Chebyshev Inequalities for Products of Random Variables

Napat Rujeerapaiboon[1], Daniel Kuhn[1], and Wolfram Wiesemann[2]

[1] *Risk Analytics and Optimization Chair, École Polytechnique Fédérale de Lausanne, Switzerland*
[2] *Imperial College Business School, Imperial College London, United Kingdom*

May 18, 2016

Abstract

We derive sharp probability bounds on the tails of a product of symmetric non-negative random variables using only information about their first two moments. If the covariance matrix of the random variables is known exactly, these bounds can be computed numerically using semidefinite programming. If only an upper bound on the covariance matrix is available, the probability bounds on the right tails can be evaluated analytically. The bounds under precise and imprecise covariance information coincide for all left tails as well as for all right tails corresponding to quantiles that are either sufficiently small or sufficiently large. We also prove that all left probability bounds reduce to the trivial bound 1 if the number of random variables in the product exceeds an explicit threshold. Thus, in the worst case, the weak-sense geometric random walk defined through the running product of the random variables is absorbed at 0 with certainty as soon as time exceeds the given threshold. The techniques devised for constructing Chebyshev bounds for products can also be used to derive Chebyshev bounds for sums, maxima and minima of non-negative random variables.

Keywords. Chebyshev inequality, probability bounds, distributionally robust optimization, convex optimization

1 Introduction

The classical one-sided Chebyshev inequality [5, 7] for a random variable $\tilde{\xi}$ with mean μ and variance σ^2 can be represented as

$$\mathbb{P}(\tilde{\xi} \geq \gamma) \leq \begin{cases} \frac{\sigma^2}{\sigma^2 + (\gamma - \mu)^2} & \text{if } \gamma \geq \mu, \\ 1 & \text{if } \gamma < \mu. \end{cases} \tag{1}$$

This inequality is sharp. Indeed, for $\gamma \neq \mu$ it is binding under the two-point distribution

$$\mathbb{P}^\star = \begin{cases} \frac{\sigma^2}{\sigma^2 + (\gamma - \mu)^2} \delta_\gamma + \frac{(\gamma - \mu)^2}{\sigma^2 + (\gamma - \mu)^2} \delta_{\mu - \sigma^2/(\gamma - \mu)} & \text{if } \gamma > \mu, \\ \frac{\sigma^2}{\sigma^2 + (\mu - \gamma)^2} \delta_\gamma + \frac{(\mu - \gamma)^2}{\sigma^2 + (\mu - \gamma)^2} \delta_{\mu + \sigma^2/(\mu - \gamma)} & \text{if } \gamma < \mu. \end{cases} \tag{2}$$

In the degenerate case $\gamma = \mu$, the inequality (1) is still sharp because the distributions

$$\mathbb{P}_\kappa = \frac{1}{1 + \kappa^2} \delta_{\gamma - \sigma\kappa} + \frac{\kappa^2}{1 + \kappa^2} \delta_{\gamma + \sigma/\kappa}$$

have mean μ and variance σ^2 for every $\kappa > 0$, while $\lim_{\kappa\uparrow\infty} \mathbb{P}_\kappa(\tilde\xi \geq \gamma) = 1$. Note, however, that no single distribution with mean $\mu = \gamma$ and variance $\sigma^2 > 0$ can satisfy $\mathbb{P}(\tilde\xi \geq \gamma) = 1$.

If we have the extra information that the random variable $\tilde\xi$ is non-negative (and without much loss of generality that $\mu > 0$), then one can strengthen the Chebyshev inequality (1) to

$$\mathbb{P}(\tilde\xi \geq \gamma) \leq \begin{cases} \frac{\sigma^2}{\sigma^2 + (\gamma-\mu)^2} & \text{if } \gamma \geq \mu + \sigma^2/\mu, \\ \frac{\mu}{\gamma} & \text{if } \mu \leq \gamma < \mu + \sigma^2/\mu, \\ 1 & \text{if } \gamma < \mu, \end{cases} \tag{3}$$

see, e.g., [12, 28]. The extremal distributions (2) are supported on the non-negative real line if either $\gamma \geq \mu + \sigma^2/\mu > \mu$ or if $\gamma < \mu$. Thus, they certify the sharpness of (3) in the respective parameter domains. For $\mu \leq \gamma < \mu + \sigma^2/\mu$ the Chebyshev inequality (3) for non-negative random variables reduces in fact to the classical Markov inequality $\mathbb{P}(\tilde\xi \geq \gamma) \leq \mu/\gamma$. In this Markov regime, the Chebyshev inequality (3) remains sharp because the distributions

$$\mathbb{P}_\kappa = \left[1 + \frac{\sigma^2}{\kappa\gamma} - \frac{\mu(\kappa-\mu)}{\gamma(\kappa-\gamma)} - \frac{\mu(\gamma-\mu)}{\kappa(\kappa-\gamma)}\right]\delta_0 + \frac{\mu(\kappa-\mu) - \sigma^2}{\gamma(\kappa-\gamma)}\delta_\gamma + \frac{\sigma^2 - \mu(\gamma-\mu)}{\kappa(\kappa-\gamma)}\delta_\kappa$$

have mean μ and variance σ^2 for every $\kappa > \mu + \sigma^2/\mu$, while $\lim_{\kappa\uparrow\infty} \mathbb{P}_\kappa(\tilde\xi \geq \gamma) = \mu/\gamma$. From the textbook proof of Markov's inequality it follows that $\mathbb{P}^\star = [1 - \mu/\gamma]\delta_0 + [\mu/\gamma]\delta_\gamma$ is the only distribution on the non-negative reals that has mean μ and satisfies $\mathbb{P}^\star(\tilde\xi \geq \gamma) = \mu/\gamma$. However, the additional requirement that the variance of $\tilde\xi$ under $\mathbb{P}^\star$ must equal σ^2 implies $\gamma = \mu + \sigma^2/\mu$. Thus, for $\mu \leq \gamma < \mu + \sigma^2/\mu$ there cannot exist any single distribution with $\mathbb{P}(\tilde\xi \geq \gamma) = \mu/\gamma$.

In the rest of the paper we consider a sequence of T random variables $\tilde\xi_1, \tilde\xi_2, \ldots, \tilde\xi_T$ and assume that the first two moments of these random variables are known and permutation symmetric. Specifically, assume that all random variables share the same mean μ and variance σ^2, respectively, while all pairs of mutually distinct random variables share the same correlation coefficient ρ. Thus, the mean vector and the covariance matrix of $\tilde{\boldsymbol\xi} = (\tilde\xi_1, \ldots, \tilde\xi_T)^\intercal$ are given by

$$\boldsymbol\mu = \begin{bmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{bmatrix} \in \mathbb{R}^T \quad \text{and} \quad \boldsymbol\Sigma = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \cdots & \rho\sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho\sigma^2 & \rho\sigma^2 & \cdots & \sigma^2 \end{bmatrix} \in \mathbb{S}^T, \tag{4}$$

respectively. Throughout the paper we assume that $\sigma > 0$ and $-\frac{1}{T-1} < \rho < 1$. These conditions are necessary and sufficient for the covariance matrix $\boldsymbol\Sigma$ to be strictly positive definite. Note that $\tilde{\boldsymbol\xi}$ constitutes a weak-sense stationary stochastic process in the sense of [20].

An elementary calculation reveals that the sum $\sum_{t=1}^T \tilde\xi_t$ has mean value $T\mu$ and variance $T\sigma^2(1 + (T-1)\rho)$. The classical Chebyshev inequality (1) applied to $\sum_{t=1}^T \tilde\xi_t$ thus implies

$$\mathbb{P}(\sum_{t=1}^T \tilde\xi_t \geq \gamma) \leq \begin{cases} \frac{T\sigma^2(1+(T-1)\rho)}{T\sigma^2(1+(T-1)\rho) + (\gamma-T\mu)^2} & \text{if } \gamma \geq T\mu, \\ 1 & \text{if } \gamma < T\mu. \end{cases} \tag{5}$$

This inequality is still sharp due to a projection property of distribution families with compatible first and second moments. Indeed, for any distribution $\mathbb{P}_\zeta$ of a random variable $\tilde{\zeta}$ with mean value $T\mu$ and variance $T\sigma^2(1 + (T-1)\rho)$ there exists a distribution $\mathbb{P}$ of the random vector $\tilde{\boldsymbol{\xi}}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ such that $\mathbb{P}_\zeta$ coincides with the marginal distribution of $\sum_{t=1}^T \tilde{\xi}_t$ under $\mathbb{P}$, that is, $\mathbb{P}_\zeta(\tilde{\zeta} \in B) = \mathbb{P}(\sum_{t=1}^T \tilde{\xi}_t \in B)$ for every Borel set $B \subseteq \mathbb{R}$ [32]. The extremal distributions (2) certifying the sharpness of (1) can therefore be used to construct multivariate extremal distributions of $\tilde{\boldsymbol{\xi}}$ certifying the sharpness of (5). This result may be unexpected. Indeed, if $\tilde{\xi}_1, \ldots, \tilde{\xi}_T$ are independent and identically distributed, then, by the central limit theorem, their sum is approximately normally distributed with mean $T\mu$ and variance $T\sigma^2$. In contrast, if $\tilde{\xi}_1, \ldots, \tilde{\xi}_T$ are only known to be uncorrelated with a common mean and variance (but not necessarily independent and identically distributed), then, by the projection theorem, their sum may follow *any* distribution with mean $T\mu$ and variance $T\sigma^2$.

Assume now that $\tilde{\xi}_t$ is non-negative for every $t = 1, \ldots, T$ (and without much loss of generality that $\mu > 0$). As we will prove in Proposition 2.1 below, a distribution $\mathbb{P}$ supported on $\mathbb{R}_+^T$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ as given in (4) exists iff $\mu^2 + \rho\sigma^2 \geq 0$. We will assume that this condition holds throughout the rest of the paper. In this setting, the generalized Chebyshev inequality (3) applied to the non-negative random variable $\sum_{t=1}^T \tilde{\xi}_t$ implies

$$\mathbb{P}(\sum_{t=1}^T \tilde{\xi}_t \geq \gamma) \leq \begin{cases} \frac{T\sigma^2(1+(T-1)\rho)}{T\sigma^2(1+(T-1)\rho)+(\gamma-T\mu)^2} & \text{if } \gamma \geq T\mu + \sigma^2(1 + (T-1)\rho)/\mu, \\ \frac{T\mu}{\gamma} & \text{if } T\mu \leq \gamma < T\mu + \sigma^2(1 + (T-1)\rho)/\mu, \\ 1 & \text{if } \gamma < T\mu. \end{cases} \quad (6)$$

Even though the multivariate extension (6) of the univariate Chebyshev inequality (3) can still be shown to be sharp, we are not aware of an elementary proof; see Theorem 6.3 below.

In this paper we aim to derive Chebyshev inequalities for *products* of non-negative random variables. Specifically, we will derive sharp upper bounds on the left and right tail probabilities $\mathbb{P}(\prod_{t=1}^T \tilde{\xi}_t \leq \gamma)$ and $\mathbb{P}(\prod_{t=1}^T \tilde{\xi}_t \geq \gamma)$, respectively. Products of random variables frequently arise in physics, statistics, finance, number theory and many other branches of science [11]. Indeed, they are at the heart of stochastic models of many complex phenomena. When rocks are crushed, for example, the size of a fragment is multiplied by a random factor (that is smaller than 1) in every single breakup event [10]. Similar multiplicative phenomena explain the distribution of body weights, stock prices, the sizes of biological populations, income, rainfall etc. [1].

Note that the stochastic process $\tilde{\boldsymbol{\pi}} = \{\tilde{\pi}_T\}_{T \in \mathbb{N}}$ defined through $\tilde{\pi}_T = \prod_{t=1}^T \tilde{\xi}_t$ can be interpreted as a geometric random walk driven by the weak-sense stationary process $\tilde{\boldsymbol{\xi}} = \{\tilde{\xi}_t\}_{t \in \mathbb{N}}$. Chebyshev inequalities for the products of the $\tilde{\xi}_t$ thus provide tight bounds on the quantiles of a geometric random walk when there is limited distributional information. Consequently, they are potentially relevant for the many applications in economics and operations research, where geometric Brownian motions are traditionally used to model the prices of assets [17]. An improved understanding of weak-sense geometric random walks may also stimulate new research directions in distributionally robust optimziation [8, 13, 31] and optimal uncertainty quantification [14, 24].

Remark 1.1 (Chebyshev in Log-Space) *It seems natural to reduce Chebyshev inequalities for products of non-negative random variables to Chebyshev inequalities for their logarithms. Assume thus that the first two moments of the logarithmic random variables $\tilde{\eta}_t = \log(\tilde{\xi}_t)$, $1, \ldots, T$, are known and permutation symmetric. Specifically, denote by μ_η, σ_η^2 and ρ_η the mean, variance and correlation coefficient in log-space. Then, the Chebyshev inequality* (5) *for sums implies*

$$\mathbb{P}(\textstyle\prod_{t=1}^T \tilde{\xi}_t \geq \gamma) = \mathbb{P}(\textstyle\sum_{t=1}^T \tilde{\eta}_t \geq \log \gamma) \leq \begin{cases} \frac{T\sigma_\eta^2(1+(T-1)\rho_\eta)}{T\sigma_\eta^2(1+(T-1)\rho_\eta)+(\log\gamma - T\mu_\eta)^2} & \text{if } \log\gamma \geq T\mu_\eta, \\ 1 & \text{if } \log\gamma < T\mu_\eta. \end{cases} \quad (7)$$

Note that (7) *is sharp because* (5) *is sharp. However, there is no one-to-one correspondence between the moments of the original and the logarithmic random variables. Even worse, it is possible that μ is finite while $\mu_\eta = -\infty$ (e.g., if $\xi_t = 0$ with positive probability), or that μ_η is finite while $\mu = +\infty$ (e.g., if $\tilde{\xi}_t$ follows a Pareto distribution with unit shape parameter). In this work we focus on the case where the $\tilde{\xi}_t$ have known finite first and second moments, and we explicitly allow the event $\tilde{\xi}_t = 0$ to have positive probability. This assumption can be crucial for truthfully capturing the bankruptcy risks in financial applications, for instance.*

The starting point of this paper is the intriguing observation that modern optimization theory provides powerful tools for constructing and analyzing probability inequalities [4]. Assume for instance that we aim to find a sharp probability inequality for a target event characterized through finitely many polynomial inequalities on a random vector $\tilde{\boldsymbol{\xi}}$. Assume further that the desired inequality should hold for all distributions of $\tilde{\boldsymbol{\xi}}$ satisfying finitely many polynomial support and moment constraints. In the special case of the Chebyshev inequality (1), the target event corresponds to the set $\{\xi \in \mathbb{R} : \xi \geq \gamma\}$, while the relevant distribution family corresponds to the class of all distributions on $\mathbb{R}$ with mean μ and variance σ^2. Constructing the desired probability inequality is thus tantamount to maximizing the probability of the target event over the given distribution family. This leads to a generalized moment problem over probability measures. Under a mild regularity condition, this moment problem admits a strong dual linear program subject to polynomially parameterized semi-infinite constraints [15, 16, 18]. A key insight of [4] is that this dual problem can be approximated systematically by tractable semidefinite programs. The resulting approximations are safe (*i.e.*, they are guaranteed to provide *upper* bounds on the probability of the semialgebraic event). Moreover, these approximations are always tight in the univariate case but generically loose in the multivariate setting.

