**. *We have that* $f(\boldsymbol{x}) \geqslant \underline{f}(\boldsymbol{x})$ *for all* $\boldsymbol{x} \in \mathbb{R}^T_+$.

*Proof.* This is an immediate consequence of Theorem 1 as we can interpret $f(\boldsymbol{x})$ as the optimal objective value of Problem (5) with the uncertain vector $\boldsymbol{e} \in \mathcal{E}$ and $\underline{f}(\boldsymbol{x})$ as the optimal objective value of the same robust program with, however, a smaller uncertainty set $\{\check{\boldsymbol{e}}^0, \ldots, \check{\boldsymbol{e}}^T\} \subseteq \mathcal{E}$. □

To motivate the rationale behind this progressive approximation, we consider a stylised example of a two-period distributionally robust uncapacitated inventory problem where the inventory is initially empty and show that optimizing $\underline{f}(\boldsymbol{x})$ and $f(\boldsymbol{x})$ attain the same optimal objective value.

Theorem 7. *When $T = 2$ and $y_0 = 0$, we have that* $\min_{\boldsymbol{x}\in\mathbb{R}^T_+} f(\boldsymbol{x}) = \min_{\boldsymbol{x}\in\mathbb{R}^T_+} \underline{f}(\boldsymbol{x})$.

As a progressive approximation, its optimal solution $\underline{\boldsymbol{x}}^\star$ is not necessarily optimal in Problem (2). Even when $T = 2$, Theorem 7 could only imply that $\arg\min_{\boldsymbol{x}\in\mathbb{R}^2_+} f(\boldsymbol{x}) \subseteq \arg\min_{\boldsymbol{x}\in\mathbb{R}^2_+} \underline{f}(\boldsymbol{x})$ and this subset relationship could be strict. We thus propose to use a conservative approximation which is due to Theorem 4 or Theorem 5 to obtain an approximate optimal solution $\overline{\boldsymbol{x}}^\star$ and approximate the associated worst-case cost $f(\overline{\boldsymbol{x}}^\star)$ by $[\underline{f}(\overline{\boldsymbol{x}}^\star), \overline{f}(\overline{\boldsymbol{x}}^\star)]$ as well as upper bound the optimality gap $f(\overline{\boldsymbol{x}}^\star) - f(\boldsymbol{x}^\star)$ by $\overline{f}(\overline{\boldsymbol{x}}^\star) - \underline{f}(\underline{\boldsymbol{x}}^\star)$. To avoid overloading the notation, we will henceforth use $\overline{\boldsymbol{x}}^\star_{\mathtt{L}}$ and $\overline{\boldsymbol{x}}^\star_{\mathtt{Q}}$ to denote the L- and the Q-conservative solutions, respectively, in the remainder of the paper.

## 4. Plug-and-play conic extensions

We now consider how to reduce the conservatism of the proposed distributionally robust inventory model and its approximations by injecting additional distributional information besides the mean and the variance. The results in this section primarily rely on the conic representations of our base and approximate models, which are the consequences of Theorems 1, 4, and 6. The main idea is that when the ambiguity set is shrunk, a new cone is derived and used in lieu of $\mathcal{K}_T$ in our earlier results. We refer to this cone replacement as a *plug-and-play* feature of the studied inventory model.

### 4.1. Non-negative support

When returns are not allowed, naturally demands are non-negative with probability one, and to incorporate this information, we may consider the shrunk ambiguity set

$$\mathcal{P}^\dagger = \left\{ \mathbb{P} \in \mathcal{M}_+(\mathbb{R}^T_+) : \mathbb{P}\left(\boldsymbol{\xi} \in \mathbb{R}^T_+\right) = 1,\ \mathbb{E}_{\mathbb{P}}[\xi_t] = \mu,\ \mathbb{E}_{\mathbb{P}}[\xi_t^2] = \mu^2 + \sigma^2 \quad \forall t = 1, \ldots, T \right\},$$

and the less conservative cost function

$$f^\dagger(\boldsymbol{x}) = \max_{\mathbb{P}\in\mathcal{P}^\dagger} \mathbb{E}_{\mathbb{P}} \left[ \sum_{t=1}^T \left( c x_t + \max\left\{ h\left( y_0 + \sum_{\tau=1}^t (x_\tau - \xi_\tau) \right), -b\left( y_0 + \sum_{\tau=1}^t (x_\tau - \xi_\tau) \right) \right\} \right) \right].$$

In this case, Theorems 1, 4 and 6 readily extend by replacing the original cone $\mathcal{K}_T$ therein with

$$\mathcal{K}^\dagger_T = \left\{ (\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}) \in \mathbb{R}_+ \times \mathbb{R}^T \times \mathbb{R}^T_+ : \exists (\boldsymbol{\delta}, \boldsymbol{\theta}) \in \mathbb{R}^T_+ \times \mathbb{R}^T_+,\ \alpha \geqslant \mathbf{1}^\top \boldsymbol{\theta},\ (\theta_t, \beta_t - \delta_t, \gamma_t) \in \mathcal{K}_1\ \forall t = 1, \ldots, T \right\}.$$

One could verify that $\mathcal{K}_T \subset \mathcal{K}^\dagger_T$, which is a sign of reduced conservatism. We establish the correctness of this new cone in view of the exact model below.

THEOREM 8. *We have that*

$$\begin{aligned} f^{\dagger}(\boldsymbol{x}) = \text{ minimize } \quad & c\mathbf{1}^\top \boldsymbol{x} + \alpha + \mu \mathbf{1}^\top \boldsymbol{\beta} + (\mu^2 + \sigma^2)\mathbf{1}^\top \boldsymbol{\gamma} \\ \text{subject to } \quad & \alpha \in \mathbb{R},\ \boldsymbol{\beta} \in \mathbb{R}^T,\ \boldsymbol{\gamma} \in \mathbb{R}^T \\ & (\alpha(\boldsymbol{x}, \boldsymbol{e}), \boldsymbol{\beta}(\boldsymbol{e}), \boldsymbol{\gamma}) \succeq_{\mathcal{K}_T^{\dagger}} \mathbf{0} \quad \forall \boldsymbol{e} \in \mathcal{E}, \end{aligned}$$

*where* $\alpha(\boldsymbol{x}, \boldsymbol{e})$ *and* $\boldsymbol{\beta}(\boldsymbol{e})$ *are defined as in Theorem 1, for all* $\boldsymbol{x} \in \mathbb{R}_+^T$.

Though, it appears in our experiments that the difference between $\min_{\boldsymbol{x} \in \mathcal{X}} f(\boldsymbol{x})$ and $\min_{\boldsymbol{x} \in \mathcal{X}} f^{\dagger}(\boldsymbol{x})$ is not significant. Hence, the non-negativity of $\boldsymbol{\xi}$ may not be a crucial factor to consider.

### 4.2. Uncorrelated demands

Next, we consider a scenario of uncorrelated demands, which itself could be an approximation of independence. Analogously to Section 4.1, we focus on the following restricted ambiguity set

$$\mathcal{P}^{\perp} = \left\{ \mathbb{P} \in \mathcal{M}_+(\mathbb{R}^T) : \begin{array}{ll} \mathbb{P}\left(\boldsymbol{\xi} \in \mathbb{R}^T\right) = 1 & \\ \mathbb{E}_{\mathbb{P}}[\xi_t] = \mu,\ \mathbb{E}_{\mathbb{P}}[\xi_t^2] = \mu^2 + \sigma^2 & \forall t = 1, \ldots, T \\ \mathbb{E}_{\mathbb{P}}[\xi_s \xi_t] = \mu^2 & \forall (s,t) : 1 \leqslant s < t \leqslant T \end{array} \right\}$$

and the corresponding less conservative cost function

$$f^{\perp}(\boldsymbol{x}) = \max_{\mathbb{P} \in \mathcal{P}^{\perp}} \mathbb{E}_{\mathbb{P}} \left[ \sum_{t=1}^{T} \left( c x_t + \max\left\{ h\left( y_0 + \sum_{\tau=1}^{t} (x_\tau - \xi_\tau) \right), -b\left( y_0 + \sum_{\tau=1}^{t} (x_\tau - \xi_\tau) \right) \right\} \right) \right].$$

For the subsequent results, we use the following proper cone

$$\mathcal{K}_T^{\perp} = \left\{ (\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Theta}) \in \mathbb{R}_+ \times \mathbb{R}^T \times \mathbb{R}_+^T \times \mathbb{S}^T : \begin{bmatrix} \operatorname{diag}(\boldsymbol{\gamma}) + \boldsymbol{\Theta} & \frac{1}{2}\boldsymbol{\beta} \\ \frac{1}{2}\boldsymbol{\beta}^\top & \alpha \end{bmatrix} \succeq \mathbf{0} \right\}.$$

THEOREM 9. *We have that*

$$\begin{aligned} f^{\perp}(\boldsymbol{x}) = \text{ minimize } \quad & c\mathbf{1}^\top \boldsymbol{x} + \alpha + \mu \mathbf{1}^\top \boldsymbol{\beta} + (\mu^2 + \sigma^2)\mathbf{1}^\top \boldsymbol{\gamma} + \mu^2 \mathbf{1}^\top \boldsymbol{\Theta} \mathbf{1} \\ \text{subject to } \quad & \alpha \in \mathbb{R},\ \boldsymbol{\beta} \in \mathbb{R}^T,\ \boldsymbol{\gamma} \in \mathbb{R}^T,\ \boldsymbol{\Theta} \in \mathbb{S}^T \\ & (\alpha(\boldsymbol{x}, \boldsymbol{e}), \boldsymbol{\beta}(\boldsymbol{e}), \boldsymbol{\gamma}, \boldsymbol{\Theta}) \succeq_{\mathcal{K}_T^{\perp}} \mathbf{0} \quad \forall \boldsymbol{e} \in \mathcal{E} \\ & \operatorname{diag}(\boldsymbol{\Theta}) = \mathbf{0}, \end{aligned}$$

*where* $\alpha(\boldsymbol{x}, \boldsymbol{e})$ *and* $\boldsymbol{\beta}(\boldsymbol{e})$ *are defined as in Theorem 1, for all* $\boldsymbol{x} \in \mathbb{R}_+^T$.

The above exact formulation is a correlation-aware counterpart of Theorem 1, which readily leads us to a progressive approximation akin to Theorem 6. For the conservative approximation, we can have the following result, which is a counterpart of Theorem 4. We could also note from the Schur Complement Lemma that $(\alpha,\boldsymbol{\beta},\boldsymbol{\gamma},\mathbf{0})\in\mathcal{K}_T^\perp$ if and only if $(\alpha,\boldsymbol{\beta},\boldsymbol{\gamma})\in\mathcal{K}_T$.

THEOREM 10. *We have that*

$$
\begin{aligned}
f^\perp(\boldsymbol{x}) \leqslant \text{minimize}\quad & c\mathbf{1}^\top\boldsymbol{x}+\alpha+\mu\mathbf{1}^\top\boldsymbol{\beta}+(\mu^2+\sigma^2)\mathbf{1}^\top\boldsymbol{\gamma}+\mu^2\mathbf{1}^\top\boldsymbol{\Theta}\mathbf{1}\\
\text{subject to}\quad & \alpha\in\mathbb{R},\ \boldsymbol{\beta}\in\mathbb{R}^T,\ \boldsymbol{\gamma}\in\mathbb{R}^T,\ \boldsymbol{\Theta}\in\mathbb{S}^T,\ \boldsymbol{\kappa}\in\mathbb{R}_+^{2T+1},\\
& \qquad (\pi_t,\boldsymbol{\pi}_t^u,\boldsymbol{\pi}_t^v,\boldsymbol{\Pi}_t^w)\in\mathbb{R}\times\mathbb{R}^T\times\mathbb{R}^T\times\mathbb{S}^T\ \ \forall t=1,\dots,T\\
& (\pi_t-\kappa_t,\boldsymbol{\pi}_t^u,\boldsymbol{\pi}_t^v,\boldsymbol{\Pi}_t^w)\succeq_{\mathcal{K}_T^\perp}\mathbf{0}\quad \forall t=1,\dots,T\\
& \left(\pi_t+y_0+\sum_{\tau=1}^t x_\tau-\kappa_{T+t},\boldsymbol{\pi}_t^u-(1,\dots,1,0,\dots,0)^\top,\boldsymbol{\pi}_t^v,\boldsymbol{\Pi}_t^w\right)\succeq_{\mathcal{K}_T^\perp}\mathbf{0}\\
& \qquad \forall t=1,\dots,T\\
& \Bigg(\alpha-hTy_0-h\sum_{t=1}^T x_t(T-t+1)-(b+h)\sum_{t=1}^T\pi_t-\kappa_{2T+1},\\
& \qquad \boldsymbol{\beta}-(b+h)\sum_{t=1}^T\boldsymbol{\pi}_t^u+h(T,T-1,\dots,1)^\top,\\
& \qquad \boldsymbol{\gamma}-(b+h)\sum_{t=1}^T\boldsymbol{\pi}_t^v,\boldsymbol{\Theta}-(b+h)\sum_{t=1}^T\boldsymbol{\Pi}_t^w\Bigg)\succeq_{\mathcal{K}_T^\perp}\mathbf{0}\\
& \operatorname{diag}(\boldsymbol{\Theta})=\mathbf{0},
\end{aligned}
$$

*where the vector* $(1,\dots,1,0,\dots,0)^\top$ *in the third line of constraints has* $t$ *components equal to one and* $T-t$ *components equal to zero, for all* $\boldsymbol{x}\in\mathbb{R}_+^T$.

It is observed that both the progressive and the L-conservative approximation of $\min_{\boldsymbol{x}\in\mathcal{X}} f^\perp(\boldsymbol{x})$ are still close and that $\min_{\boldsymbol{x}\in\mathcal{X}} f^\perp(\boldsymbol{x})$ could be significantly smaller than $\min_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x})$. Therefore, if the demands appear to be serially uncorrelated, adopting the extended models with the cone $\mathcal{K}^\perp$ lessens conservatism of the robust model.

# 5. Numerical experiments

In this section, we will compare the solutions of our distributionally robust inventory model with those of the state-of-the-art benchmarks which have access to the (possibly incorrect) demand distribution. In addition, we also assess the quality of our conservative and progressive approximations. Finally, we perform a sensitivity analysis with respect to the cost parameters: $c$, $b$, and $h$.

## 5.1. Comparison with benchmarking approaches

We compare the exact minimizer $\boldsymbol{x}^\star$ of $\min_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x})$, which due to Theorem 1 is representable as a finite second-order cone program, with the solutions of two stochastic inventory problems. First, we refer to the minimizer of

$$\begin{aligned} \text{minimize} \quad & \mathbb{E}_{\mathbb{P}}\left[\sum_{t=1}^{T}\left(cx_t + \max\left\{h\left(y_0 + \sum_{\tau=1}^{t}(x_\tau - \xi_\tau)\right), -b\left(y_0 + \sum_{\tau=1}^{t}(x_\tau - \xi_\tau)\right)\right\}\right)\right] \\ \text{subject to} \quad & \boldsymbol{x} \in \mathcal{X} \end{aligned}$$

as a *stochastic solution* and that of

$$\begin{aligned} \text{minimize} \quad & \mathbb{E}_{\mathbb{P}}\left[\sum_{t=1}^{T}\left(cx_t(\boldsymbol{\xi}^{t-1}) + \max\left\{h\left(y_0 + \sum_{\tau=1}^{t}(x_\tau(\boldsymbol{\xi}^{\tau-1}) - \xi_\tau)\right),\right.\right.\right. \\ & \qquad \left.\left.\left. -b\left(y_0 + \sum_{\tau=1}^{t}(x_\tau(\boldsymbol{\xi}^{\tau-1}) - \xi_\tau)\right)\right\}\right)\right] \\ \text{subject to} \quad & \left(x_1, x_2(\boldsymbol{\xi}^1), \ldots, x_T(\boldsymbol{\xi}^{T-1})\right)^\top \in \mathcal{X} \quad \mathbb{P}\text{-a.s.} \end{aligned}$$

as an *adaptive solution*. In this experiment, we set aside computational tractability and are only concerned with the quality of different solutions. Note that tractability will be the focus of our next experiment concerning the proposed approximations. Throughout, it is assumed that $\mathcal{X} = \mathbb{R}_+^T$ and the demands are serially independent and identically distributed with

$$\mathbb{P}\left(\xi_t = \xi^{\mathrm{L}}\right) = p^{\mathrm{L}} \quad \text{and} \quad \mathbb{P}\left(\xi_t = \xi^{\mathrm{H}}\right) = p^{\mathrm{H}} \quad \forall t \in \{1, \ldots, T\}$$

for some $(p^{\mathrm{L}}, \xi^{\mathrm{L}}, p^{\mathrm{H}}, \xi^{\mathrm{H}})$ such that $\xi^{\mathrm{H}} > \xi^{\mathrm{L}}$, $p^{\mathrm{L}}$ and $p^{\mathrm{H}}$ are both non-negative, and they sum up to one. It is known from (Scarf 1960) that the adaptive solution follows a base-stock policy denoted by $\boldsymbol{S}^\star \in \mathbb{R}^T$, whereas the stochastic solution denoted by $\tilde{\boldsymbol{x}}^\star$ has been recently studied by Basciftci et al. (2021). When $\mathbb{P}$ is the true demand distribution, the base-stock policy $\boldsymbol{S}^\star$ yields a lower expected total cost than the stochastic solution $\tilde{\boldsymbol{x}}^\star$, and the latter yields a lower expected total cost than the proposed robust solution $\boldsymbol{x}^\star$. As this premise however is likely untrue, we perform a stress test by determining the expected total cost corresponding to $\boldsymbol{x}^\star$ and $\tilde{\boldsymbol{x}}^\star$ under the contaminated probability distribution $\mathbb{P}_\lambda = (1-\lambda)\mathbb{P} + \lambda\mathbb{P}^{\rm wc}$, where $\lambda \in [0,1]$ represents the contamination level and $\mathbb{P}^{\rm wc}$ is the demand distribution that (approximately) attains the worst-case expected total cost when $\boldsymbol{x} = \tilde{\boldsymbol{x}}^\star$ ($\varepsilon = 10^{-4}$ is used). Note that, unlike $\mathbb{P}$, $\mathbb{P}^{\rm wc}$ does not generate serially independent demands.

