Fourth-order Marginal Moment Model: Reformulations and Applications

Mingyang Fu
Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore, fumingyang@u.nus.edu,

Xiaobo Li
Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore, iselix@nus.edu.sg,

This paper investigates the bounds on the expectation of combinatorial optimization given moment information for each individual random variable. A popular approach to solving this problem, known as the marginal moment model (MMM), is to reformulate it as a semidefinite program (SDP). In this paper, we investigate the structure of MMM with up to fourth-order marginal moments and reformulate them as second-order cone programs (SOCP). Additionally, we establish that this SOCP formulation is equivalent to a convex optimization problem over the convex hull of the feasible region of the original combinatorial optimization, presenting closed-form expressions for both the objective function and its derivative. These reformulations enable more efficient computation of the bounds and persistency value. In addition, we explore two types of ambiguity sets characterized by incomplete moment information. We further discuss the relationship between the aforementioned MMMs and another widely-used model, the marginal distribution model (MDM). Beyond bounding the worst-case expectations, our approaches can be modified to bound the worst-case conditional value at risk (CVaR) of the combinatorial optimization.

Building on this theoretical advancement, we explore two applications. First, we consider the project crashing problem, wherein both the means and variances of activity durations can be controlled through effort. We demonstrate that the distributionally robust project crashing problem, incorporating up to fourth-order moment information, can be reformulated as either an SOCP or a convex minimization over a simple polytope. Numerical analysis reveals that MMM with fourth moment information yields tighter bounds on expected delays and requires a significantly smaller budget than the mean-variance model for a fixed delay guarantee. Second, we apply our reformulations to solve the distributionally robust newsvendor problem with moment information, extending the well-known Scarf's model. We derive several new closed-form solutions and explore how the optimal order quantities depend on skewness and kurtosis. Numerically, we show that incorporating additional moment information can lead to better performance, especially in the high service level regime.

Key words: marginal moment model, persistency, skewness, kurtosis, project management, newsvendor problem

1. Introduction

The study of combinatorial optimization problems under uncertainty has been an active and challenging area of research in operations research and related fields. These problems encompass a broad range of real-world applications, including supply chain, transportation, scheduling, and network design, where uncertainty is an inherent aspect of the decision-making process.

In general, a combinatorial optimization problem under uncertainty can be written as:

$$Z_{\max}(\tilde{\boldsymbol{c}}) = \max\left\{\tilde{\boldsymbol{c}}^T\boldsymbol{x} : \boldsymbol{x} \in \mathcal{X}\right\}. \tag{1}$$

where the feasible region is $\mathcal{X} \subseteq \{0,1\}^n$ and random objective coefficients $\tilde{\boldsymbol{c}}$. The most common approach in the literature is the stochastic approach, which assumes that $\tilde{\boldsymbol{c}}$ adheres to a known distribution, θ. Numerous applications have thoroughly explored the problem of evaluating the expected objective value, i.e., $E_\theta\left[Z_{\max}(\tilde{\boldsymbol{c}})\right]$. Examples of such applications include estimating the expected project completion time (Bertsimas et al. 2006a), calculating the expected order statistics (Bertsimas et al. 2006b), and determining the conditional value at risk (Rockafellar et al. 2000). Nonetheless, calculating the expected optimal objective value based on the distribution of $\boldsymbol{c}$ is generally challenging. For example, the longest path problem in a directed acyclic graph has been demonstrated by Hagstrom (1988) to be a $\#P$-complete problem.

In addition to determining the expected objective value, another important problem involves calculating the *persistency* of binary variables (see, e.g., Bertsimas et al. 2006a). The persistency of a binary variable x_i refers to the probability that the variable takes a value of one in the optimal solution, expressed as $Pr\left[x_i^*(\tilde{\boldsymbol{c}}) = 1\right]$ or, equivalently, $E\left[x_i^*(\tilde{\boldsymbol{c}})\right]$, where

$$\boldsymbol{x}^*(\boldsymbol{c}) \in \arg\max_{\boldsymbol{x}\in\mathcal{X}}\left\{\boldsymbol{c}^T\boldsymbol{x}\right\}. \tag{2}$$

Persistency extends the concepts of criticality index in project networks and choice probability in discrete choice models. As expected, computing persistency values could be more difficult. Even in cases where each objective coefficient has only two potential values, Bertsimas et al. (2006a) showed that the problem can be NP-hard.

Besides the computational complexity, the stochastic approach faces another challenge in real-world scenarios: the lack of knowledge about the distribution of $\tilde{\boldsymbol{c}}$. Instead, we often have access to samples of random variables. To address this distributional ambiguity, the *moment model* is commonly employed. It estimates moment information, such as mean and variance, from historical samples. By doing so, the moment model (as discussed in works like Bertsimas et al. 2006a, Natarajan et al. 2009, Li et al. 2014) aims to find an upper bound for $E[Z_{\max}(\tilde{\boldsymbol{c}})]$ across all distributions that are consistent with these estimated moments. The moment model encompasses two key variants: the marginal moment model (MMM) (Bertsimas et al. 2006a), which considers the first k-th marginal moment information of each random variable, and the cross moment model (CMM) (Mishra et al. 2012), which incorporates the mean and covariances of the random vector. Under certain conditions, the distributionally robust bounds can be reformulated as semi-definite programs (SDPs). Moreover, by solving these SDPs, we can compute the persistency under the worst-case distribution.

In most applications, although SDP formulation of the general marginal moment model (MMM) is available, the mean-variance model remains the most widely used approach. This is primarily due to the tractability of the mean-variance model, which can be formulated as an SOCP or a simple convex optimization problem without incorporating moment information into the constraints. However, it is widely recognized that the inclusion of third and fourth order moments provides additional valuable information beyond mean and variance alone. In particular, Skewness captures the asymmetry of a distribution, while kurtosis reveals its tailedness. Even in the fundamental newsvendor problem, the asymmetry and tailedness of the demand distribution play crucial roles (as demonstrated in works like Natarajan et al. 2018, Das et al. 2021). It is therefore acknowledged that considering skewness and kurtosis, rather than relying solely on mean and variance, can provide a more comprehensive understanding of the underlying distribution.

This paper attempts to address this dilemma by providing two types of reformulation techniques for the marginal moment model with up to fourth moment. The first approach reformulate the problem into an SOCP, while the second approach leads to minimizing a convex function over the convex hull of $\mathcal{X}$. Consequently, solving fourth-order MMM is as easy as solving the corresponding mean-variance models. These reformulations facilitate the adoption of the fourth-order MMM in applications compared to the SDP form.

Building on the theoretical advancement mentioned, we delve into two specific applications that leverage its potential. Firstly, we focus on the project crashing problem, which involves the ability to control both the means and variances of activity durations through dedicated efforts. For this problem, we showcase that the distributionally robust project crashing problem can be effectively reformulated using second-order cone programming (SOCP) or convex minimization over a simple polytope, when considering up to fourth-order moment information. Through comprehensive numerical analysis, we discover that employing the third and fourth moment in the distributionally robust project crashing problem yields tighter bounds on expected delays. Furthermore, it demands a considerably smaller budget compared to the mean-variance model, while ensuring a fixed delay guarantee.

Secondly, we employ the reformulations to address the distributionally robust newsvendor problem with moment information, which expands upon the well-known Scarf's mean-variance model.

By incorporating additional moment information, we derive several novel closed-form solutions. The closed-form solutions enable us to analytically explore the dependence of the optimal order quantity on the critical ratio, skewness, and kurtosis. Through extensive numerical simulations, we demonstrate that integrating higher moments' information can result in improved performance, particularly when the requirement for service level is high. This signifies the potential benefits of considering a wider range of statistical moments beyond mean and variance.

The structure of the remaining sections in this paper is outlined as follows. Section 2 provides a review of the relevant literature and covers the necessary preliminaries, including discussions on moments, skewness, kurtosis, moment problems, and a summary of existing results for the MMM. In Section 3.1, we present a new SOCP reformulation of 4-MMM which incorporating the first to the fourth order moments. We demonstrate its equivalence to a convex optimization problem over the convex hull of the feasible region with a closed-form objective function. In Section 3.2 we explore two extensions: the 124-MMM and 14-MMM, which account for scenarios where certain moments may be missing. Addition, we explored the relationship between the previously mentioned MMMs and MDM in Section 3.3 and the extend our approach to computing of CVaR in Section 3.4. Section 4 and 5 focus on two applications of our reformulations to two optimization problems: the project crashing problem and the newsvendor problem. Lastly, in Section 6, we conclude the paper.

2. Preliminaries and Related Literature

In this section, we review the relevant literature related to marginal moment models, including some technical details. The methodologies of these models will be thoroughly explored, and essential preliminaries will be provided to facilitate a comprehensive understanding of our results.

2.1. Moments and Moment Problems

Let $\tilde{c}$ be a random variable and α be a real number, the α-th moment of $\tilde{c}$ is defined as $m_\alpha = E\left[\tilde{c}^\alpha\right]$. For a d-dimensional random vector $\tilde{\boldsymbol{c}} = (\tilde{c}_1, \cdots, \tilde{c}_d)$ and a real power vector $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_d)$, the $\boldsymbol{\alpha}$-moment of $\tilde{\boldsymbol{c}}$ is defined as $m_{\boldsymbol{\alpha}} = E\left[\prod_{i=1}^d \tilde{c}_i^{\alpha_i}\right]$. Typically, the powers α or $\boldsymbol{\alpha}$ are assumed to be positive integers. In this work, we are particularly interested in the cases where $\alpha = 1, 2, 3, 4$.

It is clear that when $\alpha = 1$, the first moment is the mean of the random variable: $E[\tilde{c}] = \mu$. The α-th central (standardized) moment is defined as the α-th moment with respect to the mean, that is, $m'_\alpha = E\left[(\tilde{c} - \mu)^\alpha\right]$. The second-order central moment is the variance of the random variable: $V[\tilde{c}] = \sigma^2$. The skewness and the (excess) kurtosis of the random variable are defined as follows.

$$\text{Skewness: } S[\tilde{c}] = \frac{E\left[(\tilde{c} - \mu)^3\right]}{\sigma^3} = \gamma; \qquad \text{Kurtosis: } \quad K[\tilde{c}] = \frac{E\left[(\tilde{c} - \mu)^4\right]}{\sigma^4} - 3 = \kappa. \tag{3}$$

It is not difficult to see that $\mu, \sigma^2, \sigma^3\gamma$, and $\sigma^4\kappa$, are polynomials of the first to fourth moments. Conversely, for $n = 1, \cdots, 4$, n-th moment can also be presented as polynomials of $\mu, \sigma^2, \sigma^3\gamma$, and $\sigma^4\kappa$.

$$\begin{aligned}
\mu &= m_1, & m_1 &= \mu, \\
\sigma^2 &= m_2 - m_1^2, & m_2 &= \sigma^2 + \mu^2, \\
\sigma^3\gamma &= m_3 - 3m_1m_2 + 2m_1^3, & m_3 &= \sigma^3\gamma + 3\mu\sigma^2 + \mu^3, \\
\sigma^4\kappa &= m_4 - 4m_3m_1 - 3m_2^2 + 12m_2m_1^2 - 6m_1^4; & m_4 &= \sigma^4\kappa + 4\mu\sigma^3\gamma + 3\sigma^4 + 6\mu^2\sigma^2 + \mu^4.
\end{aligned}$$

When $\sigma = 0$, the distribution is a degenerate distribution and the random variable becomes deterministic. The skewness and kurtosis satisfy $\kappa - \gamma^2 + 2 \geq 0$, which is known as Pearson's inequality (see, e.g., Pearson 1905). When $\kappa - \gamma^2 + 2 = 0$, the distribution is a two point distribution. To avoid triviality, in the remainder of this paper, we assume that $\sigma > 0$ and $\kappa - \gamma^2 + 2 > 0$.

One important topic related to moments is the *moment problem*, which aims to determine whether a probability distribution exists for a given sequence of moments[1] (Schmüdgen 2017). If such a distribution exists, the sequence of moments is referred to as *feasible*; otherwise, it is termed *infeasible*. Depending on the sample spaces involved, moment problems can be classified into various types. The Hamburger moment problem deals with the sample space $\mathbb{R}$, the Stieltjes moment problem focuses on the sample space $\mathbb{R}^+$, and the Hausdorff moment problem pertains to the sample space $[0,1]$. Additionally, when only the first k-th moment is provided, the problem is known as the k-th order truncated moment problem. This paper primarily focuses on the fourth-order truncated Hamburger moment problem.

It is not difficult to see that all feasible moment sequence forms a convex cone. Therefore, we aim to characterize $\overline{\mathbb{M}_k(\Omega)}$, the closure of the k-th order truncated moment cone, which is defined as:

$$\mathbb{M}_k(\Omega) = \left\{ \boldsymbol{y} \in \mathbb{R}^{k+1} : \boldsymbol{y} = y_0 \left(1, E[\tilde{c}], \cdots, E\left[\tilde{c}^k\right]\right) \text{ for some } \tilde{c} \text{ with sample space } \Omega \text{ and } y_0 \geq 0 \right\}. \tag{4}$$

It is well known that the moment cones for all previous three types of moment problems can be represented using positive semidefinite constraints (see, e.g., Lasserre 2002, Bertsimas et al. 2006a). In particular, the positive semidefinite representation of the closure of $\mathbb{M}_{2k}(\mathbb{R})$ is

$$\overline{\mathbb{M}_{2k}(\mathbb{R})} = \{\boldsymbol{y} : \boldsymbol{H}_k(\boldsymbol{y}) \succeq 0\}$$

where

$$\boldsymbol{H}_k(\boldsymbol{y}) = \begin{pmatrix} y_0 & y_1 & \cdots & y_k \\ y_1 & y_2 & \cdots & y_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_k & y_{k+1} & \cdots & y_{2k} \end{pmatrix} \tag{5}$$

[1] The distribution can also be the limit of a sequence of distributions that satisfy the given moments

is called the Hankel matrix of $\boldsymbol{y}$.

In the n-dimensional case, the (n, k, Ω)-moment problem aims to determine the existence of a probability distribution on the sample space Ω that follow a specified series of moments $m_{\boldsymbol{\alpha}}$, where $\boldsymbol{\alpha}$ belongs to the set $\{\boldsymbol{\alpha} : \alpha_i \in \mathbb{Z}_+, \sum_{i=1}^n \alpha_i \leq k\}$. When $k = 2$, the closure of the $(n, 2, \mathbb{R}^n)$-moment cone can be represented as a positive semidefinite cone. Similarly, the closure of the $(n, 2, \mathbb{R}^n_+)$-moment cone can be represented using a completely positive program (refer to Natarajan et al. 2011 for more details).

2.2. Moment Models

Besides the moment problem, another important problem regarding moments is the *moment model*, which aims to bound the expectation of Z_{max} with moments. Among the moment models, the most popular one is the *marginal moment model* (MMM), which posits that only marginal moment information is available. In the context of the 0-1 integer problem 1, where the optimal solution $\boldsymbol{x}^*(\tilde{\boldsymbol{c}})$ is assumed to be unique almost surely, Bertsimas et al. (2006a) proposed the following k-MMM:

$$\begin{aligned} & Z^*_{MMM} = \sup_{\theta \in \Theta_k} E_\theta \left[Z_{\max}(\tilde{\boldsymbol{c}}) \right] \\ & \text{where } \Theta_k = \left\{ \theta : \theta_i \subset \mathcal{P}\left(\Omega_i\right), \left(E_\theta[\tilde{c}_i], \cdots, E_\theta\left[\tilde{c}_i^k\right] \right) = \left(m_{i,1}, \cdots, m_{i,k}\right), \forall i \in \{1, \cdots, n\} \right\}, \end{aligned} \tag{6}$$

where θ represents the joint distribution of $\tilde{\boldsymbol{c}}$, and θ_i represents the marginal distribution $\tilde{c}_i$. The distribution set Θ_k encompasses all possible distributions whose marginal distributions satisfy the given moment information. Assume that the convex hull of $\mathcal{X}$ is characterized by a set of linear constraints $\boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}$. According to Theorem 1 in Bertsimas et al. (2006a), the problem can be expressed as:

$$\begin{aligned} Z^*_{MMM} = \max \quad & \sum_{i=1}^n w_{i1} \\ \text{s.t.} \quad & \boldsymbol{w_i} + \boldsymbol{v_i} = \boldsymbol{m_i}, \quad \forall i \in \{1, \cdots, n\} \\ & \boldsymbol{A}\left(w_{1,0}, w_{2,0}, \cdots, w_{n,0}\right) \leq \boldsymbol{b} \\ & \boldsymbol{w_i}, \boldsymbol{v_i} \in \overline{\mathbb{M}_k\left(\mathbb{R}\right)}, \quad \forall i \in \{1, \cdots, n\}, \end{aligned} \tag{7}$$

where $\boldsymbol{w_i} = (w_{i,0}, w_{i,1}, \cdots, w_{i,k})$, $\boldsymbol{v_i} = (v_{i,0}, v_{i,1}, \cdots, v_{i,k})$. By leveraging the results of the Hamberger moment problem, the moment cones $\overline{\mathbb{M}_k\left(\mathbb{R}\right)}$ can be represented using positive semidefinite constraints on Hankel matrices, enabling the reformulation of the problem as an SDP. Here, $\boldsymbol{w_i}$ and $\boldsymbol{v_i}$ represent the conditional moments of the distribution of variable c_i conditional on $x_i^* = 1$ and 0. Moreover, the persistency of variable c_i, (i.e., the expectation of optimal $x_i^*(\boldsymbol{c})$ in (2)) is exactly $w_{i,0}$ in the optimal solution.

Natarajan et al. (2009) conducted further studies on the $k = 2$ case where only the first and second marginal moments are given. In this situation, the marginal moment vector is denoted as $\boldsymbol{m}_i = (1, m_{i,1}, m_{i,2})$. This 2-MMM is commonly referred to as the mean-variance model (MV), characterized by a mean $m_{i,1}$ and a variance $\sigma_i^2 = m_{i,2} - m_{i,1}^2$. Notably, the constraints $\boldsymbol{w_i}, \boldsymbol{v_i} \in \overline{\mathbb{M}_2(\mathbb{R})}$ can also be expressed as SOC constraints, as demonstrated below:

$$\begin{pmatrix} w_{i,0} & w_{i,1} \\ w_{i,1} & w_{i,2} \end{pmatrix} \succeq 0 \quad \Leftrightarrow \quad w_{i,1}^2 \leq w_{i,0} w_{i,2}.$$

Therefore, MV can be further reduced to following SOCP:

$$\begin{aligned} Z^*_{\max} = \max \quad & \sum_{i=1}^{n} \left(m_{i,1} x_i + \sigma_i y_i \right). \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}, \\ & y_i^2 \leq x_i (1 - x_i) \quad \forall i \in \{1, \cdots, n\}. \end{aligned} \tag{8}$$

Moreover, it is clear that the SOCP reformulation (8) can be further transformed into the following concave maximization problem with only the convex hull constraint:

$$\begin{aligned} Z^*_{\max} = \max \quad & \sum_{i=1}^{n} \left(m_{i,1} x_i + \sigma_i \sqrt{x_i (1 - x_i)} \right) \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}. \end{aligned} \tag{9}$$

Due to the simplicity of both formulations, the mean-variance model is the most popular momemt model. Until now, the literature has been unclear about the existence of such convenient formulations for higher marginal moment cases. In this paper, however, we take a step forward and demonstrate that these convenient formulations do exist for marginal moment information up to the fourth order. Moreover, our results can be extended to scenarios where some orders of moments are missing (see section 3.2). These findings significantly enhance the modeling analysis when higher-order moment information is available.