Stronger statements are available for probability inequalities that rely exclusively on first- and second-order moments. Specifically, if the support of the random vector $\tilde{\boldsymbol{\xi}}$ is unrestricted, the best upper bound on the probability of a *convex* target event is given by $1/(1 + d^2)$, where d represents the distance of the target event from the mean vector of $\tilde{\boldsymbol{\xi}}$ under the Mahalanobis norm induced by the covariance matrix of $\tilde{\boldsymbol{\xi}}$ [22]. More generally, if the target event constitutes a *union* of finitely many convex sets, over each of which convex quadratic optimization problems can be solved in polynomial time, then the best Chebyshev bound can be computed by an efficient algorithm reminiscent of the ellipsoid method of convex optimization [4]. Recently it has

been observed that if the target event is defined by quadratic inequalities, the best Chebyshev bound coincides exactly with the optimal value of a single tractable semidefinite program [30]. In spite of these encouraging results, the computation of Chebyshev bounds becomes hard in the presence of support constraints. Specifically, if $\tilde{\boldsymbol{\xi}}$ is supported on the non-negative orthant, it is already NP-hard to find sharp Chebyshev bounds for convex polyhedral target events [4].

For a random vector $\tilde{\boldsymbol{\xi}}$ with zero mean and unrestricted support, the above methods have been used to derive a sharp Chebyshev bound on $\mathbb{P}(\prod_{t=1}^{T} \tilde{\xi}_t \geq 1, \ \tilde{\xi}_t > 0 \ \forall t)$, which is expressed in terms of the solution of a tractable convex program [22]. As the $\tilde{\xi}_t$ are allowed to adopt negative values, however, we believe that the practical relevance of this bound is limited. In this paper we aim to derive sharp Chebyshev bounds on $\mathbb{P}(\prod_{t=1}^{T} \tilde{\xi}_t \geq \gamma)$ and $\mathbb{P}(\prod_{t=1}^{T} \tilde{\xi}_t \leq \gamma)$ under the explicit assumption that $\tilde{\boldsymbol{\xi}}$ is supported on the non-negative orthant. Note that the second target event $\{\boldsymbol{\xi} \in \mathbb{R}_+^T : \prod_{t=1}^{T} \xi_t \leq \gamma\}$ is neither convex nor representable as a finite union of convex sets, nor representable through finitely many quadratic constraints in $\boldsymbol{\xi}$. Thus, none of the existing techniques could be used to bound its probability even if there were no support constraints. As support constraints generically lead to intractability [4], we focus here on the special case where the first- and second-order moments are permutation-symmetric.

The main results of this paper can be summarized as follows.

(i) If the distribution $\mathbb{P}$ of the non-negative random variables has mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ as given in (4), then the sharp upper Chebyshev bounds on $\mathbb{P}(\prod_{t=1}^{T} \tilde{\xi}_t \geq \gamma)$ and $\mathbb{P}(\prod_{t=1}^{T} \tilde{\xi}_t \leq \gamma)$ can both be expressed as the optimal values of explicit semidefinite programs, which are amenable to efficient numerical solution via interior point algorithms.

(ii) If the distribution $\mathbb{P}$ of the non-negative random variables has mean $\boldsymbol{\mu}$ and a covariance matrix bounded above by $\boldsymbol{\Sigma}$ in a positive semidefinite sense, then we obtain an explicit analytical formula for the sharp upper Chebyshev bound on $\mathbb{P}(\prod_{t=1}^{T} \tilde{\xi}_t \geq \gamma)$.

(iii) The Chebyshev bound in (ii) coincides with the corresponding bound in (i) for all values of γ that are either sufficiently small or sufficiently large. For intermediate values of γ the numerical bound in (i) may be strictly smaller than the analytical bound in (ii).

(iv) If the distribution $\mathbb{P}$ of the non-negative random variables has mean $\boldsymbol{\mu}$ and a covariance matrix bounded above by $\boldsymbol{\Sigma}$ in a positive semidefinite sense, then the sharp upper Chebyshev bound on $\mathbb{P}(\prod_{t=1}^{T} \tilde{\xi}_t \leq \gamma)$ coincides with the corresponding numerical bound in (i). Thus, there is a distribution that makes this bound sharp and has covariance matrix $\boldsymbol{\Sigma}$.

(v) The Chebyshev bound in (iv) reduces to the trivial bound 1 for every $\gamma > 0$ if T exceeds an explicit threshold T_0. Thus, in the worst case, the weak-sense geometric random walk $\tilde{\boldsymbol{\pi}} = \{\tilde{\pi}_T\}_{T \in \mathbb{N}}$ defined through $\tilde{\pi}_T = \prod_{t=1}^{T} \tilde{\xi}_t$ is absorbed at 0 *with certainty* if $T \geq T_0$.

(vi) The techniques devised for constructing Chebyshev bounds for products of random variables can also be used to derive Chebyshev bounds on sums, maxima and minima (and possibly other permutation-symmetric functionals) of non-negative random variables.

The rest of the paper is structured as follows. In Section 2 we formalize the connection between probability inequalities and convex optimization. Left- and right-sided Chebyshev inequalities for products of random variables are then derived in Sections 3 and 4, respectively, while generalized Chebyshev inequalities that account for imprecise knowledge of the covariances are discussed in Section 5. Chebyshev inequalities for other permutation-symmetric functionals of the random variables are presented in Section 6, and examples are given in Section 7.

Notation The symbol $\mathbb{I}$ stands for the identity matrix, $\mathbf{1}$ for the vector of all ones, and $\mathbf{e}_i$ for the i-th standard basis vector. Their dimensions will always be clear from the context. The space of symmetric $T \times T$ matrices is denoted by $\mathbb{S}^T$, and its subset of all positive (negative) semidefinite matrices is denoted by $\mathbb{S}_+^T$. For $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{S}^T$, the statements $\boldsymbol{A} \succeq \boldsymbol{B}$ and $\boldsymbol{B} \preceq \boldsymbol{A}$ both mean that $\boldsymbol{A} - \boldsymbol{B} \in \mathbb{S}_+^T$. The indicator function $1_{\mathcal{E}}$ of a logical statement $\mathcal{E}$ is defined through $1_{\mathcal{E}} = 1$ if $\mathcal{E}$ holds true; $= 0$ otherwise. Random variables are denoted by tilde signs, while their realizations are denoted by the same symbols without tildes. The Dirac distribution concentrating unit mass at $\boldsymbol{\xi}$ is denoted by $\delta_{\boldsymbol{\xi}}$. For any closed set $\mathcal{S} \subseteq \mathbb{R}^T$, we let $\mathcal{M}_+(\mathcal{S})$ be the cone of all non-negative Borel measures supported on $\mathcal{S}$.

2 Optimization Perspective on Chebyshev Inequalities

To analyze probability bounds using tools from optimization, we first introduce an *ambiguity set* $\mathcal{P}$, that is, a family of distributions for which the desired probability bound should hold. In this paper we mainly focus on the ambiguity set of all distributions supported on $\mathbb{R}_+^T$ that share the permutation-symmetric mean and covariance matrix defined in (4), that is, we set

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{M}_+(\mathbb{R}_+^T) \; : \; \mathbb{P}\left(\tilde{\boldsymbol{\xi}} \geq \mathbf{0} \right) = 1, \; \mathbb{E}_{\mathbb{P}}\left(\tilde{\boldsymbol{\xi}} \right) = \boldsymbol{\mu}, \; \mathbb{E}_{\mathbb{P}}\left(\tilde{\boldsymbol{\xi}}\tilde{\boldsymbol{\xi}}^{\mathsf{T}} \right) = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^{\mathsf{T}} \right\}. \qquad (8)$$

We highlight that $\mathcal{P}$ is characterized by only four parameters: T, μ, σ, ρ. Without much loss of generality, we assume henceforth that $\mu > 0$, $\sigma > 0$ and $-\frac{1}{T-1} < \rho < 1$. The last two conditions are equivalent to $\boldsymbol{\Sigma} \succ \mathbf{0}$. To rule out trivial special cases, we further restrict attention to $T \geq 2$. However, all of these conditions do not yet guarantee that $\mathcal{P}$ is non-empty. Proposition 2.1 below provides a necessary and sufficient condition for the non-emptiness of $\mathcal{P}$.

Proposition 2.1 (Non-emptiness of $\mathcal{P}$) *The ambiguity set $\mathcal{P}$ is non-empty iff $\mu^2 + \rho\sigma^2 \geq 0$.*

Proof: If $\mathcal{P}$ is non-empty, then any $\mathbb{P} \in \mathcal{P}$ satisfies

$$\mathbf{0} \leq \mathbb{E}_{\mathbb{P}}\left(\tilde{\boldsymbol{\xi}}\tilde{\boldsymbol{\xi}}^{\mathsf{T}} \right) = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^{\mathsf{T}} \iff \begin{cases} \mu^2 + \sigma^2 \geq 0 \\ \mu^2 + \rho\sigma^2 \geq 0 \end{cases} \iff \mu^2 + \rho\sigma^2 \geq 0,$$

where the equivalences follow from the definition of $\boldsymbol{\Sigma}$ and the assumption that $\rho < 1$.

Assume now that $\mu^2 + \rho\sigma^2 \geq 0$. We show that $\mathcal{P}$ contains a discrete distribution $\mathbb{P}$ satisfying

$$\mathbb{P}\left(\tilde{\boldsymbol{\xi}} = y\mathbf{1} + (x - y)\mathbf{e}_i \right) = \frac{p}{T}, \quad i = 1, \ldots, T, \quad \text{and} \quad \mathbb{P}\left(\tilde{\boldsymbol{\xi}} = z\mathbf{1} \right) = 1 - p \qquad (9)$$

for $x \geq y \geq 0$, $z \geq 0$ and $p \in [0,1]$. For this distribution to be contained in $\mathcal{P}$, it must also satisfy the following moment conditions:

(i) $\mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\xi}}] = \boldsymbol{\mu}$ $\quad\quad\quad\quad \Longleftrightarrow \quad \dfrac{p}{T}(x + (T-1)y) + (1-p)z = \mu;$

(ii) $\mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\xi}}\tilde{\boldsymbol{\xi}}^{\mathsf{T}}] = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^{\mathsf{T}} \quad \Longleftrightarrow \quad \dfrac{p}{T}(x^2 + (T-1)y^2) + (1-p)z^2 = \mu^2 + \sigma^2,$

$$\dfrac{p}{T}(2xy + (T-2)y^2) + (1-p)z^2 = \mu^2 + \rho\sigma^2.$$

To construct $\mathbb{P}$, it is notationally convenient to perform the change of variables $m_1 \leftarrow \frac{1}{T}(x + (T-1)y)$ and $m_2 \leftarrow \frac{1}{T}(x^2 + (T-1)y^2)$. For a given (m_1, m_2), we can then recover (x,y) via

$$x = m_1 + \sqrt{(T-1)(m_2 - m_1^2)} \quad \text{and} \quad y = m_1 - \sqrt{(m_2 - m_1^2)/(T-1)}.$$

Note that the correspondence between (x,y) and (m_1, m_2) is one-to-one and onto over $\{(x,y) \in \mathbb{R}_+^2 : x \geq y\}$ and $\{(m_1, m_2) \in \mathbb{R}_+^2 : m_1^2 \leq m_2 \leq Tm_1^2\}$. Now, for $\mathbb{P}$ to be in $\mathcal{P}$, we require that

(i') $\mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\xi}}] = \boldsymbol{\mu}$ $\quad\quad\quad\quad \Longleftrightarrow \quad pm_1 + (1-p)z = \mu;$

(ii') $\mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\xi}}\tilde{\boldsymbol{\xi}}^{\mathsf{T}}] = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^{\mathsf{T}} \quad \Longleftrightarrow \quad pm_2 + (1-p)z^2 = \mu^2 + \sigma^2,$

$$\dfrac{p}{T-1}(Tm_1^2 - m_2) + (1-p)z^2 = \mu^2 + \rho\sigma^2.$$

In the remainder of the proof, we thus need to show that there is $m_1, m_2, z \geq 0$, $m_1^2 \leq m_2 \leq Tm_1^2$, and $p \in [0,1]$ satisfying (i') and (ii'). To this end, consider the choice

$$p = \begin{cases} \min\left\{ \dfrac{T\mu^2}{T\mu^2 + (1+(T-1)\rho)\sigma^2}, \dfrac{\rho T}{1+(T-1)\rho} \right\} & \text{if } \rho > 0, \\[2mm] \dfrac{T\mu^2}{T\mu^2 + \sigma^2} & \text{if } \rho = 0, \\[2mm] \dfrac{-\rho T}{1-\rho} & \text{if } \rho < 0, \end{cases} \quad\quad (10)$$

which satisfies $p \in [0,1]$ by construction, as well as

$$m_1 = \mu + \sigma\sqrt{\dfrac{(1-p)(1+(T-1)\rho)}{pT}}, \quad m_2 = m_1^2 + \dfrac{(1-p)(T-1)\sigma^2}{pT}, \quad z = \mu - \sigma\sqrt{\dfrac{p(1+(T-1)\rho)}{(1-p)T}}.$$

Note that the terms inside the square roots are non-negative since $\rho > -1/(T-1)$.

Step 1: We show that $m_1, m_2, z \geq 0$. The non-negativity of m_1 and m_2 holds by construction. To check that $z \geq 0$, we distinguish the cases $\rho > 0$, $\rho = 0$ and $\rho < 0$. For $\rho > 0$, we obtain $z = 0$ for $p = \frac{T\mu^2}{T\mu^2 + (1+(T-1)\rho)\sigma^2}$. Since the square root term in the expression for z is increasing in p, we thus conclude that $z \geq 0$. The case where $\rho = 0$ is analogous since $\frac{T\mu^2}{T\mu^2 + \sigma^2} = \frac{T\mu^2}{T\mu^2 + (1+(T-1)\rho)\sigma^2}$ for $\rho = 0$. For $\rho < 0$, on the other hand, we obtain $z = \mu - \sigma\sqrt{-\rho}$ for our choice of p. The resulting z is thus non-negative due to the assumption that $\mu^2 + \rho\sigma^2 \geq 0$.

Step 2: To check that $m_1^2 \leq m_2 \leq Tm_1^2$, we first use the definition of m_2 and the assumption that $\rho < 1$ to verify that $m_1^2 \leq m_2$. The other inequality holds if and only if

$$
\begin{aligned}
m_2 \leq Tm_1^2 \quad &\Longleftrightarrow \quad \sqrt{\frac{1-\rho}{pT}}\sigma \leq m_1 \\
&\Longleftrightarrow \quad \mu\sqrt{pT} + \left(\sqrt{(1+(T-1)\rho)(1-p)} - \sqrt{1-\rho}\right)\sigma \geq 0,
\end{aligned}
\tag{11}
$$

where the first and second equivalence follow from the definitions of m_2 and m_1, respectively. We now show that the last inequality holds by distinguishing the cases $\rho > 0$, $\rho = 0$ and $\rho < 0$.

For $\rho > 0$, we observe that the expression $\sqrt{(1+(T-1)\rho)(1-p)} - \sqrt{1-\rho}$ in (11) evaluates to 0 for $p = \frac{T\rho}{1+(T-1)\rho}$ and that it is decreasing in p. Since $\mu\sqrt{pT} \geq 0$ by construction, we thus conclude that the last inequality in (11) holds, and hence $m_2 \leq Tm_1^2$ when $\rho \geq 0$. In combination with (10) and (11), the above inequality ensures that $m_2 \leq Tm_1^2$.

For $\rho = 0$, equation (11) simplifies to

$$
\mu\sqrt{pT} + (\sqrt{1-p}-1)\sigma \geq 0 \quad \Longleftrightarrow \quad \frac{\mu\sqrt{T}}{\sigma} \geq \frac{1-\sqrt{1-p}}{\sqrt{p}} \quad \Longleftarrow \quad \frac{\mu\sqrt{T}}{\sigma} \geq \sqrt{p},
$$

where the two implications follow from algebraic manipulations and the fact that $\sqrt{p} \geq \frac{1-\sqrt{1-p}}{\sqrt{p}}$ for $p \in [0,1]$, respectively. One readily verifies that the last inequality is satisfied by $p = \frac{T\mu^2}{T\mu^2+\sigma^2}$.