We consider two plausible scenarios: (i) when it is likely for the demand to be small but there can be a rare surge and (ii) when it is likely for the demand to be large but there might be a rare drop. As a representative of scenario (i), we set $p^{\rm L} = 0.7$, $p^{\rm H} = 0.3$, $\xi^{\rm L} = 30$ and $\xi^{\rm H} = 70$. Even though it is more likely for the demands to be small, the inventory manager should still prepare for a sudden surge in demand, which may force him or her to backlog the unmet demands. This is particularly crucial when backlogging is expensive, and to investigate this effect we set $b = 3 > 1 = h$ and $c = 8$. As a representative of scenario (ii), we similarly set $p^{\rm L} = 0.3$, $p^{\rm H} = 0.7$ and retain $\xi^{\rm L} = 30$, $\xi^{\rm H} = 70$. For this case, we are particularly concerned with the possible drop in the demands and the high holding cost, and we thus set $h = 3 > 1 = b$ and $c = 3$. As none of these optimization problems are tractable, we choose to work with a relatively small $T = 6$, and we further assume that the initial inventory is $y_0 = 0$. **Managerial insight:** Results for these two scenarios are reported in Figure 1 (left) and Figure 1 (right), respectively, from which it can be observed that the robust policy $\boldsymbol{x}^\star$ is more resilient to the misspecification of demand distribution. Besides, the contamination level $\lambda$ at which the robust solution starts to outperform the stochastic solution is given by 11.85% for scenario (i) and by 34.78% for scenario (ii), respectively.

Next, we similarly compare our robust solution with the base-stock policy $\boldsymbol{S}^\star$. However, since the base-stock policy is adaptive and thus at an advantageous position, we execute our robust policy in a shrinking horizon fashion to facilitate a fairer comparison. Put differently, for the first period we solve $\min_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x})$ to determine $\boldsymbol{x}^\star$ but only implement the here-and-now decision $x_1^\star$. For the second period, we resolve the same optimization problem with the updated initial inventory and

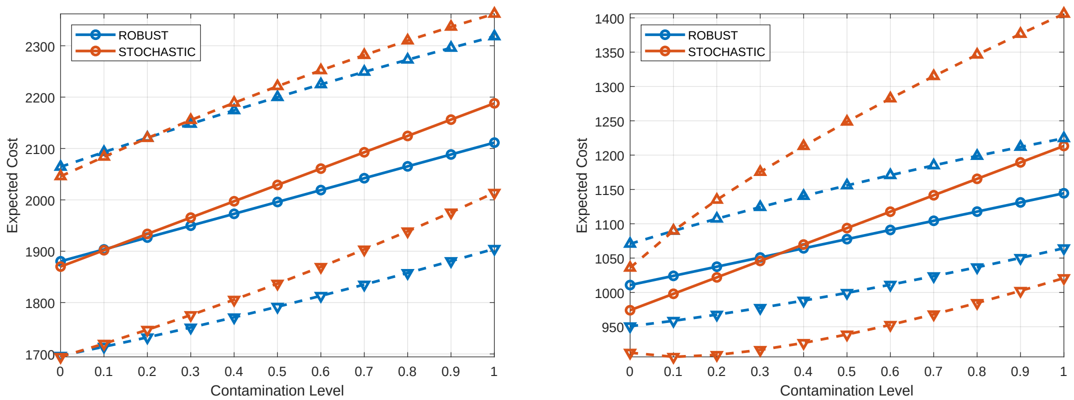


**Figure 1 Comparison between our distributionally robust and the stochastic solutions under different contaminated demand distributions in terms of the expected total cost (solid line) and the expected total cost $\pm\frac{1}{2}$ standard deviation (upper/lower dashed line)**

with the number of periods reduced by one, and we only implement the here-and-now ordering decision again, and so on and so forth; see *e.g.* Solyalı et al. (2016) and Mamani et al. (2017). We evaluate both the adaptive robust and the base-stock policies under different contaminated demand distributions $\mathbb{P}_\lambda$, $\lambda \in [0,1]$, where we substitute for $\mathbb{P}^{\text{wc}}$ the worst-case demand distribution when $\boldsymbol{x} = \mathbb{E}_{\mathbb{P}}[\tilde{\tilde{\boldsymbol{x}}}^\star]$ with $\tilde{\tilde{\boldsymbol{x}}}^\star$ representing (adaptive) ordering decisions equivalent to $\boldsymbol{S}^\star$. For both scenarios (i) and (ii), when $\lambda = 0$ (and thus $\mathbb{P}_\lambda = \mathbb{P}$), the base-stock policy (which is known to be optimal under this distributional setting) is superior to the adaptive robust policy. **Managerial insight:** As the out-of-sample distribution deviates further from the in-sample distribution, Figure 2 shows that the adaptive robust policy becomes increasingly competitive and eventually outperforms the base-stock policy, which highlights the resilience of the proposed robust policies even in the adaptive setting. We also remark that the differences between the two policies are more prominent in scenario (ii), see Figure 2 (right), where the adaptive robust policy seems to have an inadvertent variance reduction effect for the distribution of the total cost. Finally, we can also relate Figure 2 to its non-adaptive counterpart, *i.e.*, Figure 1. While we indeed observe that the adaptive robust and the base-stock solutions yield a smaller expected cost than the robust and the stochastic solutions, respectively, the reduction is not monumental. Thus, when $T$ is only moderately large, determining the adaptive robust and/or the base-stock policy may not be a worthwhile pursuit considering the other benefits of agreeing to an advance purchase, which we discuss more in Section 6.

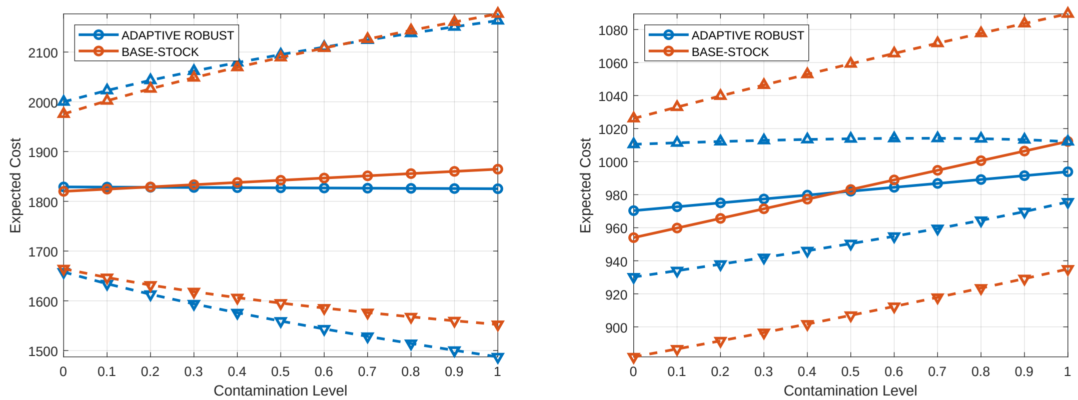


**Figure 2 Comparison between our adaptive distributionally robust and the adaptive (i.e., base-stock) solutions under different contaminated demand distributions in terms of the expected total cost (solid line) and the expected total cost $\pm\frac{1}{2}$ standard deviation (upper/lower dashed line)**

## 5.2. Quality of the proposed approximations

To validate the performance of our proposed approximations, we experiment with randomly generated problem instances. Without loss of generality, we set the unit ordering cost $c$ and the mean demand $\mu$ to be one. We then independently choose the unit holding cost $h$ and the unit backlogging cost $b$ from a uniform distribution $\mathcal{U}([0,1])$ as well as the demand's standard deviation $\sigma$ from $\mathcal{U}([0,2])$. For each $T \in \{10,20,30,40,50\}$, we set $y_0 = 0$ and run fifty random experiments in total. Note that, when $T \geqslant 20$, $\min_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x})$ cannot be solved exactly within a time limit of 12 hours using MOSEK ApS (2019) and YALMIP interface (Löfberg 2004) on a 2.90GHz i7-10700 CPU machine with 16GB RAM, which implies the need for scalable approximate solution methods.

First, we would like to show that the L-approximation due to Theorem 4 dominates the approximation from Remark 2, but it is still (marginally) inferior to the Q-approximation due to Theorem 5. For convenience, we denote the optimizers of $\min_{\boldsymbol{x}\in\mathcal{X}} \overline{f}_{\mathtt{L}}(\boldsymbol{x})$, $\min_{\boldsymbol{x}\in\mathcal{X}} \overline{f}_{\mathtt{Q}}(\boldsymbol{x})$ and $\min_{\boldsymbol{x}\in\mathcal{X}} \overline{f}_{\mathtt{MAD}}(\boldsymbol{x})$ by $\overline{\boldsymbol{x}}^\star_{\mathtt{L}}$, $\overline{\boldsymbol{x}}^\star_{\mathtt{Q}}$ and $\overline{\boldsymbol{x}}^\star_{\mathtt{MAD}}$, respectively. Figure 3 (left) shows, for each fixed $T$, the mean as well as the 10% and 90% quantiles of the *approximate worst-case expected costs*, namely $\overline{f}_{\mathtt{L}}(\overline{\boldsymbol{x}}^\star_{\mathtt{L}})$ and $\overline{f}_{\mathtt{MAD}}(\overline{\boldsymbol{x}}^\star_{\mathtt{MAD}})$. The difference between the L- and the Q-approximations is not as discernible, and as a result, we measure the *improvement of* Q *over* L (in terms of conservativeness) by

$$\frac{\overline{f}_{\mathtt{L}}(\overline{\boldsymbol{x}}^\star_{\mathtt{L}}) - \overline{f}_{\mathtt{Q}}(\overline{\boldsymbol{x}}^\star_{\mathtt{Q}})}{\overline{f}_{\mathtt{Q}}(\overline{\boldsymbol{x}}^\star_{\mathtt{Q}})},$$

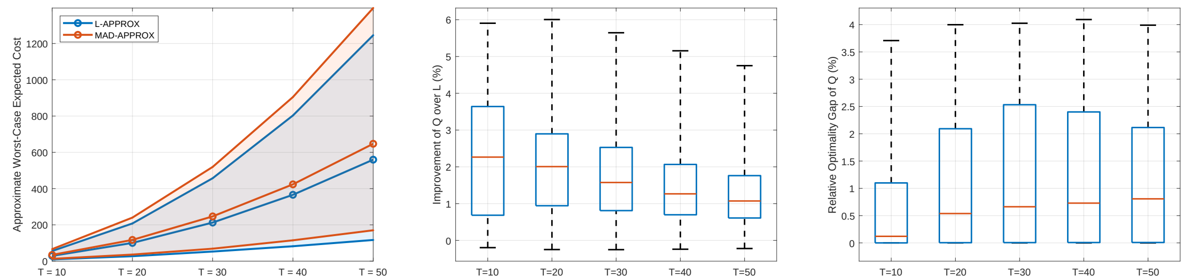


**Figure 3** **Measurements of our approximation quality: (left) the approximate worst-case expected costs ($\overline{f}_{\mathtt{L}}(\overline{\boldsymbol{x}}^\star_{\mathtt{L}})$ and $\overline{f}_{\mathtt{MAD}}(\overline{\boldsymbol{x}}^\star_{\mathtt{MAD}})$), (middle) improvement in conservativeness of $\overline{\boldsymbol{x}}^\star_{\mathtt{Q}}$ over $\overline{\boldsymbol{x}}^\star_{\mathtt{L}}$, and (right) our conservative bound on the relative optimality gap of $\overline{\boldsymbol{x}}^\star_{\mathtt{Q}}$.**

which is then visualized in Figure 3 (middle). Last but not least, we want to quantify how close $\overline{\boldsymbol{x}}^\star_{\mathtt{Q}}$ is to being optimal in view of the original problem $\min_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x})$. To this end, we define the *relative optimality gap of* $\mathtt{Q}$ as $(f(\overline{\boldsymbol{x}}^\star_{\mathtt{Q}}) - f(\boldsymbol{x}^\star))/f(\boldsymbol{x}^\star)$ and propose to upper bound it by

$$\frac{f(\overline{\boldsymbol{x}}^\star_{\mathtt{Q}}) - f(\boldsymbol{x}^\star)}{f(\boldsymbol{x}^\star)} \leqslant \frac{\overline{f}_{\mathtt{Q}}(\overline{\boldsymbol{x}}^\star_{\mathtt{Q}}) - \underline{f}(\boldsymbol{x}^\star)}{\underline{f}(\boldsymbol{x}^\star)} \leqslant \frac{\overline{f}_{\mathtt{Q}}(\overline{\boldsymbol{x}}^\star_{\mathtt{Q}}) - \underline{f}(\underline{\boldsymbol{x}}^\star)}{\underline{f}(\underline{\boldsymbol{x}}^\star)},$$

and the rightmost expression, dependent only on the approximate worst-case expected total costs, could be computed efficiently even for a large $T$. Evidently, Figure 3 (right) indicates that the optimality gap always almost vanishes highlighting the efficacy of both our conservative and progressive approximations. **Managerial insight:** On average, $\overline{\boldsymbol{x}}^\star_{\mathtt{Q}}$ is slightly less conservative than $\overline{\boldsymbol{x}}^\star_{\mathtt{L}}$, and through transitivity, it is significantly less conservative than $\overline{\boldsymbol{x}}^\star_{\mathtt{MAD}}$ especially as $T$ gets large. More importantly, $\overline{\boldsymbol{x}}^\star_{\mathtt{Q}}$ is close to being robustly optimal, which implies that our distributionally robust inventory problem can be efficiently solved almost exactly. While $\overline{\boldsymbol{x}}^\star_{\mathtt{L}}$ appears to be a (slightly) poorer substitute of $\overline{\boldsymbol{x}}^\star_{\mathtt{Q}}$, it is still useful as it can cater for additional information more readily and is thus more receptive to changes; see Section 4.