In addition to studying the information of marginal distributions, researchers have also delved into moment models that incorporate the covariance information of random variables. This is known as the cross moment model (CMM). Mishra et al. (2012) provided an SDP approach to bound the optimization problem when the feasible region is described by a finite set of points. They also constructed an extremal distribution based on the optimal solution. Natarajan et al. (2011) developed a completely positive programming reformulation for mixed 0-1 linear programs. Subsequently, Natarajan and Teo (2017) further reduced the size of the SDP, enhancing the model's computational tractability.

Beyond bounding the expectation of the objective $Z_{\max}$ in (1), related works in this domain have also investigated bounding other expected cost functions (Shapiro and Kleywegt 2002, Delage and Ye 2010), bounding probabilities (Vandenberghe et al. 2007, He et al. 2010), and bounding risk measures (Chen et al. (2011)). Recently, Guo et al. (2022) explored bounding functional expectations given generalized moment information. All these problems are special cases of the generalized moment problem (Birge and Wets 1987). In Section 3.4 of this paper, we also extend our results to bound the conditional value at risk of $Z_{\max}$ in (1), providing a solution technique to a class of generalized moment problem.

3. Fourth-Order Marginal Moment Models: Reformulation and Analysis

This section provides a comprehensive exploration of marginal moment models up to the fourth order. We start by presenting the key findings, highlighting that the 4-MMM (the MMM with the first four marginal moments) retains advantages akin to those of the mean-variance (MV) model. Specifically, the 4-MMM can be reformulated as both a second-order cone program and a convex optimization problem over the convex hull. Additionally, we demonstrate that the 4-MMM can be equivalently derived from the marginal distribution models (MDM) and delve into the properties of the marginal distributions. We also extend our analysis to the 124-MMM and 14-MMM models, where certain moment information is absent.

3.1. Complete Moment Information Case

In this subsection, we study the 4-MMM, i.e., the MMM with all the first four marginal moments are available, defined as:

$$\begin{aligned} & Z^*_{4-MMM} = \sup_{\theta\in\Theta_4} E_\theta\left(Z_{\max}(\tilde{\boldsymbol{c}})\right) \\ & \text{where } \Theta_4 = \left\{\theta : \theta_i \subset \mathcal{P}(\mathbb{R}), \left(E_\theta[\tilde{c}_i], \cdots, E_\theta\left[\tilde{c}_i^4\right]\right) = (m_{i,1}, \cdots, m_{i,4})\right\}. \end{aligned} \tag{4-MMM}$$

First, we show that 4-MMM can be reformulated as an SOCP in which the mean and variance of $\tilde{c}_i$ are the linear coefficients in the objective.

Theorem 1. *The 4-MMM problem (4-MMM) can be reformulated as the following SOCP:*

$$\begin{aligned} Z^*_{4-MMM} = \max_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}} \quad & \sum_{i=1}^n (\mu_i x_i + \sigma_i y_i) \\ s.t. \quad & y_i^2 \le x_i\left(x_i + \gamma_i y_i + z_i\sqrt{\kappa_i - \gamma_i^2 + 2}\right) && \forall i \in \{1, \cdots, n\} \\ & y_i^2 \le (1-x_i)\left(1 - x_i - \gamma_i y_i + z_i\sqrt{\kappa_i - \gamma_i^2 + 2}\right) && \forall i \in \{1, \cdots, n\} \\ & x_i^2 + y_i^2 + z_i^2 \le x_i && \forall i \in \{1, \cdots, n\} \\ & \boldsymbol{A}(x_1, x_2, \cdots, x_n) \le \boldsymbol{b}. \end{aligned} \tag{10}$$

Moreover, the persistency of variable c_i is x_i in the optimal solution.

The proof of this theorem is highly non-trivial and contains three main steps. Firstly, we introduce a substitution for $\boldsymbol{w}$, $\boldsymbol{v}$, and $\boldsymbol{m}$ in (7). Subsequently, by fixing the variable of persistency $\boldsymbol{x}$, we delve into the subproblems (40) and discuss them across several cases to incorporate degenerate distributions. In this step, most decision variables are eliminated, and the simplified subproblem becomes a univariate optimization problem. Finally, we integrate these simplified subproblems back into the primary problem (7), demonstrating that it can be concisely formulated as an SOCP problem. A detailed proof can be found in Appendix B.

It is not difficult to see that if the skewness γ_i and kurtosis κ_i are free variables, we can always let $\gamma_i = \frac{1-2x_i}{y_i}$ and $z_i = 0$ for any positive y_i. Then the problem precisely becomes the SOCP form of MV in formulation (8), which is consistent with the result of the mean-variance model.

Another noteworthy aspect of Theorem 1 is that the objective function in the optimization problem incorporates the means and standard deviations of random variables as linear coefficients. Meanwhile, the constraints involve only skewness and kurtosis. This characteristic presents the possibility of further optimizing the mean and variance by manipulating the location and scale of the random coefficient $\tilde{c}_i$ after applying duality. In Section 4, we utilize this technique in the context of the robust project crashing problem, where the manager has the ability to simultaneously influence the mean and variance of activities.

Based on the SOCP formulation, we further show that 4-MMM can also be reformulated as a concave maximization problem, where the feasible region is the convex hull of $\mathcal{X}$.

Theorem 2. *Assume that $\kappa_i - \gamma_i^2 + 2 > 0$, the 4-MMM problem (7) can be reduced to the following convex optimization problem:*

$$
\begin{aligned}
Z^*_{4-MMM} = \max \quad & \sum_{i=1}^{n} \left(\mu_i x_i + \sigma_i y_{\gamma_i,\kappa_i}(x_i)\right) \\
s.t. \quad & \boldsymbol{A}\left(x_1, x_2, \cdots, x_n\right) \leq \boldsymbol{b},
\end{aligned} \tag{11}
$$

where $y_{\gamma_i,\kappa_i}(x)$ is defined as:

$$
y_{\gamma_i,\kappa_i}(x) = \begin{cases} f_{\kappa_i}(x,\gamma_i), & x \leq \dfrac{1}{2}\left(1 - \dfrac{\gamma_i}{\sqrt{\gamma_i^2+4}}\right), \\ f_{\kappa_i}(1-x,-\gamma_i), & x > \dfrac{1}{2}\left(1 - \dfrac{\gamma_i}{\sqrt{\gamma_i^2+4}}\right), \end{cases} \tag{12}
$$

and $f_\kappa(x,\gamma)$ is the largest real root[2] *of the following quartic function:*

$$
Q(y) = \left(y^2 - \gamma x y - x^2\right)^2 + x^2\left(\kappa + 2 - \gamma^2\right) y^2 - x^3(1-x)\left(\kappa + 2 - \gamma^2\right).
$$

In addition, the persistency of variable c_i is x_i in the optimal solution.

Proof of Theorem 2: We only need to prove that, for any fixed x_i, the function $y_{\gamma_i,\kappa_i}(x_i)$ matches the objective to the optimization problem:

[2] The closed form expression of f_κ is presented in Appendix A.

$$
\begin{aligned}
\max_{y_i, z_i} \quad & y_i \\
\text{s.t.} \quad & y_i^2 \le x_i \left(x_i + \gamma_i y_i + \sqrt{\kappa_i - \gamma_i^2 + 2} z_i \right) \\
& y_i^2 \le (1 - x_i) \left(1 - x_i - \gamma_i y_i + \sqrt{\kappa_i - \gamma_i^2 + 2} z_i \right) \\
& x_i^2 + y_i^2 + z_i^2 \le x_i.
\end{aligned} \tag{13}
$$

For brevity, we omit the subscript of the function $y_{\gamma_i, \kappa_i}(x_i)$ and the index i of all variables in subsequent discussions. The constraints can be rephrased as follows:

$$
\begin{aligned}
\sqrt{x(1-x) - y^2} &\ge \frac{y^2 - x\gamma y - x^2}{x\sqrt{\kappa + 2 - \gamma^2}} = Q_1(y) \\
\sqrt{x(1-x) - y^2} &\ge \frac{y^2 + (1-x)\gamma y - (1-x)^2}{(1-x)\sqrt{\kappa + 2 - \gamma^2}} = Q_2(y).
\end{aligned} \tag{14}
$$

Here, Q_1 and Q_2 are upward-opening quadratic functions, and the left hand side of each constraint is a circular curve centered at $(0,0)$ with radius $\sqrt{x(1-x)}$. It is clear that both $Q_1(0)$ and $Q_2(0)$ are negative. However, at $\sqrt{x(1-x)}$, we find that,

$$
\begin{aligned}
Q_1\left(\sqrt{x(1-x)}\right) &= \frac{1 - 2x - \gamma\sqrt{x(1-x)}}{\sqrt{\kappa + 2 - \gamma^2}}; \\
Q_2\left(\sqrt{x(1-x)}\right) &= \frac{2x - 1 + \gamma\sqrt{x(1-x)}}{\sqrt{\kappa + 2 - \gamma^2}} = -Q_1\left(\sqrt{x(1-x)}\right).
\end{aligned}
$$

This indicates that one of the constraints is redundant given x.

Case 1: $x \le \frac{1}{2}\left(1 - \frac{\gamma}{\sqrt{\gamma^2+4}}\right)$

When $\gamma \ge 0$, we have

$$
Q_1\left(\sqrt{x(1-x)}\right) \ge Q_1\left(\frac{1}{\sqrt{\gamma^2+4}}\right) = 0.
$$

When $\gamma < 0$, we have

$$
Q_1\left(\sqrt{x(1-x)}\right) \ge \begin{cases} 0, & x \le \dfrac{1}{2} \\ Q_1\left(\dfrac{1}{\sqrt{\gamma^2+4}}\right) = 0, & x \ge \dfrac{1}{2}. \end{cases}
$$

Therefore, in this case, $Q_1\left(\sqrt{x(1-x)}\right) \ge 0$ while $Q_2\left(\sqrt{x(1-x)}\right) = -Q_1\left(\sqrt{x(1-x)}\right) \le 0$, which means the second constraint in (14) is redundant. Consequently, the optimal y^* is the solution of equation

$$
Q_1(y) = \sqrt{x(1-x) - y^2}
$$

which is exactly the largest real root of the following quartic equation that aligns with $y(x) = f(x, \gamma)$ as defined in Theorem 2.

$$
Q_1(y)^2 + y^2 - x(1-x) = 0.
$$

Case 2: $x > \frac{1}{2}\left(1 - \frac{\gamma}{\sqrt{\gamma^2+4}}\right)$.

The analysis parallels Case 1, and the final quartic equation will be

$$Q_2(y)^2 + y^2 - x(1-x) = 0.$$

Noticing that this equation is equivalent to that in Case 1 by substituting x with $1-x$ and γ with $-\gamma$, we have $y(x) = f(1-x, -\gamma)$.

Lastly, we establish the concavity of $y(x)$. Since x becomes variable in following discussion, we denote the right hand side of each constraint in (14) as $Q_1(x,y)$ and $Q_2(x,y)$, respectively and denote the right hand side of each constraint in (14) as $C(x,y)$. It is not hard to see that Q_1 and Q_2 are convex with respect to (x,y) and C is concave with respect to (x,y). Therefore, for any $\alpha \in (0,1)$ and two different x_1 and x_2, $(x_0, \alpha y(x_1) + (1-\alpha)y(x_2))$ is also feasible for constraints (14). Thus $y(x_0) \geq \alpha y(x_1) + (1-\alpha)y(x_2)$, which means $y(x)$ is concave. The initial objective function's concavity results from its separability. □

The proof utilizes the reformulation from Theorem 1. The key idea of the analysis is to justify that only one of the first two constraints in subproblem with respect to index i is active. For the detailed expression of the function $y_{\gamma_i,\kappa_i}(x)$, please refer to the Appendix A. Furthermore, this function is both continuous and differentiable, with its derivative also admitting a closed form expression, as demonstrated below. To simplify notation, we shall abbreviate $y_{\gamma_i,\kappa_i}(x)$ as $y_i(x)$ and suppress the subscript i.

PROPOSITION 1 **(Differentiability of the Objective Function)**. *When $\kappa - \gamma^2 + 2 > 0$, the function $y(x)$, as defined in Theorem 2, is continuous and differentiable on the interval $(0,1)$. Moreover, the derivative has the following closed-form expression*

$$y'(x) = \begin{cases} H(x,\gamma) & x \leq \dfrac{1}{2}\left(1 - \dfrac{\gamma}{\sqrt{\gamma^2+4}}\right), \\ -H(1-x,-\gamma), & x > \dfrac{1}{2}\left(1 - \dfrac{\gamma}{\sqrt{\gamma^2+4}}\right), \end{cases} \tag{15}$$

where

$$H(x,\gamma) = \frac{y(x)}{x} + \frac{y(x)^2 - \gamma x y(x) - x^2}{2\gamma(1-x)x^2 - 2y(x)(x(2-3x) - y(x)^2)}.$$

In addition, $\lim_{x\to 0} y'(x) = +\infty$ *and* $\lim_{x\to 1} y'(x) = -\infty$.

The formulation in (11) shares several similarities with the MV formulation (9). First, the objective functions in both formulations are convex with respect to $\boldsymbol{x}$. Second, the constraints in both cases are identical, corresponding to the convex hull of $\mathcal{X}$. Third, the coefficients μ_i and σ_i appear exclusively in the objective function. Fourth, the objective functions possess closed-form derivatives, making them suitable for first-order optimization algorithms, such as gradient descent (Boyd and Vandenberghe 2004). Finally, the derivatives in both cases tend to infinity as $\boldsymbol{x}$ approaches the boundary, implying that the optimal solutions lie within the interior of the feasible region.

3.2. Incomplete Moment Information

In this subsection, we extend the results for 4-MMM to encompass scenarios where certain lower-order moment information is absent. Specifically, we examine two types of moment ambiguity sets: one considering the marginal mean, second-order central moment, and fourth-order central moment (referred to as Θ_{124}^C), and another with only the marginal mean and the fourth-order central moment (referred to as Θ_{14}^C). The definitions of these two ambiguity sets are presented below:

$$\begin{aligned}
\Theta_{124}^C &= \left\{\theta : \theta_i \subset \mathcal{P}(\mathbb{R}), E[\tilde{c}_i] = \mu_i, E\left[(\tilde{c}_i - \mu_i)^2\right] = \sigma_i^2, E\left[(\tilde{c}_i - \mu_i)^4\right] = \sigma_i^4(\kappa_i + 3)\right\}, \\
\Theta_{14}^C &= \left\{\theta : \theta_i \subset \mathcal{P}(\mathbb{R}), E[\tilde{c}_i] = \mu_i, E\left[(\tilde{c}_i - \mu_i)^4\right] = m'_{i,4}\right\}.
\end{aligned} \tag{16}$$

Similar ambiguity sets has been studied in literature. He et al. (2010) has studied the moment ambiguity sets knowing first, second and fourth moment. Das et al. (2021) and Guo et al. (2022) study moment ambiguity set prescribed by the mean and fourth order moment. Here we highlight that Θ_{124}^C and Θ_{14}^C in this section differ from the literature by studying the centralized moment instead of the raw moment. We refer to the moment models considering the ambiguity sets Θ_{124}^C or Θ_{14}^C as *124-MMM* or *14-MMM*, respectively. We start with the results on *124-MMM*.

THEOREM 3. *The* 124*-MMM problem can be reformulated as the following SOCP:*

$$\begin{aligned}
Z_{124-MMM}^* = \max_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}} \quad & \sum_{i=1}^n (\mu_i x_i + \sigma_i y_i) \\
s.t. \quad & y_i^2 \le x_i \left(x_i + \sqrt{\kappa_i + 2} z_i\right) && \forall i \in \{1, \cdots, n\}, \\
& y_i^2 \le (1 - x_i)\left(1 - x_i + \sqrt{\kappa_i + 2} z_i\right) && \forall i \in \{1, \cdots, n\}, \\
& x_i^2 + z_i^2 \le x_i && \forall i \in \{1, \cdots, n\}, \\
& y_i \le z_i && \forall i \in \{1, \cdots, n\} \\
& \boldsymbol{A}(x_1, x_2, \cdots, x_n) \le \boldsymbol{b}.
\end{aligned} \tag{17}$$

Furthermore, it is equivalent to the following convex optimization problem:

$$\begin{aligned}
Z_{124-MMM}^* = \max_{\boldsymbol{x}} \quad & \sum_{i=1}^n (\mu_i x_i + \sigma_i y_{\kappa_i}(x_i)) \\
s.t. \quad & \boldsymbol{A}(x_1, x_2, \cdots, x_n) \le \boldsymbol{b},
\end{aligned} \tag{18}$$

where $y_{\kappa_i}(x_i)$ is a continuous and concave function defined as follows

$$y_\kappa(x) = \begin{cases}
\sqrt{x^2 + x\sqrt{x(1-x)(\kappa+2)}}, & x < \frac{1}{2}\left(1 - \sqrt{\frac{\kappa+2}{\kappa+6}}\right); \\
\sqrt{x(1-x)}, & \frac{1}{2}\left(1 - \sqrt{\frac{\kappa+2}{\kappa+6}}\right) \le x \le \frac{1}{2}\left(1 + \sqrt{\frac{\kappa+2}{\kappa+6}}\right); \\
\sqrt{(1-x)^2 + (1-x)\sqrt{x(1-x)(\kappa+2)}}, & x > \frac{1}{2}\left(1 + \sqrt{\frac{\kappa+2}{\kappa+6}}\right).
\end{cases}$$

In addition, the persistency of variable c_i is x_i in the optimal solution.

Theorem 3 provides both an SOCP reformulation and a convex optimization interpretation for 124-MMM, analogous to those of 4-MMM. The following theorem establishes the convex optimization interpretation for 14-MMM.

THEOREM 4. *The 14-MMM problem can be reformulated as the following convex optimization problem*

$$\begin{aligned} Z^*_{14-MMM} = \max \quad & \sum_{i=1}^{n}\left(\mu_i x_i + \left(\frac{m'_{i,4}x^3(1-x)^3}{1-3x(1-x)}\right)^{\frac{1}{4}}\right) \\ s.t. \quad & \boldsymbol{A}\left(x_1, x_2, \cdots, x_n\right) \leq \boldsymbol{b}. \end{aligned} \tag{19}$$

In addition, the persistency of variable c_i is x_i in the optimal solution.

Theorem 4 is derived based on Theorem 3 and the fact $\sigma^4(\kappa+3) = m'_4$. When $m'_4 > 0$, it is not hard to verify that the derivative of y_κ and $\left(\frac{m'_{i,4}x^3(1-x)^3}{1-3x(1-x)}\right)^{\frac{1}{4}}$ approaches $+\infty$ or $-\infty$ when x approaches 0 or 1, which is consistent with 4-MMM and MV.