For $\rho < 0$, substituting p in (11) with its definition from (10) yields

$$
\begin{aligned}
\mu\sqrt{pT} + (\sqrt{(1+(T-1)\rho)(1-p)} - \sqrt{1-\rho})\sigma &= \frac{T\mu\sqrt{-\rho}}{\sqrt{1-\rho}} + \left(\frac{1+(T-1)\rho}{\sqrt{1-\rho}} - \sqrt{1-\rho}\right)\sigma \\
&\geq \frac{-T\rho\sigma}{\sqrt{1-\rho}} + \left(\frac{1+(T-1)\rho}{\sqrt{1-\rho}} - \sqrt{1-\rho}\right)\sigma \\
&= 0,
\end{aligned}
$$

where the equalities follow from direct calculations and the inequality holds since $\mu^2 + \rho\sigma^2 \geq 0$. We thus conclude that $m_2 \leq Tm_1^2$ whenever $\rho < 0$ as postulated.

Step 3: We show that our choice of m_1, m_2 and z meets the requirements (i') and (ii'), regardless of the value of p. First, a direct calculation shows that requirement (i') follows from the definitions of m_1 and z. Next, the first requirement in (ii') follows from

$$
\begin{aligned}
pm_2 + (1-p)z^2 &= pm_2 + (1-p)z^2 - (pm_1 + (1-p)z)^2 + \mu^2 \\
&= p(m_2 - m_1^2) + \left(pm_1^2 + (1-p)z^2\right) - (pm_1 + (1-p)z)^2 + \mu^2 \\
&= p(m_2 - m_1^2) + p(1-p)(m_1 - z)^2 + \mu^2 \\
&= \tfrac{1}{T}(1-\rho)(T-1)\sigma^2 + \tfrac{1}{T}(1+(T-1)\rho)\sigma^2 + \mu^2 \\
&= \sigma^2 + \mu^2,
\end{aligned}
$$

where the first equality holds since the requirement (i') is met, and the fourth equality follows from the definitions of m_1, m_2 and z.

Finally, to prove the second requirement in (ii'), we first observe that

$$pm_2 - \frac{p}{T-1}(Tm_1^2 - m_2) = \frac{pT}{T-1}(m_2 - m_1^2) = (1-\rho)\sigma^2,$$

where the second equality follows from the definition of m_2. Note that the term on the left (right) side of this equality constitutes the difference between the left (right) sides of the requirements in (ii'). The second requirement in (ii') and the claim thus follow. ∎

In order to establish Chebyshev bounds for products of random variables, we will formulate generalized moment problems that optimize over the probability measures in the ambiguity set $\mathcal{P}$. We can then leverage powerful duality results from convex optimization to reformulate these moment problems as explicit semidefinite programs that are amenable to efficient solution via interior point methods. The *weak duality* principle, which holds true for every optimization problem, states that the optimal value of a (primal) minimization problem is bounded from below by the optimal value of its associated dual (maximization) problem. To establish tight probability bounds, we need to invoke the *strong duality* principle, which states that under certain conditions the optimal values of the primal and dual optimization problems coincide. In our setting, strong duality holds whenever $\mu^2 + \rho\sigma^2 > 0$.

Theorem 2.1 (Slater Condition) *If $\mu^2 + \rho\sigma^2 > 0$, then the moment vector $(1, \boldsymbol{\mu}, \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^{\mathsf{T}})$ is contained in the interior of the moment cone $\mathcal{K}$ defined through*

$$\mathcal{K} = \left\{ \left(\int_{\mathbb{R}_+^T} \mathbb{P}(\mathrm{d}\boldsymbol{\xi}), \int_{\mathbb{R}_+^T} \boldsymbol{\xi}\,\mathbb{P}(\mathrm{d}\boldsymbol{\xi}), \int_{\mathbb{R}_+^T} \boldsymbol{\xi}\boldsymbol{\xi}^{\mathsf{T}}\,\mathbb{P}(\mathrm{d}\boldsymbol{\xi}) \right) : \mathbb{P} \in \mathcal{M}_+(\mathbb{R}_+^T) \right\}.$$

Proof: We first show that $\mathcal{P}$ contains a distribution of the form (9) where the inequalities $x \geq y \geq 0$, $z \geq 0$ and $p \in [0, 1]$ hold *strictly*, as well as $x + (T-1)y > Tz$ (Step 1). This distribution allows us to show that $(1, \boldsymbol{\mu}, \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^{\mathsf{T}})$ is in the relative interior of $\mathcal{K}_1 = \mathcal{K} \cap (\{1\} \times \mathbb{R}_+^T \times \mathbb{S}_+^T)$ (Step 2), from which the result follows directly by re-scaling the measures in $\mathcal{K}_1$ (Step 3).

Step 1: We distinguish the cases $\rho < 0$ and $\rho \geq 0$. For $\rho < 0$, one readily verifies that the choice of p, x, y and z in the proof of Proposition 2.1 satisfies $x > y > 0$, $z > 0$, $p \in (0, 1)$ and $x + (T-1)y > Tz$ by construction. Moreover, these inequalities are also satisfied strictly for $\rho \geq 0$ if we replace p in (10) with any value from the open interval $(0, p)$.

Step 2: To prove that $(1, \boldsymbol{\mu}, \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^{\mathsf{T}}) \in \mathrm{rel\,int}\,\mathcal{K}_1$, we show that all perturbed ambiguity sets

$$\mathcal{P}(\boldsymbol{\mu}^\epsilon, \boldsymbol{\Omega}^\epsilon) = \left\{ \mathbb{P} \in \mathcal{M}_+(\mathbb{R}_+^T) \ : \ \mathbb{P}\left(\tilde{\boldsymbol{\xi}} > \boldsymbol{0} \right) = 1, \ \mathbb{E}_{\mathbb{P}}\left(\tilde{\boldsymbol{\xi}} \right) = \boldsymbol{\mu}^\epsilon, \ \mathbb{E}_{\mathbb{P}}\left(\tilde{\boldsymbol{\xi}}\tilde{\boldsymbol{\xi}}^{\mathsf{T}} \right) = \boldsymbol{\Omega}^\epsilon \right\}$$

with $\boldsymbol{\mu}^\epsilon \in \mathcal{B}_\epsilon(\boldsymbol{\mu})$ and $\boldsymbol{\Omega}^\epsilon \in \mathcal{B}_\epsilon(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^{\mathsf{T}})$ are non-empty for sufficiently small ϵ, where $\mathcal{B}_\epsilon(\boldsymbol{x})$ denotes the ϵ-ball around $\boldsymbol{x}$ in the respective space. Note that the covariance matrix of any

distribution in $\mathcal{P}(\boldsymbol{\mu}^\epsilon, \boldsymbol{\Omega}^\epsilon)$ is positive definite for small ϵ since $\boldsymbol{\Sigma} \succ \mathbf{0}$ and the eigenvalues are continuous functions of the second-order moment matrix. In the following, we construct a discrete distribution $\mathbb{P}^\epsilon \in \mathcal{P}(\boldsymbol{\mu}^\epsilon, \boldsymbol{\Omega}^\epsilon)$ with

$$\mathbb{P}^\epsilon \left(\tilde{\boldsymbol{\xi}} = \boldsymbol{\xi}^{\epsilon,i} \right) = \frac{p}{T} \quad i = 1, \ldots, T \quad \text{and} \quad \mathbb{P}^\epsilon \left(\tilde{\boldsymbol{\xi}} = \boldsymbol{\xi}^{\epsilon,T+1} \right) = 1 - p, \tag{12}$$

where p is the constant chosen in Step 1. The moment conditions for $\mathbb{P}^\epsilon$ then simplify to:

(i) $\mathbb{E}_{\mathbb{P}^\epsilon}[\tilde{\boldsymbol{\xi}}] = \boldsymbol{\mu}^\epsilon \quad \Longleftrightarrow \quad \dfrac{p}{T} \sum_{i=1}^{T} \xi_t^{\epsilon,i} + (1-p)\xi_t^{\epsilon,T+1} = \mu_t^\epsilon \qquad \forall t = 1, \ldots, T;$

(ii) $\mathbb{E}_{\mathbb{P}^\epsilon}[\tilde{\boldsymbol{\xi}}\tilde{\boldsymbol{\xi}}^\intercal] = \boldsymbol{\Omega}^\epsilon \quad \Longleftrightarrow \quad \dfrac{p}{T} \sum_{i=1}^{T} \left(\xi_t^{\epsilon,i} \right)^2 + (1-p) \left(\xi_t^{\epsilon,T+1} \right)^2 = \Omega_{tt}^\epsilon \quad \forall t = 1, \ldots, T,$

$$\dfrac{p}{T} \sum_{i=1}^{T} \xi_s^{\epsilon,i} \xi_t^{\epsilon,i} + (1-p)\xi_s^{\epsilon,T+1}\xi_t^{\epsilon,T+1} = \Omega_{st}^\epsilon \quad \forall 1 \le s < t \le T.$$

These moment conditions represent a system of nonlinear equations $\boldsymbol{F}(\boldsymbol{\mu}^\epsilon, \boldsymbol{\Omega}^\epsilon; \{\boldsymbol{\xi}^{\epsilon,i}\}_{i=1}^{T+1}) = \mathbf{0}$ in the moments $\boldsymbol{\mu}^\epsilon$ and $\boldsymbol{\Omega}^\epsilon$ as well as the atoms $\boldsymbol{\xi}^{\epsilon,i}$, $i = 1, \ldots, T+1$, of the distribution $\mathbb{P}^\epsilon$. From Step 1 we know that $\boldsymbol{F}(\boldsymbol{\mu}, \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal; \{\boldsymbol{\xi}^i\}_{i=1}^{T+1}) = \mathbf{0}$ for $\boldsymbol{\xi}^i = y\mathbf{1} + (x-y)\mathbf{e}_i$, $i = 1, \ldots, T$, $\boldsymbol{\xi}^{T+1} = z\mathbf{1}$ and for some $x, y, z \in \mathbb{R}_+$ satisfying $x > y > 0$, $z > 0$ and $x + (T-1)y > Tz$. Moreover, the implicit function theorem proves the existence of continuously differentiable functions $\boldsymbol{g}^i$: $\mathbb{R}_+^T \times \mathbb{S}_+^T \to \mathbb{R}^T$, $i = 1, \ldots, T+1$, such that $\boldsymbol{F}(\boldsymbol{\mu}^\epsilon, \boldsymbol{\Omega}^\epsilon; \{\boldsymbol{g}^i(\boldsymbol{\mu}^\epsilon, \boldsymbol{\Omega}^\epsilon)\}_{i=1}^{T+1}) = \mathbf{0}$ for all $\boldsymbol{\mu}^\epsilon \in \mathcal{B}_\epsilon(\boldsymbol{\mu})$ and $\boldsymbol{\Omega}^\epsilon \in \mathcal{B}_\epsilon(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal)$, provided that ϵ is sufficiently small, $\boldsymbol{F}$ is continuously differentiable, and the Jacobian of $\boldsymbol{F}$ with respect to $\boldsymbol{\xi}^{\epsilon,i}$ has full row rank at $(\boldsymbol{\mu}^\epsilon, \boldsymbol{\Omega}^\epsilon, \{\boldsymbol{\xi}^{\epsilon,i}\}_{i=1}^{T+1}) = (\boldsymbol{\mu}, \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal, \{\boldsymbol{\xi}^i\}_{i=1}^{T+1})$. Thus, the functions $\boldsymbol{g}^i$ allow us to construct distributions of the form (12) that satisfy the moment conditions of the perturbed ambiguity sets $\mathcal{P}(\boldsymbol{\mu}^\epsilon, \boldsymbol{\Omega}^\epsilon)$ for all $\boldsymbol{\mu}^\epsilon \in \mathcal{B}_\epsilon(\boldsymbol{\mu})$ and $\boldsymbol{\Omega}^\epsilon \in \mathcal{B}_\epsilon(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal)$. Since each $\boldsymbol{g}^i$ is continuous, we have $\boldsymbol{g}^i(\boldsymbol{\mu}^\epsilon, \boldsymbol{\Omega}^\epsilon) > \mathbf{0}$ for all $\boldsymbol{\mu}^\epsilon \in \mathcal{B}_\epsilon(\boldsymbol{\mu})$ and $\boldsymbol{\Omega}^\epsilon \in \mathcal{B}_\epsilon(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal)$ when ϵ is sufficiently small, that is, the support of $\mathbb{P}^\epsilon$ is contained in $\mathbb{R}_+^T$, and thus $\mathbb{P}^\epsilon$ is indeed contained in $\mathcal{P}(\boldsymbol{\mu}^\epsilon, \boldsymbol{\Omega}^\epsilon)$.

The moment function $\boldsymbol{F}$ is continuously differentiable by construction. To apply the implicit function theorem, we therefore only need to show that the Jacobian $\mathbf{J}$ of $\boldsymbol{F}$ with respect to $\boldsymbol{\xi}^{\epsilon,1}, \ldots, \boldsymbol{\xi}^{\epsilon,T+1}$ has full row rank at $(\boldsymbol{\mu}^\epsilon, \boldsymbol{\Omega}^\epsilon, \{\boldsymbol{\xi}^{\epsilon,i}\}_{i=1}^{T+1}) = (\boldsymbol{\mu}, \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal, \{\boldsymbol{\xi}^i\}_{i=1}^{T+1})$. For ease of exposition, we divide the first T^2 and the last T columns of $\mathbf{J}$ by $\frac{p}{T}$ and $1 - p$, respectively, and we divide the rows corresponding to the first requirement in (ii) by 2. We then obtain

$$\mathbf{J} = \left[\begin{array}{c|c|c|c|c} \mathbb{I} & \mathbb{I} & \cdots & \mathbb{I} & \mathbb{I} \\ \hline y\mathbb{I} + (x-y)\mathbf{e}_1\mathbf{e}_1^\intercal & y\mathbb{I} + (x-y)\mathbf{e}_2\mathbf{e}_2^\intercal & \cdots & y\mathbb{I} + (x-y)\mathbf{e}_T\mathbf{e}_T^\intercal & z\mathbb{I} \\ \hline \mathbf{C}^1 & \mathbf{C}^2 & \cdots & \mathbf{C}^T & \mathbf{C}^{T+1} \end{array} \right],$$

where for $i = 1, \ldots, T$, the matrix $\mathbf{C}^i \in \mathbb{R}^{\binom{T}{2} \times T}$ satisfies

$$
C^i_{st,j} = \begin{cases} x & \text{if } (s,t) \in \{(i,j), (j,i)\}, \\ y & \text{if } (s,t) \in \{(j,\tau) : \tau \neq i\} \cup \{(\tau, j) : \tau \neq i\}, \\ 0 & \text{otherwise.} \end{cases}
$$

Here, the indices s and t, $1 \leq s < t \leq T$, encode the row and the index j refers to the column of $\mathbf{C}^i$, respectively. The matrix $\mathbf{C}^{T+1}$ is defined analogously with x and y replaced by z.