Given the quality of the proposed progressive approximation, we can use the underlying scenario-reduction idea to construct an approximate worst-case distribution by solving a variant of Problem (6) with the exponentially-sized uncertainty set $\mathcal{E}$ replaced by $\breve{\mathcal{E}} = \{\breve{\boldsymbol{e}}^0, \ldots, \breve{\boldsymbol{e}}^T\}$. In this case, the construction of the approximate worst-case distribution is dependent on the optimally chosen $\overline{\alpha}^\star : \breve{\mathcal{E}} \mapsto \mathbb{R}$, $\overline{\boldsymbol{\beta}}^\star : \breve{\mathcal{E}} \mapsto \mathbb{R}^T$, $\overline{\boldsymbol{\gamma}}^\star : \breve{\mathcal{E}} \mapsto \mathbb{R}^T$. While the construction of the distribution described

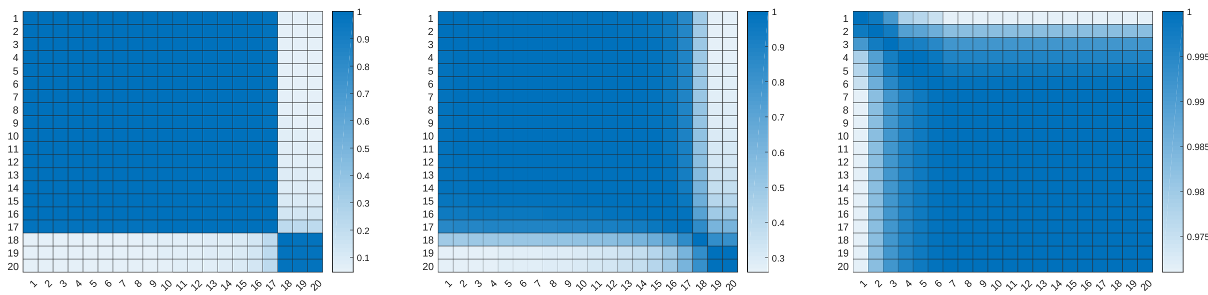

**Figure 4 Correlation matrices of the (approximate) worst-case demand distributions.**

in (7) and (8) indicates that demands should be statistically dependent in the worst-case, we provide evidence asserting a stronger statement that they are in fact highly correlated (*i.e.,* they are strongly linearly dependent). As an illustrative example, we set $c = 1$, $h = 1$, $b = 1/3$, $T = 20$, $y_0 = 0$, $\mu = 1$ and consider different $\sigma \in \{\sqrt{0.001}, \sqrt{0.005}, \sqrt{0.025}\}$. For each of the parameter settings, we compute the $\overline{\boldsymbol{x}}^\star_{\mathtt{Q}}$ and the corresponding (approximate) worst-case distribution accordingly to the described procedure. **Managerial insight:** We visualize the resultant correlation matrices with heatmaps in Figure 4 and find that there exists a breakpoint $t^\star \in \{1, \ldots, T\}$ such that the demands $\xi_1, \ldots, \xi_{t^\star}$ are almost perfectly positively correlated and so are the remaining demands $\xi_{t^\star+1}, \ldots, \xi_T$. One could think of $t^\star$ as the period with a transitional shift, *e.g.*, a sudden surge or a sudden drop in the demands over time, which could be due to emergency or seasonal fluctuations as well as the availability of substitute products. Recent examples include Covid-19 vaccines and energy supplies. This observation remains unchanged with different parameter settings, and it sheds light on the potential benefit of using the uncorrelated ambiguity set $\mathcal{P}^\perp$ instead of $\mathcal{P}$ when justified.

### 5.3. Ordering pattern and sensitivity analysis

In this experiment, we are interested in comparing the *stochastic solution* with our *robust solution*, but this time we consider a demand distribution $\mathbb{P}$ which is serially independent and is characterized by $\xi_t \sim \mathcal{N}(\mu, \sigma^2)$, $\forall t$, as opposed to the two-point distribution previously considered. The objective here is to understand the ordering patterns of both. The computation of the stochastic solution $\tilde{\boldsymbol{x}}^\star$ becomes more challenging but could still be achieved by using a projected gradient method (Polyak 1987), which iteratively constructs a new solution by following the gradient descent direction, derived in Appendix B, and projecting it on the feasible set $\mathcal{X} = \mathbb{R}^T_+$ when necessary.

For tractability purposes, we approximate our robust solution $\boldsymbol{x}^\star$ by its conservative estimate $\overline{\boldsymbol{x}}^\star_{\mathtt{Q}}$. Figure 5 (left) shows the total order of both solutions, *i.e.*, $\mathbf{1}^\top\tilde{\boldsymbol{x}}^\star$ and $\mathbf{1}^\top\overline{\boldsymbol{x}}^\star_{\mathtt{Q}}$ when $T = 20$, $c = 1$, $b = 0.5$, $y_0 = 0$, $\mu = 1$ and $\sigma = 0.5$ for varying $h \in \{0, \ldots, 1\}$. Similarly, Figure 5 (right) shows the total order of both solutions for the same set of parameters with the exception that now $h$ is fixed at 0.5 and $b$ is varied in $\{0, \ldots, 1\}$. As expected, when the holding cost increases, both solutions order less, and when it is expensive to backlog, both solutions order more. From the figure, in comparison to the stochastic solution, it can be seen that the robust solution is more receptive to the changes in $h$ and $b$, which we attribute to the fact that the change in $h$ and $b$ has an impact on the worst-case demand distribution but not on the nominal normal distribution. Note that each bar consists of two parts: the darker-shaded part represents the total order made in the first half of the planning horizon ($t \in \{1, \ldots, T/2\}$), whereas the lighter-shaded part represents that made in the second half ($t \in \{T/2+1, \ldots, T\}$). **Managerial insight:** For both solutions, at least half of the total order is made in the first half of the planning horizon. The ratio between the first-half and the second-half order of the robust solution appears significantly greater than that of the stochastic solution. Hence, a stakeholder who supplies products to an inventory manager adopting our robust policy can better utilize its production capacities in the second half for other purposes including manufacturing the same products and selling them to other stores at a higher price.

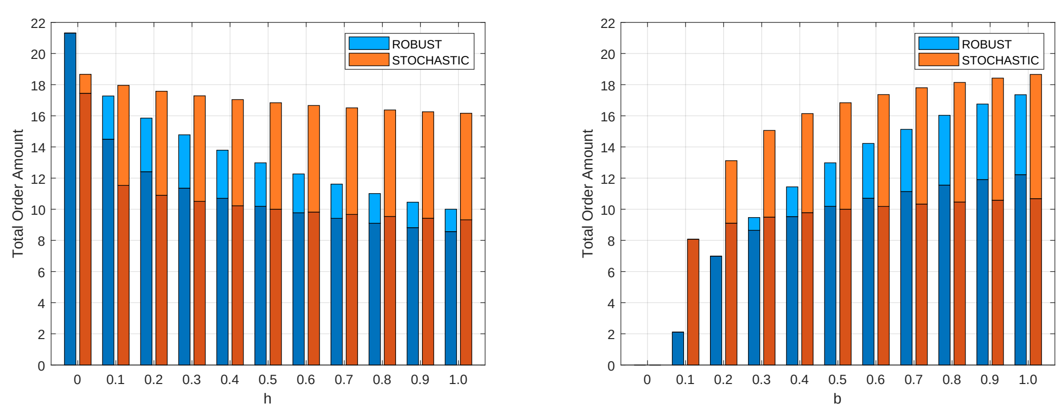


**Figure 5 Total order comparison between our robust and the stochastic solutions under the normality and independence assumption.**

## 6. Closing the loop: Is there an advantage to commit early?

An *advance purchase discount* (Tang et al. 2004) is often used by a supplier to incentivize some inventory managers to commit their purchase orders early. In this case study, we aim to answer the question: could *both* the supplier and the inventory manager simultaneously benefit from striking an early deal with a discount? Perhaps surprisingly, the answer is positive.

In our case study, we use a dataset from Babyonlinedress, a fast-fashion e-retailer that through Amazon sells wedding gowns and full dresses, which was made available in (Sun et al. 2021). Partly, this dataset is concerned with the distribution of products from Amazon, which we could regard as our inventory, for overseas deliveries. Within the dataset, there are 79 products whose demand appears to follow a two-point distribution. For each of such products, the demand distribution $\mathbb{P}$ satisfies $\mathbb{P}(\xi_t = \xi^{\mathrm{L}}) = p^{\mathrm{L}}$ and $\mathbb{P}(\xi_t = \xi^{\mathrm{H}}) = p^{\mathrm{H}},\ \forall t \in \{1, \ldots, T\}$, where $T = 6$, $\xi^{\mathrm{L}} = 0$ and $\xi^{\mathrm{H}}$ could be normalized to one for numerical convenience and we estimate $p^{\mathrm{L}}$ and $p^{\mathrm{H}}$ using a sample estimator. We then sort the products into four equally-sized groups, with the last group containing just 19 products, based on the value of $p^{\mathrm{H}}$ in an ascending order. While the unit purchasing and the holding cost, $c$ and $h$, can be readily extracted, the dataset does not provide the unit backlogging cost. To circumvent, we choose the average of "returned commission fee" as our $b$.

As an inventory manager, we consider two possibilities of a purchase order. First, we consider an online policy, namely the base-stock policy $\boldsymbol{S}^\star$, and we also consider an advance purchase agreement whose order quantities $\boldsymbol{x}^\star$ are robustly optimal. We further assume that the supplier is willing to offer us a discount from $\{0\%, 5\%, \ldots, 50\%\}$ for the unit purchasing cost provided that we are committed to an advance purchase. As a result, for each product, $\boldsymbol{x}^\star$ depends on the discounted $c$.

From the supplier side, we compare its earning from $\boldsymbol{x}^\star$ to the expected total revenue generated by the base-stock policy, where the expectation is taken with respect to the in-sample two-point distribution $\mathbb{P}$. In other words, we intentionally put the proposed robust policy $\boldsymbol{x}^\star$ at a disadvantage because here the robustness is uncalled for. Table 1 then reports the improvement in terms of the supplier's earning of the robust policy relative to that of the base-stock policy:

$$\frac{(1 - \mathtt{discount})\, \mathbf{1}^\top \boldsymbol{x}^\star - \mathbb{E}_{\mathbb{P}}[\mathbf{1}^\top \tilde{\tilde{\boldsymbol{x}}}^\star]}{\mathbb{E}_{\mathbb{P}}[\mathbf{1}^\top \tilde{\tilde{\boldsymbol{x}}}^\star]} \times 100\%;$$

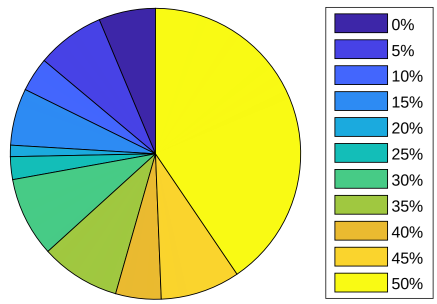


**Figure 6 Distribution of the optimal discount levels for all products when adopting our robust solution.**

see Section 5 for the definition of $\tilde{\tilde{\boldsymbol{x}}}^\star$. In Table 1, the numbers are averaged across different products within the same group. Note that, with discounts, the supplier could earn more because the inventory manager (*i.e.,* Babyonlinedress) may be tempted to order more; however, such benefit diminishes if the discount level is too high. For visualization, we use boldface numbers to indicate where the robust policy outperforms the base-stock policy even when there is no misspecification of the demand distribution. On the contrary, one could observe that the products of Group 4 do not generate significant additional revenue to the supplier when the discount is offered. This is because when $p^{\mathrm{H}}$ is high, Babyonlinedress should prepare to face the demand regardless of the discount level. In this case, the supplier could earn maximally by offering little to no discount. For the sake of comprehensiveness, Figure 6 shows the optimal discount levels for all 79 products.

| Discount level | 0% | 5% | 10% | 15% | 20% | 25% | 30% | 35% | 40% | 45% | 50% |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Group 1 | −16.00 | −0.07 | **15.91** | **32.86** | **46.17** | **55.70** | **62.76** | **68.62** | **73.50** | **78.78** | **83.03** |
| Group 2 | −6.72 | **5.74** | **15.85** | **22.29** | **26.59** | **30.49** | **35.63** | **41.42** | **46.09** | **49.15** | **50.29** |
| Group 3 | −7.07 | −1.67 | **2.16** | **4.62** | **6.99** | **9.88** | **12.93** | **15.15** | **15.85** | **15.13** | **13.17** |
| Group 4 | −16.81 | −15.25 | −14.55 | −14.66 | −14.53 | −14.90 | −15.86 | −17.42 | −19.24 | −21.41 | −24.18 |

**Table 1 Supplier's revenue improvement (%) of the robust solution relative to the base-stock policy for different product groups at different discount levels.**

Similarly, we also analyze the impact of discounts on the inventory management cost. Table 2 reports the improvement in the expected purchasing, holding and backlogging cost of the robust policy (at different discount levels) relative to that of the base-stock policy (which receives no discount). We again use boldface fonts to indicate where the robust inventory policy is less costly

| Discount level | 0% | 5% | 10% | 15% | 20% | 25% | 30% | 35% | 40% | 45% | 50% |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Group 1 | −31.08 | −33.37 | −35.10 | −36.49 | −36.90 | −36.17 | −34.43 | −31.94 | −28.66 | −24.76 | −19.98 |
| Group 2 | −21.64 | −21.15 | −19.95 | −17.99 | −15.41 | −12.32 | −9.21 | −5.55 | −1.18 | **3.76** | **8.45** |
| Group 3 | −11.89 | −9.62 | −7.10 | −4.29 | −1.32 | **1.76** | **5.15** | **8.95** | **12.89** | **16.50** | **20.18** |
| Group 4 | −5.99 | −2.94 | **0.43** | **3.81** | **7.50** | **11.15** | **15.05** | **19.07** | **22.89** | **26.85** | **31.26** |

**Table 2 Inventory manager's cost improvement (%) of the robust solution relative to the base-stock policy for different product groups at different discount levels.**

than the base-stock policy. It is intuitive that, with a higher discount, the robust inventory policy becomes more cost-efficient. Though, we remark that this is not necessarily the case for products in Group 1 because the robust policy minimizes the expected cost under the worst-case distribution, which could greatly deviates from the nominal demand distribution $\mathbb{P}$. Indeed, when $p^{\mathrm{H}}$ is small, the base-stock policy may order little whereas the robust policy may order relatively more to avoid backlogging. Incidentally, this also explains the exorbitant numbers seen in the top row of Table 1.

**Managerial insight:** Finally, we note that, for 33 out of 79 products under consideration, there exists a discount level such that, through the advance purchase agreement, the inventory manager pays less and simultaneously the supplier earns more in comparison to the base-stock policy, which has access to the true demand distribution.

## 7. Conclusions

Concerning with advance purchase contracts which are prevalent for acquiring pharmaceutical, gadget and fashion products, we study a new robust inventory model under the mean-variance ambiguity set whose objective is to minimize the worst-case expected total cost. Innately, this problem does not appear tractable; hence, several high-quality approximations (both conservative and progressive) which are amendable to additional distributional information are proposed. Our numerical experiments and case studies show that not only the proposed inventory policies are efficiently-computable and resilient but they can also save costs and generate higher revenues for the suppliers.

## References

Andres, G. 2022. Novavax's Nuvaxovid COVID-19 vaccine granted interim authorisation in Singapore. `https://www.channelnewsasia.com/singapore/novavax-nuvaxovid-covid-19-vaccine-interim-authorisation-singapore-hsa-2496476`. Accessed on 2nd May 2022.

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

Basciftci, B., S. Ahmed, N. Gebraeel. 2021. Adaptive two-stage stochastic programming with an application to capacity expansion planning. Available on arXiv #1906.03513.

Ben-Tal, A., B. Golany, A. Nemirovski, J. Vial. 2005. Retailer-supplier flexible commitments contracts: A robust optimization approach. *Manufacturing & Service Operations Management* **7**(3) 248–271.

Ben-Tal, A., A. Goryashko, E. Guslitzer, A. Nemirovski. 2004. Adjustable robust solutions of uncertain linear programs. *Mathematical programming* **99**(2) 351–376.

Ben-Tal, A., E. Hochman. 1976. Stochastic programs with incomplete information. *Operations Research* **24**(2) 336–347.

Bertsekas, D. 1995. *Dynamic programming and optimal control*, vol. 1. Athena Scientific.

Bertsimas, D., X. Doan, K. Natarajan, C. Teo. 2010. Models for minimax stochastic linear optimization problems with risk aversion. *Mathematics of Operations Research* **35**(3) 580–602.

Bertsimas, D., M. Sim, M. Zhang. 2019. Adaptive distributionally robust optimization. *Management Science* **65**(2) 604–618.

Bertsimas, D., A. Thiele. 2006. A robust optimization approach to inventory theory. *Operations Research* **54**(1) 150–168.

Bienstock, D., N. Özbay. 2008. Computing robust basestock levels. *Discrete Optimization* **5**(2) 389–414.

Chen, Y., G. Iyengar, C. Wang. 2022. Robust inventory management: A cycle-based approach. *Manufacturing & Service Operations Management* **In press**.

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

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

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

Federgruen, A., P. Zipkin. 1984. An efficient algorithm for computing optimal $(s, S)$ policies. *Operations Research* **32**(6) 1268–1285.

Fu, M. 1994. Sample path derivatives for $(s, S)$ inventory systems. *Operations Research* **42**(2) 351–364.

Fu, M., X. Li, L. Zhang. 2021. Data-driven feature-based newsvendor: A distributionally robust approach. Available at SSRN #3885663.

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

Glaisher, J. 1871. XXXII. On a class of definite integrals. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **42**(280) 294–302.

Gorissen, B., D. den Hertog. 2013. Robust counterparts of inequalities containing sums of maxima of linear functions. *European Journal of Operational Research* **227**(1) 30–43.

Hadjiyiannis, M., P. Goulart, D. Kuhn. 2011. A scenario approach for estimating the suboptimality of linear decision rules in two-stage robust optimization. *In Proceedings of the* $50^{th}$ *IEEE Conference on Decision and Control and European Control Conference*. Florida, United States.

Halman, N., D. Klabjan, M. Mostagir, J. Orlin, D. Simchi-Levi. 2009. A fully polynomial-time approximation scheme for single-item stochastic inventory control with discrete demand. *Mathematics of Operations Research* **34**(3) 674–685.