Note that for all aforementioned ambiguity sets, the objective function with respect to index i in the convex optimization formulations (9), (2), (18), and (19) shares a common form $x_i\mu_i + Y_{\Theta,i}(x_i)$. Specifically, for the MV, 4-MMM, 124-MMM, and 14-MMM models, $Y_{\Theta,i}$ assumes the forms $\sigma_i\sqrt{x_i(1-x_i)}$, $\sigma_i y_{\gamma_i,\kappa_i}$, $\sigma_i y_{\gamma_i}$, and $\left(\frac{m'_{i,4}x^3(1-x)^3}{1-3x(1-x)}\right)^{\frac{1}{4}}$, respectively. To compare different models, Figure 1 depicts the function $Y_{\Theta,i}$ for the 4-MMM, 124-MMM, 14-MMM, and MV models when $\mu = 0$, $\sigma = 1$, $\gamma = 1$, $\kappa = 0$, and $m'_{i,4} = 3$. The $Y_{\Theta,i}$ curves for the 4-MMM and 124-MMM models lie below those of the MV and 14-MMM, indicating tighter bounds when additional moments are included. Additionally, the function $Y_{\Theta_4,i}$ is no longer symmetric, unlike the other models, due to the incorporation of skewness information. Moreover, the objective curve in the 124-MMM model is not always continuously differentiable, differing from the other cases.

3.3. Connection with the Marginal Distribution Model

A model closely related to the moment models is the *marginal distribution model* (MDM) (see, for instance, Natarajan et al. 2009, Mishra et al. 2014), wherein the assumption is that we have knowledge of the marginal distributions of random variables while allowing for an arbitrary correlation structure. MDM is a special case of the well-known Fréchet problem (Hoeffding 1940, Fréchet 1951), which concerns the aggregation of several random variables with information only on the marginal distribution of each individual random variable. The earliest investigations into MDM were conducted by Nadas (1979), Meilijson and Nádas (1979), Klein Haneveld (1986) in the context of project management problems. They proposed to determine the worst-case expected completion time through the solution of a convex optimization problem.

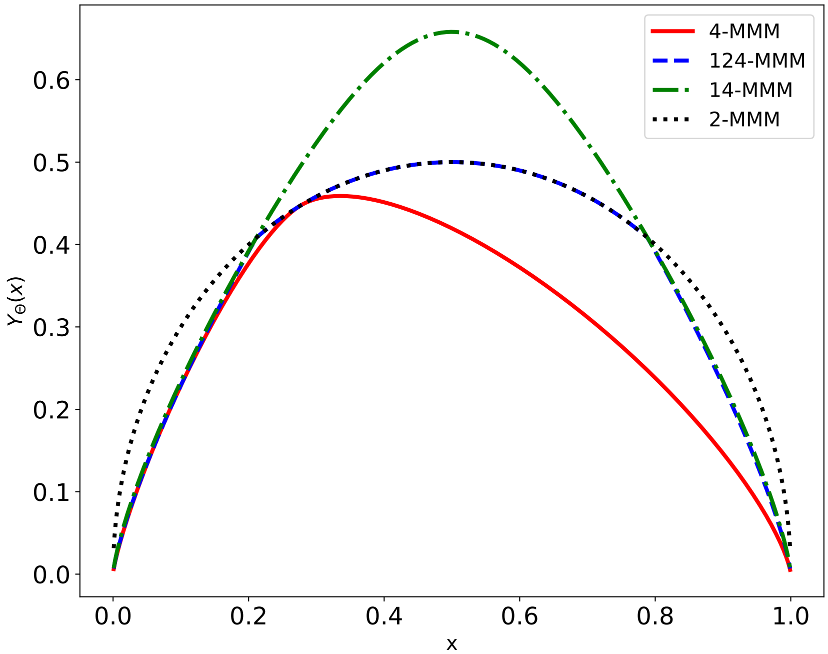


Figure 1 Objectives of different marginal moment models

If the marginal distribution is known precisely, the MDM can be represented as:

$$\begin{aligned} Z^*_{MDM} &= \sup_{\theta\in\Theta} E_\theta\left(Z_{\max}(\tilde{\boldsymbol{c}})\right) \\ \text{where } \Theta &= \left\{\theta : \tilde{c}_i \sim F_i(c), \forall i = 1, ..., n\right\}, \end{aligned} \tag{20}$$

where F_i denotes the marginal distribution function of $\tilde{c}_i$. The distribution set Θ encompasses all conceivable multivariate joint distributions with prescribed marginal distributions. Natarajan et al. (2009) demonstrated that the MDM is equivalent to a concave maximization problem expressed as follows:

$$\begin{aligned} Z^* = \max & \sum_{i=1}^{n}\left(\int_{1-x_i}^{1} F_i^{-1}(t)\mathrm{d}t\right) \\ \text{s.t. } & \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}. \end{aligned} \tag{21}$$

From the modeling standpoint, MDM requires complete information on the marginal distributions, while MMM necessitates only the marginal moments. However, Natarajan et al. (2009) demonstrated that the MV is equivalent to a special case of MDM (see 21) when the marginal distributions are t-distributions. This result establishes the first connection between MDM and MMM and provides an alternative interpretation for the solution produced by MV.

Here, building upon Theorem 2, we demonstrate that 4-MMM, 124-MMM, 14-MMM can all be regarded as special cases of MDM with specific marginal distributions.

Proposition 2. *The aforementioned three marginal moment models are equivalent to MDM (20) with the marginal inverse CDF of random variable $\tilde{c}_i$ defined as follows.*

- *For 4-MMM:*

$$F_i^{-1}(x) = \mu_i + \sigma_i y'_{\gamma_i,\kappa_i}(1-x).$$

- *For 124-MMM:*

$$F_i^{-1}(x) = \begin{cases} \mu - \dfrac{\sigma y_\kappa(x)}{x} + \sigma \dfrac{y_\kappa(x)^2 - x^2}{4x(1-x)y_\kappa(x)}, & x < \dfrac{1-\rho_\kappa}{2}, \\ \mu + \dfrac{2x-1}{2\sqrt{x(1-x)}}, & \dfrac{1-\rho_\kappa}{2} \le x < \dfrac{1+\rho_\kappa}{2}, \\ \mu + \dfrac{\sigma y_\kappa(x)}{1-x} - \sigma \dfrac{y_\kappa(x)^2 - (1-x)^2}{4x(1-x)y_\kappa(x)}, & x \ge \dfrac{1+\rho_\kappa}{2}, \end{cases}$$

where $\rho_\kappa = \sqrt{(\kappa+2)/(\kappa+6)}$.

- *For 14-MMM:*

$$F_i^{-1}(x) = \mu + \frac{3(2x-1)(1-2x(1-x)){m'_4}^{\frac{1}{4}}}{4(1-x)^{\frac{1}{4}}x^{\frac{1}{4}}(1-3x(1-x))^{\frac{5}{4}}}.$$

Proof of Proposition 2: Given the separability of objectives in both MMM (in convex optimization forms) and MDM, alongside identical constraints, it suffices to show that for any specific i and feasible x_i, the objective's component corresponding to x_i aligns in both models.

As the objectives of both 4-MMM and 14-MMM in the convex optimization form are continuously differentiable, a direct comparison of the formulations in (11) and (19) with (21) illustrates the equivalence. Specifically, the inverse CDF in MDM mirrors the derivative of MMM's objective at $1-x_i$.

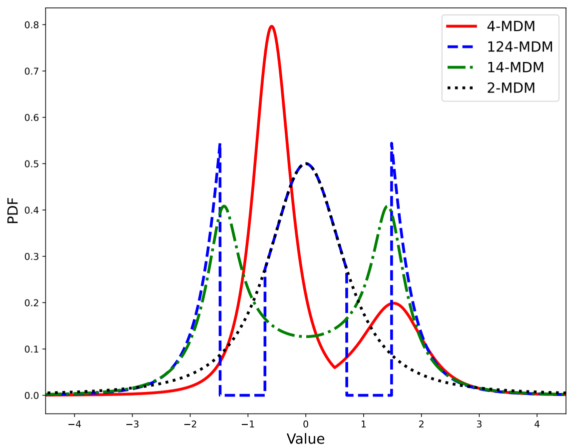
For 124-MMM model, the objective of (18) is continuously differentiable over three intervals divided by the points at $\frac{1}{2}\left(1 \pm \sqrt{\frac{\kappa_i+2}{\kappa_i+6}}\right)$ albeit subdifferentiable at these two points. Therefore, we have

$$\mu_i x_i + \sigma_i y_{\kappa_i}(x_i) = \int_{1-x_i}^{1} \left(\mu_i + \sigma_i \hat{y}'_{\kappa_i}(t)\right) \mathrm{d}t,$$

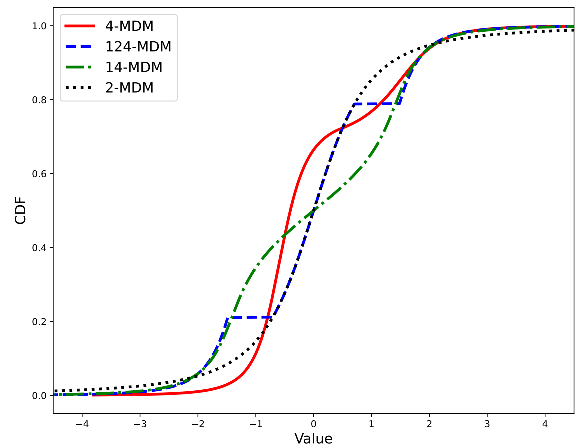
where $\hat{y}'_{\kappa_i}(t)$ can take any value at $\frac{1}{2}\left(1 \pm \sqrt{\frac{\kappa_i+2}{\kappa_i+6}}\right)$ and $\hat{y}'_{\kappa_i}(t) = y'_{\kappa_i}(t)$ at any other points in $(0,1)$. Since the CDF of a random variable is a right-continuous, monotone increasing function, we let the value of $\hat{y}'_{\kappa_i}$ to be the left derivative of y_κ at points $\frac{1}{2}\left(1 - \sqrt{\frac{\kappa_i+2}{\kappa_i+6}}\right)$ such that $\hat{y}'_{\kappa_i}(t)$ is a valid inverse CDF.

The final results are available by computing the derivative and sub-derivative. □

With a slight abuse of notation, we note the MDM corresponding to 4-MMM, 124-MMM, 14-MMM, and MV as 4-MDM 124-MDM, 14-MDM, and 2-MDM, respectively. Figure 2 shows the probability density function (PDF) and the cumulative distribution function (CDF) of the marginal distribution in aforementioned MDMs when the choice of μ, σ, γ and κ is the same as that in Figure 1.



(a)



(b)

Figure 2 PDFs and CDFs of marginal distributions in different MDMs

As shown in the figure, the distribution in 4-MDM is an asymmetric bimodal distribution and the CDF of 14-MDM is a symmetric bimodal distribution. The CDF of 124-MDM is a symmetric but discontinuous distribution, encompassing three continuous sections. Due to the incorporation of higher order moment information, 4-MDM, 124-MDM and 14-MDM have thinner tails than 2-MDM.

It is worth noting that Proposition 2 does not imply that the worst-case distribution in MMM coincides with the distribution outlined in Proposition 2. In fact, these distributions do not even belong to the ambiguity sets for the corresponding MMM; that is, the marginal moments of these distributions are likely to differ from the prescribed marginal moments. Nevertheless, for each pair of MMM and MDM, they yield the same optimal objectives and solutions, regardless of the feasible region $\mathcal{X}$.

3.4. Extention to the Analysis of Conditional Value at Risk

In many instances, statistical characteristics of the objective value $Z_{\max}(\tilde{\boldsymbol{c}})$, beyond mere expectation, are important. A prevalent strategy in this context involves minimizing the risk measure, commonly linked to the likelihood of unfavorable outcomes. The conditional value at risk (CVaR) is one of the most popular risk measures in risk management. CVaR quantifies the mean value over designated intervals in scenarios worse than a specific threshold. This section will illustrate how, with a slight modification, our model can be extended to minimize the worst-case CVaR.

Definition 1. The β-VaR and β-CVaR of a random variable $\tilde{z}$ which follows distribution θ are defined as:

$$\mathrm{VaR}_\beta^\theta(\tilde{z}) \stackrel{\text{def}}{=} \inf\{z \in \mathbb{R}; P_\theta[\tilde{z} \le z] > \beta\};$$
$$\mathrm{CVaR}_\beta^\theta(\tilde{z}) \stackrel{\text{def}}{=} \frac{1}{1-\beta}\int_{1-\beta}^1 \mathrm{VaR}_t^\theta(\tilde{z})dt.$$

Given the aforementioned moment based ambiguity set Θ, we would like to consider the worst-case CVaR, that is,

$$\sup_{\theta \in \Theta} \mathrm{CVaR}_\beta^\theta(Z_{\max}(\tilde{\boldsymbol{c}})). \tag{22}$$

The following proposition introduces a optimization approach to assess the worst-case CVaR (22).

Proposition 3. *The worst-case β-CVaR of $Z_{\max}(\tilde{\boldsymbol{c}})$ with ambiguity set $\Theta \in \{\Theta_4, \Theta_{124}^C, \Theta_{14}^C\}$ is given by the following optimization problem.*

$$\begin{aligned} \max \quad & \sum_{i=1}^n \left(x_i\mu_i + \frac{Y_{\Theta,i}((1-\beta)x_i)}{1-\beta} \right) \\ s.t. \quad & \boldsymbol{Ax} \le \boldsymbol{b}. \end{aligned} \tag{23}$$

The proof of this proposition is presented in Appendix B. This proposition illustrates that the only distinction between computing CVaR and calculating the expectation lies in the rescaling of the objective and the convex hull constraint. Consequently, the benefits of our reformulation will be largely applicable to CVaR scenarios as well.

4. Application: Project Crashing

In this section, we apply the theoretical results developed in Section 3 to project crashing, an important problem in project management. In project management, estimating the completion time (or makespan) of a project, which consists of multiple activities each with random durations, is crucial (Elmaghraby 1977). Even when the durations are assumed to follow simple distributions, the expected completion time may not be easy to compute. For instance, Hagstrom (1988) demonstrated that computing the expected project makespan is #P-hard when the activity durations are independent discrete random variables. A prevalent method to compute this expectation is through the use of Monte Carlo simulation methods (Van Slyke 1963, Burt Jr and Garman 1971).

Moreover, in practical scenarios, the underlying distribution for project duration is typically not available. Instead, estimating the moment information from historical data can be straightforward (see, e.g., Williams 1992, Chapman and Ward 2003). Hence, distributionally robust optimization has been embraced to examine the worst-case expected makespan based on moment information in project management. References such as Birge and Maddox (1995), Bertsimas et al. (2004, 2006a), Natarajan et al. (2011), and Natarajan and Teo (2017) have leveraged these approaches, addressing both the intractability of the problem and the ambiguity of the project duration distribution.

In addition to evaluating the makespan, another important question is how to shorten the project makespan by allocating resources to reduce the time of some activities. This task is called the project crashing. In the literature, robust optimization methods were typically applied to solve such problems (see, e.g., Chen et al. 2007, Cohen et al. 2007, Wiesemann et al. 2012) where the activity durations are assumed in some uncertainty sets. However, robust optimization may be overly conservative as it focuses on the worst-case scenario only. Also, estimating the uncertain parameters in the uncertainty set is usually challenging.

Instead of planning based on the worst-case makespan, several attempts have been made to allocate resources such that the expectation of the makespan is reduced under the worst-case distribution in a distributional ambiguity set prescribed by some statistical characteristics of the random variables. This is referred to as the distributionally robust project crashing problem. Goh and Hall (2013) proposed an ambiguity set based on the support, mean, and correlation of the project durations, and approximated the corresponding distributionally robust project crashing problem using linear and piecewise linear decision rules. Ahipasaoglu et al. (2019) studied the distributionally robust project crashing problem with mean, variance, and partial correlation information.

While prior literature primarily focuses on the covariance of activities, in this section, we study the distributionally robust project crashing problem with higher-order marginal moments. This approach is particularly relevant in scenarios where individual activity information is more readily accessible than covariance information, a situation frequently encountered when activity data are collected from different projects or at varied frequencies or time scale.

4.1. Solution approaches

Assume that the project is represented by an activity-on-arc network $\mathcal{G}(\mathcal{V},\mathcal{E})$ where $V=\{1,\cdots,n\}$ is the set of nodes denoting the events and $\mathcal{E}\in\{(j,k): j,k\in\mathcal{V}\}$ is the set of edges denoting the activities. The total number of edges is $|\mathcal{E}|=m$. Each edge (j,k) has a random length $\tilde{t}_{j,k}$, which denotes the duration of the corresponding activity. The completion time of the project is equal to the length of the longest path (also referred to as the critical path) of the project network $\mathcal{G}(\mathcal{V},\mathcal{E})$ from node 1 to node n. We use $\boldsymbol{x}=(x_{j,k})_{(j,k)\in\mathcal{E}}\in\{0,1\}^m$ to denote any possible path with $x_{j,k}$ equal to 1 if and only if the activity is included in the critical path. The set of feasible $\boldsymbol{x}$ can be represented as $\mathcal{X}=CH(\mathcal{X})\cap\{0,1\}^m$ where $CH(\mathcal{X})$ is the convex hull of the feasible region $\mathcal{X}$, defined as (see Natarajan et al. 2009):

$$CH(\mathcal{X})=\left\{\boldsymbol{x}: \sum_{k:(j,k)\in\mathcal{E}} x_{j,k}-\sum_{k:(k,j)\in\mathcal{E}} x_{k,j}=\left\{\begin{array}{l} \ \ 1,\ j=1 \\ \ \ 0,\ j=2,3,\cdots,n-1, \\ -1,\ j=n \end{array}\right. \quad x_{j,k}\in[0,1],\forall(j,k)\in\mathcal{E}\right\}. \tag{24}$$

Assuming that the mean, variance, skewness, and kurtosis of each $\tilde{t}_{j,k}$ are known as $\mu_{j,k}$, $\sigma_{j,k}$, $\gamma_{j,k}$, and $\kappa_{j,k}$, we denote the fourth-order moment ambiguity set as

$$\Theta(\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\gamma},\boldsymbol{\kappa}) = \left\{ \theta : \begin{array}{rl} E_\theta\left[\tilde{t}_{ij}\right] &= \mu_{j,k}, \\ V_\theta\left[\tilde{t}_{j,k}\right] &= \sigma_{j,k}^2, \\ S_\theta\left[\tilde{t}_{j,k}\right] &= \gamma_{j,k}, \\ K_\theta\left[\tilde{t}_{j,k}\right] &= \kappa_{j,k}, \end{array} \quad \forall (j,k) \in \mathcal{E} \right\}. \tag{25}$$

Now suppose that the project manager can influence activity completion time by allocating additional resources such as equipment, budget, and manpower. We assume that the resources can influence the mean and variance of each activities $\tilde{t}_{ij}$, which is a reasonable assumption in practical setting. For example, the manager can let the staff work on holidays or allocate more manpower to an activity by increasing staff expenditure. In the first case, working overtime on one unit time will decrease the completion time by a certain amount. In the latter case, assigning more staffs to the activity can improve efficiency proportionately and make the new completion time become $\frac{\alpha}{\alpha'}\tilde{t}_i$, where α and α' are the initial and new efficiency, respectively. In both cases, the skewness and kurtosis would not be affected.