Consider the linear combination $(\boldsymbol{m}^\intercal, \boldsymbol{v}^\intercal, \boldsymbol{c}^\intercal) \mathbf{J}$ of all rows of $\mathbf{J}$ with the coefficients m_t $(t = 1, \ldots, T)$ for the first block of T rows, v_t $(t = 1, \ldots, T)$ for the second block of T rows, and c_{st} for the third block of $\binom{T}{2}$ rows. For notational convenience, we define $c_{st} = c_{ts}$ for $s > t$. To prove that $\mathbf{J}$ has full row rank, we need to show that $(\boldsymbol{m}^\intercal, \boldsymbol{v}^\intercal, \boldsymbol{c}^\intercal) \mathbf{J}$ evaluates to $\mathbf{0}^\intercal$ only if $\boldsymbol{m}$, $\boldsymbol{v}$ and $\boldsymbol{c}$ vanish. To this end, consider the first and the $(T+1)$th element (i.e., the first elements of the first two column blocks) of the equation $(\boldsymbol{m}^\intercal, \boldsymbol{v}^\intercal, \boldsymbol{c}^\intercal) \mathbf{J} = \mathbf{0}^\intercal$, which are equivalent to

$$
m_1 + x v_1 + y \sum_{t=2}^T c_{1t} = 0 \quad \text{and} \quad m_1 + y v_1 + x c_{12} + y \sum_{t=3}^T c_{1t} = 0.
$$

Subtracting the two equations implies that $(x - y)(v_1 - c_{12}) = 0$, which in turn yields $v_1 = c_{12}$ since $x \neq y$. Generalizing this observation to the tth columns in each pair of column blocks s and t, we find that all v_t and c_{st} must be equal to a single variable v. Next, consider the $(T^2 + 1)$th and $(T^2 + 2)$th columns (i.e., the first two elements of the last column block) of the equation $(\boldsymbol{m}^\intercal, \boldsymbol{v}^\intercal, \boldsymbol{c}^\intercal) \mathbf{J} = \mathbf{0}^\intercal$, which are equivalent to

$$
m_1 + z v_1 + z \sum_{t=2}^T c_{1t} = 0 \quad \text{and} \quad m_2 + z v_2 + z \left(c_{21} + \sum_{t=3}^T c_{2t} \right) = 0.
$$

However, since $v_t = c_{st} = v$ for all s and t, we conclude that $m_1 = m_2$. Again, generalizing this observation to each pair of columns in the last column block, we can identify all m_t by a single number m. Replacing v_t and c_{st} by v and m_t by m, the previous two equations simplify to

$$
m + (x + (T-1)y)v = 0 \quad \text{and} \quad m + Tzv = 0,
$$

and we conclude that $m = v = 0$ since we established earlier that $x + (T-1)y \neq Tz$. Hence, the Jacobian $\mathbf{J}$ indeed has full row rank, which concludes Step 2.

Step 3: We have shown in Step 2 that $\mathcal{P}(\boldsymbol{\mu}^\epsilon, \boldsymbol{\Omega}^\epsilon) \neq \emptyset$ for all $\boldsymbol{\mu}^\epsilon \in \mathcal{B}_\epsilon(\boldsymbol{\mu})$ and $\boldsymbol{\Omega}^\epsilon \in \mathcal{B}_\epsilon(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal)$, which implies that $(1, \boldsymbol{\mu}, \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal) \in \operatorname{rel int} \mathcal{K}_1$. Since $\{\lambda \mathcal{K}_1 : \lambda \in \mathbb{R}_+\} \subseteq \mathcal{K}$, we have $\lambda \mathcal{P}(\boldsymbol{\mu}^\epsilon, \boldsymbol{\Omega}^\epsilon) \subseteq \mathcal{K}$ for all $\lambda \geq 0$. As the moments are linear in the measure, we thus conclude that $(1, \boldsymbol{\mu}, \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal) \in \operatorname{int} \mathcal{K}$ as desired. ∎

Theorem 2.1 will allow us to use the strong duality theorem of [27, Proposition 3.4], which states that a linear optimization problem over the distributions in $\mathcal{P}$ has the same optimal value

as its associated dual problem. In the remainder of the paper, we will make extensive use of this insight, and we therefore assume from now on that $\mu^2 + \rho\sigma^2 > 0$.

3   Left-Sided Chebyshev Bounds

In this section we study *left-sided Chebyshev bounds* of the form

$$\mathrm{L}(\gamma) = \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\left(\prod_{t=1}^{T} \tilde{\xi}_t \le \gamma\right),$$

where the ambiguity set $\mathcal{P}$ is defined in (8). We begin with the main result of this section.

Theorem 3.1 (Left-Sided Chebyshev Bound) *Let $\gamma > 0$. For all $T \ge 3$, the left-sided Chebyshev bound $\mathrm{L}(\gamma)$ coincides with the optimal objective value of the semidefinite program*

$$
\begin{aligned}
\inf \quad & \alpha + T\mu\beta + T(\mu^2 + \sigma^2)\gamma_1 + T\left[T\mu^2 + \sigma^2 + (T-1)\rho\sigma^2\right]\gamma_2 \\
\mathrm{s.t.} \quad & \alpha, \beta, \gamma_1, \gamma_2 \in \mathbb{R}, \ \lambda_1, \lambda_2, \lambda_3 \ge 0, \ \boldsymbol{p} \in \mathbb{R}^{2T+1}, \ \boldsymbol{P} \in \mathbb{S}_+^{T+1}, \ \boldsymbol{q} \in \mathbb{R}^{2T-1}, \ \boldsymbol{Q} \in \mathbb{S}_+^T \\
& \alpha \ge 1, \quad \gamma_1 + \gamma_2 \ge 0, \quad \gamma_1 + T\gamma_2 \ge 0 \\
& \gamma_2 + \frac{\gamma_1}{T} + \alpha \ge \left\|\left(\beta - \lambda_1, \gamma_2 + \frac{\gamma_1}{T} - \alpha\right)\right\|_2 \\
& \gamma_2 + \gamma_1 + \alpha - 1 \ge \|(\beta - \lambda_2, \gamma_2 + \gamma_1 - \alpha + 1)\|_2 \\
& \gamma_2 + \frac{\gamma_1}{T} + \lambda_3 + \alpha - 1 \ge \left\|\left(\beta - \lambda_3 T\gamma^{1/T}, \gamma_2 + \frac{\gamma_1}{T} + \lambda_3 - \alpha + 1\right)\right\|_2 \\
& p_0 = (T-1)\gamma_1\gamma^{\frac{2}{T-1}} + (T-1)^2\gamma_2\gamma^{\frac{2}{T-1}}, \quad p_1 + q_0 = (T-1)\beta\gamma^{\frac{1}{T-1}} \\
& p_2 + q_1 = \alpha - 1, \quad p_T + q_{T-1} = 2(T-1)\gamma_2\gamma^{\frac{1}{T-1}}, \quad p_{T+1} + q_T = \beta \\
& p_{2T} = \gamma_1 + \gamma_2, \quad p_t + q_{t-1} = 0 \quad \forall t = 3, \ldots, T-1, T+2, \ldots, 2T-1 \\
& p_t = \sum_{i+j=t} P_{i,j} \quad \forall t = 0, \ldots, 2T, \quad q_t = \sum_{i+j=t} Q_{i,j} \quad \forall t = 0, \ldots, 2T-2,
\end{aligned}
\tag{13}
$$

where we use the convention that the entries of $\boldsymbol{p}$, $\boldsymbol{P}$, $\boldsymbol{q}$ and $\boldsymbol{Q}$ are numbered starting from 0. For $T = 2$, $\mathrm{L}(\gamma)$ is given by a variant of (13) where the constraints $p_2 + q_1 = \alpha - 1$ and $p_T + q_{T-1} = 2(T-1)\gamma_2\gamma^{\frac{1}{T-1}}$ are combined to $p_2 + q_1 = \alpha - 1 + 2(T-1)\gamma_2\gamma^{\frac{1}{T-1}}$.

Proof: We first reformulate the maximum probability of the left tail of the product $\prod_{t=1}^{T}\tilde{\xi}_t$ falling below γ as the generalized moment problem

$$
\begin{aligned}
\mathrm{L}(\gamma) = \sup \quad & \int_{\mathbb{R}_+^T} \mathbb{1}_{\{\prod_{t=1}^{T}\xi_t \le \gamma\}} \, \mathbb{P}(\mathrm{d}\boldsymbol{\xi}) \\
\mathrm{s.t.} \quad & \mathbb{P} \in \mathcal{M}_+(\mathbb{R}_+^T) \\
& \int_{\mathbb{R}_+^T} \mathbb{P}(\mathrm{d}\boldsymbol{\xi}) = 1 \\
& \int_{\mathbb{R}_+^T} \boldsymbol{\xi} \, \mathbb{P}(\mathrm{d}\boldsymbol{\xi}) = \boldsymbol{\mu} \\
& \int_{\mathbb{R}_+^T} \boldsymbol{\xi}\boldsymbol{\xi}^{\mathsf{T}} \, \mathbb{P}(\mathrm{d}\boldsymbol{\xi}) = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^{\mathsf{T}}.
\end{aligned}
\tag{14}
$$

This moment problem admits a strong conic dual in the Lagrange multipliers $\alpha \in \mathbb{R}$, $\boldsymbol{\beta} \in \mathbb{R}^T$ and $\boldsymbol{\Gamma} \in \mathbb{S}^T$ corresponding to the normalization, mean and covariance constraints in (14), respectively, see Theorem 2.1 and [27, Proposition 3.4]. Recalling that $\boldsymbol{\mu} = \mu\mathbf{1}$ and $\boldsymbol{\Sigma} = (1-\rho)\sigma^2\mathbb{I} + \rho\sigma^2\mathbf{1}\mathbf{1}^\intercal$, the dual problem can be expressed as

$$
\begin{aligned}
\mathrm{L}(\gamma) = \ \inf \quad & \alpha + \mu\mathbf{1}^\intercal\boldsymbol{\beta} + \left\langle (1-\rho)\sigma^2\mathbb{I} + \left(\mu^2 + \rho\sigma^2\right)\mathbf{1}\mathbf{1}^\intercal, \boldsymbol{\Gamma} \right\rangle \\
\mathrm{s.\,t.} \quad & \alpha \in \mathbb{R}, \ \boldsymbol{\beta} \in \mathbb{R}^T, \ \boldsymbol{\Gamma} \in \mathbb{S}^T \\
& \alpha + \boldsymbol{\xi}^\intercal\boldsymbol{\beta} + \boldsymbol{\xi}^\intercal\boldsymbol{\Gamma}\boldsymbol{\xi} \geq 0 \quad \forall \boldsymbol{\xi} \geq \mathbf{0} \\
& \alpha + \boldsymbol{\xi}^\intercal\boldsymbol{\beta} + \boldsymbol{\xi}^\intercal\boldsymbol{\Gamma}\boldsymbol{\xi} \geq 1 \quad \forall \boldsymbol{\xi} \geq \mathbf{0} : \ \textstyle\prod_{t=1}^{T}\xi_t \leq \gamma.
\end{aligned}
\tag{15}
$$

By Lemma 3.1 below, the symmetry of problem (15) implies that we may restrict attention to permutation-symmetric solutions of the form $(\alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma})$ with $\boldsymbol{\beta} = \beta\mathbf{1}$ and $\boldsymbol{\Gamma} = \gamma_1\mathbb{I} + \gamma_2\mathbf{1}\mathbf{1}^\intercal$ for some $\beta, \gamma_1, \gamma_2 \in \mathbb{R}$. Thus, problem (15) simplifies to

$$
\begin{aligned}
\mathrm{L}(\gamma) = \ \inf \quad & \alpha + T\mu\beta + T(\mu^2 + \sigma^2)\gamma_1 + T\left[T\mu^2 + \sigma^2 + (T-1)\rho\sigma^2\right]\gamma_2 \\
\mathrm{s.\,t.} \quad & \alpha, \beta, \gamma_1, \gamma_2 \in \mathbb{R} \\
& \alpha + \beta\|\boldsymbol{\xi}\|_1 + \gamma_1\|\boldsymbol{\xi}\|_2^2 + \gamma_2\|\boldsymbol{\xi}\|_1^2 \geq 0 \quad \forall \boldsymbol{\xi} \geq \mathbf{0} \\
& \alpha + \beta\|\boldsymbol{\xi}\|_1 + \gamma_1\|\boldsymbol{\xi}\|_2^2 + \gamma_2\|\boldsymbol{\xi}\|_1^2 \geq 1 \quad \forall \boldsymbol{\xi} \geq \mathbf{0} : \ \textstyle\prod_{t=1}^{T}\xi_t \leq \gamma.
\end{aligned}
\tag{16}
$$

Lemma 3.2 then implies that (16) can be reduced to

$$
\begin{aligned}
\mathrm{L}(\gamma) = \ \inf \quad & \alpha + T\mu\beta + T(\mu^2 + \sigma^2)\gamma_1 + T\left[T\mu^2 + \sigma^2 + (T-1)\rho\sigma^2\right]\gamma_2 \\
\mathrm{s.\,t.} \quad & \alpha, \beta, \gamma_1, \gamma_2 \in \mathbb{R} \\
& \inf_{s \geq 0} \ \alpha + \beta s + \gamma_2 s^2 + \frac{\gamma_1}{T}s^2 \geq 0 \\
& \inf_{s \geq 0} \ \alpha + \beta s + \gamma_2 s^2 + \gamma_1 s^2 \geq 1 \\
& \inf_{s \geq 0} \ \alpha + \beta s + \gamma_2 s^2 + \gamma_1 s^2 f_T\left(0, \frac{\gamma}{s^T}\right) \geq 1.
\end{aligned}
\tag{17}
$$

By assigning a Lagrange multiplier $\lambda_1 \geq 0$ to the constraint $s \geq 0$ and using the $\mathcal{S}$-lemma [25], the first constraint in (17) can be reformulated as the linear matrix inequality

$$
\begin{bmatrix} \gamma_2 + \frac{\gamma_1}{T} & \frac{\beta - \lambda_1}{2} \\ \frac{\beta - \lambda_1}{2} & \alpha \end{bmatrix} \succeq \mathbf{0}
\iff
\begin{cases} \alpha \geq 0 \\ \gamma_2 + \frac{\gamma_1}{T} \geq 0 \\ (\gamma_2 + \frac{\gamma_1}{T})\alpha \geq \frac{1}{4}(\beta - \lambda_1)^2 \end{cases}
$$

$$
\iff
\begin{cases} \alpha \geq 0 \\ \gamma_1 + T\gamma_2 \geq 0 \\ \gamma_2 + \frac{\gamma_1}{T} + \alpha \geq \left\| \left(\beta - \lambda_1, \gamma_2 + \frac{\gamma_1}{T} - \alpha\right) \right\|_2, \end{cases}
$$

where the first equivalence follows from the observation that a 2×2-matrix is positive semidefinite iff it has non-negative diagonal elements as well as a non-negative determinant, while the second equivalence uses a well-known reformulation of hyperbolic constraints as second-order cone constraints [6, p. 197]. Similarly, the second constraint in (17) holds iff there exists

$\lambda_2 \geq 0$ with

$$\begin{bmatrix} \gamma_2 + \gamma_1 & \frac{\beta - \lambda_2}{2} \\ \frac{\beta - \lambda_2}{2} & \alpha - 1 \end{bmatrix} \succeq \mathbf{0} \iff \begin{cases} \alpha \geq 1 \\ \gamma_2 + \gamma_1 \geq 0 \\ \gamma_2 + \gamma_1 + \alpha - 1 \geq \|(\beta - \lambda_2, \gamma_2 + \gamma_1 - \alpha + 1)\|_2 \,. \end{cases}$$

Lemma 3.3 below further allows us to decompose the third constraint in (17) into two simpler semi-infinite constraints.

$$\inf_{s \in [0, T\gamma^{1/T}]} \alpha + \beta s + \gamma_2 s^2 + \gamma_1 \frac{s^2}{T} \geq 1 \tag{18a}$$

$$\inf_{s \geq T\gamma^{1/T}} \left\{ \alpha + \beta s + \gamma_2 s^2 + \gamma_1 \min_{\underline{\xi}, \overline{\xi} \geq 0} \left\{ \underline{\xi}^2 + (T-1)\overline{\xi}^2 : \underline{\xi} + (T-1)\overline{\xi} = s, \ \underline{\xi}\,\overline{\xi}^{T-1} = \gamma \right\} \right\} \geq 1 \tag{18b}$$