Jalelah, B. 2021. Singapore made advance purchases for COVID-19 vaccines, including Sinovac. `https://www.channelnewsasia.com/singapore/covid-19-vaccines-singapore-sinovac-advance-purchases-277091`. Accessed on $24^{st}$ February 2022.

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

Löfberg, J. 2004. YALMIP : A toolbox for modeling and optimization in MATLAB. *In Proceedings of the CACSD Conference*. Taipei, Taiwan.

Lu, H., B. Sturt. 2022. On the sparsity of optimal linear decision rules in robust inventory management. Available on arXiv #2203.10661.

Mak, H., Y. Rong, J. Zhang. 2015. Appointment scheduling with limited distributional information. *Management Science* **61**(2) 316–334.

Mamani, H., S. Nassiri, M. Wagner. 2017. Closed-form solutions for robust inventory management. *Management Science* **63**(5) 1625–1643.

MOSEK ApS. 2019. *The MOSEK optimization toolbox for MATLAB manual. Version 9.3*. URL `http://docs.mosek.com/9.3/toolbox/index.html`.

Özer, Ö., W. Wei. 2006. Strategic commitments for an optimal capacity decision under asymmetric forecast information. *Management science* **52**(8) 1238–1257.

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

Polyak, B. 1987. *Introduction to Optimization*. Translations series in mathematics and engineering, Optimization Software, Publications Division.

Postek, K., A. Ben-Tal, D. den Hertog, B. Melenberg. 2018. Robust optimization with ambiguous stochastic constraints under mean and dispersion information. *Operations Research* **66**(3) 814–833.

Price, N., R. Sachs, J. Sherkow, L. Ouellette. 2020. COVID-19 vaccine advance purchases explained. `https://blog.petrieflom.law.harvard.edu/2020/08/11/covid19-vaccine-advance-purchases-explained/`. Accessed on 29$^{\text{th}}$ Jan 2023.

Rodrigues, F., A. Agra, C. Requejo, E. Delage. 2021. Lagrangian duality for robust problems with decomposable functions: The case of a robust inventory problem. *INFORMS Journal on Computing* **33**(2) 685–705.

Scarf, H. 1958. A min-max solution of an inventory problem. *Studies in the mathematical theory of inventory and production* 201–209.

Scarf, H. 1960. The optimality of $(S, s)$ policies in the dynamic inventory problem. K. Arrow, S. Sarlin, P. Suppes, eds., *Mathematical Models in the Social Sciences*.

See, C., M. Sim. 2010. Robust approximation to multiperiod inventory management. *Operations Research* **58**(3) 583–594.

Shapiro, A. 2001. On duality theory of conic linear problems. M. Goberna, M López, eds., *Semi-Infinite Programming: Recent Advances*. Springer, 135–165.

Shapiro, A., D. Dentcheva, A. Ruszczyński. 2014. *Lectures on Stochastic Programming: Modeling and Theory*. Society for Industrial and Applied Mathematics.

Solyalı, O., J.-F. Cordeau, G. Laporte. 2016. The impact of modeling on robust inventory management under demand uncertainty. *Management Science* **62**(4) 1188–1201.

Sun, L., G. Lyu, Y. Yu, C. Teo. 2021. Cross-border e-commerce data set: Choosing the right fulfillment option. *Manufacturing & Service Operations Management* **23**(5) 1297–1313.

Tang, C., K. Rajaram, A. Alptekinoğlu, J. Ou. 2004. The benefits of advance booking discount programs: Model and analysis. *Management Science* **50**(4) 465–478.

Veinott, A., H. Wagner. 1965. Computing optimal $(s, S)$ inventory policies. *Management Science* **11**(5) 525–552.

Zhang, X., Z. Ye, W. Haskell. 2021. Asymptotic analysis for data-driven inventory policies. Available on arXiv #2008.08275.

Zheng, Y., A. Federgruen. 1991. Finding optimal $(s, S)$ policies is about as simple as evaluating a single policy. *Operations research* **39**(4) 654–665.

Zymler, S., D. Kuhn, B. Rustem. 2013. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming* **137**(1) 167–198.

# Appendix

## A. Comparison to Mak et al. (2015) and Padmanabhan et al. (2021)

Theorem 1 reveals that, for any fixed $y_0 \in \mathbb{R}$ and $\boldsymbol{x} \in \mathcal{X}$ at optimality,

$$\alpha = \max_{\boldsymbol{e} \in \mathcal{E}} \left\{ \max_{\boldsymbol{\xi}} \ y_0 \mathbf{1}^\top \boldsymbol{e} + \sum_{t=1}^T e_t \sum_{\tau=1}^t x_\tau - \sum_{t=1}^T \beta_t \xi_t - \sum_{t=1}^T e_t \sum_{\tau=1}^t \xi_\tau - \sum_{t=1}^T \gamma_t \xi_t^2 \right\}, \tag{11}$$

where the optimization problem over $\boldsymbol{e} \in \mathcal{E}$ on the right-hand side constitutes a convex maximization problem. Mak et al. (2015) and Padmanabhan et al. (2021) recently encountered a similar optimization problem when analyzing a robust appointment scheduling model, where they have

$$\alpha = \max_{\boldsymbol{e} \in \mathcal{E}} \left\{ \sum_{t=1}^T \max_{\xi_t} g_t(e_t, \xi_t, \beta_t, \gamma_t) \right\} = \max_{\boldsymbol{e} \in \mathrm{conv}(\mathcal{E})} \left\{ \sum_{t=1}^T \max_{\xi_t} g_t(e_t, \xi_t, \beta_t, \gamma_t) \right\} \tag{12}$$

for suitably defined functions $g_1, \ldots, g_T$ that are convex in $\boldsymbol{e}$. Subsequently, they express (12) as a mixed-integer binary program with a totally unimodular constraint matrix so that the binarity requirement can be lifted. We cannot however directly express (11) as an instance of (12) because of the term $\sum_{t=1}^T e_t \sum_{\tau=1}^t \xi_\tau = \sum_{t=1}^T \xi_t \sum_{\tau=t}^T e_\tau$. Although we can use a variable transformation $\eta_t = \sum_{\tau=t}^T e_\tau$, $t \in \{1, \ldots, T\}$, and $\eta_{T+1} = 0$ and replace each $e_t$ with $\eta_t - \eta_{t+1}$ so that

$$\alpha = \max_{\boldsymbol{\eta}} \left\{ \sum_{t=1}^T \max_{\xi_t} \bar{g}_t(\eta_t, \xi_t, \beta_t, \gamma_t) : -b \leqslant \eta_t - \eta_{t+1} \leqslant +h, \ \eta_{T+1} = 0 \right\}$$

for suitable functions $\bar{g}_1, \ldots, \bar{g}_T$ that are convex in $\boldsymbol{e}$, such a transformation appears to prevent us from following the steps in Mak et al. (2015) to obtain a polynomially-sized exact reformulation of Problem (5).

## B. Gradient of the total expected cost under normal demands

To supplement the experiment details outlined in Section 5.3, we consider the case where the demands $\{\xi_t\}_{t=1}^T$ are independent and normally distributed. For the ease of exposition, we introduce the cumulative demands $\zeta_t = \sum_{\tau=1}^t \xi_\tau$, $\forall t$. It follows that $\zeta_t$ is a normal random variable with the following density function

$$f_t(z) = \frac{1}{\sigma\sqrt{2t\pi}} \exp\left( -\frac{(z - t\mu)^2}{2t\sigma^2} \right).$$

Besides, we denote the cumulative distribution function of $\zeta_t$ by $F_t$. By its definition, the stochastic solution $\tilde{\boldsymbol{x}}^\star$ minimizes an objective $g(\boldsymbol{x})$ which equals

$$\sum_{t=1}^T \left\{ c x_t + h \int_{-\infty}^{y_0 + \sum_{\tau=1}^t x_\tau} \left( y_0 + \sum_{\tau=1}^t x_\tau - \zeta_t \right) f_t(\zeta_t) \ \mathrm{d}\zeta_t - b \int_{y_0 + \sum_{\tau=1}^t x_\tau}^{+\infty} \left( y_0 + \sum_{\tau=1}^t x_\tau - \zeta_t \right) f_t(\zeta_t) \ \mathrm{d}\zeta_t \right\}.$$

Note that, even though the above objective function $g$ is convex in $\boldsymbol{x}$, its explicit characterization involves the Gauss error function $\texttt{erf}(z) = \frac{2}{\sqrt{\pi}} \int_{-\infty}^{z} \exp(-u^2)\ \mathrm{d}u$; see *e.g.* Glaisher (1871). Despite this complication, however, the gradient of $g$ can be easily determined as

$$\begin{aligned}
\frac{\partial g}{\partial x_t} &= c + h \sum_{\tau=t}^{T} \frac{\partial}{\partial x_t} \int_{-\infty}^{y_0+x_1+\ldots+x_\tau} (y_0 + x_1 + \ldots + x_\tau - \zeta_\tau)\, f_\tau(\zeta_\tau)\ \mathrm{d}\zeta_\tau \\
&\quad - b \sum_{\tau=t}^{T} \frac{\partial}{\partial x_t} \int_{y_0+x_1+\ldots+x_\tau}^{+\infty} (y_0 + x_1 + \ldots + x_\tau - \zeta_\tau)\, f_\tau(\zeta_\tau)\ \mathrm{d}\zeta_\tau \\
&= c + h \sum_{\tau=t}^{T} \left\{ \frac{\partial}{\partial x_t} (y_0 + x_1 + \ldots + x_\tau)\, F_\tau(y_0 + x_1 + \ldots + x_\tau) - \frac{\partial}{\partial x_t} \int_{-\infty}^{y_0+x_1+\ldots+x_\tau} \zeta_\tau f_\tau(\zeta_\tau)\ \mathrm{d}\zeta_\tau \right\} \\
&\quad - b \sum_{\tau=t}^{T} \left\{ \frac{\partial}{\partial x_t} (y_0 + x_1 + \ldots + x_\tau)\, (1 - F_\tau(y_0 + x_1 + \ldots + x_\tau)) - \frac{\partial}{\partial x_t} \int_{y_0+x_1+\ldots+x_\tau}^{+\infty} \zeta_\tau f_\tau(\zeta_\tau)\ \mathrm{d}\zeta_\tau \right\} \\
&= c + h \sum_{\tau=t}^{T} F_\tau(y_0 + x_1 + \ldots + x_\tau) - b \sum_{\tau=t}^{T} (1 - F_\tau(y_0 + x_1 + \ldots + x_\tau)) \\
&= c - b(T - t + 1) + (h + b) \sum_{\tau=t}^{T} F_\tau(y_0 + x_1 + \ldots + x_\tau),
\end{aligned}$$

where the third equality holds because of the Fundamental Theorem of Calculus.

## C. Omitted proofs and auxiliary technical results

*Proof of Theorem 1.* Starting from its definition in (3), we may characterize the function $f$ as the optimal objective value of a generalized moment problem, *i.e.*,

$$\begin{aligned}
f(\boldsymbol{x}) - c\mathbf{1}^\top \boldsymbol{x} \;=\; \max \quad & \int_{\mathbb{R}^T} \max_{\boldsymbol{e}\in\mathcal{E}} \sum_{t=1}^{T} e_t \left( y_0 + \sum_{\tau=1}^{t} (x_\tau - \xi_\tau) \right) \mathbb{P}(\mathrm{d}\boldsymbol{\xi}) \\
\text{s.t.} \quad & \mathbb{P} \in \mathcal{M}_+(\mathbb{R}^T) \\
& \int_{\mathbb{R}^T} \mathbb{P}(\mathrm{d}\boldsymbol{\xi}) = 1 \\
& \int_{\mathbb{R}^T} \xi_t \mathbb{P}(\mathrm{d}\boldsymbol{\xi}) = \mu \qquad \forall t = 1, \ldots, T \\
& \int_{\mathbb{R}^T} \xi_t^2 \mathbb{P}(\mathrm{d}\boldsymbol{\xi}) = \mu^2 + \sigma^2 \qquad \forall t = 1, \ldots, T.
\end{aligned} \tag{13}$$

By assigning the dual variables $\alpha$ to the normalization constraint, each $\beta_t$ to each of the first-order moment constraints, and each $\gamma_t$ to each of the second-order moment constraints and invoking strong duality due to Shapiro (2001, Proposition 3.4), we find a minimization problem that attains the same optimal objective value as the above maximization problem:

$$\begin{aligned}
\min \quad & \alpha + \mu \mathbf{1}^\top \boldsymbol{\beta} + (\mu^2 + \sigma^2)\mathbf{1}^\top \boldsymbol{\gamma} \\
\text{s.t.} \quad & \alpha \in \mathbb{R},\ \boldsymbol{\beta} \in \mathbb{R}^T,\ \boldsymbol{\gamma} \in \mathbb{R}^T \\
& \alpha + \sum_{t=1}^{T} \beta_t \xi_t + \sum_{t=1}^{T} \gamma_t \xi_t^2 \;\geqslant\; \sum_{t=1}^{T} e_t \left( y_0 + \sum_{\tau=1}^{t} (x_\tau - \xi_\tau) \right) \quad \forall \boldsymbol{e} \in \mathcal{E}\ \forall \boldsymbol{\xi} \in \mathbb{R}^T.
\end{aligned}$$

Note that the inequality constraint in the above minimization problem could be rearranged as

$$\begin{aligned}
& \left[ \alpha - y_0 \mathbf{1}^\top \boldsymbol{e} - \sum_{t=1}^{T} e_t \sum_{\tau=1}^{t} x_\tau \right] + \left[ \sum_{t=1}^{T} \beta_t \xi_t + \sum_{t=1}^{T} e_t \sum_{\tau=1}^{t} \xi_\tau \right] + \sum_{t=1}^{T} \gamma_t \xi_t^2 \;\geqslant\; 0 \quad \forall \boldsymbol{e} \in \mathcal{E}\ \forall \boldsymbol{\xi} \in \mathbb{R}^T \\
\iff \quad & \alpha(\boldsymbol{x}, \boldsymbol{e}) + \sum_{t=1}^{T} \beta_t(\boldsymbol{e}) \xi_t + \sum_{t=1}^{T} \gamma_t \xi_t^2 \;\geqslant\; 0 \quad \forall \boldsymbol{e} \in \mathcal{E}\ \forall \boldsymbol{\xi} \in \mathbb{R}^T.
\end{aligned}$$

For any fixed $\boldsymbol{e} \in \mathcal{E}$, as the above quadratic inequality holds for every $\boldsymbol{\xi} \in \mathbb{R}^T$, it follows that $\gamma_t$ must be non-negative. Furthermore, if $\gamma_t$ vanishes, then so does $\beta_t(\boldsymbol{e})$. Denoting by $\mathcal{T}$ the index set $\{t \in \{1,\ldots,T\} : \gamma_t > 0\}$, we can re-express the considered quadratic inequality as

$$\begin{aligned} & \alpha(\boldsymbol{x},\boldsymbol{e}) + \sum_{t\in\mathcal{T}} \beta_t(\boldsymbol{e})\xi_t + \sum_{t\in\mathcal{T}} \gamma_t \xi_t^2 \ \geqslant \ 0 \quad \forall \boldsymbol{\xi} \in \mathbb{R}^T \\ \iff & \alpha(\boldsymbol{x},\boldsymbol{e}) + \sum_{t\in\mathcal{T}} \gamma_t \left[ \left( \xi_t + \frac{\beta_t(\boldsymbol{e})}{2\gamma_t} \right)^2 - \frac{\beta_t(\boldsymbol{e})^2}{4\gamma_t^2} \right] \geqslant 0 \quad \forall \boldsymbol{\xi} \in \mathbb{R}^T \\ \iff & 4\alpha(\boldsymbol{x},\boldsymbol{e}) \geqslant \sum_{t\in\mathcal{T}} \frac{\beta_t(\boldsymbol{e})^2}{\gamma_t} = \sum_{t=1}^{T} \frac{\beta_t(\boldsymbol{e})^2}{\gamma_t}, \end{aligned}$$

and the proof is completed by recalling the definition of the cone $\mathcal{K}_T$. □