We denote the feasible region for the means and variances of random lengths as $(\boldsymbol{\mu},\boldsymbol{\sigma}) \in \mathcal{M}$. One possible formulation for the distributionally robust project crashing problem can be to select the means and variances to minimize the total project makespans, which is stated as follows:

$$\min_{(\boldsymbol{\mu},\boldsymbol{\sigma})\in\mathcal{M}} \quad \sup_{\theta\in\Theta(\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\gamma},\boldsymbol{\kappa})} E_\theta\left[\max_{\boldsymbol{x}\in\mathcal{X}} \tilde{\boldsymbol{t}}^T\boldsymbol{x}\right]. \tag{26}$$

Alternatively, assuming that each targeted means and variances will incur a certain cost $c(\boldsymbol{\mu},\boldsymbol{\sigma})$, we can also consider a budget minimization problem subject to the worst-case expected completion time is less than T:

$$\begin{array}{ll} \min\limits_{(\boldsymbol{\mu},\boldsymbol{\sigma})\in\mathcal{M}} & c(\boldsymbol{\mu},\boldsymbol{\sigma}) \\ \text{s.t.} & \sup\limits_{\theta\in\Theta(\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\gamma},\boldsymbol{\kappa})} E_\theta\left[\max\limits_{\boldsymbol{x}\in\mathcal{X}} \tilde{\boldsymbol{t}}^T\boldsymbol{x}\right] \le T. \end{array} \tag{27}$$

For convenience, from this point, let N be the total number of edges, and we reindex the edges $(j,k) \in \mathcal{E}$ as $i = 1,...,N$. Following this notation, $\boldsymbol{x}$ now becomes $\{x_1,...,x_N\}$. Assume $CH(\mathcal{X}) = \{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}\}$. The inner problem of (26) and (27) can be written as an SOCP according to Theorem 1, which satisfies weak Slater's condition since we can always let $y_i = z_i = 0$. Therefore, the strong duality holds. By taking the duality of (10), we obtain the following reformulations.

Proposition 4. *Let $\mathcal{C}$ denote an SOC-representable set containing elements $(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w},\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\boldsymbol{\xi})$ which satisfy the following constraints:*

$$\begin{aligned}
&w_i^{(1)} - 2w_i^{(2)} - 2x_i + 2y_i - z_i + (\boldsymbol{A}^T\boldsymbol{\xi})_i = \mu_i, && \forall i \in \{1,\cdots,N\},\\
&\gamma_i(v_i^{(2)} - u_i^{(2)} + y_i - x_i) - 2(u_i^{(1)} + v_i^{(1)}) - 2w_i^{(3)} = \sigma_i, && \forall i \in \{1,\cdots,N\},\\
&\sqrt{\kappa_i + 2 - \gamma_i^2}(u_i^{(2)} + v_i^{(2)} + x_i + y_i) + 2w_i^{(4)} = 0, && \forall i \in \{1,\cdots,N\},\\
&\|\boldsymbol{u}_i\| \le x_i, && \forall i \in \{1,\cdots,N\},\\
&\|\boldsymbol{v}_i\| \le y_i, && \forall i \in \{1,\cdots,N\},\\
&\|\boldsymbol{w}_i\| \le z_i, && \forall i \in \{1,\cdots,N\},\\
&\boldsymbol{\xi} \ge \boldsymbol{0}.
\end{aligned} \tag{28}$$

Then the distributionally robust project crashing problem (26) is equivalent to

$$\begin{aligned}
\min_{\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{u},\boldsymbol{v},\boldsymbol{w},\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\boldsymbol{\xi}} \quad & \sum_{i=1}^{N}\left(w_i^{(1)} + 2y_i + z_i\right) + \boldsymbol{b}^T\boldsymbol{\xi},\\
s.t. \quad & (\boldsymbol{u},\boldsymbol{v},\boldsymbol{w},\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\boldsymbol{\xi}) \in \mathcal{C},\\
& (\boldsymbol{\mu},\boldsymbol{\sigma}) \in \mathcal{M}.
\end{aligned} \tag{29}$$

Similarly, the distributionally robust crash budget minimization problem (27) is equivalent to

$$\begin{aligned}
\min_{\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{u},\boldsymbol{v},\boldsymbol{w},\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\boldsymbol{\xi}} \quad & c(\boldsymbol{\mu},\boldsymbol{\sigma}),\\
s.t. \quad & \sum_{i=1}^{N}\left(w_i^{(1)} + 2y_i + z_i\right) + \boldsymbol{b}^T\boldsymbol{\xi} \le T,\\
& (\boldsymbol{u},\boldsymbol{v},\boldsymbol{w},\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\boldsymbol{\xi}) \in \mathcal{C},\\
& (\boldsymbol{\mu},\boldsymbol{\sigma}) \in \mathcal{M}.
\end{aligned} \tag{30}$$

When $\mathcal{M}$ is a second-order cone, the problem (29) becomes an SOCP. In addition, if cost function $c(\boldsymbol{\mu},\boldsymbol{\sigma})$ has an SOCP representation, (30) is also an SOCP. Next, we show that when $\mathcal{M}$ is a convex set, the distributionally robust project crashing problems can also be reformulated as the following convex optimization problems.

Proposition 5. *The distributionally robust project crashing problem (26) is equivalent to:*

$$\begin{aligned}
\min_{\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\sigma}} \quad & \sum_{i=1}^{n}\left(\sigma_i\varphi_i\left(\frac{\mu_i - \boldsymbol{a}_i^T\boldsymbol{\lambda}}{\sigma_i}\right) - \boldsymbol{b}^T\boldsymbol{\lambda}\right),\\
s.t. \quad & \boldsymbol{\lambda} \ge 0,\\
& (\boldsymbol{\mu},\boldsymbol{\sigma}) \in \mathcal{M},
\end{aligned} \tag{31}$$

where $\varphi_i(\cdot)$ is the convex conjugate of the $-y_i(\cdot)$ and $\boldsymbol{a}_i$ is the i-th column vector of $\boldsymbol{A}$. Similarly, the distributionally robust crash budget minimization problem (27) is equivalent to

$$\begin{aligned}
\min_{\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\sigma}} \quad & c(\boldsymbol{\mu},\boldsymbol{\sigma}),\\
s.t. \quad & \sum_{i=1}^{n}\left(\sigma_i\varphi_i\left(\frac{\mu_i-\boldsymbol{a}_i^T\boldsymbol{\lambda}}{\sigma_i}\right)-\boldsymbol{b}^T\boldsymbol{\lambda}\right)\leq T\\
& \boldsymbol{\lambda}\geq 0,\\
& (\boldsymbol{\mu},\boldsymbol{\sigma})\in\mathcal{M}.
\end{aligned}\tag{32}$$

Proposition 4 and 5 demonstrate that the distributionally robust project crashing problem under the first four marginal moments are almost as easy as the corresponding mean-variance model. If some moment information is missing, similar results can be obtained based on Theorem 3 and 4.

4.2. Numerical study

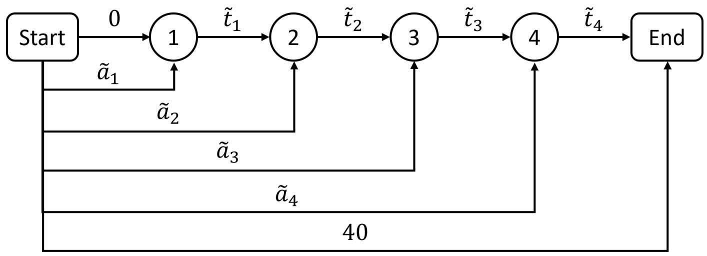
Subsequently, we consider a service system as an example to examine the performance of our proposed models. A similar problem has been studied by Mak et al. (2015), who optimize the appointment schedule to minimize total waiting and overtime costs with information on marginal means and variances. In contrast, we aim to allocate additional resources to the existing schedule to reduce the final delay given marginal moments up to fourth degree.

Assuming the system consists of four individuals who are scheduled to arrive sequentially at times 0, 10, 20, and 30, each requiring a service duration of 10 units. Any service begins upon a client's arrival and the completion of the previous client's service. Additionally, the first service can only begin after time 0. If the system adheres to the predetermined schedule, the service completion time will be 40 units, with no delay.

We now introduce random perturbations to arrival times and service times. Assume that the perturbation terms for arrival and service times are identical and independent triangular distributions. Specifically, let the perturbation term be defined as $s(\varepsilon_m-(1+m)/3)$, where s represents the scale parameter and ε_m follows a standard triangular distribution with a minimum of 0, a maximum of 1, and a mode of m. The inclusion of the term $-(1+m)/3$ serves to standardize the mean of the perturbation to 0. With the inclusion of random perturbations, the completion time, denoted by $\tilde{T}$, becomes a random variable. Our focus is now on the system's expected delay compared to the scheduled completion time, represented by $(\tilde{T}-40)_+$. The system is illustrated in Figure 3, where $\tilde{a}_i$ and $\tilde{t}_i$ represent the random arrival and service times of client i.

We first approximate the expected delay by utilizing three different methods: Monte Carlo simulation (with one million samples), 2MMM and 4MMM. Then we set the target expected delay as half of the mean delay obtained from the Monte Carlo simulation. We assume that the decision

Figure 3 Service system with four clients



maker is able to improve the effeciency with efforts. In particular, the new service time becomes $(1-\alpha_i)\tilde{t}_i$, with different cost functions presented in Table 1. Note that all these cost functions have SOCP representations. The performance is measured by the budget ratio between 4-MMM and MV (indicated as B4/B2 in Table 2) in order to accomplish the target expected completion time T.

Table 1 Cost functions

Cost function	C_1	C_2	C_4	C_I	C_{I2}
Expression	$\sum_{i=1}^4 \alpha_i$	$\sum_{i=1}^4 \alpha_i^2$	$\sum_{i=1}^4 \alpha_i^4$	$\sum_{i=1}^4 \frac{\alpha_i}{1-\alpha_i}$	$\sum_{i=1}^4 \left(\frac{\alpha_i}{1-\alpha_i}\right)^2$

We conducted tests with different scales $(2,3,4,5)$ and modes $(0.2,0.5,0.8)$. The resulting estimated expected delays and budget ratios are presented in Table 2.

Table 2 Estimated expected delays and budget ratios under different cost functions

		Estimated Delay*			B4/B2				
Mode	Scale	MCS	4-MMM	MV	C_1	C_2	C_4	C_I	C_{I2}
0.2	2	0.704	1.291	1.477	0.590	0.303	0.089	0.527	0.268
	3	1.057	1.936	2.216	0.606	0.322	0.100	0.525	0.271
	4	1.408	2.582	2.954	0.621	0.340	0.112	0.524	0.273
	5	1.760	3.227	3.693	0.634	0.356	0.123	0.524	0.275
0.5	2	0.650	1.200	1.396	0.514	0.235	0.055	0.456	0.205
	3	0.976	1.800	2.094	0.530	0.251	0.062	0.454	0.208
	4	1.300	2.400	2.792	0.544	0.267	0.070	0.453	0.210
	5	1.623	3.001	3.489	0.557	0.281	0.078	0.453	0.211
0.8	2	0.673	1.204	1.477	0.426	0.163	0.027	0.370	0.138
	3	1.010	1.806	2.216	0.442	0.176	0.031	0.368	0.139
	4	1.348	2.408	2.954	0.456	0.189	0.035	0.367	0.140
	5	1.685	3.010	3.693	0.470	0.201	0.040	0.367	0.141

* For the original system

In Table 2, the left part details the mode and scale parameters for random perturbations. For each pair of parameters, the middle section displays the estimated delays for the original system derived from different methods and the right section outlines the budget ratios associated with various cost functions. This table demonstrates that 4-MMM provides tighter bounds on expected delay than MV. Moreover, the Monte Carlo simulation results reveal that mode 0.2 incurs greater delay than mode 0.8. This property is captured by 4-MMM but not by MV, as their mean and variance remain the same. Consequently, in the crash budget minimization problem, 4-MMM requires significantly less budget than MV due to its better approximation. Even in the case of linear cost (C_1), 4-MMM achieves a budget reduction of approximately 40%. In the case of quartic cost (C_4), 4-MMM requires only 10% of the budget needed by MV.

5. Application: Newsvendor Problem

In this section, we study the distributionally robust newsvendor problem with moment information up to the fourth degree. The newsvendor problem has been extensively studied since its introduction (see, e.g., Arrow et al. 1951, Arrow et al. 1958, and Iglehart 1963). While the mainstream setting of the newsvendor problem assumes that the demand distribution is known, an important stream of literature focuses on making the newsvendor decision with only moment information in a distributionally robust optimization (DRO) setting. Scarf (1958) considered all distributions having the same mean and variance, representing the earliest DRO newsvendor framework. Ben-Tal and Hochman (1976) extended this result to develop a closed-form optimal order quantity when the mean and mean absolute deviation are known. Natarajan et al. (2018) consider all distributions with a given mean, variance, and semi-variance. Ardestani-Jaafari and Delage (2016) studied a newsvendor problem with given mean and first-order partial moments (specifically, the conditional expectation of an upside or downside tail with respect to a threshold). Das et al. (2021) extended Scarf's model to a case where the ambiguity set is determined by the first and α-th moment. Govindarajan et al. (2021) studied the multilocation inventory polling problem with an ambiguity set determined by mean and covariance.

It is worth noting that, besides optimizing worst-case profit, distributionally robust regret minimization problems with moment information are also considered in the literature. Yue et al. (2006) assumed that the demand distribution belongs to the class of distributions with given mean and variance. Perakis and Roels (2008) extended this approach by incorporating additional demand information such as median, unimodality, and symmetry.

5.1. Closed-form solutions

Given the order quantity decision q and the random demand $\tilde{d}$, the newsvendor profit function is expressed as $p\min\{d,q\} - cq$, where p and c are the selling price and purchasing cost, respectively.

The seller's objective is to maximize the expected profit by determining the optimal order quantity q, that is

$$\max_{q} \quad \mathbb{E}\left[p\min\left\{\tilde{d}, q\right\} - cq\right]. \tag{33}$$

If the demand distribution is exactly known, the optimal order quantity equals the quantile of the so-called *critical ratio* $\eta = \frac{p-c}{p}$. More specifically, assume that the cumulative distribution function (CDF) of $\tilde{d}$ is F, the optimal order quantity is $q^* = \inf_q \{q : F(q) \geq \eta\}$.

In cases where the exact demand distribution is not known, distributionally robust optimization models that consider partial distribution information are widely applied. Denote the distributional ambiguity set, which contains all possible demand distributions, as Θ. The distributionally robust newsvendor (DR-NV) problem is formulated as:

$$\max_{q} \inf_{\theta\in\Theta} \quad E_\theta\left[p\min\left\{\tilde{d}, q\right\} - cq\right]. \tag{DR-NV}$$

The seminal (DR-NV) was introduced by Scarf (1958), which assumed that only the mean and variance of the demand are known and derived a succinct, closed-form solution for the optimal order quantity. This result is presented in the following proposition.

Proposition 6 **(Scarf 1958)**. *If Θ contains all distributions with a given mean μ and standard deviation σ, then the optimal solution to* (DR-NV) *is*

$$q_2^* = \mu + \sigma\frac{1-2\eta}{2\sqrt{\eta(1-\eta)}}. \tag{34}$$

The significance of this model lies not only in its computational efficiency but also in its managerial insights. It has two important implications: First, the order quantity is larger than the mean demand if the critical ratio η is greater than half and smaller than the mean otherwise. Second, for a fixed η, the order quantity is linear in both the mean μ and standard deviation σ.

Das et al. (2021) extended Scarf's classic result by considering the mean and α-order moment, where $\alpha > 0$ is a positive real number. Their primary focus was on the heavy-tail case where the variance may not exist. Guo et al. (2022) further extended this result and proposed a new model considering the mean and the exponential moment. Both of these papers focus mainly on the distributionally robust bound and the worst-case demand distribution.

In this section, we extend Scarf's result by incorporating moments up to the fourth order and provide closed-form solutions. In particular, we consider three different DR-NV models with ambiguity sets presented as follows.

1. 4-NV: $\Theta = \left\{\theta:\ E_\theta\left[\tilde{d}\right] = \mu,\ V_\theta\left[\tilde{d}\right] = \sigma^2,\ S_\theta\left[\tilde{d}\right] = \gamma,\ K_\theta\left[\tilde{d}\right] = \kappa\right\}$.

2. 124-NV: $\Theta = \left\{\theta: \ E_\theta\left[\tilde{d}\right] = \mu, \ V_\theta\left[\tilde{d}\right] = \sigma^2, \ K_\theta\left[\tilde{d}\right] = \kappa\right\}$.
3. 14-NV: $\Theta = \left\{\theta: \ E_\theta\left[\tilde{d}\right] = \mu, \ E_\theta\left[\left(\tilde{d} - \mu\right)^4\right] = m_4'\right\}$.

Note that the newsvendor objective involves taking the minimum of demand and inventory, which is a simple combinatorial optimization problem. We can convert the inner part of the newsvendor problem into 4-MMM, 124-MMM, and 14-MMM and obtain closed-form solutions under the DRO framework based on Theorems 2, 3, and 4. It turns out that the optimal newsvendor solutions are closely related to the MDM formulation developed in Section 3.3.

PROPOSITION 7 **(Closed-form solutions)**. *The optimal solutions to 4-NV, 124-NV and 14-NV are the same as the optimal solutions of the stochastic newsvendor problem with demand distributions defined in Proposition 2. Specifically, the closed-form optimal inventory level are presented as follows.*

1. *The optimal solution of 4-NV is*

$$q_4^* = \mu + \sigma y'(1-\eta), \tag{35}$$

where the closed-form expression of y' is given in Proposition 1.

2. *The optimal solution of 124-NV is*

$$q_{124}^* = \begin{cases} \mu - \dfrac{\sigma y_\kappa(\eta)}{\eta} + \sigma \dfrac{y_\kappa(\eta)^2 - \eta^2}{4\eta(1-\eta)y_\kappa(\eta)}, & \eta < \dfrac{1-\rho_\kappa}{2}, \\ \left[\mu - \dfrac{\sigma(\rho_\kappa^2 + \rho_\kappa + 1)}{\sqrt{1-\rho_\kappa^2}(1+\rho_\kappa)}, \mu - \dfrac{\sigma\rho_\kappa}{\sqrt{1-\rho_\kappa^2}}\right],^3 & \eta = \dfrac{1-\rho_\kappa}{2}, \\ \mu + \dfrac{2\eta - 1}{2\sqrt{\eta(1-\eta)}}\sigma, & \dfrac{1-\rho_\kappa}{2} < \eta < \dfrac{1+\rho_\kappa}{2}, \\ \left[\mu + \dfrac{\sigma\rho_\kappa}{\sqrt{1-\rho_\kappa^2}}, \mu + \dfrac{\sigma(\rho_\kappa^2 + \rho_\kappa + 1)}{\sqrt{1-\rho_\kappa^2}(1+\rho_\kappa)}\right], & \eta = \dfrac{1+\rho_\kappa}{2}, \\ \mu + \dfrac{\sigma y_\kappa(\eta)}{1-\eta} - \sigma \dfrac{y_\kappa(\eta)^2 - (1-\eta)^2}{4\eta(1-\eta)y_\kappa(\eta)}, & \eta > \dfrac{1+\rho_\kappa}{2}, \end{cases} \tag{36}$$

where the function y_κ is defined in Theorem 3 and $\rho_\kappa = \sqrt{(\kappa+2)/(\kappa+6)}$.

3. *The optimal solution of 14-NV is*

$$q_{14}^* = \mu + \frac{3(2\eta-1)(1-2\eta(1-\eta))m_4'^{\frac{1}{4}}}{4(1-\eta)^{\frac{1}{4}}\eta^{\frac{1}{4}}(1-3\eta(1-\eta))^{\frac{5}{4}}}. \tag{37}$$

[3] All values in the interval are optimal.