As $s \in \left[0, T\gamma^{1/T}\right]$ iff $s(T\gamma^{1/T} - s) \geq 0$, we can once again use the $\mathcal{S}$-lemma to show that (18a) holds iff there exists $\lambda_3 \geq 0$ with

$$\begin{bmatrix} \gamma_2 + \frac{\gamma_1}{T} + \lambda_3 & \frac{\beta - \lambda_3 T\gamma^{1/T}}{2} \\ \frac{\beta - \lambda_3 T\gamma^{1/T}}{2} & \alpha - 1 \end{bmatrix} \succeq \mathbf{0} \iff \begin{cases} \alpha \geq 1 \\ \gamma_2 + \frac{\gamma_1}{T} + \lambda_3 \geq 0 \\ \gamma_2 + \frac{\gamma_1}{T} + \lambda_3 + \alpha - 1 \\ \qquad \geq \left\|\left(\beta - \lambda_3 T\gamma^{1/T}, \gamma_2 + \frac{\gamma_1}{T} + \lambda_3 - \alpha + 1\right)\right\|_2 \,. \end{cases}$$

Finally, it remains to be shown that (18b) also admits a conic reformulation. To do so, we first argue that one can replace (18b) with

$$\inf_{s \geq T\gamma^{1/T}, \underline{\xi}, \overline{\xi} \geq 0} \left\{ \alpha + \beta s + \gamma_2 s^2 + \gamma_1 \left[\underline{\xi}^2 + (T-1)\overline{\xi}^2 \right] : \underline{\xi} + (T-1)\overline{\xi} = s, \ \underline{\xi}\,\overline{\xi}^{T-1} = \gamma \right\} \geq 1 \tag{19}$$

without changing the optimal value of problem (17). If $\gamma_1 \geq 0$, then (19) is indeed equivalent to (18b). On the other hand, if $\gamma_1 < 0$, we find

$$\inf_{s \geq T\gamma^{1/T}} \left\{ \alpha + \beta s + \gamma_2 s^2 + \gamma_1 \min_{\underline{\xi}, \overline{\xi} \geq 0} \left\{ \underline{\xi}^2 + (T-1)\overline{\xi}^2 : \underline{\xi} + (T-1)\overline{\xi} = s, \ \underline{\xi}\,\overline{\xi}^{T-1} = \gamma \right\} \right\}$$

$$\geq \inf_{s \geq T\gamma^{1/T}} \left\{ \alpha + \beta s + \gamma_2 s^2 + \gamma_1 \max_{\underline{\xi}, \overline{\xi} \geq 0} \left\{ \underline{\xi}^2 + (T-1)\overline{\xi}^2 : \underline{\xi} + (T-1)\overline{\xi} = s, \ \underline{\xi}\,\overline{\xi}^{T-1} = \gamma \right\} \right\}$$

$$= \inf_{s \geq T\gamma^{1/T}, \underline{\xi}, \overline{\xi} \geq 0} \left\{ \alpha + \beta s + \gamma_2 s^2 + \gamma_1 \left[\underline{\xi}^2 + (T-1)\overline{\xi}^2 \right] : \underline{\xi} + (T-1)\overline{\xi} = s, \ \underline{\xi}\,\overline{\xi}^{T-1} = \gamma \right\}$$

$$\geq \inf_{s \geq T\gamma^{1/T}} \alpha + \beta s + \gamma_2 s^2 + \gamma_1 s^2,$$

which means that (18b) is implied by the second semi-infinite constraint in problem (17). By eliminating $s = \underline{\xi} + (T-1)\overline{\xi}$, the maximization problem on the left hand side of (19) reduces to

$$\inf_{\underline{\xi}, \overline{\xi} \geq 0, \ \underline{\xi}\,\overline{\xi}^{T-1} = \gamma} \alpha + \beta \left[\underline{\xi} + (T-1)\overline{\xi} \right] + \gamma_2 \left[\underline{\xi} + (T-1)\overline{\xi} \right]^2 + \gamma_1 \left[\underline{\xi}^2 + (T-1)\overline{\xi}^2 \right].$$

Note that the constraint $s \geq T\gamma^{1/T}$ has been dropped in the above formulation. This constraint is redundant due to the inequality of arithmetic and geometric means, which implies that

$$s = \underline{\xi} + (T-1)\overline{\xi} \geq T(\underline{\xi}\,\overline{\xi}^{T-1})^{1/T} = T\gamma^{1/T}.$$

By setting $\kappa = \underline{\xi}^{1/(T-1)}$, we can further replace $\underline{\xi}$ and $\overline{\xi}$ with κ^{T-1} and $\gamma^{1/(T-1)}/\kappa$, respectively. Using elementary manipulations, one can then show that (19) reduces to

$$\inf_{\kappa \geq 0}\ (T-1)\gamma_1\gamma^{\frac{2}{T-1}} + (T-1)^2\gamma_2\gamma^{\frac{2}{T-1}} + (T-1)\beta\gamma^{\frac{1}{T-1}}\kappa + (\alpha-1)\kappa^2$$
$$+ 2(T-1)\gamma_2\gamma^{\frac{1}{T-1}}\kappa^T + \beta\kappa^{T+1} + (\gamma_1+\gamma_2)\kappa^{2T}\ \geq\ 0. \qquad (20)$$

Note that the objective of the maximization problem on the left hand side of (20) constitutes a polynomial of degree $2T$ in κ and is therefore representable as $l(\kappa) = \sum_{i=0}^{2T} a_i\kappa^i$, where

$$a_i = \begin{cases} (T-1)\gamma_1\gamma^{\frac{2}{T-1}} + (T-1)^2\gamma_2\gamma^{\frac{2}{T-1}} & \text{if } i = 0, \\ (T-1)\beta\gamma^{\frac{1}{T-1}} & \text{if } i = 1, \\ \alpha-1 & \text{if } i = 2, \\ 2(T-1)\gamma_2\gamma^{\frac{1}{T-1}} & \text{if } i = T, \\ \beta & \text{if } i = T+1, \\ \gamma_1+\gamma_2 & \text{if } i = 2T, \\ 0 & \text{otherwise.} \end{cases} \qquad (21)$$

Here we assumed that $T > 2$. For $T = 2$, the quadratic monomial in $l(\kappa)$ would have the coefficient $\alpha - 1 + 2(T-1)\gamma_2\gamma^{\frac{1}{T-1}}$ instead of $\alpha - 1$. Thus, the case $T = 2$ could be handled via a case distinction, which we omit for the sake of brevity.

Constraint (19) thus requires the polynomial $l(\kappa)$ to be non-negative for all $\kappa \geq 0$. By the Markov-Lukacs Theorem [19], this is equivalent to postulating that $l(\kappa)$ admits a sum-of-squares representation of the form $l(\kappa) = p(\kappa) + \kappa q(\kappa)$, where $p(\kappa) = \sum_{i=0}^{2T} p_i\kappa^i$ and $q(\kappa) = \sum_{i=0}^{2T-2} q_i\kappa^i$ are sum-of-squares polynomials of degrees $2T$ and $2T-2$, respectively. By matching the coefficients of all monomials, one verifies that the identity $l(\kappa) = p(\kappa) + \kappa q(\kappa)$ holds iff

$$p_0 = a_0, \quad p_t + q_{t-1} = a_t \quad \forall t = 1, \ldots, 2T-1 \quad \text{and} \quad p_{2T} = a_{2T}, \qquad (22)$$

Moreover, by [23, Theorem 3], $p(\kappa)$ and $q(\kappa)$ are sum-of-squares polynomials iff there exist positive semidefinite matrices $\boldsymbol{P} \in \mathbb{S}_+^{T+1}$ and $\boldsymbol{Q} \in \mathbb{S}_+^T$ such that

$$p_t = \sum_{i+j=t} P_{i,j} \quad \forall t = 0, \ldots, 2T \quad \text{and} \quad q_t = \sum_{i+j=t} Q_{i,j} \quad \forall t = 0, \ldots, 2T-2. \qquad (23)$$

Thus, (19) holds iff the conic constraints (22) and (23) are satisfied. The claim now follows by replacing the three semi-infinite constraints in (17) with their explicit conic reformulations. $\blacksquare$

The proof of Theorem 3.1 relies on 4 auxiliary lemmas, which we prove next.

Lemma 3.1 *Problem (15) has a permutation symmetric minimizer $(\alpha^\star, \boldsymbol{\beta}^\star, \boldsymbol{\Gamma}^\star)$ that satisfies $\boldsymbol{\beta}^\star = \beta^\star \mathbf{1}$ and $\boldsymbol{\Gamma}^\star = \gamma_1^\star \mathbb{I} + \gamma_2^\star \mathbf{1}\mathbf{1}^\mathsf{T}$ for some $\beta^\star, \gamma_1^\star, \gamma_2^\star \in \mathbb{R}$.*

Proof: Let $\mathfrak{P}$ be the set of all permutations of the index set $\{1, \ldots, T\}$. For any $\pi \in \mathfrak{P}$ we denote by $\mathbf{P}_\pi \in \mathbb{R}^{T \times T}$ the permutation matrix defined through $(\mathbf{P}_\pi)_{ij} = 1$ if $\pi(i) = j$; $= 0$ otherwise. Let $(\alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma})$ by any optimal solution to (15), which exists by [27, Proposition 3.4]. We first show that the permuted solution $(\alpha_\pi, \boldsymbol{\beta}_\pi, \boldsymbol{\Gamma}_\pi) = (\alpha, \mathbf{P}_\pi \boldsymbol{\beta}, \mathbf{P}_\pi \boldsymbol{\Gamma} \mathbf{P}_\pi^\mathsf{T})$ is also optimal in (15). To this end, we observe that

$$
\begin{aligned}
&\alpha_\pi + \mu \mathbf{1}^\mathsf{T} \boldsymbol{\beta}_\pi + \left\langle (1-\rho)\sigma^2 \mathbb{I} + \left(\mu^2 + \rho\sigma^2\right) \mathbf{1}\mathbf{1}^\mathsf{T}, \boldsymbol{\Gamma}_\pi \right\rangle \\
&= \alpha + \mu \mathbf{1}^\mathsf{T} \mathbf{P}_\pi \boldsymbol{\beta} + \left\langle (1-\rho)\sigma^2 \mathbb{I} + \left(\mu^2 + \rho\sigma^2\right) \mathbf{1}\mathbf{1}^\mathsf{T}, \mathbf{P}_\pi \boldsymbol{\Gamma} \mathbf{P}_\pi^\mathsf{T} \right\rangle \\
&= \alpha + \mu (\mathbf{P}_\pi^\mathsf{T} \mathbf{1})^\mathsf{T} \boldsymbol{\beta} + \left\langle (1-\rho)\sigma^2 \mathbf{P}_\pi^\mathsf{T} \mathbf{P}_\pi + \left(\mu^2 + \rho\sigma^2\right) \mathbf{P}_\pi^\mathsf{T} \mathbf{1}(\mathbf{P}_\pi^\mathsf{T} \mathbf{1})^\mathsf{T}, \boldsymbol{\Gamma} \right\rangle \\
&= \alpha + \mu \mathbf{1}^\mathsf{T} \boldsymbol{\beta} + \left\langle 1-\rho)\sigma^2 \mathbb{I} + \left(\mu^2 + \rho\sigma^2\right) \mathbf{1}\mathbf{1}^\mathsf{T}, \boldsymbol{\Gamma} \right\rangle,
\end{aligned}
$$

where the first equality follows from the definition of α_π, $\boldsymbol{\beta}_\pi$ and $\boldsymbol{\Gamma}_\pi$, the second equality exploits the cyclicity property of the trace scalar product, and the third equality holds due to the permutation symmetry of $\mathbf{1}$ and the fact that $\mathbf{P}_\pi^\mathsf{T} = \mathbf{P}_{\pi^{-1}} = \mathbf{P}_\pi^{-1}$. Thus, $(\alpha_\pi, \boldsymbol{\beta}_\pi, \boldsymbol{\Gamma}_\pi)$ has the same objective value as $(\alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma})$. To show that $(\alpha_\pi, \boldsymbol{\beta}_\pi, \boldsymbol{\Gamma}_\pi)$ is feasible in (15), we note that

$$
\begin{aligned}
&\alpha_\pi + \boldsymbol{\xi}^\mathsf{T} \boldsymbol{\beta}_\pi + \boldsymbol{\xi}^\mathsf{T} \boldsymbol{\Gamma}_\pi \boldsymbol{\xi} \geq 1_{\{\prod_{t=1}^T \xi_t \leq \gamma\}} && \forall \boldsymbol{\xi} \geq \mathbf{0} \\
\iff & \alpha + (\mathbf{P}_{\pi^{-1}} \boldsymbol{\xi})^\mathsf{T} \boldsymbol{\beta} + (\mathbf{P}_{\pi^{-1}} \boldsymbol{\xi})^\mathsf{T} \boldsymbol{\Gamma}(\mathbf{P}_{\pi^{-1}} \boldsymbol{\xi}) \geq 1_{\{\prod_{t=1}^T \xi_t \leq \gamma\}} && \forall \boldsymbol{\xi} \geq \mathbf{0} \\
\iff & \alpha + \boldsymbol{\xi}^\mathsf{T} \boldsymbol{\beta} + \boldsymbol{\xi}^\mathsf{T} \boldsymbol{\Gamma} \boldsymbol{\xi} \geq 1_{\{\prod_{t=1}^T \xi_{\pi(t)} \leq \gamma\}} && \forall \boldsymbol{\xi} \geq \mathbf{0} \\
\iff & \alpha + \boldsymbol{\xi}^\mathsf{T} \boldsymbol{\beta} + \boldsymbol{\xi}^\mathsf{T} \boldsymbol{\Gamma} \boldsymbol{\xi} \geq 1_{\{\prod_{t=1}^T \xi_t \leq \gamma\}} && \forall \boldsymbol{\xi} \geq \mathbf{0},
\end{aligned}
$$

where the first equivalence follows from the definition of α_π, $\boldsymbol{\beta}_\pi$ and $\boldsymbol{\Gamma}_\pi$ and because $\mathbf{P}_\pi^\mathsf{T} = \mathbf{P}_\pi^{-1}$, the second equivalence holds because permutations are bijective, and the third equivalence relies on the permutation symmetry of the non-negative orthant. Thus, $(\alpha_\pi, \boldsymbol{\beta}_\pi, \boldsymbol{\Gamma}_\pi)$ satisfies the semi-infinite constraints in (15) whenever $(\alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma})$ does. We conclude that $(\alpha_\pi, \boldsymbol{\beta}_\pi, \boldsymbol{\Gamma}_\pi)$ is feasible and thus optimal in (15) for every $\pi \in \mathfrak{P}$.