*Proof of Theorem 2.* Leveraging the exact characterization of the worst-case expected cost $f(\boldsymbol{x})$ from Theorem 1, we assign a set of dual variables $\left(\overline{\alpha}(\boldsymbol{e}), \overline{\boldsymbol{\beta}}(\boldsymbol{e}), \overline{\boldsymbol{\gamma}}(\boldsymbol{e})\right) \in \mathcal{K}_T^\star$ to the conic constraint in Problem (5) that is associated with each $\boldsymbol{e} \in \mathcal{E}$:

$$\begin{aligned} & (\alpha(\boldsymbol{x},\boldsymbol{e}), \boldsymbol{\beta}(\boldsymbol{e}), \boldsymbol{\gamma}) \succeq_{\mathcal{K}_T} \mathbf{0} \\ \iff & \left( \alpha - y_0 \mathbf{1}^\top \boldsymbol{e} - \sum_{t=1}^{T} x_t \sum_{\tau=t}^{T} e_\tau, \beta_1 + \sum_{\tau=1}^{T} e_\tau, \beta_2 + \sum_{\tau=2}^{T} e_\tau, \ldots, \beta_T + e_T, \gamma_1, \gamma_2, \ldots, \gamma_T \right) \succeq_{\mathcal{K}_T} \mathbf{0}, \end{aligned}$$

and obtain the following dual formulation of Problem (5):

$$\begin{aligned} f(\boldsymbol{x}) = \text{maximize} \quad & c\mathbf{1}^\top \boldsymbol{x} + \sum_{\boldsymbol{e}\in\mathcal{E}} \overline{\alpha}(\boldsymbol{e}) \left( y_0 \mathbf{1}^\top \boldsymbol{e} + \sum_{t=1}^{T} x_t \sum_{\tau=t}^{T} e_\tau \right) - \sum_{\boldsymbol{e}\in\mathcal{E}} \sum_{t=1}^{T} \overline{\beta}_t(\boldsymbol{e}) \sum_{\tau=t}^{T} e_\tau \\ \text{subject to} \quad & \overline{\alpha} : \mathcal{E} \mapsto \mathbb{R}, \ \overline{\boldsymbol{\beta}} : \mathcal{E} \mapsto \mathbb{R}^T, \ \overline{\boldsymbol{\gamma}} : \mathcal{E} \mapsto \mathbb{R}^T \\ & \left(\overline{\alpha}(\boldsymbol{e}), \overline{\boldsymbol{\beta}}(\boldsymbol{e}), \overline{\boldsymbol{\gamma}}(\boldsymbol{e})\right) \succeq_{\mathcal{K}_T^\star} \mathbf{0} \quad \forall \boldsymbol{e} \in \mathcal{E} \\ & \sum_{\boldsymbol{e}\in\mathcal{E}} \overline{\alpha}(\boldsymbol{e}) = 1 \\ & \sum_{\boldsymbol{e}\in\mathcal{E}} \overline{\boldsymbol{\beta}}(\boldsymbol{e}) = \mu \mathbf{1} \\ & \sum_{\boldsymbol{e}\in\mathcal{E}} \overline{\boldsymbol{\gamma}}(\boldsymbol{e}) = (\mu^2 + \sigma^2)\mathbf{1}. \end{aligned}$$

By noting that the objective function of the above optimization problem, when evaluated at any feasible $(\overline{\alpha}, \overline{\boldsymbol{\beta}}, \overline{\boldsymbol{\gamma}})$, could be re-expressed as

$$\begin{aligned} & c\mathbf{1}^\top \boldsymbol{x} + \sum_{\substack{\boldsymbol{e}\in\mathcal{E}: \\ \overline{\alpha}(\boldsymbol{e})\neq 0}} \overline{\alpha}(\boldsymbol{e}) \left( y_0 \mathbf{1}^\top \boldsymbol{e} + \sum_{t=1}^{T} \left( x_t - \frac{\overline{\beta}_t(\boldsymbol{e})}{\overline{\alpha}(\boldsymbol{e})} \right) \sum_{\tau=t}^{T} e_\tau \right) - \sum_{\substack{\boldsymbol{e}\in\mathcal{E}: \\ \overline{\alpha}(\boldsymbol{e}) = 0}} \sum_{t=1}^{T} \overline{\beta}_t(\boldsymbol{e}) \sum_{\tau=t}^{T} e_\tau \\ = & c\mathbf{1}^\top \boldsymbol{x} + \sum_{\substack{\boldsymbol{e}\in\mathcal{E}: \\ \overline{\alpha}(\boldsymbol{e})\neq 0}} \overline{\alpha}(\boldsymbol{e}) \left( y_0 \mathbf{1}^\top \boldsymbol{e} + \sum_{t=1}^{T} e_t \sum_{\tau=1}^{t} \left( x_\tau - \frac{\overline{\beta}_\tau(\boldsymbol{e})}{\overline{\alpha}(\boldsymbol{e})} \right) \right) \\ = & c\mathbf{1}^\top \boldsymbol{x} + \sum_{\substack{\boldsymbol{e}\in\mathcal{E}: \\ \overline{\alpha}(\boldsymbol{e})\neq 0}} \overline{\alpha}(\boldsymbol{e}) \left( \sum_{t=1}^{T} e_t \left( y_0 + \sum_{\tau=1}^{t} \left( x_\tau - \frac{\overline{\beta}_\tau(\boldsymbol{e})}{\overline{\alpha}(\boldsymbol{e})} \right) \right) \right), \end{aligned}$$

where the first equality holds because $\overline{\boldsymbol{\beta}}(\boldsymbol{e}) = \mathbf{0}$ whenever $\overline{\alpha}(\boldsymbol{e}) = 0$ (see Proposition 1), the proof is completed. □

The characterization of Problem (6) involves the cone $\mathcal{K}_T^\star$ that is dual to $\mathcal{K}_T$. In order to solve this problem, we need the exact description of the dual cone, which is given next.

PROPOSITION 1. *We have that*

$$\mathcal{K}_t^\star = \left\{ (\overline{\alpha}, \overline{\boldsymbol{\beta}}, \overline{\boldsymbol{\gamma}}) \in \mathbb{R}_+ \times \mathbb{R}^t \times \mathbb{R}_+^t : \overline{\alpha} \geqslant \max_{\tau \in \{1,\ldots,t\}} \frac{\overline{\beta}_\tau^2}{\overline{\gamma}_\tau} \right\}. \tag{14}$$

*Proof of Proposition 1.* We denote the set given on the right-hand side of (14) by $\mathcal{K}_t'$. So as to show that $\mathcal{K}_t^\star = \mathcal{K}_t'$, it suffices to prove that $\mathcal{K}_t' \subseteq \mathcal{K}_t^\star$ (step 1) and $\mathcal{K}_t^\star \subseteq \mathcal{K}_t'$ (step 2).

For the first part, we consider an arbitrary $(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}) \in \mathcal{K}_t$ and $(\overline{\alpha}, \overline{\boldsymbol{\beta}}, \overline{\boldsymbol{\gamma}}) \in \mathcal{K}_t'$. It follows that

$$\alpha\overline{\alpha} + \boldsymbol{\gamma}^\top \overline{\boldsymbol{\gamma}} \;\geqslant\; \frac{1}{4}\overline{\alpha} \sum_{\tau=1}^t \frac{\beta_\tau^2}{\gamma_\tau} + \sum_{\tau=1}^t \gamma_\tau \overline{\gamma}_\tau \;\geqslant\; \sum_{\tau=1}^t \left( \frac{\beta_\tau^2 \overline{\beta}_\tau^2}{4\gamma_\tau \overline{\gamma}_\tau} + \gamma_\tau \overline{\gamma}_\tau \right) \;\geqslant\; -\sum_{\tau=1}^t \beta_\tau \overline{\beta}_\tau \;=\; -\boldsymbol{\beta}^\top \overline{\boldsymbol{\beta}},$$

where the last inequality holds because $\frac{\beta_\tau^2 \overline{\beta}_\tau^2}{4\gamma_\tau \overline{\gamma}_\tau} + \gamma_\tau \overline{\gamma}_\tau \geqslant |\beta_\tau \overline{\beta}_\tau|$, $\forall \tau$. Indeed, due to the relationship between arithmetic and geometric means this latter inequality holds whenever $\gamma_\tau$ and $\overline{\gamma}_\tau$ are both strictly positive. If $\gamma_\tau = 0$ then $\beta_\tau = 0$, and if $\overline{\gamma}_\tau = 0$ then $\overline{\beta}_\gamma = 0$. In both cases, the same inequality is still valid. The above derivation implies that $\mathcal{K}_t^\star$ contains $\mathcal{K}_t'$ inside. Conversely, consider an arbitrary $(\overline{\alpha}, \overline{\boldsymbol{\beta}}, \overline{\boldsymbol{\gamma}}) \in \mathcal{K}_t^\star$. It holds that

$$\alpha\overline{\alpha} + \boldsymbol{\beta}^\top \overline{\boldsymbol{\beta}} + \boldsymbol{\gamma}^\top \overline{\boldsymbol{\gamma}} \geqslant 0 \quad \forall (\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}) \in \mathcal{K}_t.$$

As $(1, \mathbf{0}, \mathbf{0}) \in \mathcal{K}_t$, it necessarily follows from the above inequality that $\overline{\alpha} \geqslant 0$. Similarly as $(0, \mathbf{0}, \mathbf{1}_\tau) \in \mathcal{K}_t$, $\forall \tau \in \{1, \ldots, t\}$, it necessarily follows that $\overline{\gamma}_\tau \geqslant 0$. Moreover, for any fixed $\tau \in \{1, \ldots, t\}$, if $\overline{\beta}_\tau \neq 0$, then $\overline{\gamma}_\tau > 0$. Suppose otherwise for the sake of a contradiction that there exists $(\overline{\alpha}, \overline{\boldsymbol{\beta}}, \overline{\boldsymbol{\gamma}}) \in \mathcal{K}_t^\star$ such that $\overline{\beta}_\tau \neq 0$ and $\overline{\gamma}_\tau = 0$, then as

$$\left( \alpha, -\overline{\beta}_\tau \mathbf{1}_\tau, \frac{\overline{\beta}_\tau^2}{4\alpha} \mathbf{1}_\tau \right) \in \mathcal{K}_t \quad \forall \alpha > 0,$$

it must follow that $\alpha\overline{\alpha} - \overline{\beta}_\tau^2 \geqslant 0$ for any $\alpha > 0$. This resultant inequality however cannot hold true because $\alpha$ could be arbitrarily close to zero. Assuming now that $\overline{\beta}_\tau \neq 0$ and $\overline{\gamma}_\tau > 0$, we hence find

$$\left( \frac{\overline{\gamma}_\tau^2}{\overline{\beta}_\tau^2}, -\frac{2\overline{\gamma}_\tau}{\overline{\beta}_\tau} \mathbf{1}_\tau, \mathbf{1}_\tau \right) \in \mathcal{K}_t,$$

and consequently,

$$\overline{\alpha} \frac{\overline{\gamma}_\tau^2}{\overline{\beta}_\tau^2} - \overline{\beta}_\tau \frac{2\overline{\gamma}_\tau}{\overline{\beta}_\tau} + \overline{\gamma}_\tau \geqslant 0 \quad \Longrightarrow \quad \overline{\alpha} \geqslant \frac{\overline{\beta}_\tau^2}{\overline{\gamma}_\tau}.$$

As $\overline{\alpha} \geqslant 0$, the same conclusion can also be reached even if $\overline{\beta}_\tau = 0$. Since the above inequality holds for any $\tau \in \{1, \ldots, t\}$, we may conclude that $\mathcal{K}_t^\star$ is contained in $\mathcal{K}_t'$. Combining both halves of the argument yields the desired result. $\square$

The proof of Theorem 3 relies heavily on the following two technical lemmas.

LEMMA 1. *There exists an optimal solution* $(\overline{\alpha}^\star, \overline{\boldsymbol{\beta}}^\star, \overline{\boldsymbol{\gamma}}^\star)$ *of Problem* (6) *such that* $\overline{\boldsymbol{\beta}}^\star(\boldsymbol{e}) = \overline{\boldsymbol{\gamma}}^\star(\boldsymbol{e}) = \mathbf{0}$, $\forall \boldsymbol{e} \in \mathcal{E} : \overline{\alpha}^\star(\boldsymbol{e}) = 0$.

*Proof of Lemma 1.* Denote by $(\overline{\alpha}^\dagger, \overline{\boldsymbol{\beta}}^\dagger, \overline{\boldsymbol{\gamma}}^\dagger)$ an arbitrary optimal solution of Problem (6) and by $\mathcal{E}^0$ a subset of $\mathcal{E}$ which contains all scenarios $\boldsymbol{e}$ such that $\overline{\alpha}^\dagger(\boldsymbol{e}) = 0$. Note that $\mathcal{E}^0$ must be strictly contained in $\mathcal{E}$ and there must exist a scenario $\boldsymbol{e}^\dagger \in \mathcal{E} \backslash \mathcal{E}^0$ such that $\overline{\alpha}^\dagger(\boldsymbol{e}^\dagger)$ is strictly positive. Then, we construct a new solution $(\overline{\alpha}^\star, \overline{\boldsymbol{\beta}}^\star, \overline{\boldsymbol{\gamma}}^\star)$ with $\overline{\alpha}^\star = \overline{\alpha}^\dagger$, $\overline{\boldsymbol{\beta}}^\star = \overline{\boldsymbol{\beta}}^\dagger$, and

$$\overline{\boldsymbol{\gamma}}^\star(\boldsymbol{e}) = \begin{cases} \overline{\boldsymbol{\gamma}}^\dagger(\boldsymbol{e}) + \sum_{\boldsymbol{e}' \in \mathcal{E}^0} \overline{\boldsymbol{\gamma}}^\dagger(\boldsymbol{e}') & \text{if } \boldsymbol{e} = \boldsymbol{e}^\dagger, \\ \mathbf{0} & \text{if } \boldsymbol{e} \in \mathcal{E}^0, \\ \overline{\boldsymbol{\gamma}}^\dagger(\boldsymbol{e}) & \text{if } \boldsymbol{e} \in \mathcal{E} \backslash (\mathcal{E}^0 \cup \{\boldsymbol{e}^\dagger\}). \end{cases}$$

It is readily seen that $\sum_{e\in\mathcal{E}}\overline{\boldsymbol{\gamma}}^\star(\boldsymbol{e})=\sum_{e\in\mathcal{E}}\overline{\boldsymbol{\gamma}}^\dagger(\boldsymbol{e})=(\mu^2+\sigma^2)\mathbf{1}$, $(\overline{\alpha}^\star(\boldsymbol{e}^\dagger),\overline{\boldsymbol{\beta}}^\star(\boldsymbol{e}^\dagger),\overline{\boldsymbol{\gamma}}^\star(\boldsymbol{e}^\dagger))\in\mathcal{K}_T^\star$ and that all other constraints remain satisfied by this newly constructed solution. Since $\overline{\boldsymbol{\gamma}}$ does not appear in the objective function of Problem (6), we may thus conclude that $(\overline{\alpha}^\star,\overline{\boldsymbol{\beta}}^\star,\overline{\boldsymbol{\gamma}}^\star)$ is feasible and optimal. The proof is then completed by noting that $\overline{\boldsymbol{\beta}}^\star(\boldsymbol{e})$ automatically vanishes whenever $\overline{\alpha}^\star(\boldsymbol{e})$ does. □

LEMMA 2. *Any optimal solution* $(\overline{\alpha}^\star,\overline{\boldsymbol{\beta}}^\star,\overline{\boldsymbol{\gamma}}^\star)$ *of Problem* (6) *satisfies*

$$e_t\left(y_0+\sum_{\tau=1}^{t}\left(x_\tau-\frac{\overline{\beta}_\tau^\star(\boldsymbol{e})}{\overline{\alpha}^\star(\boldsymbol{e})}\right)\right)=\max\left\{h\left(y_0+\sum_{\tau=1}^{t}\left(x_\tau-\frac{\overline{\beta}_\tau^\star(\boldsymbol{e})}{\overline{\alpha}^\star(\boldsymbol{e})}\right)\right),-b\left(y_0+\sum_{\tau=1}^{t}\left(x_\tau-\frac{\overline{\beta}_\tau^\star(\boldsymbol{e})}{\overline{\alpha}^\star(\boldsymbol{e})}\right)\right)\right\}$$

*for any* $t\in\{1,\dots,T\}$ *and any* $\boldsymbol{e}\in\mathcal{E}$ *such that* $\overline{\alpha}^\star(\boldsymbol{e})>0$.