Proof of Proposition 7: We first proof the result for 4-NV. We assume, without loss of generality, that $p=1$, $p-c=\eta$. According to Theorem 2, the 4-NV problem can be reformulated as

$$\begin{aligned}
&\min_q \sup_{\theta\in\Theta} \quad E_\theta\left[\max\left\{-\tilde{d},-q\right\}\right]+(1-\eta)q \\
=&\min_q \max_{x\in[0,1]} \quad -(1-x)\mu+\sigma y_{-\gamma,\kappa}(1-x)-xq+(1-\eta)q \\
=&\min_q \max_{x\in[0,1]} \quad (\mu-q)x+\sigma y_{\gamma,\kappa}(x)+(1-\eta)q-\mu \\
=&\min_q \quad \sigma\varphi\left(\frac{\mu-q}{\sigma}\right)-(1-\eta)(\mu-q)-\eta\mu,
\end{aligned} \tag{38}$$

where φ represents the convex conjugate of $-y(x)$. The optimal q^* satisfies $\varphi'\left(\frac{\mu-q^*}{\sigma}\right)=1-\eta$. According to the property of convex conjugate, we have

$$\frac{\mu-q^*}{\sigma}=-y'\left(\varphi'\left(\frac{\mu-q^*}{\sigma}\right)\right)=-y'\left(1-\eta\right).$$

Therefore, $q_4^*=\mu+\sigma y'\left(1-\eta\right)$.

The results of q_{124}^* and q_{14}^* follow the same argument. For 124-NV, we only need to replace function $y(x)$ with $y_\kappa(x)$ and use the sub-derivative instead of the derivative. For 14-NV, we only need to replace $\sigma y(x)$ with $\left(\frac{x^3(1-x)^3m_4'}{1-3x(1-x)}\right)^{\frac{1}{4}}$.

Finally, it is not hard to verify that the optimal distributionally robust Newsvendor solutions are the same as the optimal solution of the stochastic newsvendor problem with demand distributions defined in Proposition 2. □

Proposition 7 provides closed-form solutions for the newsvendor problem under different ambiguity sets, enabling the derivation of new managerial insights. In Scarf's model, the order quantity exceeds the mean demand μ if the critical ratio is greater than 0.5, and is less if the ratio is smaller. This property also holds true for 124-NV and 14-NV. However, for 4-NV, when incorporating skewness information, this may not necessarily apply.

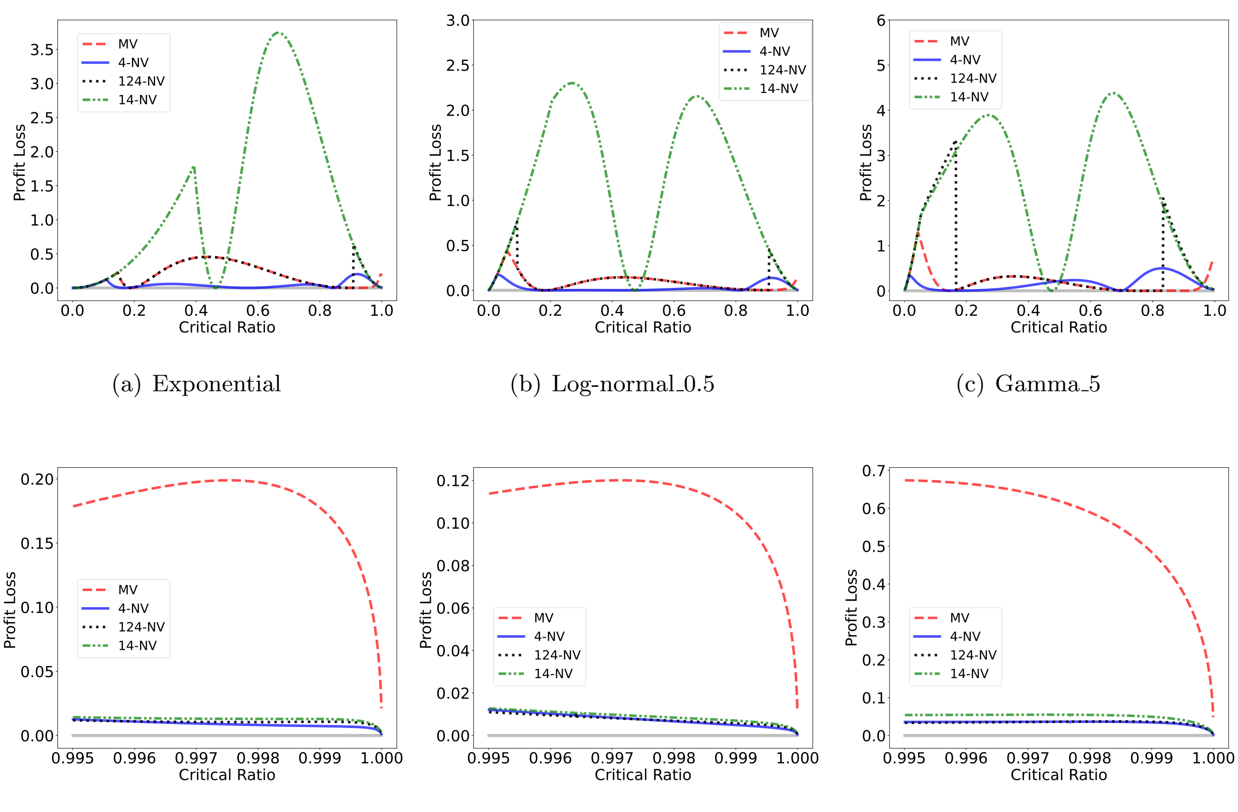
Let $\eta_0(\gamma,\kappa)$ represent the critical ratio where μ is exactly an optimal solution for the 4-NV, given skewness γ and kurtosis κ. Given that the function y in (35) is concave, the optimal order quantity in 4-NV increases with the critical ratio η. Clearly, η_0 serves as the threshold at which the optimal order quantity either exceeds or falls below μ. We now explore some theoretical properties related to η_0.

Proposition 8. *Given skewness $\gamma\le 0$, $\eta_0(\gamma,\kappa)$ increase from $\frac{1}{2}+\frac{\gamma}{2\sqrt{\gamma^2+4}}$ to $\frac{1}{2}$ as kurtosis κ increase from γ^2-2 to $+\infty$. Given skewness $\gamma\ge 0$, $\eta_0(\gamma,\kappa)$ decrease from $\frac{1}{2}+\frac{\gamma}{2\sqrt{\gamma^2+4}}$ to $\frac{1}{2}$ as kurtosis κ increase from γ^2-2 to $+\infty$.*

This proposition implies that regardless of the kurtosis, if the skewness $\gamma\le 0$ and the critical ratio $\eta\ge 0.5$, then $q_4^*\ge\mu$. Conversely, when the skewness $\gamma\ge 0$ and the critical ratio $\eta\le 0.5$, it follows that $q_4^*\le\mu$.

5.2. Numerical study

To demonstrate the benefits of incorporating higher-order moments information, we conduct numerical experiments to evaluate different distributionally robust newsvendor solutions against several ground-truth demand distributions. We assess our 4-NV, 124-NV, and 14-NV solutions in comparison with Scarf's MV model (34). The demand distributions selected for testing include exponential, log-normal (with shape parameter 0.5), and gamma (with shape parameter 5) distributions. If a model suggests a negative order quantity, we simply set the order quantity to 0. The product's price p is constant, and the cost varies from 0 to p, causing the critical ratio to shift from 1 to 0. The results are depicted in Figure 4.



(a) Exponential (b) Log-normal_0.5 (c) Gamma_5

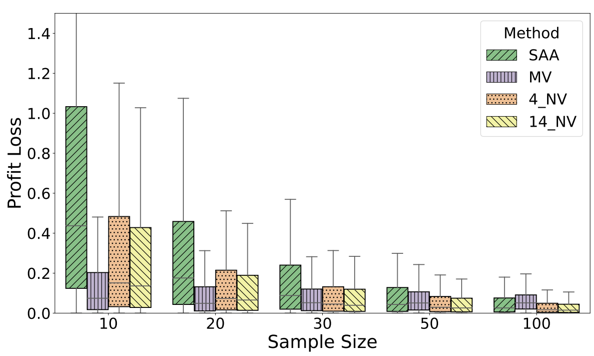
(d) Exponential (tail) (e) Log-normal_0.5 (tail) (f) Gamma_5 (tail)

Figure 4 Expected Profit loss under different moment based newsvendor solution

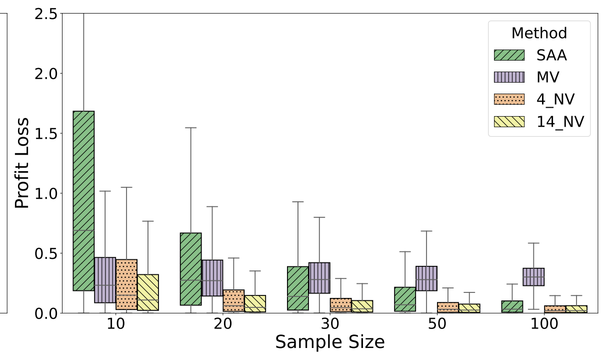
The upper part of Figure 4 depicts the profit loss relative to the optimal solution across the entire range of the critical ratio η, whereas the lower part focuses on the high service-level regime, namely when η approaches 1. In terms of overall performance, the 4-NV model excels compared to other models, likely due to its inclusion of higher-order moment information. It is notable that MV performs better than 14-NV in the mid-range of the critical ratio, despite both models incorporating information from two moments. This highlights the significance of incorporating variance information. The 124-NV model's solutions coincide with those of MV within the mid-range of the critical ratio; however, the performance of 124-NV is comparable to that of 14-NV as the critical ratio nears 0 or 1.

As the critical ratio approaches to one, the models 4-NV, 124-NV, and 14-NV significantly surpass MV in performance. This enhanced performance may be due to their ability to capture the tail behavior of distributions, which is more accurately captured by higher-order moments.

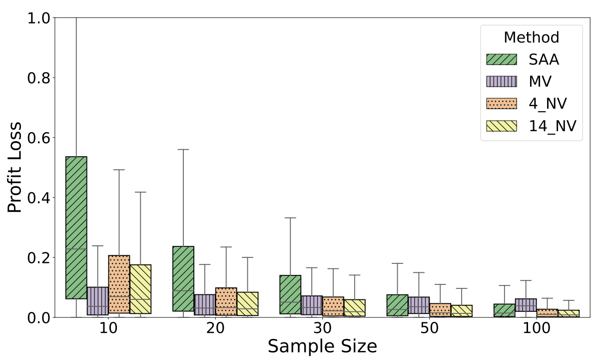
We also evaluate our proposed 4-NV solution in a data-driven context, comparing it with the MV model and the sample average approximation (SAA) method. Our focus is on the high service level regime where the critical ratio is set to 0.99. The assessment encompasses varying sample sizes from 10 to 100, with each scenario replicated 2000 times. For these evaluations, the underlying demand distributions are assumed to be exponential, log-normal (with shape parameter 0.5 and 1), and gamma (with shape parameter 5) distributions. The results of these tests are illustrated in the subsequent figure.



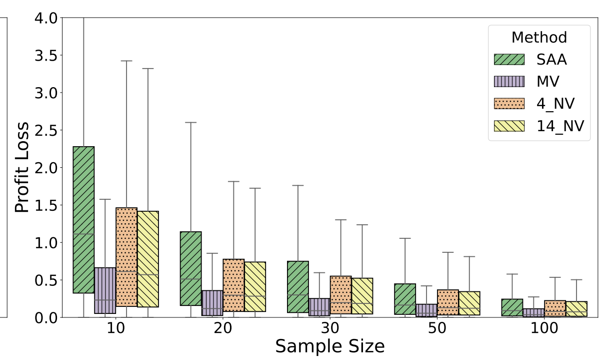
(a) Exponential



(b) Gamma_5



(c) Lognormal_0.5



(d) Log-normal_1

Figure 5 Expected Profit loss under different newsvendor solution

In Figure 5, the box plot illustrates the 25th percentile, median, and 75th percentile of the profit loss relative to the true optimal solution. This figure provides several insights. First, moment-based methods outperform the SAA method in high service level regimes. This advantage arises because estimating tail quantiles may require more samples than those near the distribution's mode (Zielinski 2004). The MV method shows benefits with small sample sizes, likely due to the relative simplicity of estimating the mean and variance. However, its performance becomes erratic with increasing sample sizes, alternating between the best and worst outcomes among the methods. Conversely, methods that incorporate the fourth-order moment, such as 4-NV and 14-NV, consistently outperform SAA across a range of sample sizes and generally surpass the MV method. An intriguing observation is that 14-NV often outperforms 4-NV, despite using fewer moments. This might be attributed to the increased estimation errors introduced by employing four estimated moments compared to two, indicating a potential trade-off between the number of utilized moments and the precision of moment estimation. In conclusion, within a high service level regime, the MV model is suitable for small data samples. For larger sample sizes, decision-makers should consider the 4-NV or 14-NV methods to better capture the tail behavior of the demand distribution. In scenarios with ample data, although not depicted in the figure, the SAA method becomes preferable, consistently converging to the optimal decision with sufficiently large sample sizes.

6. Conclusion

This paper introduces reformulations of the fourth-order marginal moment model into a as an SOCP and a convex optimization problem. These new formulations are notably more succinct than the existing SDP formulation, enhancing their ease of implementation and suitability for analytical exploration. Leveraging these reformulations, we further analytically study scenarios where information on the mean, variance, and fourth-order central moment is available, as well as cases where only the mean and fourth-order central moment are known. Through applications of project crashing and the newsvendor problem, we demonstrate the practicality and insightful nature of our reformulations. Specifically, in the project crashing problem, we illustrate that incorporating information on marginal higher-order moments such as skewness and kurtosis can be almost as easy as applying the mean-variance model. In the newsvendor problem, we derive closed-form solutions, which are challenging to obtain with existing methodologies.

References

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

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

Arrow KJ, Harris T, Marschak J (1951) Optimal inventory policy. *Econometrica: Journal of the Econometric Society* 250–272.

Arrow KJ, Karlin S, Scarf HE, et al. (1958) *Studies in the mathematical theory of inventory and production* (Stanford University Press).

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

Bertsimas D, Natarajan K, Teo CP (2004) Probabilistic combinatorial optimization: Moments, semidefinite programming, and asymptotic bounds. *SIAM Journal on Optimization* 15(1):185–209.

Bertsimas D, Natarajan K, Teo CP (2006a) Persistence in discrete optimization under data uncertainty. *Mathematical programming* 108(2):251–274.

Bertsimas D, Natarajan K, Teo CP (2006b) Tight bounds on expected order statistics. *Probability in the Engineering and Informational Sciences* 20(4):667–686.

Birge JR, Maddox MJ (1995) Bounds on expected project tardiness. *Operations Research* 43(5):838–850.

Birge JR, Wets RJB (1987) Computing bounds for stochastic programming problems by means of a generalized moment problem. *Mathematics of Operations Research* 12(1):149–162.

Boyd SP, Vandenberghe L (2004) *Convex optimization* (Cambridge university press).

Burt Jr JM, Garman MB (1971) Conditional monte carlo: A simulation technique for stochastic network analysis. *Management Science* 18(3):207–217.

Chapman C, Ward S (2003) *Project risk management processes, techniques and insights* (John Wiley & Sons Ltd,).

Chen L, He S, Zhang S (2011) Tight bounds for some risk measures, with applications to robust portfolio selection. *Operations Research* 59(4):847–865.

Chen X, Sim M, Sun P (2007) A robust optimization perspective on stochastic programming. *Operations research* 55(6):1058–1071.

Cohen I, Golany B, Shtub A (2007) The stochastic time–cost tradeoff problem: a robust optimization approach. *Networks: An International Journal* 49(2):175–188.

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

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

Elmaghraby S (1977) Project planning and control by network models.

Fréchet M (1951) Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon, 3ˆ e serie, Sciences, Sect. A* 14:53–77.

Goh J, Hall NG (2013) Total cost control in project management via satisficing. *Management Science* 59(6):1354–1372.

Govindarajan A, Sinha A, Uichanco J (2021) Distribution-free inventory risk pooling in a multilocation newsvendor. *Management Science* 67(4):2272–2291.

Guo J, He S, Jiang B, Wang Z (2022) A unified framework for generalized momentproblems: a novel primal-dual approach. *arXiv preprint arXiv:2201.01445* .

Hagstrom JN (1988) Computational complexity of pert problems. *Networks* 18(2):139–147.

He S, Zhang J, Zhang S (2010) Bounding probability of small deviation: A fourth moment approach. *Mathematics of Operations Research* 35(1):208–232.

Hoeffding W (1940) Masstabinvariante korrelationtheorie, schriften des mathematis chen instituts und des instituts für angewandte mathematik der universität berlin 5, 181# 233.(translated in fisher, ni and pk sen (1994). the collected works of wassily hoeffding, new york.

Iglehart DL (1963) Optimality of (s, s) policies in the infinite horizon dynamic inventory problem. *Management science* 9(2):259–267.

Klein Haneveld WK (1986) Robustness against dependence in pert: An application of duality and distributions with known marginals. *Stochastic Programming 84 Part I*, 153–182 (Springer).

Lasserre JB (2002) Bounds on measures satisfying moment conditions. *The Annals of Applied Probability* 12(3):1114–1137.

Li X, Natarajan K, Teo CP, Zheng Z (2014) Distributionally robust mixed integer linear programs: Persistency models with applications. *European Journal of Operational Research* 233(3):459–473.

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

Meilijson I, Nádas A (1979) Convex majorization with an application to the length of critical paths. *Journal of Applied Probability* 16(3):671–677.

Mishra VK, Natarajan K, Padmanabhan D, Teo CP, Li X (2014) On theoretical and empirical aspects of marginal distribution choice models. *Management Science* 60(6):1511–1531.

Mishra VK, Natarajan K, Tao H, Teo CP (2012) Choice prediction with semidefinite optimization when utilities are correlated. *IEEE Transactions on Automatic Control* 57(10):2450–2463.

Nadas A (1979) Probabilistic pert. *IBM Journal of Research and Development* 23(3):339–347.

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

Natarajan K, Song M, Teo CP (2009) Persistency model and its applications in choice modeling. *Management Science* 55(3):453–469.

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

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

Pearson K (1905) Mathematical contributions to the theory of evolution: Xiv. *On the general theory of skew correlations and nonlinear regression* .

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

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

Scarf H (1958) A min-max solution of an inventory problem. *Studies in the mathematical theory of inventory and production* .

Schmüdgen K (2017) *The moment problem*, volume 9 (Springer).

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

Smith DE (2012) *A source book in mathematics* (Courier Corporation).

Van Slyke RM (1963) Monte carlo methods and the pert problem. *Operations Research* 11(5):839–860.

Vandenberghe L, Boyd S, Comanor K (2007) Generalized chebyshev bounds via semidefinite programming. *SIAM review* 49(1):52–64.

Wiesemann W, Kuhn D, Rustem B (2012) Robust resource allocations in temporal networks. *Mathematical programming* 135(1):437–471.

Williams T (1992) Practical use of distributions in network analysis. *Journal of the operational research society* 43(3):265–270.

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

Zielinski R (2004) Effective wlln, slln and clt in statistical models. *Applicationes Mathematicae* 31(1):117–125.