Due to the convexity of the (semi-infinite) linear program (15), the equally weighted average $(\alpha^\star, \boldsymbol{\beta}^\star, \boldsymbol{\Gamma}^\star) = \frac{1}{T!} \sum_{\pi \in \mathfrak{P}} (\alpha_\pi, \boldsymbol{\beta}_\pi, \boldsymbol{\Gamma}_\pi)$ constitutes another optimal solution. It is now clear that $\mathbf{P}_\pi \boldsymbol{\beta}^\star = \boldsymbol{\beta}^\star$ and $\mathbf{P}_\pi \boldsymbol{\Gamma}^\star \mathbf{P}_\pi^\mathsf{T} = \boldsymbol{\Gamma}^\star$ for any $\pi \in \mathfrak{P}$ since $\pi(\mathfrak{P}) = \mathfrak{P}$. Thus, the claim follows. ∎

Lemma 3.2 *For $\alpha, \beta, \gamma_1, \gamma_2, \Delta \in \mathbb{R}$ and $\underline{\gamma}, \overline{\gamma} \in \mathbb{R}_+ \cup \{\infty\}$, $\underline{\gamma} \leq \overline{\gamma}$, we have*

$$
\begin{aligned}
&\inf_{\boldsymbol{\xi} \geq \mathbf{0}} \left\{ \alpha + \beta \|\boldsymbol{\xi}\|_1 + \gamma_1 \|\boldsymbol{\xi}\|_2^2 + \gamma_2 \|\boldsymbol{\xi}\|_1^2 \ : \ \prod_{t=1}^T \xi_t \in [\underline{\gamma}, \overline{\gamma}] \right\} \geq \Delta \\
&\iff \begin{cases} \displaystyle\inf_{s \geq T\underline{\gamma}^{1/T}} \alpha + \beta s + \gamma_2 s^2 + \gamma_1 s^2 \ f_T(\underline{\gamma}/s^T, \overline{\gamma}/s^T) \geq \Delta \\ \displaystyle\inf_{s \geq T\underline{\gamma}^{1/T}} \alpha + \beta s + \gamma_2 s^2 + \gamma_1 s^2 \ g_T(\underline{\gamma}/s^T, \overline{\gamma}/s^T) \geq \Delta, \end{cases}
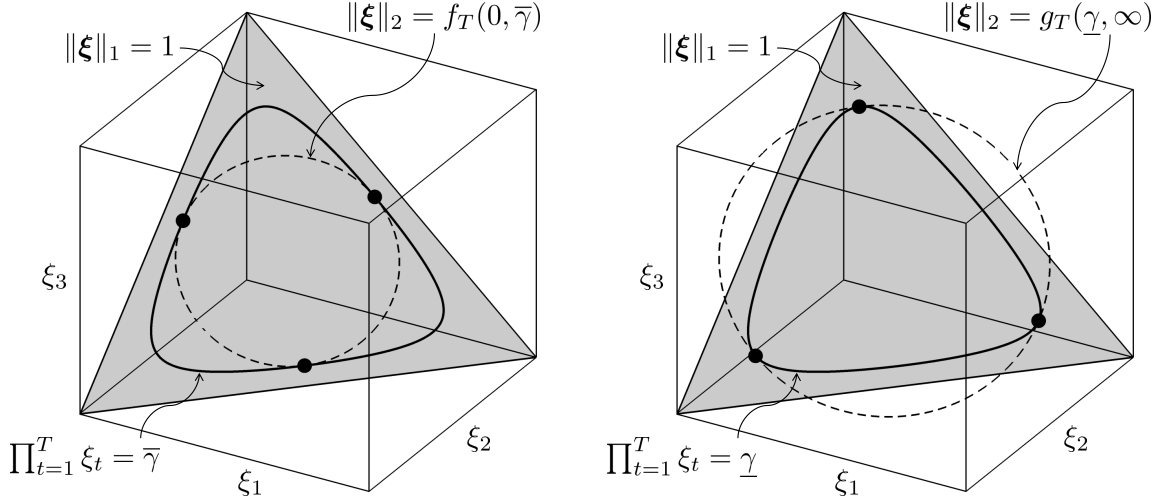\end{aligned}
\tag{24}
$$

Figure 1: The subproblems (25a) (left) and (25b) (right) determine the smallest and the largest spheres centered at the origin that intersect with the hyperplane $\|\boldsymbol{\xi}\|_1 = 1$ (shaded areas) and the hyperbola $\prod_{t=1}^T \xi_t = \overline{\gamma}, \underline{\gamma}$ (solid lines). The dashed circles represent level sets of the objective function $\|\boldsymbol{\xi}\|_2^2$. Both graphs illustrate the case where $T = 3$.

where

$$f_T(\underline{\gamma}, \overline{\gamma}) = \inf_{\boldsymbol{\xi} \geq \mathbf{0}} \left\{ \|\boldsymbol{\xi}\|_2^2 : \|\boldsymbol{\xi}\|_1 = 1, \ \prod_{t=1}^T \xi_t \in [\underline{\gamma}, \overline{\gamma}] \right\} \tag{25a}$$

$$and \quad g_T(\underline{\gamma}, \overline{\gamma}) = \sup_{\boldsymbol{\xi} \geq \mathbf{0}} \left\{ \|\boldsymbol{\xi}\|_2^2 : \|\boldsymbol{\xi}\|_1 = 1, \ \prod_{t=1}^T \xi_t \in [\underline{\gamma}, \overline{\gamma}] \right\}. \tag{25b}$$

Moreover, we have $f_T(\underline{\gamma}, \infty) = 1/T$ for $\underline{\gamma} \leq T^{-T}$ and $g_T(0, \overline{\gamma}) = 1$ for $\overline{\gamma} \in \mathbb{R}_+ \cup \{\infty\}$.

Figure 1 visualizes the two parametric subproblems (25a) and (25b). Note that both problems are non-convex whenever $\overline{\gamma} < \infty$ as their last constraints are equivalent to $(\prod_{t=1}^T \xi_t)^{1/T} \in [\underline{\gamma}^{1/T}, \overline{\gamma}^{1/T}]$ and because geometric means are concave [6, § 3.1]. Moreover, the subproblem (25b) remains non-convex for $\overline{\gamma} = \infty$ since it maximizes a convex objective function.

Proof of Lemma 3.2: The first constraint in (24) can be reduced to

$$\inf_{s \geq T\underline{\gamma}^{1/T}} \alpha + \beta s + \gamma_2 s^2 + \inf_{\boldsymbol{\xi} \geq \mathbf{0}} \left\{ \gamma_1 \|\boldsymbol{\xi}\|_2^2 : \|\boldsymbol{\xi}\|_1 = s, \ \prod_{t=1}^T \xi_t \in [\underline{\gamma}, \overline{\gamma}] \right\} \geq \Delta \tag{26}$$

by decomposing the maximization over all $\boldsymbol{\xi} \geq \mathbf{0}$ into two nested maximization problems over all $s \geq T\underline{\gamma}^{1/T}$ and over all $\boldsymbol{\xi} \geq \mathbf{0}$ with $\|\boldsymbol{\xi}\|_1 = s$, respectively. Here, the lower bound on s is owed to the fact that there is $\boldsymbol{\xi} \geq \mathbf{0}$ satisfying $\|\boldsymbol{\xi}\|_1 = s$ and $\prod_{t=1}^T \xi_t \in [\underline{\gamma}, \overline{\gamma}]$ if and only if $s \geq T\underline{\gamma}^{1/T}$. A case distinction on the sign of γ_1 shows that constraint (26) holds if and only if

$$\begin{cases} \displaystyle\inf_{s \geq T\underline{\gamma}^{1/T}} \alpha + \beta s + \gamma_2 s^2 + \gamma_1 \inf_{\boldsymbol{\xi} \geq \mathbf{0}} \left\{ \|\boldsymbol{\xi}\|_2^2 : \|\boldsymbol{\xi}\|_1 = s, \ \prod_{t=1}^T \xi_t \in [\underline{\gamma}, \overline{\gamma}] \right\} \geq \Delta \\[2mm] \displaystyle\inf_{s \geq T\underline{\gamma}^{1/T}} \alpha + \beta s + \gamma_2 s^2 + \gamma_1 \sup_{\boldsymbol{\xi} \geq \mathbf{0}} \left\{ \|\boldsymbol{\xi}\|_2^2 : \|\boldsymbol{\xi}\|_1 = s, \ \prod_{t=1}^T \xi_t \in [\underline{\gamma}, \overline{\gamma}] \right\} \geq \Delta \end{cases}$$

17

is satisfied. The change of variables $\boldsymbol{\xi} \leftarrow s\boldsymbol{\xi}$ shows that this constraint system is equivalent to the second constraint system in (24). Finally, we have $f_T(\underline{\gamma}, \infty) = 1/T$ for $\underline{\gamma} \leq T^{-T}$ and $g_T(0, \overline{\gamma}) = 1$ for $\overline{\gamma} \in \mathbb{R}_+ \cup \{\infty\}$ since the inequalities $\frac{1}{T}\|\boldsymbol{\xi}\|_1^2 \leq \|\boldsymbol{\xi}\|_2^2 \leq \|\boldsymbol{\xi}\|_1^2$ are tight for $\boldsymbol{\xi} = \mathbf{1}$ and $\boldsymbol{\xi} = \mathbf{e}_i$, respectively. ∎

Lemma 3.3 *For $T \geq 2$, $\underline{\gamma} = 0$ and $\overline{\gamma} \geq 0$, the optimal value $f_T(0, \overline{\gamma})$ of (25a) equals*

$$f_T(0, \overline{\gamma}) = \begin{cases} \min_{\underline{\xi} \geq 0, \overline{\xi} \geq 0} \left\{ \underline{\xi}^2 + (T-1)\overline{\xi}^2 : \underline{\xi} + (T-1)\overline{\xi} = 1, \ \underline{\xi}\, \overline{\xi}^{T-1} = \overline{\gamma} \right\} & \text{if } 0 \leq \overline{\gamma} \leq T^{-T}, \\ \frac{1}{T} & \text{if } \overline{\gamma} > T^{-T}. \end{cases} \tag{27}$$

Proof: We first observe that the non-convex optimization problem (25a) is bounded below by its relaxation $\min_{\|\boldsymbol{\xi}\|_1 = 1} \|\boldsymbol{\xi}\|_2^2$. Note, however, that the optimal solution $\boldsymbol{\xi} = \frac{1}{T}\mathbf{1}$ of this relaxation is feasible and thus optimal in (25a) whenever $\overline{\gamma} \geq T^{-T}$. Thus, we have $f_T(0, \overline{\gamma}) = \frac{1}{T}$ for $\overline{\gamma} \geq T^{-T}$. For $0 \leq \overline{\gamma} < T^{-T}$, on the other hand, the product constraint $\prod_{t=1}^T \xi_t \leq \overline{\gamma}$ must be binding, for otherwise convex combinations of the optimal solution $\boldsymbol{\xi}$ with $\frac{1}{T}\mathbf{1}$ would improve the objective function of $f_T(0, \overline{\gamma})$, which is a contradiction. In summary, we thus find

$$f_T(0, \overline{\gamma}) = \begin{cases} \inf_{\boldsymbol{\xi} \geq \mathbf{0}} \left\{ \|\boldsymbol{\xi}\|_2^2 : \|\boldsymbol{\xi}\|_1 = 1, \ \prod_{t=1}^T \xi_t = \overline{\gamma} \right\} & \text{if } 0 \leq \overline{\gamma} < T^{-T}, \\ \frac{1}{T} & \text{if } \overline{\gamma} \geq T^{-T}. \end{cases} \tag{28}$$

When $\overline{\gamma} = 0$, the product constraint in the first line of (28) can only be satisfied if $\xi_t = 0$ for at least one t. By permutation symmetry, we may assume without loss of generality that $\xi_T = 0$. Then, the product constraint is automatically satisfied and may be disregarded, implying that the minimization problem in the first line of (28) is solved by $\xi_1 = \xi_2 = \cdots = \xi_{T-1} = \frac{1}{T-1}$ and $\xi_T = 0$. We thus conclude that $f_T(0, 0) = \frac{1}{T-1}$ and therefore

$$f_T(0, \overline{\gamma}) = \begin{cases} \frac{1}{T-1} & \text{if } \overline{\gamma} = 0, \\ \inf_{\boldsymbol{\xi} > \mathbf{0}} \left\{ \|\boldsymbol{\xi}\|_2^2 : \|\boldsymbol{\xi}\|_1 = 1, \ \prod_{t=1}^T \xi_t = \overline{\gamma} \right\} & \text{if } 0 < \overline{\gamma} < T^{-T}, \\ \frac{1}{T} & \text{if } \overline{\gamma} \geq T^{-T}. \end{cases} \tag{29}$$

We now study the non-convex parametric optimization problem

$$\min_{\boldsymbol{\xi} > \mathbf{0}} \left\{ \|\boldsymbol{\xi}\|_2^2 : \|\boldsymbol{\xi}\|_1 = 1, \ \prod_{t=1}^T \xi_t = \overline{\gamma} \right\} \tag{30}$$

on the domain $0 < \overline{\gamma} < T^{-T}$. Observe that (30) has a non-empty compact feasible set for any admissible $\overline{\gamma}$ and is therefore solvable. Assigning Lagrange multipliers a and b to the norm and product constraints, respectively, we find that any optimal solution to (30) must satisfy the stationarity conditions

$$2\xi_t + a + \frac{b}{\xi_t} \prod_{t'=1}^T \xi_{t'} = 0 \quad \forall t = 1, \ldots, T \quad \Longleftrightarrow \quad 2\xi_t^2 + a\xi_t + b\overline{\gamma} = 0 \quad \forall t = 1, \ldots, T,$$

where the equivalence follows from primal feasibility. Note that each ξ_t needs to satisfy an identical quadratic equation, which must have two distinct positive real roots[1] $\underline{\xi}$ and $\overline{\xi}$. The roots depend on a, b and $\overline{\gamma}$, but this dependence is notationally suppressed to avoid clutter. At optimality, the decision variables $\xi_1, \xi_2 \ldots, \xi_T$ can thus be partitioned into two groups, where all variables in the first group are equal to $\underline{\xi}$, and all variables in the second group are equal to $\overline{\xi}$. This structural insight allows us to simplify problem (30). Indeed, by permutation symmetry, it is sufficient to consider only solutions that satisfy $\xi_1 = \cdots = \xi_k = \underline{\xi}$ and $\xi_{k+1} = \cdots = \xi_T = \overline{\xi}$ for some $\underline{\xi}, \overline{\xi} > 0$ and for some $k \in \{1, \ldots, \lfloor \frac{T}{2} \rfloor\}$. Thus, the optimal value of (30) coincides with

$$\min_{k \in \{1, \ldots, \lfloor \frac{T}{2} \rfloor\}} f_{T,k}(\overline{\gamma}), \tag{31}$$

where the functions $f_{T,k} : (0, T^{-T}) \to \mathbb{R}$ for $k = 1, 2, \ldots, \lfloor \frac{T}{2} \rfloor$ are defined through

$$f_{T,k}(\overline{\gamma}) = \min_{\underline{\xi} > 0, \overline{\xi} > 0} \left\{ k\underline{\xi}^2 + (T-k)\overline{\xi}^2 : k\underline{\xi} + (T-k)\overline{\xi} = 1, \ \underline{\xi}^k \overline{\xi}^{T-k} = \overline{\gamma} \right\}. \tag{32}$$

By Lemma 3.4 below, the optimal value of (31) is given by $f_{T,1}(\overline{\gamma})$. Hence, if we replace the minimization problem in (29) with $f_{T,1}(\overline{\gamma})$, we obtain

$$f_T(0, \overline{\gamma}) = \begin{cases} \min_{\underline{\xi} \geq 0, \overline{\xi} \geq 0} \left\{ \underline{\xi}^2 + (T-1)\overline{\xi}^2 : \underline{\xi} + (T-1)\overline{\xi} = 1, \ \underline{\xi} \overline{\xi}^{T-1} = \overline{\gamma} \right\} & \text{if } 0 \leq \overline{\gamma} < T^{-T}, \\ \frac{1}{T} & \text{if } \overline{\gamma} \geq T^{-T}. \end{cases}$$

The statement of the lemma now follows since the minimization problem in the equation above evaluates to $1/T$ at $\overline{\gamma} = T^{-T}$. Indeed, the minimization problem is bounded below by $\min_{\|\xi\|_1 = 1} \|\xi\|_2^2$, and the optimal value $1/T$ of this bound is achieved by the feasible solution $\underline{\xi} = \overline{\xi} = 1/T$ of the minimization problem at $\overline{\gamma} = T^{-T}$. ∎

Lemma 3.4 *For $T \geq 2$ and $0 < \overline{\gamma} < T^{-T}$, the optimal value of (31) is given by $f_{T,1}(\overline{\gamma})$.*

Proof: The statement holds trivially true when $\lfloor \frac{T}{2} \rfloor = 1$, that is, for $T \in \{2, 3\}$. Next, we show that $f_{4,1}(\overline{\gamma}) < f_{4,2}(\overline{\gamma})$ for any $\overline{\gamma} \in (0, 4^{-4})$. This inequality not only implies that the statement holds true for $T = 4$ but will also be instrumental for proving the statement for $T > 4$.