*Proof of Lemma 2.* Consider any scenario $\boldsymbol{e}\in\mathcal{E}$ such that $\overline{\alpha}^\star(\boldsymbol{e})>0$. To simplify the exposition, we abbreviate (and slightly abuse the notation) the dual optimal solution $\overline{\alpha}^\star(\boldsymbol{e}),\overline{\boldsymbol{\beta}}^\star(\boldsymbol{e}),\overline{\gamma}^\star(\boldsymbol{e})$ by $\overline{\alpha}^\star$, $\overline{\beta}^\star$, $\overline{\gamma}^\star$, respectively.[1] Besides, we denote by $(\alpha^\star,\boldsymbol{\beta}^\star,\boldsymbol{\gamma}^\star)$ an optimal solution of the primal problem (5). The KKT optimality condition of this primal-dual pair includes a complementary slackness condition which necessarily holds and reads

$$\begin{aligned}
&\overline{\alpha}^\star\left(\alpha^\star-y_0\mathbf{1}^\top\boldsymbol{e}-\sum_{t=1}^{T}x_t\sum_{\tau=t}^{T}e_\tau\right)+\sum_{t=1}^{T}\overline{\beta}_t^\star\left(\beta_t^\star+\sum_{\tau=t}^{T}e_\tau\right)+\sum_{t=1}^{T}\overline{\gamma}_t^\star\gamma_t^\star=0\\
\iff\quad&\overline{\alpha}^\star\left(y_0\mathbf{1}^\top\boldsymbol{e}+\sum_{t=1}^{T}x_t\sum_{\tau=t}^{T}e_\tau\right)-\sum_{t=1}^{T}\overline{\beta}_t^\star\sum_{\tau=t}^{T}e_\tau=\alpha^\star\overline{\alpha}^\star+(\boldsymbol{\beta}^\star)^\top\overline{\boldsymbol{\beta}}^\star+(\boldsymbol{\gamma}^\star)^\top\overline{\boldsymbol{\gamma}}^\star\\
\iff\quad&\overline{\alpha}^\star\sum_{t=1}^{T}e_t\left(y_0+\sum_{\tau=1}^{t}\left(x_\tau-\frac{\overline{\beta}_\tau^\star}{\overline{\alpha}^\star}\right)\right)=\alpha^\star\overline{\alpha}^\star+(\boldsymbol{\beta}^\star)^\top\overline{\boldsymbol{\beta}}^\star+(\boldsymbol{\gamma}^\star)^\top\overline{\boldsymbol{\gamma}}^\star.
\end{aligned}$$

For any time period $t'\in\{1,\dots,T\}$, it thus follows that

$$\begin{aligned}
&\overline{\alpha}^\star e_{t'}\left(y_0+\sum_{\tau=1}^{t'}\left(x_\tau-\frac{\overline{\beta}_\tau^\star}{\overline{\alpha}^\star}\right)\right)\\
&=\alpha^\star\overline{\alpha}^\star+(\boldsymbol{\beta}^\star)^\top\overline{\boldsymbol{\beta}}^\star+(\boldsymbol{\gamma}^\star)^\top\overline{\boldsymbol{\gamma}}^\star-\overline{\alpha}^\star\sum_{t\neq t'}e_t\left(y_0+\sum_{\tau=1}^{t}\left(x_\tau-\frac{\overline{\beta}_\tau^\star}{\overline{\alpha}^\star}\right)\right)\\
&=\overline{\alpha}^\star\left(\alpha^\star-y_0\sum_{t\neq t'}e_t-\sum_{t=1}^{T}x_t\sum_{\substack{\tau\geqslant t\\ \tau\neq t'}}e_\tau\right)+\sum_{t=1}^{T}\overline{\beta}_t^\star\left(\beta_t^\star+\sum_{\substack{\tau\geqslant t\\ \tau\neq t'}}e_\tau\right)+\sum_{t=1}^{T}\overline{\gamma}_t^\star\gamma_t^\star.
\end{aligned}\tag{15}$$

Next, we collect the coefficients of the dual solution on the right-hand side of the above equation

$$\boldsymbol{\zeta}=\left(\alpha^\star-y_0\sum_{t\neq t'}e_t-\sum_{t=1}^{T}x_t\sum_{\substack{\tau\geqslant t\\ \tau\neq t'}}e_\tau,\beta_1^\star+\sum_{\substack{\tau\geqslant 1\\ \tau\neq t'}}e_\tau,\dots,\beta_T^\star+\sum_{\substack{\tau\geqslant T\\ \tau\neq t'}}e_\tau,\gamma_1^\star,\dots,\gamma_T^\star\right)^\top$$

and observe that $\boldsymbol{\zeta}$ can be expressed as a convex combination of the following two vectors

$$\begin{aligned}
\overline{\boldsymbol{\zeta}}&=\left(\alpha^\star-y_0\mathbf{1}^\top\overline{\boldsymbol{e}}-\sum_{t=1}^{T}x_t\sum_{\tau=t}^{T}\overline{e}_\tau,\beta_1^\star+\sum_{\tau=1}^{T}\overline{e}_\tau,\beta_2^\star+\sum_{\tau=2}^{T}\overline{e}_\tau,\dots,\beta_T^\star+\overline{e}_T,\gamma_1^\star,\gamma_2^\star,\dots,\gamma_T^\star\right)^\top\\
\underline{\boldsymbol{\zeta}}&=\left(\alpha^\star-y_0\mathbf{1}^\top\underline{\boldsymbol{e}}-\sum_{t=1}^{T}x_t\sum_{\tau=t}^{T}\underline{e}_\tau,\beta_1^\star+\sum_{\tau=1}^{T}\underline{e}_\tau,\beta_2^\star+\sum_{\tau=2}^{T}\underline{e}_\tau,\dots,\beta_T^\star+\underline{e}_T,\gamma_1^\star,\gamma_2^\star,\dots,\gamma_T^\star\right)^\top,
\end{aligned}$$

[1] That is we drop their dependency on $\boldsymbol{e}$.

where the scenarios $\overline{\boldsymbol{e}}$ and $\underline{\boldsymbol{e}}$ are characterized by

$$\overline{e}_t = \underline{e}_t = e_t \quad \forall t \in \{1, \ldots, T\} \backslash \{t'\} \quad \text{and} \quad \overline{e}_{t'} = +h \quad \text{as well as} \quad \underline{e}_{t'} = -b.$$

In particular, we can express $\boldsymbol{\zeta}$ as $\frac{b}{h+b}\overline{\boldsymbol{\zeta}} + \frac{h}{h+b}\underline{\boldsymbol{\zeta}}$. Note that the feasibility of $(\alpha^\star, \boldsymbol{\beta}^\star, \boldsymbol{\gamma}^\star)$ in view of Problem (5) implies that $\overline{\boldsymbol{\zeta}}, \underline{\boldsymbol{\zeta}} \in \mathcal{K}_T$. As $\mathcal{K}_T$ constitutes a convex cone, we have that $\boldsymbol{\zeta} \in \mathcal{K}_T$ and

$$e_{t'} \left( y_0 + \sum_{\tau=1}^{t'} \left( x_\tau - \frac{\overline{\beta}_\tau^\star}{\overline{\alpha}^\star} \right) \right) \geqslant 0$$

because $(\overline{\alpha}^\star, \overline{\boldsymbol{\beta}}^\star, \overline{\boldsymbol{\gamma}}^\star) \in \mathcal{K}_T^\star$ and because of (15). The above inequality and the fact that $e_{t'} \in \{+h, -b\}$, together, finally completes the proof. □

We are now ready to utilize the Lemma 1 and Lemma 2 to establish the proof of Theorem 3.

*Proof of Theorem 3.* To establish that the mixture distribution $\mathbb{P}^\varepsilon$ belongs to the ambiguity set $\mathcal{P}$, we first compute from (7) the first two moments of each probability distribution component

$$\mathbb{E}_{\mathbb{P}^{\boldsymbol{e},\varepsilon}}\left[\xi_t\right] = \frac{\overline{\beta}_t^\star(\boldsymbol{e})}{\overline{\alpha}^\star(\boldsymbol{e})} \quad \text{and} \quad \mathbb{E}_{\mathbb{P}^{\boldsymbol{e},\varepsilon}}\left[\xi_t^2\right] = \frac{\overline{\gamma}_t^\star(\boldsymbol{e})}{\overline{\alpha}^\star(\boldsymbol{e})} \quad \forall t \in \{1, \ldots, T\}.$$

Subsequently, from (8) we obtain

$$\mathbb{E}_{\mathbb{P}^\varepsilon}\left[\xi_t\right] = \sum_{\substack{\boldsymbol{e} \in \mathcal{E}: \\ \overline{\alpha}^\star(\boldsymbol{e}) > 0}} \overline{\beta}_t^\star(\boldsymbol{e}) = \sum_{\boldsymbol{e} \in \mathcal{E}} \overline{\beta}_t^\star(\boldsymbol{e}) = \mu \quad \text{and} \quad \mathbb{E}_{\mathbb{P}^\varepsilon}\left[\xi_t^2\right] = \sum_{\substack{\boldsymbol{e} \in \mathcal{E}: \\ \overline{\alpha}^\star(\boldsymbol{e}) > 0}} \overline{\gamma}_t^\star(\boldsymbol{e}) = \sum_{\boldsymbol{e} \in \mathcal{E}} \overline{\gamma}_t^\star(\boldsymbol{e}) = \mu^2 + \sigma^2$$

for any $t \in \{1, \ldots, T\}$ because of Lemma 1, and we can conclude that $\mathbb{P}^\varepsilon \in \mathcal{P}$.

By construction, it similarly follows that

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}^{\boldsymbol{e},\varepsilon}}\left[ \sum_{t=1}^T \max\left\{ h\left( y_0 + \sum_{\tau=1}^t (x_\tau - \xi_\tau) \right), -b\left( y_0 + \sum_{\tau=1}^t (x_\tau - \xi_\tau) \right) \right\} \right] \\
&= (1-\varepsilon) \sum_{t=1}^T \max\left\{ h\left( y_0 + \sum_{\tau=1}^t (x_\tau - \underline{\xi}_\tau^{\boldsymbol{e},\varepsilon}) \right), -b\left( y_0 + \sum_{\tau=1}^t (x_\tau - \underline{\xi}_\tau^{\boldsymbol{e},\varepsilon}) \right) \right\} + \\
&\quad \varepsilon \sum_{t=1}^T \max\left\{ h\left( y_0 + \sum_{\tau=1}^t (x_\tau - \overline{\xi}_\tau^{\boldsymbol{e},\varepsilon}) \right), -b\left( y_0 + \sum_{\tau=1}^t (x_\tau - \overline{\xi}_\tau^{\boldsymbol{e},\varepsilon}) \right) \right\}.
\end{aligned}$$

Since $\lim_{\varepsilon \downarrow 0} \varepsilon \overline{\xi}_\tau^{\boldsymbol{e},\varepsilon} = 0$ for all $\tau \in \{1, \ldots, T\}$, we further find

$$\begin{aligned}
&\lim_{\varepsilon \downarrow 0} \mathbb{E}_{\mathbb{P}^{\boldsymbol{e},\varepsilon}}\left[ \sum_{t=1}^T \max\left\{ h\left( y_0 + \sum_{\tau=1}^t (x_\tau - \xi_\tau) \right), -b\left( y_0 + \sum_{\tau=1}^t (x_\tau - \xi_\tau) \right) \right\} \right] \\
&= \lim_{\varepsilon \downarrow 0} \sum_{t=1}^T \max\left\{ h\left( y_0 + \sum_{\tau=1}^t (x_\tau - \underline{\xi}_\tau^{\boldsymbol{e},\varepsilon}) \right), -b\left( y_0 + \sum_{\tau=1}^t (x_\tau - \underline{\xi}_\tau^{\boldsymbol{e},\varepsilon}) \right) \right\} \\
&= \sum_{t=1}^T \max\left\{ h\left( y_0 + \sum_{\tau=1}^t \left( x_\tau - \frac{\overline{\beta}_\tau^\star(\boldsymbol{e})}{\overline{\alpha}^\star(\boldsymbol{e})} \right) \right), -b\left( y_0 + \sum_{\tau=1}^t \left( x_\tau - \frac{\overline{\beta}_\tau^\star(\boldsymbol{e})}{\overline{\alpha}^\star(\boldsymbol{e})} \right) \right) \right\} \\
&= \sum_{t=1}^T e_t \left( y_0 + \sum_{\tau=1}^t \left( x_\tau - \frac{\overline{\beta}_\tau^\star(\boldsymbol{e})}{\overline{\alpha}^\star(\boldsymbol{e})} \right) \right),
\end{aligned}$$

where the last equality is due to Lemma 2. Utilizing the fact that $\mathbb{P}^\varepsilon$ is a mixture of $\mathbb{P}^{\boldsymbol{e},\varepsilon}$ and invoking Theorem 2 complete the proof. □

Similarly, the proof of Theorem 7 builds on two technical lemmas, which we present below.

Lemma 3. *Given $a \in \mathbb{R}_+$ and $b, c \in \mathbb{R}$, it holds that*

$$\min_{z \in \mathbb{R}} \left\{ az^2 + bz + c \right\} \; = \; c - \frac{b^2}{4a}.$$

*Proof of Lemma 3.* When $a = 0$ and $b \neq 0$, the minimization problem is unbounded. When $a = 0$ but $b = 0$, the optimal objective value trivially evaluates to $c$ which coincides with the right-hand side because of our convention that $0/0 = 0$. Finally, when $a > 0$, it is a routine exercise to verify that the minimizer $z^\star$ is $-\frac{b}{2a}$ and hence the statement follows again. □

Lemma 4. *For $\boldsymbol{a}, \boldsymbol{q} \in \mathbb{R}^n$ such that $\mathbf{1}^\top \boldsymbol{q} = 1$, it holds that*

$$\left(\sum_{i=1}^n q_i a_i^2\right) - \left(\sum_{i=1}^n q_i a_i\right)^2 \; = \sum_{1 \leqslant i < j \leqslant n} q_i q_j (a_i - a_j)^2.$$

*Proof of Lemma 4.* The statement can be verified algebraically as

$$\begin{aligned}
\left(\sum_{i=1}^n q_i a_i^2\right) - \left(\sum_{i=1}^n q_i a_i\right)^2 &= \left(\sum_{i=1}^n q_i a_i^2\right)\left(\sum_{i=1}^n q_i\right) - \left(\sum_{i=1}^n q_i a_i\right)^2 \\
&= \left[\sum_{i=1}^n q_i^2 a_i^2 + \sum_{1 \leqslant i < j \leqslant n} q_i q_j (a_i^2 + a_j^2)\right] - \left[\sum_{i=1}^n q_i^2 a_i^2 + \sum_{1 \leqslant i < j \leqslant n} 2 q_i q_j a_i a_j\right] \\
&= \sum_{1 \leqslant i < j \leqslant n} q_i q_j (a_i^2 + a_j^2 - 2a_i a_j) = \sum_{1 \leqslant i < j \leqslant n} q_i q_j (a_i - a_j)^2.
\end{aligned}$$

The proof is then completed. □

We are now ready to utilize Lemmas 3 and 4 to validate Theorem 7.

*Proof of Theorem 7.* Thanks to Theorem 1, when $T = 2$ and $y_0 = 0$, $\min_{\boldsymbol{x} \in \mathbb{R}_+^T} f(\boldsymbol{x})$ could be expanded to

$$\begin{aligned}
\text{minimize} \quad & c(x_1 + x_2) + \alpha + \mu(\beta_1 + \beta_2) + (\mu^2 + \sigma^2)(\gamma_1 + \gamma_2) && \text{(16a)} \\
\text{subject to} \quad & x_1 \in \mathbb{R}_+,\ x_2 \in \mathbb{R}_+,\ \alpha \in \mathbb{R},\ \beta_1 \in \mathbb{R},\ \beta_2 \in \mathbb{R},\ \gamma_1 \in \mathbb{R}_+,\ \gamma_2 \in \mathbb{R}_+ && \text{(16b)} \\
& \alpha - 2hx_1 - hx_2 - (\beta_1 + 2h)^2/4\gamma_1 - (\beta_2 + h)^2/4\gamma_2 \geqslant 0 && \text{(16c)} \\
& \alpha + 2bx_1 + bx_2 - (\beta_1 - 2b)^2/4\gamma_1 - (\beta_2 - b)^2/4\gamma_2 \geqslant 0 && \text{(16d)} \\
& \alpha + (b - h)x_1 - hx_2 - (\beta_1 + h - b)^2/4\gamma_1 - (\beta_2 + h)^2/4\gamma_2 \geqslant 0 && \text{(16e)} \\
& \alpha + (b - h)x_1 + bx_2 - (\beta_1 + h - b)^2/4\gamma_1 - (\beta_2 - b)^2/4\gamma_2 \geqslant 0, && \text{(16f)}
\end{aligned}$$

where the four constraints correspond to $\boldsymbol{e} = (h, h)^\top = \check{\boldsymbol{e}}^2$, $\boldsymbol{e} = (-b, -b)^\top = \check{\boldsymbol{e}}^0$, $\boldsymbol{e} = (-b, h)^\top$ and $\boldsymbol{e} = (h, -b)^\top = \check{\boldsymbol{e}}^1$, respectively. To prove that $\min_{\boldsymbol{x} \in \mathbb{R}_+^T} f(\boldsymbol{x}) = \min_{\boldsymbol{x} \in \mathbb{R}_+^T} \underline{f}(\boldsymbol{x})$, it suffices to show that the constraint (16e) that is corresponding to the scenario $\boldsymbol{e} = (-b, h)^\top$ is redundant and can be removed without any impact on the optimal objective value. To achieve this, we will derive the dual problem of Problem (16) by letting $\boldsymbol{q} \in \mathbb{R}_+^4$ be a collection of Lagrange multipliers of the four constraints in Problem (16) respectively, and we will show that at optimality $q_3^\star$, which is the multiplier of the constraint (16e), vanishes.