Appendix

A. Closed Form of Function f_κ

According to the well known result of roots of a quartic equation (see, e.g., Smith 2012), the function f_κ takes the largest real value of the following complex expression.

$$f_\kappa(x,\gamma) = \frac{\gamma x}{2} + \frac{1}{2}\sqrt{g_1 + \frac{g_2}{g_3} + g_3} + \sqrt{2g_1 - \frac{g_2}{g_3} - g_3 - \frac{2\gamma P x^3}{\sqrt{g_1 + \frac{g_2}{g_3} + g_3}}},$$

$$\begin{aligned}
P =& \kappa - \gamma^2 + 2, \\
g_1 =& \left(\gamma^2 - \frac{2\kappa}{3}\right) x^2, \\
g_2 =& \frac{x^3}{9}\left(x\left(\kappa+6\right)^2 - 12P\right). \\
g_3 =& \frac{1}{3}\Big(P^3x^6 + \left(4+\gamma^2\right)^3 x^6 + 3P^2\left(12+\gamma^2 x - 14x\right)x^5 + 3P\left(4+\gamma^2\right)\left(4x+\gamma^2 x - 6\right)x^5 \\
&+ 6\sqrt{3x}Px^4\sqrt{P^3(1-x) + P^2x\left(8-3x\left(8-\gamma^2(1-x)-5x\right)\right) - \left(4+\gamma^2\right)^2x^5 + \left(4+\gamma^2\right)^3x^6(1-x) \overline{+Px^2\left(3(1-x)\left((4+\gamma^2)x-3\right)^2 + 19x - 2\gamma^2x - 11\right)}}\Big)^{\frac{1}{3}}.
\end{aligned}$$

Here, the square root and the cubic root in g_3 can be arbitrarily chosen. The two square-root operators take all possible values, while the value of $\sqrt{g_1 + \frac{g_2}{g_3} + g_3}$ that appears twice in the formula is always consistent. Therefore, there are 4 different complex values given by f_κ. As shown in the proof of Theorem 2, we will always take the largest real value.

B. Proofs

Proof of Theorem 1:

Step 1:

Firstly, we introduce a mapping f:

$$\left\{(1,m_1,m_2,m_3,m_4) \in \overline{\mathbb{M}_4(\mathbb{R})}\right\} \xrightarrow{f} \left\{(\zeta_1,\zeta_2,\zeta_3,\zeta_4) \in \mathbb{R}^4 : \zeta_2 \ge 0, \zeta_4 \ge 0\right\}.$$

The rule for mapping is defined as follows:

$$\begin{aligned}
&\zeta_1 = m_1, \\
&\zeta_2 = m_2 - m_1^2 = \begin{vmatrix} 1 & m_1 \\ m_1 & m_2 \end{vmatrix}, \\
&\zeta_3 = m_3 - 3m_1m_2 + 2m_1^3, \\
&\zeta_4 = \left(m_2 - m_1^2\right)m_4 - m_2^3 + 2m_1m_2m_3 - m_3^2 = \begin{vmatrix} 1 & m_1 & m_2 \\ m_1 & m_2 & m_3 \\ m_2 & m_3 & m_4 \end{vmatrix}.
\end{aligned} \tag{39}$$

We now show that this mapping becomes a bijection by slightly modifying the domain and codomain as:

$$\left\{(1,m_1,m_2,m_3,m_4)\in\overline{\mathbb{M}_4\left(\mathbb{R}\right)},m_2>m_1^2\right\}\xrightarrow{f}\left\{(\zeta_1,\zeta_2,\zeta_3,\zeta_4)\in\mathbb{R}^4:\zeta_2>0,\zeta_4\geq 0\right\}.$$

We can see that m_i can also be represented by ζ_i

$$\begin{aligned}
m_1 &= \zeta_1 = g_{1,i}(\zeta_1),\\
m_2 &= \zeta_2+\zeta_1^2 = g_{2,i}(\zeta_1,\zeta_2),\\
m_3 &= \zeta_1^3+3\zeta_1\zeta_2+\zeta_3 = g_{3,i}(\zeta_1,\zeta_2,\zeta_3),\\
m_4 &= \frac{1}{\zeta_2}\left(\zeta_1^4\zeta_2+6\zeta_1^2\zeta_2^2+\zeta_2^3+4\zeta_1\zeta_2\zeta_3+\zeta_3^2+\zeta_4\right) = g_4(\zeta_1,\zeta_2,\zeta_3,\zeta_4).
\end{aligned}$$

By checking all principal minors of $\begin{vmatrix} 1 & g_{1,i}(\zeta_1) & g_{2,i}(\zeta_1,\zeta_2)\\ g_{1,i}(\zeta_1) & g_{2,i}(\zeta_1,\zeta_2) & g_{3,i}(\zeta_1,\zeta_2,\zeta_3)\\ g_{2,i}(\zeta_1,\zeta_2) & g_{3,i}(\zeta_1,\zeta_2,\zeta_3) & g_4(\zeta_1,\zeta_2,\zeta_3,\zeta_4)\end{vmatrix}$, We can find that this matrix is positive semidefinite. So in this case, f is a one-to-one correspondence.

Step 2:

Given $w_{i0}=x\in[0,1]$, consider the following subproblem:

$$\begin{aligned}
w_{i1}^* = \sup_{\boldsymbol{w_i},\boldsymbol{v_i}} \quad & w_{i1},\\
\text{s.t.} \quad & \boldsymbol{w_i}+\boldsymbol{v_i}=\boldsymbol{m_i},\\
& \boldsymbol{w_i},\boldsymbol{v_i}\in\overline{\mathbb{M}_k\left(\mathbb{R}\right)},\\
& w_{i0}=x.
\end{aligned}\tag{40}$$

In this step, we would like to show that the subproblem 40 is equivalent to the following problem:

$$\begin{aligned}
w_{i1}^* = \sup \quad & x\zeta_1+\sqrt{x(1-x)\zeta_2}t\\
\text{s.t.,} \quad & (1-x)\zeta_2^2t^2-x\zeta_2^2-\zeta_3\sqrt{x(1-x)\zeta_2}t\leq\sqrt{x(1-x)\zeta_2\zeta_4\left(1-t^2\right)},\\
& x\zeta_2^2t^2-(1-x)\zeta_2^2+\zeta_3\sqrt{x(1-x)\zeta_2}t\leq\sqrt{(1-x)x\zeta_2\zeta_4\left(1-t^2\right)},\\
& 0\leq t\leq 1,
\end{aligned}\tag{41}$$

where $\zeta_1,\zeta_2,\zeta_3,\zeta_4$ are defined in (39).

We first assume that $m_{i,2}-m_{i1}^2>0$ and $w_{i0}=x\in(0,1)$. we will discuss the problem in several cases.

Case 1: $w_{i1}^2<w_{i0}w_{i,2}$ and $v_{i1}^2<v_{i0}v_{i,2}$.

Let $(\zeta_1, \zeta_2, \zeta_3, \zeta_4) = f(1, m_{i1}, m_{i,2}, m_{i,3}, m_{i4})$, $(\omega_1, \omega_2, \omega_3, \omega_4) = f\left(1, \frac{w_{i1}}{x}, \frac{w_{i,2}}{x}, \frac{w_{i,3}}{x}, \frac{w_{i4}}{x}\right)$ and $(\nu_1, \nu_2, \nu_3, \nu_4) = f\left(1, \frac{v_{i1}}{1-x}, \frac{v_{i,2}}{1-x}, \frac{v_{i,3}}{1-x}, \frac{v_{i4}}{1-x}\right)$. The problem becomes

$$
\begin{aligned}
\sup \quad & x\omega_1, \\
\text{s.t.} \quad & \zeta_1 = x\omega_1 + (1-x)\nu_1 \\
& \zeta_2 + \zeta_1^2 = x(\omega_2 + \omega_1^2) + (1-x)(\nu_2 + \nu_1^2), \\
& \zeta_1^3 + 3\zeta_1\zeta_2 + \zeta_3 = x(\omega_1^3 + 3\omega_1\omega_2 + \omega_3) + (1-x)(\nu_1^3 + 3\nu_1\nu_2 + \nu_3), \\
& \frac{1}{\zeta_2}\left(\zeta_1^4\zeta_2 + 6\zeta_1^2\zeta_2^2 + \zeta_2^3 + 4\zeta_1\zeta_2\zeta_3 + \zeta_3^2 + \zeta_4\right) = \frac{x}{\omega_2}\left(\omega_1^4\omega_2 + 6\omega_1^2\omega_2^2 + \omega_2^3 + 4\omega_1\omega_2\omega_3 + \omega_3^2 + \omega_4\right), \\
& + \frac{(1-x)}{\nu_2}\left(\nu_1^4\nu_2 + 6\nu_1^2\nu_2^2 + \nu_2^3 + 4\nu_1\nu_2\nu_3 + \nu_3^2 + \nu_4\right), \\
& \omega_2, \nu_2 > 0, \\
& \omega_4, \nu_4 \geq 0.
\end{aligned}
\tag{42}
$$

It is not hard to see that we can always get an objective value of $x\zeta_1$ by assigning $\boldsymbol{\omega}_i = x\boldsymbol{m}_i$ and $\boldsymbol{\nu} = \mathbf{0}$. Therefore, the optimal objective is large or equal to with out loss of generality, let

$$
\begin{aligned}
&\omega_1 = \zeta_1 + \sqrt{\frac{(1-x)\zeta_2}{x}}t, \quad t \geq 0, \\
&\omega_2 = \varepsilon > 0.
\end{aligned}
$$

Notice that each of the four equality constraints in (42) contains only a linear term of $\nu_1, \cdots, \nu_4$, respectively. We can solve each ν_i successively and represent them with $\zeta_1, \cdots, \zeta_i$ and $\omega_1, \cdots, \omega_i$. Therefore, the feasibility is equivalent to finding t, ε, ω_3 and ω_4 such that $\nu_2 > 0$ and $\nu_4 \geq 0$.

Substituting ω_1 and ω_2 into the first two constraints of (42), we have

$$\nu_2 = \frac{(1-t^2)\zeta_2 - x\varepsilon}{1-x}.$$

Since $\omega_2 = \varepsilon > 0$ and $\nu_2 > 0$, we have

$$\varepsilon < \frac{\zeta_2(1-t^2)}{x} \text{ and } t < 1. \tag{43}$$

Then, represent ν_4 with $\zeta_1, \zeta_2, \zeta_3, \zeta_4, t, \varepsilon, \omega_3$ and ω_4. There is only one term that contains ω_4, which is $-\frac{x\nu_2\omega_4}{(1-x)\varepsilon}$. Therefore, we can write ν_4 in the form of

$$\nu_4 = h(t, \varepsilon, \omega_3) - \frac{x\nu_2\omega_4}{(1-x)\varepsilon}. \tag{44}$$

For a given pair of (t, ε) satisfying (43), it is feasible if and only if there exists a pair of (ω_3, ω_4) such that $\omega_4 \geq 0$ and $h(t, \varepsilon, \omega_3) - \frac{x\nu_2\omega_4}{(1-x)\varepsilon} \geq 0$, which is equivalent to

$$\max_{\omega_3\in\mathbb{R},\omega_4\geq 0}\left\{h(t,\varepsilon,\omega_3)-\frac{x\nu_2\omega_4}{(1-x)\varepsilon}\right\}=\max_{\omega_3\in\mathbb{R}}h(t,\varepsilon,\omega_3)-\min_{\omega_4\geq 0}\frac{x\nu_2\omega_4}{(1-x)\varepsilon}\geq 0 \tag{45}$$

Because $\frac{x\nu_2}{(1-x)\varepsilon}>0$, $\frac{x\nu_2\omega_4}{(1-x)\varepsilon}$ takes the minimum value 0 when $\omega_4=0$. On the other hand, we can find that h is a quadratic function of ω_3 where the coefficient of ω_3^2 is $-\frac{1-t^2}{(1-x)^2\varepsilon}<0$. So $\max_{\omega_3\in\mathbb{R}}h(t,\varepsilon,\omega_3)$ takes the maximum value when ω_3 is on the axis of symmetry of h. Then we have the following equations by computing the optimal ω_3 (noted as $\omega_3^*(t,\varepsilon)$) and substituting it back to h.

$$\omega_3^*(t,\varepsilon)=\frac{(1-x)\zeta_3-\sqrt{\frac{(1-x)\zeta_2}{x}}t\left(x\varepsilon+2\zeta_2-3x\zeta_2-(1-x)\zeta_2t^2\right)}{(1-x)(1-t^2)\zeta_2}\varepsilon$$

$$h(t,\varepsilon,\omega_3^*(t,\varepsilon))=C_0\left[x(1-x)\left(1-t^2\right)\zeta_2\zeta_4-\left((1-x)\zeta_2^2t^2+x\varepsilon\zeta_2-x\zeta_2^2-\zeta_3\sqrt{x(1-x)\zeta_2}t\right)^2\right].$$

Here,

$$C_0=\frac{\nu_2}{x\zeta_2(1-x)^2\left(x\varepsilon+(1-x)\nu_2\right)}>0.$$

Therefore, any pair of (t,ε) satisfying (43) is feasible if and only if

$$-\sqrt{x(1-x)\zeta_2\zeta_4\left(1-t^2\right)}\leq(1-x)\zeta_2^2t^2+x\zeta_2\varepsilon-x\zeta_2^2-\zeta_3\sqrt{x(1-x)\zeta_2}t\leq\sqrt{x(1-x)\zeta_2\zeta_4\left(1-t^2\right)}.$$

In addition, since $0<\varepsilon<\frac{\zeta_2(1-t^2)}{x}$, $t\in[0,1)$ is feasible if and only if

$$(1-x)\zeta_2^2t^2-x\zeta_2^2-\zeta_3\sqrt{x(1-x)\zeta_2}t<\sqrt{x(1-x)\zeta_2\zeta_4\left(1-t^2\right)};$$
$$x\zeta_2^2t^2-(1-x)\zeta_2^2+\zeta_3\sqrt{x(1-x)\zeta_2}t<\sqrt{(1-x)x\zeta_2\zeta_4\left(1-t^2\right)}.$$

It is not difficult to see that above two constraints are convex. Therefore, we can replace the supreme with maximum and let all constraints be closed. Then, we obtain the expression (41).

Case 2: $w_{i1}^2=w_{i0}w_{i,2}$ and $v_{i1}^2=v_{i0}v_{i,2}$.

Let $(\zeta_1,\zeta_2,\zeta_3,\zeta_4)=f\left(1,m_{i1},m_{i,2},m_{i,3},m_{i4}\right)$. Assume $w_{i1}=x\zeta_1+\sqrt{x(1-x)\zeta_2}t$, $v_{i1}=(1-x)\zeta_1-\sqrt{x(1-x)\zeta_2}t$.

Due to the fact that

$$m_2=\zeta_2+\zeta_1^2=w_{i,2}+v_{i,2}=\frac{w_{i1}^2}{x}+\frac{v_{i1}^2}{1-x}=\zeta_2t^2+\zeta_1^2$$

We can get solution $t=\pm1$. Then we can show that this case only happens when

$$\begin{aligned}
m_3&=\frac{w_{i1}^3}{x^2}+\frac{v_{i1}^3}{(1-x)^2}\\
&=\zeta_1^3+3\zeta_1\zeta_2+\frac{(\pm1\mp2x)\zeta_2^2}{\sqrt{x(1-x)\zeta_2}}\\
m_4&=\frac{w_{i1}^4}{x^3}+\frac{v_{i1}^4}{(1-x)^3}\\
&=\zeta_1^4+6\zeta_1\zeta_2^2+(\frac{1}{x(1-x)}-3)\zeta_2^3+\frac{4(\pm1\mp2x)\zeta_1\zeta_2^2}{\sqrt{x(1-x)\zeta_2}}
\end{aligned}$$

So we must have

$$\zeta_3 = \frac{(\pm 1 \mp 2x)\zeta_2^2}{\sqrt{x(1-x)\zeta_2}}$$
$$\zeta_4 = 0$$

If $t=-1$, it cannot be the optimal solution because there is always a feasible solution that has an objective value $x\zeta_1$. If $t=1$, both sides of the first and second constraint in (41) to be 0, which means that the constraints still hold thus this case is incorporated by (41).

Case 3: $w_{i1}^2 = w_{i0}w_{i,2}$ but $v_{i1}^2 > v_{i0}v_{i,2}$.

Firstly we must have $w_{ij} = \frac{w_{i1}^j}{x^{j-1}}$ for $j = 2,3,4$. We still let $(\zeta_1,\zeta_2,\zeta_3,\zeta_4) = f\left(1, m_{i1}, m_{i,2}, m_{i,3}, m_{i4}\right)$ and $(\nu_1,\nu_2,\nu_3,\nu_4) = f\left(1, \frac{v_{i1}}{x}, \frac{v_{i,2}}{x}, \frac{v_{i,3}}{x}, \frac{v_{i4}}{x}\right)$. The problem 40 becomes

$$\begin{aligned}
\sup \quad & x\omega_1 \\
\text{s.t.} \quad & \zeta_1 = w_{i1} + (1-x)\nu_1 \\
& \zeta_2 + \zeta_1^2 = \frac{w_{i1}^2}{x} + (1-x)(\nu_2 + \nu_1^2) \\
& \zeta_1^3 + 3\zeta_1\zeta_2 + \zeta_3 = \frac{w_{i1}^3}{x^2} + (1-x)(\nu_1^3 + 3\nu_1\nu_2 + \nu_3) \\
& \frac{1}{\zeta_2}\left(\zeta_1^4\zeta_2 + 6\zeta_1^2\zeta_2^2 + \zeta_2^3 + 4\zeta_1\zeta_2\zeta_3 + \zeta_3^2 + \zeta_4\right) = \frac{w_{i1}^4}{x^3} + \frac{(1-x)}{\nu_2}\left(\nu_1^4\nu_2 + 6\nu_1^2\nu_2^2 + \nu_2^3 + 4\nu_1\nu_2\nu_3 + \nu_3^2 + \nu_4\right) \\
& \nu_2 > 0 \\
& \nu_4 \geq 0.
\end{aligned} \tag{46}$$

Assume $w_{i1} = x\zeta_1 + \sqrt{x(1-x)\zeta_2}t$ and represent $\nu_1,\cdots,\nu_4$ with $\zeta_1,\cdots,\zeta_4$ and t. t is feasible if and only if

$$\nu_2 = \frac{1-t^2}{1-x}\zeta_2 > 0$$
$$\nu_4 = \frac{x(1-x)\left(1-t^2\right)\zeta_2\zeta_4 - \left((1-x)\zeta_2^2t^2 - x\zeta_2^2 - \zeta_3\sqrt{x(1-x)\zeta_2}t\right)^2}{x\zeta_2(1-x)^3} \geq 0$$

Similarly, we have

$$\begin{aligned}
w_{i1}^* = \max \quad & x\zeta_1 + \sqrt{x(1-x)\zeta_2}t \\
\text{s.t.} \quad & (1-x)\zeta_2^2t^2 - x\zeta_2^2 - \zeta_3\sqrt{x(1-x)\zeta_2}t \leq \sqrt{x(1-x)\zeta_2\zeta_4\left(1-t^2\right)} \\
& -(1-x)\zeta_2^2t^2 + x\zeta_2^2 + \zeta_3\sqrt{x(1-x)\zeta_2}t \leq \sqrt{(1-x)x\zeta_2\zeta_4\left(1-t^2\right)} \\
& 0 \leq t \leq 1.
\end{aligned} \tag{47}$$

The only difference between the optimization problem (47) and (41) is the second constraint. Notice that

$$\begin{aligned}
&\left[-(1-x)\zeta_2^2t^2 + x\zeta_2^2 + \zeta_3\sqrt{x(1-x)\zeta_2}t\right] - \left[x\zeta_2^2t^2 - (1-x)\zeta_2^2 + \zeta_3\sqrt{x(1-x)\zeta_2}t\right] \\
&= (1-t^2)\zeta_2^2 \geq 0
\end{aligned}$$

We can claim that the optimal objective value of (40) with $a \in (0,1)$ when $w_{i1}^2 = w_{i0}w_{i,2}$ will not exceed the value in the case when $w_{i1}^2 < w_{i0}w_{i,2}$.