Fix $\overline{\gamma} \in (0, 4^{-4})$ and note that

$$\begin{aligned} f_{4,2}(\overline{\gamma}) &= \min_{\underline{\xi} > 0, \overline{\xi} > 0} \left\{ 2\underline{\xi}^2 + 2\overline{\xi}^2 : 2\underline{\xi} + 2\overline{\xi} = 1, \ \underline{\xi}^2 \overline{\xi}^2 = \overline{\gamma} \right\} \\ &= \frac{1}{2} \min_{\underline{\xi} > 0, \overline{\xi} > 0} \left\{ \underline{\xi}^2 + \overline{\xi}^2 : \underline{\xi} + \overline{\xi} = 1, \ \underline{\xi}\overline{\xi} = 4\sqrt{\overline{\gamma}} \right\} \\ &= \frac{1}{2} f_{2,1}(4\sqrt{\overline{\gamma}}) \ = \ \frac{1}{2} - 4\sqrt{\overline{\gamma}}, \end{aligned}$$

where the second equality follows from the substitution $\underline{\xi} \leftarrow 2\underline{\xi}$ and $\overline{\xi} \leftarrow 2\overline{\xi}$, and the last equality

---

[1] The existence of at least one real root is guaranteed because (30) is solvable and because any optimal solution must satisfy the stationarity conditions. In fact, the stationarity conditions must admit *two* distinct positive real roots because otherwise $\xi = \frac{1}{T}\mathbf{1}$ would be the only conceivable optimal solution, which is impossible for $\overline{\gamma} < T^{-T}$.

holds because $f_{2,1}(\overline{\gamma}) = 1 - 2\overline{\gamma}$ for any $\overline{\gamma} \in (0, 2^{-2})$, which can be verified by direct calculation. Thus, we need to show that $f_{4,1}(\overline{\gamma}) < \frac{1}{2} - 4\sqrt{\overline{\gamma}}$, where

$$
\begin{aligned}
f_{4,1}(\overline{\gamma}) &= \min_{\underline{\xi} > 0, \overline{\xi} > 0} \left\{ \underline{\xi}^2 + 3\overline{\xi}^2 : \underline{\xi} + 3\overline{\xi} = 1, \ \underline{\xi}\,\overline{\xi}^3 = \overline{\gamma} \right\} \\
&= \min_{\overline{\xi} > 0} \left\{ (1 - 3\overline{\xi})^2 + 3\overline{\xi}^2 : (1 - 3\overline{\xi})\overline{\xi}^3 = \overline{\gamma} \right\}.
\end{aligned}
\tag{33}
$$

It is therefore sufficient to find $\overline{\xi}^\star$ feasible in (33) with

$$
(1 - 3\overline{\xi}^\star)^2 + 3(\overline{\xi}^\star)^2 < 1/2 - 4\sqrt{\overline{\gamma}} \quad \Longleftrightarrow \quad 12(\overline{\xi}^\star)^2 - 6\overline{\xi}^\star + (1/2 + 4\sqrt{\overline{\gamma}}) < 0
$$
$$
\Longleftrightarrow \quad \overline{\xi}^\star \in \left(\zeta^-, \zeta^+\right),
$$

where $\zeta^{\pm} = (3 \pm \sqrt{3 - 48\sqrt{\overline{\gamma}}})/12$ are the roots of $12(\overline{\xi}^\star)^2 - 6\overline{\xi}^\star + (1/2 + 4\sqrt{\overline{\gamma}})$. Equivalently, we should demonstrate the existence of some $\overline{\xi}^\star \in (\zeta^-, \zeta^+)$ with $(1 - 3\overline{\xi}^\star)(\overline{\xi}^\star)^3 - \overline{\gamma} = 0$. By the intermediate value theorem, this holds if

$$
(1 - 3\zeta^-)(\zeta^-)^3 - \overline{\gamma} > 0 \quad \text{and} \quad (1 - 3\zeta^+)(\zeta^+)^3 - \overline{\gamma} < 0.
\tag{34}
$$

But these inequalities are automatically satisfied under the assumption that $\overline{\gamma} \in (0, 4^{-4})$. Indeed, recalling the definition of ζ^- and defining $z^- = 12\zeta^- - 3 = -\sqrt{3 - 48\sqrt{\overline{\gamma}}}$, we have

$$
(1 - 3\zeta^-)(\zeta^-)^3 - \overline{\gamma} = \left(1 - \frac{3 + z^-}{4}\right)\left(\frac{3 + z^-}{12}\right)^3 - \left(\frac{3 - (z^-)^2}{48}\right)^2 = -\frac{1}{12^3}(z^-)^3(z^- + 2) > 0,
$$

where the inequality holds because $z^- \in (-\sqrt{3}, 0)$ for $\overline{\gamma} \in (0, 4^{-4})$. Similarly, defining $z^+ = 12\zeta^+ - 3 = \sqrt{3 - 48\sqrt{\overline{\gamma}}}$, we can prove that $(1 - 3\zeta^+)(\zeta^+)^3 - \overline{\gamma} < 0$. Thus, we have shown that $f_{4,1}(\overline{\gamma}) < f_{4,2}(\overline{\gamma})$ for any $\overline{\gamma} \in (0, 4^{-4})$, which establishes the assertion for $T = 4$.

Fix now some $T \geq 5$ and assume for the sake of argument that there exist $k \in \{2, \ldots, \lfloor \frac{T}{2} \rfloor\}$ and $\overline{\gamma} \in (0, T^{-T})$ with $f_T(0, \overline{\gamma}) = f_{T,k}(\overline{\gamma}) < f_{T,1}(\overline{\gamma})$. Hence, there are some $\underline{\xi} > 0$ and $\overline{\xi} > 0$ with $\underline{\xi} \neq \overline{\xi}$ such that the minimum of $f_T(0, \overline{\gamma})$ in (25) is attained by the solution $\xi_1 = \cdots = \xi_k = \underline{\xi}$ and $\xi_{k+1} = \cdots = \xi_T = \overline{\xi}$. Fixing $\xi_1, \ldots, \xi_{k-2}$ and $\xi_{k+3}, \ldots, \xi_T$ at their optimal values and optimizing only over the remaining four decision variables in $f_T(0, \overline{\gamma})$ yields

$$
\begin{aligned}
f_T(0, \overline{\gamma}) \ = \ \min_{\xi_{k-1}, \xi_k, \xi_{k+1}, \xi_{k+2} \geq 0} \quad & (k - 2)\underline{\xi}^2 + (T - k - 2)\overline{\xi}^2 + \sum_{t=k-1}^{k+2} \xi_t^2 \\
\text{s.t.} \quad & (k - 2)\underline{\xi} + (T - k - 2)\overline{\xi} + \sum_{t=k-1}^{k+2} \xi_t = 1 \\
& \underline{\xi}^{k-2}\overline{\xi}^{T-k-2} \prod_{t=k-1}^{k+2} \xi_t \leq \overline{\gamma}.
\end{aligned}
$$

Defining the strictly positive constant $c = 1 - (k - 2)\underline{\xi} - (T - k - 2)\overline{\xi} = 2\underline{\xi} + 2\overline{\xi}$ and using the

substitution $y_t \leftarrow \xi_{k-2+t}/c$ for $t = 1, \ldots, 4$ further yields

$$
\begin{aligned}
f_T(0, \overline{\gamma}) &= (k-2)\underline{\xi}^2 + (T - k - 2)\overline{\xi}^2 + \\
&\quad \min_{y_1, y_2, y_3, y_4 \geq 0} \left\{ \sum_{t=1}^4 c^2 y_t^2 : \sum_{t=1}^4 y_t = 1, \ \prod_{t=1}^4 y_t \leq \frac{\overline{\gamma}}{c^4 \underline{\xi}^{k-2} \overline{\xi}^{T-k-2}} \right\} \quad (35) \\
&= (k-2)\underline{\xi}^2 + (T - k - 2)\overline{\xi}^2 + c^2 f_4 \left(\frac{\overline{\gamma}}{c^4 \underline{\xi}^{k-2} \overline{\xi}^{T-k-2}} \right),
\end{aligned}
$$

where the second equality follows from the definition of $f_4(0, \overline{\gamma})$ in (25). By construction, the minimization problem in (35) must be solved by $y_1 = y_2 = \underline{\xi}$ and $y_3 = y_4 = \overline{\xi}$. However, this contradicts our previous results. In fact, we know that the solution of $f_4(0, \overline{\gamma})$ must have the following properties for $T = 4$. If $\overline{\gamma}/[c^4 \underline{\xi}^{k-2} \overline{\xi}^{T-k-2}] < 4^{-4}$, then three out of the four ξ_t variables must be equal at optimality. Conversely, if $\overline{\gamma}/[c^4 \underline{\xi}^{k-2} \overline{\xi}^{T-k-2}] \geq 4^{-4}$, then all four ξ_t variables must be equal. This contradicts our assumption that there exist $k \in \{2, \ldots, \lfloor \frac{T}{2} \rfloor\}$ and $\overline{\gamma} \in (0, T^{-T})$ with $f_T(0, \overline{\gamma}) = f_{T,k}(\overline{\gamma}) < f_{T,1}(\overline{\gamma})$. Thus, the assertion holds for all $T > 4$. ∎

We now show that in the worst case, the weak-sense geometric random walk $\tilde{\boldsymbol{\pi}} = \{\tilde{\pi}_T\}_{T \in \mathbb{N}}$ defined through $\tilde{\pi}_T = \prod_{t=1}^T \tilde{\xi}_t$ is absorbed at 0 with certainty if T exceeds a threshold T_0.

Theorem 3.2 (Certainty of Absorption) *For $T > \frac{\mu^2 + \sigma^2}{(1-\rho)\sigma^2} + 1$ we have $\mathrm{L}(\gamma) = 1$ for every $\gamma > 0$.*

Proof: From the proof of Proposition 2.1 we know that there exists a discrete distribution $\mathbb{P}_0 = \sum_{k \in \mathcal{K}} p_k \delta_{\boldsymbol{\xi}^k} \in \mathcal{P}$ with scenarios $\boldsymbol{\xi}^k$ and associated probabilities $p_k > 0$, where k ranges over a finite index set $\mathcal{K}$ of cardinality $T+1$. By the permutation symmetry, any discrete distribution of the form $\mathbb{P}_0 \in \mathcal{P}$ can be used to construct a corresponding symmetric distribution

$$
\mathbb{P} = \frac{1}{T!} \sum_{\pi \in \mathfrak{P}} \sum_{k \in \mathcal{K}} p_k \delta_{\mathbf{P}_\pi \boldsymbol{\xi}^k}, \quad (36)
$$

which is also an element of $\mathcal{P}$. Here, $\mathfrak{P}$ denotes the group of all permutations of $\{1, \ldots, T\}$, while $\mathbf{P}_\pi \in \mathbb{R}^{T \times T}$ denotes the permutation matrix induced by $\pi \in \mathfrak{P}$; see also Lemma 3.1. Next, we define $m_1^k = \frac{1}{T} \sum_{t=1}^T \xi_t^k$ and $m_2^k = \frac{1}{T} \sum_{t=1}^T (\xi_t^k)^2$ as the arithmetic and quadratic means of scenario $\boldsymbol{\xi}^k$, respectively. It turns out that the first two moments of $\tilde{\boldsymbol{\xi}}$ can be expressed in terms of m_1^k and m_2^k. Note, for instance, that for any $t \neq s$ we have

$$
\begin{aligned}
\mathbb{E}_{\mathbb{P}} \left(\tilde{\xi}_t \tilde{\xi}_s \right) &= \frac{1}{T!} \sum_{\pi \in \mathfrak{P}} \sum_{k \in \mathcal{K}} p_k \xi_{\pi(t)}^k \xi_{\pi(s)}^k = \sum_{k \in \mathcal{K}} \frac{p_k}{T!} \sum_{r=1}^T \xi_r^k \sum_{\pi \in \mathfrak{P}: \, \pi(s) = r} \xi_{\pi(t)}^k \\
&= \sum_{k \in \mathcal{K}} \frac{p_k}{T!} \sum_{r=1}^T \xi_r^k (T - 2)! \left(T m_1^k - \xi_r^k \right) = \sum_{k \in \mathcal{K}} \frac{p_k}{T - 1} \left(T (m_1^k)^2 - m_2^k \right),
\end{aligned}
$$

where the first equality follows from the definition of $\mathbb{P}$ and because the t-th component of $\mathbf{P}_\pi \boldsymbol{\xi}^{(k)}$ is given by $\xi_{\pi(t)}^k$, while the third equality holds because there are $(T - 2)!$ permutations

21

that map s to r and t to any fixed index different from r. Similarly, one can show that

$$\mathbb{E}_{\mathbb{P}}\left(\tilde{\xi}_t\right) = \sum_{k \in \mathcal{K}} p_k m_1^k \quad \text{and} \quad \mathbb{E}_{\mathbb{P}}\left(\tilde{\xi}_t^2\right) = \sum_{k \in \mathcal{K}} p_k m_2^k.$$

The moment conditions in the definition of $\mathcal{P}$ thus reduce to

$$\sum_{k \in \mathcal{K}} p_k \qquad\qquad = 1 \qquad\qquad\qquad (37a)$$

$$\sum_{k \in \mathcal{K}} p_k m_1^k \qquad\qquad = \mu \qquad\qquad\qquad (37b)$$

$$\sum_{k \in \mathcal{K}} p_k m_2^k \qquad\qquad = \mu^2 + \sigma^2 \qquad\qquad (37c)$$

$$\sum_{k \in \mathcal{K}} \frac{p_k}{T-1}\left(T(m_1^k)^2 - m_2^k\right) = \mu^2 + \rho\sigma^2. \qquad (37d)$$

In the following we will update the scenarios $\boldsymbol{\xi}^k$ of the distribution $\mathbb{P}$ iteratively in finitely many steps, always ensuring that $\mathbb{P}$ remains within $\mathcal{P}$ after each update. The terminal distribution will have the property that $\prod_{t=1}^T \xi_t^k = 0$ for every $k \in \mathcal{K}$, which means that we will have constructed a distribution $\mathbb{P} \in \mathcal{P}$ with $\mathbb{P}(\prod_{t=1}^T \tilde{\xi}_t = 0) = 1$. This will establish the claim.

Step 1: Keeping the scenario probabilities as well as the scenario-wise arithmetic and quadratic means constant, we first replace each $\boldsymbol{\xi}^k$ with a minimizer of the problem

$$\inf_{\boldsymbol{\xi} \geq \mathbf{0}} \left\{ \prod_{t=1}^T \xi_t : \frac{1}{T}\sum_{t=1}^T \xi_t = m_1^k, \ \frac{1}{T}\sum_{t=1}^T \xi_t^2 = m_2^k \right\}, \qquad (38)$$

which depends parametrically on m_1^k and m_2^k. By Lemma 3.5 (i) below, problem (38) is indeed solvable for every $k \in \mathcal{K}$. The new distribution with updated scenarios still belongs to $\mathcal{P}$ because we did not change p_k, m_1^k and m_2^k, implying that the moment conditions (37) remain valid. To gain a better understanding of the updated distribution, we define the disjoint index sets

$$\mathcal{K}^+ = \left\{ k \in \mathcal{K} : T \geq \frac{m_2^k}{(m_1^k)^2} \geq \frac{T}{T-1} \right\} \quad \text{and} \quad \mathcal{K}^- = \left\{ k \in \mathcal{K} : 1 \leq \frac{m_2^k}{(m_1^k)^2} < \frac{T}{T-1} \right\},$$

and note that $\mathcal{K} = \mathcal{K}^+ \cup \mathcal{K}^-$ by Lemma 3.5 (i) below. Lemma 3.5 (ii) further implies that

$$k \in \mathcal{K}^+ \iff T \geq \frac{m_2^k}{(m_1^k)^2} \geq \frac{T}{T-1} \iff \frac{1}{T} \leq \frac{m_2^k - (m_1^k)^2}{m_2^k} \leq \frac{T-1}{T} \iff \prod_{t=1}^T \xi_t^k = 0 \quad (39a)$$

and

$$k \in \mathcal{K}^- \iff 1 \leq \frac{m_2^k}{(m_1^k)^2} < \frac{T}{T-1} \iff 0 \leq \frac{m_2^k - (m_1^k)^2}{m_2^k} < \frac{1}{T} \iff \prod_{t=1}^T \xi_t^k > 0. \qquad (39b)$$

We will henceforth say that $\mathcal{K}^+$ ($\mathcal{K}^-$) is the index set of the *absorbing* (*non-absorbing*) scenarios. If all scenarios are absorbing (that is, if $\mathcal{K}^+ = \mathcal{K}$), then $\mathbb{P}(\prod_{t=1}^{T} \tilde{\xi}_t = 0) = 1$, and we are done.