The Lagrangian function associated with Problem (16) reads

$$\begin{aligned}
\mathcal{L}(\boldsymbol{x}, \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{q}) \; = \; & c(x_1 + x_2) + \alpha + \mu(\beta_1 + \beta_2) + (\mu^2 + \sigma^2)(\gamma_1 + \gamma_2) - \\
& q_1 \left(\alpha - 2hx_1 - hx_2 - (\beta_1 + 2h)^2/4\gamma_1 - (\beta_2 + h)^2/4\gamma_2\right) - \\
& q_2 \left(\alpha + 2bx_1 + bx_2 - (\beta_1 - 2b)^2/4\gamma_1 - (\beta_2 - b)^2/4\gamma_2\right) - \\
& q_3 \left(\alpha + (b - h)x_1 - hx_2 - (\beta_1 + h - b)^2/4\gamma_1 - (\beta_2 + h)^2/4\gamma_2\right) - \\
& q_4 \left(\alpha + (b - h)x_1 + bx_2 - (\beta_1 + h - b)^2/4\gamma_1 - (\beta_2 - b)^2/4\gamma_2\right).
\end{aligned}$$

The dual problem is thus

$$\max_{\boldsymbol{q}\geqslant\boldsymbol{0}} \min_{\boldsymbol{x}\geqslant\boldsymbol{0},\boldsymbol{\gamma}\geqslant\boldsymbol{0},\alpha,\boldsymbol{\beta}} \mathcal{L}(\boldsymbol{x},\alpha,\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{q}).$$

For the inner minimization to be bounded from below with respect to the choice of $x_1$, the inequality $c+2hq_1-2bq_2+(h-b)q_3+(h-b)q_4\geqslant 0$ must hold. With respect to the choice of $x_2$ and $\alpha$, the inequality $c+hq_1-bq_2+hq_3-bq_4\geqslant 0$ and the equality $q_1+q_2+q_3+q_4=1$ must similarly hold, respectively. These three conditions then show up explicitly as constraints in the dual problem of Problem (16).

After eliminating $\boldsymbol{x}$ and $\alpha$, the Lagrangian $\mathcal{L}$ can be decomposed into two parts:

$$\frac{1}{4\gamma_1}\left[4\mu\beta_1\gamma_1+q_1(\beta_1+2h)^2+q_2(\beta_1-2b)^2+q_3(\beta_1+h-b)^2+q_4(\beta_1+h-b)^2\right]+(\mu^2+\sigma^2)\gamma_1,$$

which depends only on $\beta_1,\gamma_1,\boldsymbol{q}$ and is thus abbreviated as $\mathcal{L}_1(\beta_1,\gamma_1,\boldsymbol{q})$, and

$$\frac{1}{4\gamma_2}\left[4\mu\beta_2\gamma_2+q_1(\beta_2+h)^2+q_2(\beta_2-b)^2+q_3(\beta_2+h)^2+q_4(\beta_2-b)^2\right]+(\mu^2+\sigma^2)\gamma_2,$$

which depends solely on $\beta_2,\gamma_2,\boldsymbol{q}$ and we shall abbreviate it as $\mathcal{L}_2(\beta_2,\gamma_2,\boldsymbol{q})$.

Next, we invoke Lemma 3 to minimize $\mathcal{L}_1(\beta_1,\gamma_1,\boldsymbol{q})$ over $\beta_1$ and find that the minimum objective is

$$\begin{aligned}(\mu^2+\sigma^2)\gamma_1+\frac{1}{4\gamma_1}\Bigg[&4q_1h^2+4q_2b^2+q_3(h-b)^2+q_4(h-b)^2-\\ &\frac{1}{4}\left(4\mu\gamma_1+4q_1h-4q_2b+2q_3(h-b)+2q_4(h-b)\right)^2\Bigg].\end{aligned} \tag{17}$$

Utilizing Lemma 3 again to minimize $\mathcal{L}_2(\beta_2,\gamma_2,\boldsymbol{q})$ over $\beta_2$ yields the minimum objective of

$$(\mu^2+\sigma^2)\gamma_2+\frac{1}{4\gamma_2}\left[q_1h^2+q_2b^2+q_3h^2+q_4b^2-\frac{1}{4}\left(4\mu\gamma_2+2q_1h-2q_2b+2q_3h-2q_4b\right)^2\right]. \tag{18}$$

We subsequently leverage Lemma 4 with $\boldsymbol{a}=(2h,-2b,h-b,h-b)^\top$ to simplify $\min_{\beta_1}\mathcal{L}_1(\beta_1,\gamma_1,\boldsymbol{q})$ from (17) further to

$$\gamma_1\sigma^2-\mu\left(2q_1h-2q_2b+q_3(h-b)+q_4(h-b)\right)+\frac{1}{4\gamma_1}(h+b)^2\left(4q_1q_2+q_1q_3+q_1q_4+q_2q_3+q_2q_4\right), \tag{19}$$

and with $\boldsymbol{a}=(h,-b,h,-b)^\top$ to simplify $\min_{\beta_2}\mathcal{L}_2(\beta_2,\gamma_2,\boldsymbol{q})$ from (18) further to

$$\gamma_2\sigma^2-\mu\left(q_1h-q_2b+q_3h-q_4b\right)+\frac{1}{4\gamma_2}(h+b)^2\left(q_1q_2+q_1q_4+q_2q_3+q_3q_4\right). \tag{20}$$

From (19) and (20), we respectively use the arithmetic mean–geometric mean inequality to argue that $\min_{\gamma_1\geqslant 0,\beta_1}\mathcal{L}_1(\beta_1,\gamma_1,\boldsymbol{q})$ is equal to

$$\sigma(h+b)\sqrt{4q_1q_2+q_1q_3+q_1q_4+q_2q_3+q_2q_4}-\mu\left(2q_1h-2q_2b+q_3(h-b)+q_4(h-b)\right)$$

and that $\min_{\gamma_2\geqslant 0,\beta_2}\mathcal{L}_2(\beta_2,\gamma_2,\boldsymbol{q})$ is equal to

$$\sigma(h+b)\sqrt{q_1q_2+q_1q_4+q_2q_3+q_3q_4}-\mu\left(q_1h-q_2b+q_3h-q_4b\right).$$

The summation of $\min_{\gamma_1\geqslant 0,\beta_1}\mathcal{L}_1(\beta_1,\gamma_1,\boldsymbol{q})$ and $\min_{\gamma_2\geqslant 0,\beta_2}\mathcal{L}_2(\beta_2,\gamma_2,\boldsymbol{q})$ gives rise to the dual objective, and as a consequence the dual problem of Problem (16) reads

$$\text{maximize} \quad (h+b)\sigma\left(\sqrt{4q_1q_2+q_1q_3+q_1q_4+q_2q_3+q_2q_4}+\sqrt{q_1q_2+q_1q_4+q_2q_3+q_3q_4}\right)-$$

$$\mu\,(3hq_1 - 3bq_2 + (2h-b)q_3 + (h-2b)q_4) \tag{21a}$$

$$\text{subject to} \quad q_1,\ q_2,\ q_3,\ q_4 \geqslant 0 \tag{21b}$$

$$q_1 + q_2 + q_3 + q_4 = 1 \tag{21c}$$

$$c + 2hq_1 - 2bq_2 + (h-b)q_3 + (h-b)q_4 \geqslant 0 \tag{21d}$$

$$c + hq_1 - bq_2 + hq_3 - bq_4 \geqslant 0. \tag{21e}$$

To show that $q_3^\star = 0$, we divide our analysis into two cases depending on the value of $q_4^\star$.

*Case 1 ($q_4^\star > 0$):* Suppose for the sake of a contradiction that $q_3^\star > 0$ and consider an alternative solution $\boldsymbol{q}^\dagger = (q_1^\star + \delta, q_2^\star + \delta, q_3^\star - \delta, q_4^\star - \delta)$ for some sufficiently small $\delta > 0$ such that $\boldsymbol{q}^\dagger$ remains in the nonnegative orthant. It can be directly verified that $\boldsymbol{q}^\dagger$ is feasible in Problem (21). Besides, it holds that

$$\begin{aligned} q_1^\dagger q_2^\dagger + q_1^\dagger q_4^\dagger + q_2^\dagger q_3^\dagger + q_3^\dagger q_4^\dagger &= (q_1^\star+\delta)(q_2^\star+\delta) + (q_1^\star+\delta)(q_4^\star-\delta) + (q_2^\star+\delta)(q_3^\star-\delta) + (q_3^\star-\delta)(q_4^\star-\delta) \\ &= q_1^\star q_2^\star + q_1^\star q_4^\star + q_2^\star q_3^\star + q_3^\star q_4^\star \end{aligned}$$

and that

$$\begin{aligned} 4q_1^\dagger q_2^\dagger + q_1^\dagger q_3^\dagger + q_1^\dagger q_4^\dagger + q_2^\dagger q_3^\dagger + q_2^\dagger q_4^\dagger &= 4(q_1^\star+\delta)(q_2^\star+\delta) + (q_1^\star+\delta)(q_3^\star-\delta) + (q_1^\star+\delta)(q_4^\star-\delta) + \\ &\qquad (q_2^\star+\delta)(q_3^\star-\delta) + (q_2^\star+\delta)(q_4^\star-\delta) \\ &= 4q_1^\star q_2^\star + q_1^\star q_3^\star + q_1^\star q_4^\star + q_2^\star q_3^\star + q_2^\star q_4^\star + 2\delta \end{aligned}$$

as well as that

$$\begin{aligned} 3hq_1^\dagger - 3bq_2^\dagger + (2h-b)q_3^\dagger + (h-2b)q_4^\dagger &= 3h(q_1^\star+\delta) - 3b(q_2^\star+\delta) + (2h-b)(q_3^\star-\delta) + (h-2b)(q_4^\star-\delta) \\ &= 3hq_1^\star - 3bq_2^\star + (2h-b)q_3^\star + (h-2b)q_4^\star. \end{aligned}$$

Together, the above three observations imply that $\boldsymbol{q}^\dagger$ attains an objective function value that is strictly larger than $\boldsymbol{q}^\star$, which in turn contradicts with the supposed optimality of $\boldsymbol{q}^\star$ and therefore renders $q_3^\star > 0$ impossible to materialize.

*Case 2 ($q_4^\star = 0$):* Next, we consider the other case with vanishing $q_4^\star$. Under this condition, Problem (21) simplifies to a trivariate optimization problem:

$$\begin{aligned} \text{maximize} \quad & (h+b)\sigma\left(\sqrt{4q_1q_2 + q_1q_3 + q_2q_3} + \sqrt{q_1q_2 + q_2q_3}\right) - \mu\,(3hq_1 - 3bq_2 + (2h-b)q_3) \\ \text{subject to} \quad & q_1,\ q_2,\ q_3 \geqslant 0 \\ & q_1 + q_2 + q_3 = 1 \\ & c + 2hq_1 - 2bq_2 + (h-b)q_3 \geqslant 0 \\ & c + hq_1 - bq_2 + hq_3 \geqslant 0 \end{aligned}$$

and subsequently to a bivariate optimization problem (barring the shifting and positive scaling of the objective function):

$$\text{maximize} \quad \sigma\left(\sqrt{q_1 + q_2 - (q_1 - q_2)^2} + \sqrt{q_2(1-q_2)}\right) + \mu\,(2q_2 - q_1) \tag{22a}$$

$$\text{subject to} \quad q_1,\ q_2 \geqslant 0 \tag{22b}$$

$$q_1 + q_2 \leqslant 1 \tag{22c}$$

$$q_2 - q_1 \leqslant \frac{c+h-b}{h+b} \tag{22d}$$

$$q_2 \leqslant \frac{c+h}{h+b}. \tag{22e}$$

We remark that Problem (22) constitutes a concave maximization problem because square root function is concave and increasing and a quadratic function with a non-positive leading coefficient is also concave. In view of Problem (22), our original goal to establish that $q_3^\star = 0$ is tantamount to showing that the constraint (22c) is binding at optimality. We begin by arguing that the constraint (22e) cannot be binding because otherwise it would hold that

$$q_1^\star + q_2^\star \;\geqslant\; 2q_2^\star - \frac{c+h-b}{h+b} \;=\; \frac{2(c+h)}{h+b} - \frac{c+h-b}{h+b} \;>\; 1,$$

where the first inequality is due to (22d) and hence a contradiction. Therefore, the constraint (22e) can be omitted without any loss of optimality. Suppose next that the constraint (22d) is binding, Problem (22) simplifies (barring again the shifting and positive scaling of the objective function) to the following univariate concave maximization problem:

$$\begin{array}{ll} \text{maximize} & \sqrt{2q_2 - \dfrac{c+h-b}{h+b} - \left(\dfrac{c+h-b}{h+b}\right)^2} + \sqrt{q_2(1-q_2)} + rq_2 \\ \text{subject to} & \max\left\{\dfrac{c+h-b}{h+b}, 0\right\} \leqslant q_2 \leqslant \dfrac{2h+c}{2h+2b}, \end{array} \tag{23}$$

where $r$ is a positive constant equal to $\frac{\mu}{\sigma}$. In Problem (23), the lower bound of $q_2$ ensures that both $q_1$ and $q_2$ are nonnegative, whereas the upper bound ensures that $q_1 + q_2 \leqslant 1$, which is an explicit constraint in Problem (22). To show that (22c) is a binding constraint, it suffices to show that the upper bound $\frac{2h+c}{2h+2b}$ is attained by $q_2^\star$, and to achieve this we will show that the objective function of (23) is increasing in $q_2$ over its admissible range. Observe that Problem (23) can only be feasible when $2b \geqslant c$, which we shall assume, and the derivative of the objective function denoted by $g_{(23)}$ is:

$$\begin{aligned} \frac{\partial g_{(23)}}{\partial q_2} &= \frac{1}{\sqrt{2q_2 - \frac{c+h-b}{h+b} - \left(\frac{c+h-b}{h+b}\right)^2}} + \frac{1-2q_2}{2\sqrt{q_2(1-q_2)}} + r \\ &= \frac{h+b}{\sqrt{2q_2(h+b)^2 - (c+h-b)(2h+c)}} + \frac{1-2q_2}{2\sqrt{q_2(1-q_2)}} + r. \end{aligned}$$

When $q_2 = \frac{2h+c}{2h+2b}$,

$$\begin{aligned} \left.\frac{\partial g_{(23)}}{\partial q_2}\right|_{q_2 = \frac{2h+c}{2h+2b}} &= \frac{h+b}{\sqrt{(h+b)(2h+c) - (c+h-b)(2h+c)}} + \frac{(b-h-c)/(h+b)}{\sqrt{(2h+c)(2b-c)/(h+b)^2}} + r \\ &= \frac{h+b}{\sqrt{(2h+c)(2b-c)}} + \frac{b-h-c}{\sqrt{(2h+c)(2b-c)}} + r \\ &= \sqrt{\frac{2b-c}{2h+c}} + r \\ &> 0. \end{aligned}$$

By its concavity, $g_{(23)}$ is increasing in $q_2 \in \left[\max\left\{\frac{c+h-b}{h+b}, 0\right\}, \frac{2h+c}{2h+2b}\right]$ and as a result $q_3^\star$ indeed vanishes.

Henceforth, we can assume without any loss of optimality that constraints (22d) and (22e) are not binding. If at optimality the constraint (22c) is also not binding (*i.e.*, if $q_3^\star > 0$), it must be possible to identify $\boldsymbol{q}^\star$ that is optimal in Problem (22) by solving

$$\begin{array}{ll} \text{maximize} & \sqrt{q_1 + q_2 - (q_1 - q_2)^2} + \sqrt{q_2(1-q_2)} + r\,(2q_2 - q_1) \\ \text{subject to} & 0 \leqslant q_1 \leqslant 1,\; 0 \leqslant q_2 \leqslant 1. \end{array} \tag{24}$$

Note that the square roots are well-defined over the feasible region as $q_1 + q_2 \geqslant |q_1 - q_2| \geqslant (q_1 - q_2)^2$. By a familiar argument that we use to analyze Problem (22) earlier, Problem (24) remains a concave maximization problem. Then, by expressing its objective function $g_{(24)}$ as

$$g_{(24)}(q_1, q_2) \;=\; \sqrt{q_1(1-q_1) + q_2(1-q_2) + 2q_1q_2} + \sqrt{q_2(1-q_2)} + r\,(2q_2 - q_1)$$

and noting that the mapping $q_2 \mapsto q_2(1-q_2)$ is strictly increasing over the interval $[0, \frac{1}{2}]$, it necessarily holds that $q_2^\star \geqslant \frac{1}{2}$. Next, we derive the partial derivative of $g_{(24)}$ with respect to $q_1$:

$$\frac{\partial g_{(24)}}{\partial q_1} \;=\; \frac{1 - 2q_1 + 2q_2}{2\sqrt{q_1 + q_2 - (q_1 - q_2)^2}} - r.$$

When $q_2$ is chosen optimally as $q_2^\star$, we can determine the values of $q_1$ such that the above partial derivative vanishes as follows:

$$\begin{aligned}
& (1 - 2q_1 + 2q_2^\star)^2 \;=\; 4r^2\left(q_1 + q_2^\star - (q_1 - q_2^\star)^2\right) \\
\iff\;& 4(q_2^\star - q_1)^2 + 4(q_2^\star - q_1) + 1 \;=\; 8r^2 q_2^\star - 4r^2(q_2^\star - q_1) - 4r^2(q_2^\star - q_1)^2 \\
\iff\;& 4(1+r^2)(q_2^\star - q_1)^2 + 4(1+r^2)(q_2^\star - q_1) + 1 - 8r^2 q_2^\star = 0 \\
\iff\;& q_2^\star - q_1 = -\frac{1}{2} \pm \frac{1}{2}\sqrt{\frac{r^2(1+8q_2^\star)}{1+r^2}} \\
\iff\;& q_1 = q_2^\star + \frac{1}{2} \mp \frac{1}{2}\sqrt{\frac{r^2(1+8q_2^\star)}{1+r^2}}.
\end{aligned}$$

The above two values of $q_1$ together with the boundary points (*i.e.,* 0 and 1) are the candidate values for $q_1^\star$. However, as $q_2^\star \geqslant \frac{1}{2}$ and as $q_1^\star + q_2^\star \leqslant 1$, we can rule out the values of $q_1^\star$ that are greater than $\frac{1}{2}$. Hence, we find that $q_1^\star \in \{0, q_1^\dagger\}$, where $q_1^\dagger = q_2^\star + \frac{1}{2} - \frac{1}{2}\sqrt{\frac{r^2(1+8q_2^\star)}{1+r^2}}$. We can now divide our remaining analysis into two subcases.