Case 4: $v_{i1}^2 = v_{i0}v_{i,2}$ but $w_{i1}^2 > w_{i0}w_{i,2}$.

Let $(\zeta_1,\zeta_2,\zeta_3,\zeta_4) = f(1, m_{i1}, m_{i,2}, m_{i,3}, m_{i4})$ and $(\omega_1,\omega_2,\omega_3,\omega_4) = f\left(1, \frac{w_{i1}}{x}, \frac{w_{i,2}}{x}, \frac{w_{i,3}}{x}, \frac{w_{i4}}{a}\right)$. Assume $w_{i1} = x\zeta_1 + \sqrt{x(1-x)\zeta_2}t$. Then $v_{i1} = (1-x)\zeta_1 - \sqrt{x(1-x)\zeta_2}t$. Similar with previous proof, we have

$$
\begin{aligned}
w_{i1}^* = \max \quad & x\zeta_1 + \sqrt{x(1-x)\zeta_2}t \\
\text{s.t.} \quad & -x\zeta_2^2t^2 + (1-x)\zeta_2^2 - \zeta_3\sqrt{x(1-x)\zeta_2}t \le \sqrt{(1-x)x\zeta_2\zeta_4(1-t^2)} \\
& x\zeta_2^2t^2 - (1-x)\zeta_2^2 + \zeta_3\sqrt{x(1-x)\zeta_2}t \le \sqrt{x(1-x)\zeta_2\zeta_4(1-t^2)} \\
& 0 \le t \le 1.
\end{aligned}
\tag{48}
$$

The only difference between the optimization problem (48) and (41) is the first constraint. Similarly, because

$$
\begin{aligned}
&\left[-x\zeta_2^2t^2 + (1-x)\zeta_2^2 - \zeta_3\sqrt{x(1-x)\zeta_2}t\right] - \left[(1-x)\zeta_2^2t^2 - x\zeta_2^2 - \zeta_3\sqrt{x(1-x)\zeta_2}t\right] \\
&= (1-t^2)\zeta_2^2 \ge 0
\end{aligned}
$$

We can also conclude that the optimal objective value of (40) with $a \in (0,1)$ when $v_{i1}^2 = v_{i0}v_{i,2}$ will not exceed the case when $v_{i1}^2 < v_{i0}w_{i,2}$.

So far, we have shown that when $m_{i,2} - m_{i1}^2 > 0$, for any $x \in (0,1)$, the optimization problem (40) is equivalent to (41). When $x = 0$ or 1, we must have $w_{i1} = 0$ or m_{i1}. When $m_{i,2} = m_{i1}^2$ ($i.e.\zeta_2 = 0$), the distribution set only contains a degenerate distribution. We must have $w_{i1} = x\zeta_1$. In these cases, the inequality still holds. Therefore, the optimization problem (40) is equivalent to (41) for all $x \in [0,1]$ with any valid input $\boldsymbol{m}_i$.

Step 3: Then we show that Z_{max}^* can be reformulated as an SOCP. Let $(1, \zeta_{i,1}, \zeta_{i,2}, \zeta_{i,3}, \zeta_{i,4}) = f(1, m_{i,1}, m_{i,2}, m_{i,3}, m_{i,4})$. Using the result we proved above, we have

$$
\begin{aligned}
Z_{max}^* = \max_{x_i, t_i} \quad & \sum_{i=1}^{n} \left(x_i\zeta_{i1} + \sqrt{x_i(1-x_i)\zeta_{i,2}}t_i \right), \\
\text{s.t.} \quad & (1-x_i)\zeta_{i,2}^2t_i^2 - x_i\zeta_{i,2}^2 - \zeta_{i,3}\sqrt{x_i(1-x_i)\zeta_{i,2}}t_i \le \sqrt{x_i(1-x_i)\zeta_{i,2}\zeta_{i4}(1-t_i^2)}, && \forall i \in \{1,\cdots,n\}, \\
& x\zeta_{i,2}^2t_i^2 - (1-x_i)\zeta_{i,2}^2 + \zeta_{i,3}\sqrt{x_i(1-x_i)\zeta_{i,2}}t_i \le \sqrt{(1-x_i)x_i\zeta_{i,2}\zeta_{i4}(1-t_i^2)}, && \forall i \in \{1,\cdots,n\}, \\
& \boldsymbol{A}(x_1, x_2, \cdots, x_n) \le \boldsymbol{b}, \\
& 0 \le t_i \le 1, && \forall i \in \{1,\cdots,n\}.
\end{aligned}
$$

Let $y_i = \sqrt{x_i(1-x_i)}t_i$ and $z_i = \sqrt{x_i(1-x_i)(1-t_i^2)}$. When $x_i = 0$ or 1, $t_i = 0$ coincides with the constraints of the optimization problem. When $x_i \in (0,1)$, constraint $0 \le t_i \le 1$ is equivalent to $y_i^2 + z_i^2 = x_i - x_i^2$ and $y_i, z_i \ge 0$. Then time x and $1-x$ at both side of the first and second constraint.

We can check that this will not affect the result of the problem when $x = 0$ or 1. So the problem is equivalent to

$$\begin{aligned}
Z^*_{max} = \max_{x_i, y_i, z_i} \quad & \sum_{i=1}^{n} \left(x_i \zeta_{i1} + \sqrt{\zeta_{i,2}} y_i \right) \\
\text{s.t.} \quad & \zeta_{i,2}^2 y_i^2 \le x_i \left(\zeta_{i,2}^2 x_i + \zeta_{i,3} \sqrt{\zeta_{i,2}} y_i + \sqrt{\zeta_{i,2}\zeta_{i4}} z_i \right), && \forall i \in \{1, \cdots, n\}, \\
& \zeta_{i,2}^2 y_i^2 \le (1 - x_i) \left(\zeta_{i,2}^2 (1 - x_i) - \zeta_{i,3} \sqrt{\zeta_{i,2}} y_i + \sqrt{\zeta_{i,2}\zeta_{i4}} z_i \right), && \forall i \in \{1, \cdots, n\}, \\
& x_i^2 + y_i^2 + z_i^2 = x_i, && \forall i \in \{1, \cdots, n\}, \\
& \boldsymbol{A} (x_1, x_2, \cdots, x_n) \le \boldsymbol{b}, \\
& 0 \le y_i, z_i && \forall i \in \{1, \cdots, n\}.
\end{aligned} \tag{49}$$

Then, we will show that the third equality constraint in (49) can be relaxed with "$\le$" constraint. Firstly, for any optimal solution of the relaxed version problem such that $x_i^{*2} + y_i^{*2} + z_i^{*2} < x_i^*$, we can always replace z_i^* with a larger $\hat{z}_i$ satisfying $x_i^{*2} + y_i^{*2} + \hat{z}_i^2 < x_i^*$. In this case, the objective remains the same and the first two constraints still hold because the left hand side of each constraint is unchanged while the right hand side is nondecreasing.

Finally, we show that the last non-negativity constraints of y_i and z_i are ignorable. If we relax the non-negativity constraint of y_i, it will always be nonnegative when taking the optimal value. On the other hand, if a negative z_i appears in the solution. We can replace it with its absolute value $|z_i|$ and keep other variables unchanged, and the new solution is still feasible. Since z_i does not appear in the objective function, the non-negativity of z_i is redundant. Therefore, the problem is reformulated as the SOCP problem (10). □

Proof of Proposition 1: It is not difficult to see that $y(x)$ is a continuous function. So, we only prove the differentiability here. We first assume that $0 < x < \frac{1}{2} - \frac{\gamma}{2\sqrt{\gamma^2+4}}$. According to the proof of Theorem 1, we have

$$\sqrt{x(1-x) - y(x)^2} = \frac{y(x)^2 - x\gamma y(x) - x^2}{x\sqrt{\kappa + 2 - \gamma^2}}. \tag{50}$$

For any fixed x_0 and an x_1 in the neighborhood of x_0, we have

$$\begin{aligned}
y(x_0)^2 + (\kappa + 2 - \gamma^2) \left(\frac{y(x_0)}{x_0} - \gamma y(x_0) - x_0 \right)^2 = x_0 - x_0^2; \\
y(x_1)^2 + (\kappa + 2 - \gamma^2) \left(\frac{y(x_1)}{x_1} - \gamma y(x_1) - x_1 \right)^2 = x_1 - x_1^2.
\end{aligned}$$

For simplicity, we denote $y_0 = y(x_0)$, $y_1 = y(x_1)$, and $p = \sqrt{\kappa + 2 - \gamma^2}$. Subtracting the two equations above, we will have

$$
\begin{aligned}
&(y_1-y_0)\left[y_0+y_1+p\frac{y_0+y_1}{x_0}\left(\frac{y_0^2}{x_0}+\frac{y_1^2}{x_1}-\gamma y_0-\gamma y_1-x_0-x_1\right)\right]\\
=&(x_1-x_0)\left[1-x_0-x_1+p\left(\frac{y_0^2}{x_0x_1}+1\right)\left(\frac{y_0^2}{x_0}+\frac{y_1^2}{x_1}-\gamma y_0-\gamma y_1-x_0-x_1\right)\right]
\end{aligned}
$$

According to Theorem 2, we know that $\frac{y_0^2}{x_0}-\gamma y_0-x_0$ and $\frac{y_1^2}{x_1}-\gamma y_1-x_1$ are nonnegative. So, we have

$$
\begin{aligned}
\lim_{x_1\to x_0}\frac{y(x_1)-y(x_0)}{x_1-x_0}&=\lim_{x_1\to x_0}\frac{1-x_0-x_1+p\left(\frac{y_0^2}{x_0x_1}+1\right)\left(\frac{y_0^2}{x_0}+\frac{y_1^2}{x_1}-\gamma y_0-\gamma y_1-x_0-x_1\right)}{y_0+y_1+p\frac{y_0+y_1}{x_0}\left(\frac{y_0^2}{x_0}+\frac{y_1^2}{x_1}-\gamma_1 y_0-\gamma y_1-x_0-x_1\right)}\\
&=\frac{\frac{1}{2}-x_0+p\left(\frac{y(x_0)^2}{x_0^2}+1\right)\left(\frac{y(x_0)^2}{x_0}-\gamma y(x_0)-x_0\right)}{y(x_0)+2p\frac{y(x_0)}{x_0}\left(\frac{y(x_0)^2}{x_0}-\gamma y(x_0)-x_0\right)}\\
&=H(x_0,\gamma).
\end{aligned}
$$

Therefore, the function $y(x)$ is differentiable in $\left(0,\frac{1}{2}-\frac{\gamma}{2\sqrt{\gamma^2+4}}\right)$ and the derivative is $H(x,\gamma)$. Similarly, we can prove that $y(x)$ is differentiable in $\left(\frac{1}{2}-\frac{\gamma}{2\sqrt{\gamma^2+4}},1\right)$ and the derivative is $-H(1-x,-\gamma)$.

Lastly, we prove that the left and right derivatives are the same when $x=\frac{1}{2}-\frac{\gamma}{2\sqrt{\gamma^2+4}}$. According to Theorem 2,

$$
y\left(\frac{1}{2}-\frac{\gamma}{2\sqrt{\gamma^2+4}}\right)=\frac{1}{\sqrt{\gamma^2+4}}.
$$

Then we have

$$
H(\frac{1}{2}-\frac{\gamma}{2\sqrt{\gamma^2+4}},\gamma)=-H(\frac{1}{2}+\frac{\gamma}{2\sqrt{\gamma^2+4}},-\gamma)=\gamma.
$$

Therefore, the function $y(x)$ is differentiable on $(0,1)$.

Lastly, we prove the derivative approaches to infinity when x approaching to 0 or 1. When $x\to 0$, it is equivalent to prove that $y(x)/x\to+\infty$ since $y(0)=0$. The equation 50 is equivalent to

$$
p\sqrt{\frac{1-x}{x}-\left(\frac{y(x)}{x}\right)^2}=\left(\frac{y(x)}{x}\right)^2-\gamma\frac{y(x)}{x}-1.
$$

So, $y(x)/x\to+\infty$ due to the fact that $\lim_{x\to 0^+}(1-x)/x=+\infty$. Symmetrically, we have $\lim_{x\to 1^-}y'(x)=-\infty$. □

Proof of Theorem 3: Giving the first-order moment, the second-order central moment and the fourth-order central moment is equivalent to giving the mean μ_i, the standard deviation σ_i and the kurtosis κ_i. The third-order central moment exists since the fourth-order central moment exists.

The feasible region of skewness γ_i is $\left[-\sqrt{\gamma_i+2}, \sqrt{\gamma_i+2}\right]$ due to Pearson's inequality. Therefore, directly adopting the SOCP form of 4-MMM (10) and letting $\gamma_i \in \left[-\sqrt{\gamma_i+2}, \sqrt{\gamma_i+2}\right]$ be a decision variable provides an optimization formula of 124-MMM.

To simplify the problem, we first assume that $\boldsymbol{x}$ is given. Then the problem is separable, so we can omit the index i and study the decoupled inner problem which presented as follows:

$$\begin{aligned} \max_{y,z,\gamma} \quad & y \\ \text{s.t.} \quad & y^2 \le x\left(x+\gamma y+\sqrt{\kappa-\gamma^2+2}z\right) \\ & y^2 \le (1-x)\left(1-x-\gamma y+\sqrt{\kappa-\gamma^2+2}z\right) \\ & x^2+y^2+z^2 \le x \\ & -\sqrt{\gamma+2} \le \gamma \le \sqrt{\gamma+2}. \end{aligned} \tag{51}$$

When $\left(1-\sqrt{(\kappa+2)/(\kappa+6)}\right)/2 \le x \le \left(1+\sqrt{(\kappa+2)/(\kappa+6)}\right)/2$, we can always let $\gamma = (1-2x)/\sqrt{x(1-x)} \in \left[-\sqrt{\gamma+2}, \sqrt{\gamma+2}\right]$, such that $\frac{1}{2} - \frac{\gamma}{2\sqrt{\gamma^2+4}} = x$. From the proof of Theorem 2, we know that in the convex optimization form of 4-MMM, when $x = \frac{1}{2} - \frac{\gamma}{2\sqrt{\gamma^2+4}}$, $y(x) = \sqrt{x(1-x)}$. Due to the third constraint of (51), $y = \sqrt{x(1-x)}$ is the optimal objective value of (51).

When $x < \left(1-\sqrt{(\kappa+2)/(\kappa+6)}\right)/2$, we always have $x < \frac{1}{2} - \frac{\gamma}{2\sqrt{\gamma^2+4}}$. Recall the proof of Theorem 2, the second constraint of (51) is redundant. Then the problem becomes:

$$\begin{aligned} \max_{y,z,\gamma} \quad & y \\ \text{s.t.} \quad & y^2 \le x\left(x+\gamma y+\sqrt{\kappa-\gamma^2+2}z\right) \\ & x^2+y^2+z^2 \le x \\ & -\sqrt{\gamma+2} \le \gamma \le \sqrt{\gamma+2}. \end{aligned} \tag{52}$$

Noted that

$$\begin{aligned} y^2 - x^2 &\le x\left(\gamma y+\sqrt{\kappa-\gamma^2+2}z\right) \\ &\le \frac{x}{2}\sqrt{\frac{\kappa+2}{x(1-x)}}\left(y^2+z^2\right) + \frac{x}{2}\sqrt{\frac{x(1-x)}{\kappa+2}}\left(\sqrt{\kappa-\gamma^2+2}^2+\gamma^2\right) \\ &\le \frac{x}{2}\sqrt{\frac{\kappa+2}{x(1-x)}}x(1-x) + \frac{x}{2}\sqrt{\frac{x(1-x)}{\kappa+2}}(\kappa+2) \\ &= x\sqrt{x(1-x)(\kappa+2)}. \end{aligned}$$

The first and third inequality is due to the first and second constraint of (52); the second inequality is because of the Cauchy-Schwarz inequality. Therefore, we get an upper bound of $y \le \sqrt{x^2 + x\sqrt{x(1-x)(\kappa+2)}}$. This bound is tight if we let

$$\gamma = \sqrt{\frac{x}{1-x} + \sqrt{\frac{x(\kappa+2)}{(1-x)}}\sqrt{\kappa+2}}$$
$$z = \sqrt{x(1-2x) - x\sqrt{x(1-x)(\kappa+2)}}.$$

When $x > \left(1+\sqrt{(\kappa+2)/(\kappa+6)}\right)/2$, the proof is similar to the case $x < \left(1-\sqrt{(\kappa+2)/(\kappa+6)}\right)/2$. The maximum y is $\sqrt{(1-x)^2 + (1-x)\sqrt{x(1-x)(\kappa+2)}}$.

When $x = \left(1-\sqrt{(\kappa+2)/(\kappa+6)}\right)/2$, value of $\sqrt{x^2 + x\sqrt{x(1-x)(\kappa+2)}}$ and $\sqrt{x(1-x)}$ are both $1/\sqrt{\kappa+6}$. When $x = \left(1+\sqrt{(\kappa+2)/(\kappa+6)}\right)/2$, the result is similar. Therefore, the function is continuous. The concavity of $y_\kappa(x)$ in each segment is not difficult to verify by computing the second-order derivative. When $x = \left(1-\sqrt{(\kappa+2)/(\kappa+6)}\right)/2$, we have

$$\begin{aligned}\frac{\mathrm{d}\sqrt{x^2 + (1-x)\sqrt{x(1-x)(\kappa+2)}}}{\mathrm{d}x} &= \frac{-(\kappa+2)^{\frac{3}{2}} + 4\sqrt{\kappa+6} + (\kappa+2)\sqrt{\kappa+6}}{8} \\ &\geq \frac{\sqrt{\kappa+2}}{2} = \frac{\mathrm{d}\sqrt{x(1-x)}}{\mathrm{d}x}.\end{aligned}$$

When $x = \left(1+\sqrt{(\kappa+2)/(\kappa+6)}\right)/2$, we have a symmetric result. Therefore, y_κ is concave.