*Case 2a ($q_1^\star = 0$):* In this case, as $g_{(24)}$ is concave in $q_1$ we have $q_1^\dagger \leqslant 0$. We will assume here that $q_1^\dagger < 0$ and treat the remaining possibility of $q_1^\star = q_1^\dagger = 0$ in *Case 2b*. Here, Problem (24) reduces to a univariate optimization problem in $q_2$. By a slight abuse of notation, we denote the resulting objective $g_{(24)}(0, q_2)$ by $g_{(24)}(q_2)$, and we find

$$\frac{\partial g_{(24)}}{\partial q_2} \;=\; \frac{1 - 2q_2}{\sqrt{q_2(1-q_2)}} + 2r.$$

This gradient simply evaluates to $2r > 0$ when $q_2 = \frac{1}{2}$, and it evaluates to a negative number when $q_2$ approaches one from below. By the intermediate value theorem, there exists a value of $q_2 \in \left(\frac{1}{2}, 1\right)$ such that the gradient vanishes. Since $g_{(24)}$ is concave in $q_2$, it then follows that

$$\frac{1 - 2q_2^\star}{\sqrt{q_2^\star(1-q_2^\star)}} + 2r \;=\; 0 \quad\Longrightarrow\quad r \;=\; \frac{2q_2^\star - 1}{2\sqrt{q_2^\star(1-q_2^\star)}} \quad\Longrightarrow\quad \sqrt{\frac{r^2}{1+r^2}} = 2q_2^\star - 1.$$

As a result, we obtain that

$$q_2^\star + \frac{1}{2} - \frac{1}{2}(2q_2^\star - 1)\sqrt{1 + 8q_2^\star} \;=\; q_1^\dagger \;<\; 0 \quad\Longrightarrow\quad (2q_2^\star + 1)^2 \;<\; (2q_2^\star - 1)^2(1 + 8q_2^\star).$$

The latter inequality implies that $32(q_2^\star)^2(q_2^\star - 1) > 0$, which is an impossible consequence in view of the feasibility of Problem (24).

*Case 2b ($q_1^\star = q_1^\dagger$):* Regardless of the value of $q_1$, we find

$$\frac{\partial g_{(24)}}{\partial q_2} \;=\; \frac{1+2q_1-2q_2}{2\sqrt{q_1+q_2-(q_1-q_2)^2}} + \frac{1-2q_2}{2\sqrt{q_2(1-q_2)}} + 2r.$$

Similarly to the above case, for any fixed $q_1 \leqslant \frac{1}{2}$, there exists a value of $q_2 \in \left(\frac{1}{2}, 1\right)$ such that this gradient vanishes, and thus at optimality, we have

$$\frac{1+2q_1^\star-2q_2^\star}{2\sqrt{q_1^\star+q_2^\star-(q_1^\star-q_2^\star)^2}} + \frac{1-2q_2^\star}{2\sqrt{q_2^\star(1-q_2^\star)}} + 2r \;=\; 0. \tag{25}$$

Note that, as $q_1^\star = q_1^\dagger$, it follows that

$$2\sqrt{q_1^\star+q_2^\star-(q_1^\star-q_2^\star)^2} \;=\; 2\sqrt{2q_2^\star+\frac{1}{2}-\frac{1}{2}\sqrt{\frac{r^2(1+8q_2^\star)}{1+r^2}}-\frac{1}{4}\left(1-\sqrt{\frac{r^2(1+8q_2^\star)}{1+r^2}}\right)^2} \;=\; \sqrt{\frac{1+8q_2^\star}{1+r^2}}$$

and that

$$1+2q_1^\star-2q_2^\star \;=\; 2-r\sqrt{\frac{1+8q_2^\star}{1+r^2}}.$$

Substituting these observations into (25), we obtain

$$2\sqrt{\frac{1+r^2}{1+8q_2^\star}} + \frac{1-2q_2^\star}{2\sqrt{q_2^\star(1-q_2^\star)}} + r = 0 \quad\Longrightarrow\quad r < \frac{2q_2^\star-1}{2\sqrt{q_2^\star(1-q_2^\star)}} \quad\Longrightarrow\quad \sqrt{\frac{r^2}{1+r^2}} < 2q_2^\star - 1,$$

where the rightmost implication holds because the mapping $r \mapsto \sqrt{\frac{r^2}{1+r^2}}$ is increasing in $r \in [0,\infty)$. As the inequality $q_1+q_2 \leqslant 1$ is an explicit constraint of Problem (23), which is supposed to be non-binding, it must strictly hold when $q_1 = q_1^\star$ and $q_2 = q_2^\star$. Hence,

$$2q_2^\star+\frac{1}{2}-\frac{1}{2}\sqrt{\frac{r^2(1+8q_2^\star)}{1+r^2}} = q_1^\star+q_2^\star < 1 \quad\Longrightarrow\quad (4q_2^\star-1)^2 < (2q_2^\star-1)^2(1+8q_2^\star).$$

The latter can hold only if $4q_2^\star(q_2^\star-1)(8q_2^\star-3) > 0$ and hence a contradiction.

We thus conclude that, in all cases, it is impossible for $q_3^\star$ to be a strictly positive number and complete the proof. □

*Proof of Theorem 8.* From an argument widely parallel to the proof of Theorem 1, we have that

$$\begin{aligned} f^\dagger(\boldsymbol{x}) - c\mathbf{1}^\top\boldsymbol{x} = \text{minimize} \quad & \alpha + \mu\mathbf{1}^\top\boldsymbol{\beta} + (\mu^2+\sigma^2)\mathbf{1}^\top\boldsymbol{\gamma} \\ \text{subject to} \quad & \alpha\in\mathbb{R},\ \boldsymbol{\beta}\in\mathbb{R}^T,\ \boldsymbol{\gamma}\in\mathbb{R}^T \\ & \alpha(\boldsymbol{x},\boldsymbol{e}) + \sum_{t=1}^T \beta_t(\boldsymbol{e})\xi_t + \sum_{t=1}^T \gamma_t\xi_t^2 \geqslant 0 \quad \forall\boldsymbol{e}\in\mathcal{E}\ \forall\boldsymbol{\xi}\in\mathbb{R}_+^T. \end{aligned}$$

For each fixed $\boldsymbol{e}\in\mathcal{E}$, the robust constraint holds for all $\boldsymbol{\xi}\in\mathbb{R}_+^T$ iff there exists $\boldsymbol{\theta}\in\mathbb{R}^T$ such that

$$\alpha(\boldsymbol{x},\boldsymbol{e}) \geqslant \mathbf{1}^\top\boldsymbol{\theta} \quad\text{and}\quad \theta_t + \beta_t(\boldsymbol{e})\xi_t + \gamma_t\xi_t^2 \geqslant 0 \quad \forall\xi_t\geqslant 0\ \forall t=1,\ldots,T.$$

Note that it immediately holds that each $\theta_t$ must be non-negative for otherwise the robust quadratic constraint fails to hold when $\xi_t = 0$. Furthermore, by the virtue of $\mathcal{S}$-lemma, each of these robust quadratic constraints holds iff there exists a $\delta_t \geqslant 0$, $t=1,\ldots,T$, such that

$$\begin{aligned} \begin{bmatrix} \gamma_t & \frac{1}{2}\beta_t(\boldsymbol{e}) \\ \frac{1}{2}\beta_t(\boldsymbol{e}) & \theta_t \end{bmatrix} \succeq_{\mathbb{S}_+^2} \delta_t \begin{bmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix} \quad &\Longleftrightarrow\quad \begin{cases} \gamma_t \geqslant 0, \quad \theta_t \geqslant 0 \\ 4\gamma_t\theta_t \geqslant (\beta_t(\boldsymbol{e})-\delta_t)^2 \end{cases} \\ &\Longleftrightarrow\quad \begin{cases} \gamma_t \geqslant 0, \quad \theta_t \geqslant 0 \\ (\theta_t, \beta_t(\boldsymbol{e})-\delta_t, \gamma_t) \succeq_{\mathcal{K}_1} \mathbf{0}. \end{cases} \end{aligned}$$

By leveraging the definition of $\mathcal{K}_T^\dagger$, we finally complete the proof. □

*Proof of Theorem 9.* From an argument widely parallel to the proof of Theorem 1, we have that

$$\begin{aligned} f^{\perp}(\boldsymbol{x}) - c\mathbf{1}^\top \boldsymbol{x} = \text{minimize} \quad & \alpha + \mu\mathbf{1}^\top\boldsymbol{\beta} + (\mu^2+\sigma^2)\mathbf{1}^\top\boldsymbol{\gamma} + \mu^2 \sum_{1\leqslant s<t\leqslant T} \theta_{st} \\ \text{subject to} \quad & \alpha\in\mathbb{R},\ \boldsymbol{\beta}\in\mathbb{R}^T,\ \boldsymbol{\gamma}\in\mathbb{R}^T,\ \boldsymbol{\Theta}\in\mathbb{R}^{T\times T} \\ & \alpha(\boldsymbol{x},\boldsymbol{e}) + \sum_{t=1}^T \beta_t(\boldsymbol{e})\xi_t + \sum_{t=1}^T \gamma_t\xi_t^2 + \sum_{1\leqslant s<t\leqslant T} \theta_{st}\xi_s\xi_t \geqslant 0 \quad \forall \boldsymbol{e}\in\mathcal{E}\ \forall\boldsymbol{\xi}\in\mathbb{R}^T, \end{aligned}$$

where each $\theta_{st}$ is a dual variable assigned to the constraint that ensures $\xi_s$ and $\xi_t$ are uncorrelated. Note that $\theta_{st}$ with $s\geqslant t$ does not enter the above optimization problem.

Without any loss of optimality, we replace each $\theta_{st}$, $s<t$, with $2\theta_{st}$ and assign $\theta_{ss}=0$ for all $s$ and $\theta_{st}=\theta_{ts}$ for all pairs $(s,t)$ such that $s>t$ resulting in an equivalent problem of the form

$$\begin{aligned} f^{\perp}(\boldsymbol{x}) - c\mathbf{1}^\top \boldsymbol{x} = \text{minimize} \quad & \alpha + \mu\mathbf{1}^\top\boldsymbol{\beta} + (\mu^2+\sigma^2)\mathbf{1}^\top\boldsymbol{\gamma} + \mu^2 \mathbf{1}^\top\boldsymbol{\Theta}\mathbf{1} \\ \text{subject to} \quad & \alpha\in\mathbb{R},\ \boldsymbol{\beta}\in\mathbb{R}^T,\ \boldsymbol{\gamma}\in\mathbb{R}^T,\ \boldsymbol{\Theta}\in\mathbb{S}^{T} \\ & \alpha(\boldsymbol{x},\boldsymbol{e}) + \sum_{t=1}^T \beta_t(\boldsymbol{e})\xi_t + \sum_{t=1}^T \gamma_t\xi_t^2 + \boldsymbol{\xi}^\top\boldsymbol{\Theta}\boldsymbol{\xi} \geqslant 0 \quad \forall \boldsymbol{e}\in\mathcal{E}\ \forall\boldsymbol{\xi}\in\mathbb{R}^T \\ & \operatorname{diag}(\boldsymbol{\Theta}) = \mathbf{0}. \end{aligned}$$

For any fixed $\boldsymbol{e}\in\mathcal{E}$, it can be recognized that the robust quadratic constraint (which has to hold for all $\boldsymbol{\xi}\in\mathbb{R}^T$) is equivalent to

$$\begin{bmatrix} \boldsymbol{\xi} \\ 1 \end{bmatrix}^\top \begin{bmatrix} \operatorname{diag}(\boldsymbol{\gamma}) + \boldsymbol{\Theta} & \frac{1}{2}\boldsymbol{\beta}(\boldsymbol{e}) \\ \frac{1}{2}\boldsymbol{\beta}(\boldsymbol{e})^\top & \alpha(\boldsymbol{x},\boldsymbol{e}) \end{bmatrix} \begin{bmatrix} \boldsymbol{\xi} \\ 1 \end{bmatrix} \geqslant 0 \quad \forall\boldsymbol{\xi}\in\mathbb{R}^T \quad \Longleftrightarrow \quad \begin{bmatrix} \operatorname{diag}(\boldsymbol{\gamma}) + \boldsymbol{\Theta} & \frac{1}{2}\boldsymbol{\beta}(\boldsymbol{e}) \\ \frac{1}{2}\boldsymbol{\beta}(\boldsymbol{e})^\top & \alpha(\boldsymbol{x},\boldsymbol{e}) \end{bmatrix} \succeq \mathbf{0}.$$

This latest observation together with the definition of $\mathcal{K}_T^{\perp}$ completes the proof. □

*Proof of Theorem 10.* The proof widely parallels to that of Theorem 4, where we reexpress the problem from Theorem 9 as an artificial two-stage robust optimization problem

$$\begin{aligned} f^{\perp}(\boldsymbol{x}) = \text{minimize} \quad & c\mathbf{1}^\top\boldsymbol{x} + \alpha + \mu\mathbf{1}^\top\boldsymbol{\beta} + (\mu^2+\sigma^2)\mathbf{1}^\top\boldsymbol{\gamma} + \mu^2 \mathbf{1}^\top\boldsymbol{\Theta}\mathbf{1} \\ \text{subject to} \quad & \alpha\in\mathbb{R},\ \boldsymbol{\beta}\in\mathbb{R}^T,\ \boldsymbol{\gamma}\in\mathbb{R}^T, \boldsymbol{\Theta}\in\mathbb{S}^{T},\ \boldsymbol{\phi}:\mathbb{R}^{2T+T^2}\mapsto\mathbb{R}^T_+,\ \boldsymbol{\psi}:\mathbb{R}^{2T+T^2}\mapsto\mathbb{R}^T_+ \\ & \alpha + \boldsymbol{\beta}^\top\boldsymbol{u} + \boldsymbol{\gamma}^\top\boldsymbol{v} + \langle\boldsymbol{\Theta},\mathbf{W}\rangle \geqslant b\mathbf{1}^\top\boldsymbol{\phi}(\boldsymbol{u},\boldsymbol{v},\mathbf{W}) + h\mathbf{1}^\top\boldsymbol{\psi}(\boldsymbol{u},\boldsymbol{v},\mathbf{W}) \\ & \qquad\qquad \forall(\boldsymbol{u},\boldsymbol{v},\mathbf{W}) : (1,\boldsymbol{u},\boldsymbol{v},\mathbf{W})\in(\mathcal{K}_T^{\perp})^\star \\ & y_0 + \phi_t(\boldsymbol{u},\boldsymbol{v},\mathbf{W}) - \psi_t(\boldsymbol{u},\boldsymbol{v},\mathbf{W}) = \sum_{\tau=1}^t (u_\tau - x_\tau) \quad \forall t=1,\dots,T \\ & \qquad\qquad \forall(\boldsymbol{u},\boldsymbol{v},\mathbf{W}) : (1,\boldsymbol{u},\boldsymbol{v},\mathbf{W})\in(\mathcal{K}_T^{\perp})^\star \\ & \operatorname{diag}(\boldsymbol{\Theta}) = \mathbf{0}. \end{aligned} \tag{26}$$

Restricting the second-stage decision variables with a linear decision rule, where the additional decision variables $\boldsymbol{\Pi}_t^w$, $t=1,\dots,T$, characterize the sensitivity of $\boldsymbol{\phi}$ in Problem (26) with respect to $\mathbf{W}$, and leveraging the primal-dual pair complete the proof. □