So far, we have completed the proof of the convex optimization formula. For the SOCP formula, we just need to relax $\sqrt{x(1-x)}$ as a decision varible z. It is not difficult to verify that formulation (17) and (18) give out the same result. □

Proof of Theorem 4: The existence of the fourth-order central moment guarantees the existence of variance and skewness. Similarly to the proof of Theorem (3), we assume that $\boldsymbol{x}$ is given and study the decoupled inner problem ignoring the index i. We can adopt the formula (18) of 124-MMM and let σ and κ be decision variables with extra constraints $\sigma \geq 0$, $\kappa \geq -2$, and $\sigma^4(\kappa+4) = m_4'$. we want to study the following problem.

$$\begin{aligned}\max_{\sigma,\kappa} \quad & \sigma y_\kappa(x) \\ \text{s.t.} \quad & \sigma^4(\kappa+3) = m_4' \\ & \sigma \geq 0 \\ & \kappa \geq -2\end{aligned} \tag{53}$$

where $y_\kappa(x)$ is defined the same as in Theorem (3). Noting that now x is fixed and κ is a variable, we reformulate $y_\kappa(x)$ as a function of κ, that is

$$y_x(\kappa)=\begin{cases}\sqrt{x^2+x\sqrt{x(1-x)(\kappa+2)}}, & x\le\frac{1}{2},\quad -2\le\kappa<\dfrac{1-6x(1-x)}{x(1-x)};\\ \sqrt{x(1-x)}, & \kappa\ge\dfrac{1-6x(1-x)}{x(1-x)};\\ \sqrt{(1-x)^2+(1-x)\sqrt{x(1-x)(\kappa+2)}}, & x>\frac{1}{2},\quad -2\le\kappa<\dfrac{1-6x(1-x)}{x(1-x)}.\end{cases}$$

When $x\le\frac{1}{2}$, $\kappa<\frac{1-6x(1-x)}{x(1-x)}$. Let $u=\sqrt{1/(\kappa+3)}$, $v=\sqrt{1-u^2}=\sqrt{(\kappa+2)/(\kappa+3)}$. The objective can be rewritten as follows.

$$\begin{aligned}\sigma y_\kappa(x)&=x\sqrt{\sigma^2+\sigma^2\sqrt{\frac{(1-x)}{x}(\kappa+2)}}\\&=x{m'_4}^{\frac{1}{4}}\sqrt{\sqrt{\frac{1}{\kappa+3}}+\sqrt{\frac{(1-x)(\kappa+2)}{x(\kappa+3)}}}\\&=x{m'_4}^{\frac{1}{4}}\sqrt{u+\sqrt{\frac{(1-x)}{x}}v}.\end{aligned}$$

(u,v) belongs to an arc with radius 1 and satisfies $\sqrt{x(1-x)/(1-3x(1-x))}<u\le 1$, $v\ge 0$. In addition, the derivative of v with respect to x satisfies

$$\frac{\mathrm{d}v}{\mathrm{d}u}=-\frac{u}{\sqrt{1-u^2}}<-\frac{\sqrt{\frac{x(1-x)}{1-3x(1-x)}}}{\sqrt{1-\frac{x(1-x)}{1-3x(1-x)}}}=-\frac{\sqrt{x(1-x)}}{1-2x}\le-\sqrt{\frac{x}{1-x}}.$$

Therefore, letting $u=\sqrt{x(1-x)/(1-3x(1-x))}$ provides an upper bound of the objective.

$$\sigma y_\kappa(x)<x^3{m'_4}^{\frac{1}{4}}\sqrt{\sqrt{\frac{x(1-x)}{1-3x(1-x)}}+\sqrt{\frac{(1-x)}{x}\frac{1-4x(1-x)}{1-3x(1-x)}}}=\left(\frac{x^3(1-x)^3m'_4}{1-3x(1-x)}\right)^{\frac{1}{4}}. \tag{54}$$

When $x>\frac{1}{2}$, $\kappa<\frac{1-6x(1-x)}{x(1-x)}$, the upper bound (54) also holds due to the symmetry.

On the other hand, for any $x\in(0,1)$, one can always let $\kappa=\frac{1-6x(1-x)}{x(1-x)}$, such that

$$\sigma y_\kappa(x)=\left(\frac{m'_4}{\kappa+3}\right)^{\frac{1}{4}}\sqrt{x(1-x)}=\left(\frac{x^3(1-x)^3m'_4}{1-3x(1-x)}\right)^{\frac{1}{4}}$$

Therefore, the optimal value of (53) is always $\left(\frac{x^3(1-x)^3m'_4}{1-3x(1-x)}\right)^{\frac{1}{4}}$, which completes the formula (19). In addition, the second-order derivative of $\left(\frac{x^3(1-x)^3m'_4}{1-3x(1-x)}\right)^{\frac{1}{4}}$ is

$$\frac{\mathrm{d}^2\left(\frac{x^3(1-x)^3m'_4}{1-3x(1-x)}\right)^{\frac{1}{4}}}{\mathrm{d}x^2}=-\frac{3(1-x)x(1-4(1-x)x(2-5(1-x)x))}{16(1-3(1-x)x)^3\left(\frac{(1-x)^3x^3}{1-3(1-x)x}\right)^{\frac{3}{4}}}\le 0,$$

which implies the objective is concave. □

Proof of Proposition 3: Denote $G_i(x_i) = x_i\mu_i + Y_{\Theta,i}(x_i)$. Given the combinatorial optimization problem (1) with distributional uncertainty set Θ, assume that the worst-case CVaR (22) is V. Consider the following extended problem:

$$\sup_{\theta\in\overline{\Theta}} E_\theta\left[\overline{Z}_{\max}(\tilde{\boldsymbol{c}},\tilde{c}_{n+1})\right] = \sup_{\theta\in\overline{\Theta}} E_\theta\left[\max\left\{Z_{\max}(\tilde{\boldsymbol{c}}),\tilde{c}_{n+1}\right\}\right], \tag{55}$$

where $\overline{\Theta}$ contains all possible joint distributions of $(\tilde{c}_1,\cdots,\tilde{c}_n,\tilde{c}_{n+1})$ $\tilde{c}_i$ such that $(\tilde{c}_1,\cdots,\tilde{c}_n)$ belongs to uncertainty set Θ and the marginal distribution of $\tilde{c}$ is a two point distribution which takes value $V+\varepsilon$, $\varepsilon>0$, with probability β and takes value $-A$ with probability $1-\beta$. $\overline{Z}_{\max}(\tilde{\boldsymbol{c}}) = \max\left\{\tilde{\boldsymbol{c}}^T\boldsymbol{x} + \tilde{c}_{n+1}x_{n+1} : (\boldsymbol{x},x_{n+1})\in\overline{\mathcal{X}}\right\}$ and $\overline{\mathcal{X}}$ is the new feasible region which defined as

$$\overline{\mathcal{X}} := \{(\boldsymbol{x},0):\boldsymbol{x}\in\mathcal{X}\}\cup\{(\boldsymbol{0},1)\}.$$

It is not hard to see that the convex hull of $\overline{\mathcal{X}}$ is

$$CH\left(\overline{\mathcal{X}}\right) = \{(\boldsymbol{x},x_{n+1}):\boldsymbol{A}\boldsymbol{x}\le(1-x_{n+1})\boldsymbol{b}, x_{n+1}\in[0,1]\}.$$

In addition, since the two point distribution can be characterized by giving mean, variance, skewness and kurtosis, the problem (55) is equivalent to the following problem.

$$\begin{aligned}\max\quad & \sum_{i=1}^{n+1} G_i(x_i)\\ \text{s.t.}\quad & \boldsymbol{A}\boldsymbol{x}\le(1-x_{n+1})\boldsymbol{b}\\ & x_{n+1}\in[0,1],\end{aligned}$$

where

$$G_{n+1}(x) = \begin{cases} Vx+\varepsilon x, & x\le\beta;\\ \beta V+\beta\varepsilon - A(x-\beta), & x>\beta.\end{cases} \tag{56}$$

Assume that $\boldsymbol{x}^*$ is the optimal solution of (23) and A is larger than $\sup\sum_{i=1}^n |G_i'(x_i^*)|$, where G_i' is the subderivative of G_i. We would like to show that the optimal x_{n+1} in (56) is β.

Since $(\boldsymbol{x}^*,\beta)$ is in the convex hull $CH\left(\overline{\mathcal{X}}\right)$, we can always find a series of positive a with summation $1-\beta$ such that $(\boldsymbol{x}^*,\beta) = \sum_{\boldsymbol{x}\in\mathcal{X}} a_{\boldsymbol{x}}(\boldsymbol{x},0) + \beta(\boldsymbol{0},1)$. Similarly, for any feasible $(\hat{\boldsymbol{x}},\hat{x}_{n+1})$ satisfying $\hat{x}_{n+1}>\beta$. We have a series of positive $\hat{a}$ with summation $1-\hat{x}_{n+1}$, such that $(\hat{\boldsymbol{x}},\hat{x}_{n+1}) = \sum_{\boldsymbol{x}\in\mathcal{X}} \hat{a}_{\boldsymbol{x}}(\boldsymbol{x},0) + \hat{x}_{n+1}(\boldsymbol{0},1)$.

The direction from $(\boldsymbol{x}^*,\beta)$ to $(\hat{\boldsymbol{x}},\hat{x}_{n+1})$ is $\sum_{\boldsymbol{x}\in\mathcal{X}}(\hat{a}_{\boldsymbol{x}}-a_{\boldsymbol{x}})(\boldsymbol{x},0) + (\hat{x}_{n+1}-\beta)(\boldsymbol{0},1)$ and $\sum_{\boldsymbol{x}\in\mathcal{X}}(\hat{a}_{\boldsymbol{x}}-a_{\boldsymbol{x}}) = \beta-\hat{x}_{n+1}$. Therefore, the derivative of this direction is negative due to our choice of A.

On the other hand, if there exist a optimal solution $(\hat{\boldsymbol{x}}, \hat{x}_{n+1})$ satisfying $\hat{x}_{n+1} < \beta$, according to the definition of persistency, we must have

$$\begin{aligned}\sup_{\theta\in\overline{\Theta}} E_\theta\left[\overline{Z}_{\max}\right] \leq & \hat{x}_{n+1}V + \hat{x}_{n+1}\varepsilon + (1-\hat{x}_{n+1})\sup_{\theta\in\Theta} CVaR^{\theta}_{\hat{x}_{n+1}}\left(Z_{\max}(\tilde{\boldsymbol{c}})\right) \\ \leq & \hat{x}_{n+1}V + \hat{x}_{n+1}\varepsilon + (1-\hat{x}_{n+1})\sup_{\theta\in\Theta} CVaR^{\theta}_{\beta}\left(Z_{\max}(\tilde{\boldsymbol{c}})\right) \\ = & V + \hat{x}_{n+1}\varepsilon.\end{aligned} \tag{57}$$

However, assume that the optimal distribution in (22) is θ^* and $VaR^{\theta}_{\beta}(Z_{max}(\tilde{\boldsymbol{c}}))$ is C_{θ^*}. We can let $\tilde{c}_{n+1} = V + \varepsilon$ when $Z_{max}(\tilde{\boldsymbol{c}}) > C_{\theta^*}$; $\tilde{c}_{n+1} = -A$ when $Z_{max}(\tilde{\boldsymbol{c}}) < C_{\theta^*}$. If $P\left[Z_{max}(\tilde{\boldsymbol{c}}) = C_{\theta^*}\right] \neq 0$, $\tilde{c}_{n+1} = V + \varepsilon$ with probability $(P\left[Z_{max}(\tilde{\boldsymbol{c}}) \leq C_{\theta^*}\right] - \beta)P\left[Z_{max}(\tilde{\boldsymbol{c}}) = C_{\theta^*}\right]$ and $\tilde{c}_{n+1} = -A$ otherwise. Denote this joint distribution as $\overline{\theta}^*$. It is not hard to see that in this case, $E_{\overline{\theta}^*}\left[\max\left\{Z_{\max}(\tilde{\boldsymbol{c}}), \tilde{c}_{n+1}\right\}\right] = (1-\beta)V + \beta(V+\varepsilon) = V + \beta\varepsilon$, which conflicts the optimality of $(\hat{\boldsymbol{x}}, \hat{x}_{n+1})$.

When the optimal $\hat{x}_{n+1} = \beta$, using the same argument of (57), we know that $\sup_{\theta\in\overline{\Theta}} E_\theta\left[\overline{Z}_{\max}\right] \leq V + \beta\varepsilon$, which means $V + \beta\varepsilon$ is exactly the optimal objective of (55). Finally, we have

$$\begin{aligned} V + \beta\varepsilon = \max \quad & \sum_{i=1}^{n} G_i(x_i) + \beta V + \beta\varepsilon \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} \leq (1-\beta)\boldsymbol{b}, \end{aligned}$$

which directly leaeds to (23). □

Proof of Proposition 5: Using the Theorem 2, the problem becomes the following.

$$\begin{aligned} \min_{\boldsymbol{\mu},\boldsymbol{\sigma}} \max_{\boldsymbol{x}} \quad & \sum_{i=1}^{n} \left(\mu_i x_i + \sigma_i y_i(x_i)\right) \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \\ & (\boldsymbol{\mu}, \boldsymbol{\sigma}) \in \mathcal{M}. \end{aligned}$$

The inner problem is a convex optimization with linear constraint, thus it satisfies the weak Slater's condition and the strong duality holds. The Lagrangian function of the inner problem is

$$\begin{aligned} L(\boldsymbol{x}, \boldsymbol{\lambda}) &= \sum_{i=1}^{n} \left(\mu_i x_i + \sigma_i y_i(x_i)\right) - \left(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\right)^T \boldsymbol{\lambda} \\ &= \sum_{i=1}^{n} \sigma_i \left(\left(\frac{\mu_i - \boldsymbol{a}_i^T \boldsymbol{\lambda}}{\sigma_i} \right) x_i + y_i(x_i) \right) + \boldsymbol{b}^T \boldsymbol{\lambda}. \end{aligned}$$

Therefore, the Lagrangian dual function is

$$\inf_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}) = \sum_{i=1}^{n} \left(\sigma_i \varphi_i \left(\frac{\mu_i - \boldsymbol{a}_i^T \boldsymbol{\lambda}}{\sigma_i} \right) - \boldsymbol{b}^T \boldsymbol{\lambda} \right),$$

where φ_i is the convex conjugate of $-y_i$. The Lagrangian dual function is jointly convex in $\boldsymbol{\mu}$,$\boldsymbol{\sigma}$ and $\boldsymbol{\lambda}$ due to the joint convexity of the perspective of a convex function. Replacing the inner problem with its duality, we get the convex optimization formula in Proposition 5. □

Proof of Proposition 8: We focus on the proof for $\gamma \le 0$. The case for $\gamma \ge 0$ can be derived with similar argument.

For $\kappa = \gamma^2 - 2$, $y_{\gamma,\kappa}$ degenerate to a piece-wise linear function and all points in $[\mu + \frac{\gamma - \sqrt{\gamma^2+4}}{2}, \mu + \frac{\gamma+\sqrt{\gamma^2+4}}{2}]$ is optimal, when $\eta = \frac{1}{2} + \frac{\gamma}{2\sqrt{\gamma^2+4}}$. So it is clear that $\eta_0 = \frac{1}{2} + \frac{\gamma}{2\sqrt{\gamma^2+4}}$.

When $\kappa > \gamma^2 - 2$, $y_{\gamma,\kappa}$ is concave and continuous differentiable, thus q_4^* is unique and increasing with respect to η. At $\eta = \frac{1}{2} + \frac{\gamma}{2\sqrt{\gamma^2+4}}$, $q_4^* = \mu + \frac{\sigma\gamma}{2} \le \mu$, indicating that $\eta_0 \ge \frac{1}{2} + \frac{\gamma}{2\sqrt{\gamma^2+4}}$.

On the other hand, setting $\eta = \frac{1}{2}$ yields the optimal solution:

$$q_4^* = \mu + \frac{1 - 16y(x)^4}{2y(x)(1-4y(x)^2) - \gamma},$$

which is $\ge \mu$ as $y(x) \le x(1-x) \le \frac{1}{4}$. Therefore, we have $\eta_0(\gamma,\kappa) \in [\frac{1}{2} + \frac{\gamma}{2\sqrt{\gamma^2+4}}, \frac{1}{2}]$.

We further demonstrate that given γ, η_0 will increase to $1/2$ as κ increases to infinity. Given γ and κ, denote $\eta_0(\gamma,\kappa)$ as $\hat{\eta}_0$ and $y_{\gamma,\kappa}(\hat{\eta}_0)$ as $\hat{y}$. From Proposition 1 and Proposition 7, we know that the optimal order quantity given μ, σ, γ and κ when $\eta = \hat{\eta}_0$ is

$$\hat{q}_4^* = \mu + \sigma \left[\frac{\hat{y}}{1-\hat{\eta}_0} + \frac{\hat{y}^2 - \gamma(1-\hat{\eta}_0)\hat{y} - (1-\hat{\eta}_0)^2}{2\gamma\hat{\eta}_0(1-\hat{\eta}_0)^2 - 2\hat{y}(1-\hat{\eta}_0)(3\hat{\eta}_0 - 1) + 2\hat{y}^3} \right] = \mu. \tag{58}$$

If the kurtosis increase to $\overline{\kappa} > \kappa$, the new optimal order quantity given μ, σ, γ and $\overline{\kappa}$ when $\eta = \hat{\eta}_0$ becomes

$$\overline{q}_4^* = \mu + \sigma \left[\frac{\overline{y}}{1-\hat{\eta}_0} + \frac{\overline{y}^2 - \gamma(1-\hat{\eta}_0)\overline{y} - (1-\hat{\eta}_0)^2}{2\gamma\hat{\eta}_0(1-\hat{\eta}_0)^2 - 2\overline{y}(1-\hat{\eta}_0)(3\hat{\eta}_0 - 1) + 2\overline{y}^3} \right], \tag{59}$$

where $\overline{y} = y_{\gamma,\overline{\kappa}}(\hat{\eta}_0)$.

To show that $\eta_0(\gamma,\overline{\kappa}) \ge \hat{\eta}_0$, it is sufficient to show that $\overline{q}_4^* \le \mu$. According to (59), we have

$$\begin{aligned}
& \overline{q}_4^* \le \mu \\
\iff \quad & \frac{\overline{y}}{1-\hat{\eta}_0} + \frac{\overline{y}^2 - \gamma(1-\hat{\eta}_0)\overline{y} - (1-\hat{\eta}_0)^2}{2\gamma\hat{\eta}_0(1-\hat{\eta}_0)^2 - 2\overline{y}(1-\hat{\eta}_0)(3\hat{\eta}_0 - 1) + 2\overline{y}^3} \le 0 \\
\iff \quad & 2\gamma\hat{\eta}_0(1-\hat{\eta}_0)^2\overline{y} - 2\overline{y}^2(1-\hat{\eta}_0)(3\hat{\eta}_0-1) + 2\overline{y}^4 + \overline{y}^2(1-\hat{\eta}_0) - \gamma(1-\hat{\eta}_0)^2\overline{y} - (1-\hat{\eta}_0)^3 \ge 0 \\
\iff \quad & 2\overline{y}^4 + (1-\hat{\eta}_0)(1-2\hat{\eta}_0)\left(3\overline{y} - \gamma(1-\hat{\eta}_0)\right)\overline{y} - (1-\hat{\eta}_0)^3 \ge 0.
\end{aligned} \tag{60}$$

The third inequality is because that

$$2\gamma\hat{\eta}_0(1-\hat{\eta}_0)^2 - 2\overline{y}(1-\hat{\eta}_0)(3\hat{\eta}_0-1) + 2\overline{y}^3 = 2\gamma(1-\hat{\eta}_0)^2\hat{\eta}_0 - 2\overline{y}\left(1-\hat{\eta}_0 + 3\hat{\eta}_0(1-\hat{\eta}_0) + \overline{y}^2\right) < 0.$$

Similarly, according to (58) we have

$$2\hat{y}^4 + (1-\hat{\eta}_0)(1-2\hat{\eta}_0)\left(3\hat{y} - \gamma(1-\hat{\eta}_0)\right)\hat{y} - (1-\hat{\eta}_0)^3 = 0.$$

Recall that in Theorem 4-MMM, larger κ leads to larger feasible region, we will have $\overline{y} \ge \hat{y}$ hence the last inequality in (60) holds.

In addition, when $\kappa \to \infty$, we know that $y_{\gamma,\kappa}(x) \to \sqrt{x(1-x)}$ and thus the optimal order quantity converges to that in mean and variance model, which means η_0 approaches to $\frac{1}{2}$. □