Step 2: If there exists a non-absorbing scenario $i \in \mathcal{K}^-$, we will alter both the scenarios and their quadratic means to make scenario i absorbing, while ensuring that all scenarios $k \in \mathcal{K}^+$ remain absorbing. To achieve this, we consider the following family of quadratic means parameterized in $\lambda \in [0, 1]$.

$$
m_2^k(\lambda) = \begin{cases} (1-\lambda)m_2^k + \lambda\frac{T}{T-1}(m_1^k)^2 & \text{for } k \in \mathcal{K}^+ \\ m_2^i + \lambda \sum_{k \in \mathcal{K}^+} \frac{p_k}{p_i}(m_2^k - \frac{T}{T-1}(m_1^k)^2) & \text{for } k = i \\ m_2^k & \text{for } k \in \mathcal{K}^- \backslash \{i\} \end{cases} \tag{40}
$$

By construction, p_k, m_1^k and $m_2^k = m_2^k(\lambda)$ satisfy the moment conditions (37) for every $\lambda \in [0, 1]$. As in Step 1, the scenario $\boldsymbol{\xi}^k(\lambda)$ is then chosen to be a minimizer of problem (38) with inputs m_1^k and $m_2^k = m_2^k(\lambda)$. However, (38) could fail to be solvable for $\lambda \lesssim 1$, in which case the proposed construction would fail. Indeed, Lemma 3.5 (i) shows that (38) is only solvable when $1 \le m_2^k(\lambda)/(m_1^k)^2 \le T$. In the remainder we will demonstrate that there is $\lambda^\star \in (0, 1)$ such that $\boldsymbol{\xi}^k(\lambda^\star)$ exists for every $k \in \mathcal{K}$ and such that all scenarios $k \in \mathcal{K}^+ \cup \{i\}$ are absorbing.

Subtracting (37d) from (37c) and dividing the difference by (37c) yields

$$
\frac{T \sum_{k \in \mathcal{K}} p_k \left(m_2^k - (m_1^k)^2\right)}{(T-1) \sum_{k \in \mathcal{K}} p_k m_2^k} = \frac{(1-\rho)\sigma^2}{\mu^2 + \sigma^2} > \frac{1}{T-1},
$$

where the inequality follows from the assumption that $T > \frac{\mu^2 + \sigma^2}{(1-\rho)\sigma^2} + 1$. Multiplying both sides of the inequality by $\frac{T-1}{T}$ and partitioning $\mathcal{K}$ into $\mathcal{K}^+$ and $\mathcal{K}^-$ further reveals that

$$
\frac{\sum_{k \in \mathcal{K}^+} p_k m_2^k \frac{m_2^k - (m_1^k)^2}{m_2^k} + \sum_{k \in \mathcal{K}^-} p_k m_2^k \frac{m_2^k - (m_1^k)^2}{m_2^k}}{\sum_{k \in \mathcal{K}^+} p_k m_2^k + \sum_{k \in \mathcal{K}^-} p_k m_2^k} > \frac{1}{T}. \tag{41}
$$

The expression on the left hand side of the above inequality represents a weighted average of the fractions $(m_2^k - (m_1^k)^2)/m_2^k$ across all $k \in \mathcal{K}$. Recall from (39a) and (39b) that the fractions indexed by $k \in \mathcal{K}^+$ are larger or equal to $1/T$, while those indexed by $k \in \mathcal{K}^-$ are strictly smaller than $1/T$. The inequality (41) asserts that the fractions corresponding to $k \in \mathcal{K}^+$ dominate those corresponding to $k \in \mathcal{K}^-$. Thus, (41) remains valid if we replace $\mathcal{K}^-$ with $\{i\}$, that is,

$$
\frac{\sum_{k \in \mathcal{K}^+} p_k m_2^k \frac{m_2^k - (m_1^k)^2}{m_2^k} + p_i m_2^i \frac{m_2^i - (m_1^i)^2}{m_2^i}}{\sum_{k \in \mathcal{K}^+} p_k m_2^k + p_i m_2^i} > \frac{1}{T},
$$

which is equivalent to

$$
\frac{\sum_{k \in \mathcal{K}^+} p_k \frac{(m_1^k)^2}{T-1} + p_i \left(m_2^i + \sum_{k \in \mathcal{K}^+} \frac{p_k}{p_i}\left(m_2^k - \frac{T}{T-1}(m_1^k)^2\right) - (m_1^i)^2\right)}{\sum_{k \in \mathcal{K}^+} p_k \frac{T}{T-1}(m_1^k)^2 + p_i \left(m_2^i + \sum_{k \in \mathcal{K}^+} \frac{p_k}{p_i}\left(m_2^k - \frac{T}{T-1}(m_1^k)^2\right)\right)} > \frac{1}{T}. \tag{42}
$$

Using the notation introduced in (40), the inequality (42) can be reformulated as

$$\frac{\sum_{k\in\mathcal{K}^+} p_k \, m_2^k(1) \frac{m_2^k(1)-(m_1^k)^2}{m_2^k(1)} + p_i \, m_2^i(1) \frac{m_2^i(1)-(m_1^i)^2}{m_2^i(1)}}{\sum_{k\in\mathcal{K}^+} p_k m_2^k(1) + p_i m_2^i(1)} > \frac{1}{T},$$

which constitutes a weighted average of the fractions $(m_2^k(1) - (m_1^k)^2)/m_2^k(1)$ across all $k \in \mathcal{K}^+ \cup \{i\}$. By construction, we have $(m_2^k(1) - (m_1^k)^2)/m_2^k(1) = \frac{1}{T}$ for every $k \in \mathcal{K}^+$, and thus the average on the left hand side of the above inequality can exceed $\frac{1}{T}$ only if

$$\frac{m_2^i(1) - (m_1^i)^2}{m_2^i(1)} > \frac{1}{T}.$$

As $i \in \mathcal{K}^-$, the relation (39b) further implies that

$$\frac{m_2^i(0) - (m_1^i)^2}{m_2^i(0)} = \frac{m_2^i - (m_1^i)^2}{m_2^i} < \frac{1}{T}.$$

The intermediate value theorem then guarantees the existence of $\lambda^\star \in (0, 1)$ with

$$\frac{m_2^i(\lambda^\star) - (m_1^i)^2}{m_2^i(\lambda^\star)} = \frac{1}{T} \qquad \Longleftrightarrow \qquad \frac{m_2^i(\lambda^\star)}{(m_1^i)^2} = \frac{T}{T-1}.$$

By construction, we thus have $1 \leq m_2^k(\lambda^\star)/(m_1^k)^2 \leq T$ for every $k \in \mathcal{K}$, which implies via Lemma 3.5 (i) that the corresponding scenarios $\boldsymbol{\xi}^k(\lambda^\star)$ are well-defined. Our construction also guarantees that $\frac{T}{T-1} \leq m_2^k(\lambda^\star)/(m_1^k)^2 \leq T$ for every $k \in \mathcal{K}^+ \cup \{i\}$, which implies via Lemma 3.5 (ii) that the corresponding scenarios $\boldsymbol{\xi}^k(\lambda^\star)$ are absorbing. Thus, by replacing $\boldsymbol{\xi}^k$ with $\boldsymbol{\xi}^k(\lambda^\star)$ in (36) we obtain a new distribution $\mathbb{P} \in \mathcal{P}$ with more absorbing scenarios. As the total number of scenarios is finite, we can repeat Step 2 finitely many times to construct a distribution $\mathbb{P} \in \mathcal{P}$ that has only absorbing scenarios. Thus, the claim follows. ∎

The proof of Theorem 3.2 relies on the following auxiliary result.

Lemma 3.5 *Assume that $m_1, m_2 > 0$ and consider the parametric program*

$$\inf_{\boldsymbol{\xi} \geq \mathbf{0}} \left\{ \prod_{t=1}^T \xi_t : \frac{1}{T} \sum_{t=1}^T \xi_t = m_1, \ \frac{1}{T} \sum_{t=1}^T \xi_t^2 = m_2 \right\}. \tag{43}$$

Then, the following statements hold:

  (i) *Problem (43) is feasible and solvable iff $T \geq \frac{m_2}{m_1^2} \geq 1$.*

  (ii) *The optimal value of (43) is zero iff $T \geq \frac{m_2}{m_1^2} \geq \frac{T}{T-1}$.*

Figure 2 visualizes how the feasible set of problem (43) depends on m_1 and m_2.

Proof of Lemma 3.5: As for assertion (i), assume that there is $\boldsymbol{\xi}$ feasible in (43). We then have $\frac{1}{T} \sum_{t=1}^T \xi_t = m_1$, which implies that $Tm_1^2 \geq m_2 \geq m_1^2$ since $\|\boldsymbol{\xi}\|_1 \geq \|\boldsymbol{\xi}\|_2 \geq \frac{1}{\sqrt{T}} \|\boldsymbol{\xi}\|_1$.
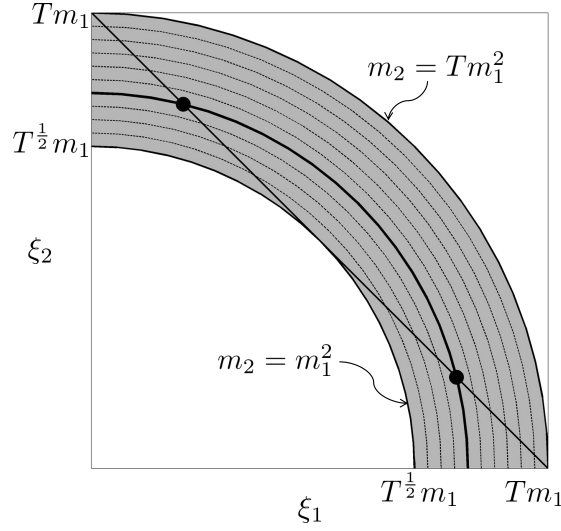
Figure 2: Feasible region of problem (43) for $T = 2$ and different values of m_1 and m_2. The diagonal line corresponds to the constraint $\frac{1}{T}\sum_{t=1}^{T}\xi_t = m_1$, and each dotted curve corresponds to the constraint $\frac{1}{T}\sum_{t=1}^{T}\xi_t^2 = m_2$ for some combination of m_1 and m_2. The innermost and the outermost curves correspond to the cases where $m_2/m_1^2 = 1$ and $m_2/m_1^2 = T$, respectively. The feasible region for the (m_1, m_2)-combination represented by the bold curve is given by the two dots.

Conversely, if $T \geq \frac{m_2}{m_1^2} \geq 1$, we may define $\boldsymbol{\xi} = (z, \frac{m_1 T - z}{T-1}, \dots, \frac{m_1 T - z}{T-1})$ for some $z \in [m_1, Tm_1]$ to be chosen later. By construction, we have $\frac{1}{T}\sum_{t=1}^{T}\xi_t = m_1$ irrespective of z, while

$$\frac{1}{T}\sum_{t=1}^{T}\xi_t^2 = \frac{z^2}{T} + \frac{T-1}{T}\left(\frac{m_1 T - z}{T-1}\right)^2$$

changes continuously from m_1^2 to Tm_1^2 when z is swept from m_1 to Tm_1. Thus, by the intermediate value theorem, we may assume that $\frac{1}{T}\sum_{t=1}^{T}\xi_t^2 = m_2 \in [m_1^2, Tm_1^2]$ for some suitably chosen $z \in [m_1, Tm_1]$. We conclude that (43) is feasible whenever $T \geq \frac{m_2}{m_1^2} \geq 1$. In that case, however, (43) is also solvable as the objective function is continuous and the feasible set is compact.

To prove assertion (ii), we observe that the optimal value of (43) vanishes iff the problem admits a minimizer $\boldsymbol{\xi}$ with $\prod_{t=1}^{T}\xi_t = 0$. More precisely, by permutation symmetry, the minimum of (43) vanishes iff there exists $\boldsymbol{\xi}$ with $\xi_T = 0$, $\frac{1}{T}\sum_{t=1}^{T-1}\xi_t = m_1$ and $\frac{1}{T}\sum_{t=1}^{T-1}\xi_t^2 = m_2$. By assertion (i), however, the last two inequalities are satisfiable iff

$$T - 1 \geq \frac{m_2\left(\frac{T}{T-1}\right)}{\left(m_1\left(\frac{T}{T-1}\right)\right)^2} \geq 1 \quad \Longleftrightarrow \quad T \geq \frac{m_2}{m_1^2} \geq \frac{T}{T-1},$$

and thus the claim follows. ∎

25

4 Right-Sided Chebyshev Bounds

We now study *right-sided Chebyshev bounds* of the form

$$\mathrm{R}(\gamma) = \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\left(\prod_{t=1}^{T} \tilde{\xi}_t \geq \gamma\right),$$

where the ambiguity set $\mathcal{P}$ is defined in (8). We first present the main result of this section.

Theorem 4.1 (Right-Sided Chebyshev Bound) *Let $\gamma > 0$. For all $T \geq 3$ the right-sided Chebyshev bound $\mathrm{R}(\gamma)$ coincides with the optimal objective value of the semidefinite program*

$$
\begin{aligned}
\inf \quad & \alpha + T\mu\beta + T(\mu^2 + \sigma^2)\gamma_1 + T(T\mu^2 + \sigma^2 + (T-1)\rho\sigma^2)\gamma_2 \\
\text{s.t.} \quad & \alpha, \beta, \gamma_1, \gamma_2 \in \mathbb{R}, \ \lambda_1, \lambda_2, \lambda_3 \geq 0, \ \boldsymbol{p} \in \mathbb{R}^{2T+1}, \ \boldsymbol{P} \in \mathbb{S}_+^{T+1}, \ \boldsymbol{q} \in \mathbb{R}^{2T-1}, \ \boldsymbol{Q} \in \mathbb{S}_+^{T} \\
& \alpha \geq 0, \quad \alpha \geq 1 - \lambda_3 T\gamma^{1/T}, \quad \gamma_1 + T\gamma_2 \geq 0, \quad \gamma_1 + \gamma_2 \geq 0 \\
& \gamma_2 + \frac{\gamma_1}{T} + \alpha \geq \left\| \left(\beta - \lambda_1, \gamma_2 + \frac{\gamma_1}{T} - \alpha\right) \right\|_2 \\
& \gamma_2 + \gamma_1 + \alpha \geq \|(\beta - \lambda_2, \gamma_2 + \gamma_1 - \alpha)\|_2 \\
& \gamma_2 + \frac{\gamma_1}{T} + \lambda_3 T\gamma^{1/T} + \alpha - 1 \geq \left\| \left(\beta - \lambda_3, \gamma_2 + \frac{\gamma_1}{T} - \lambda_3 T\gamma^{1/T} - \alpha + 1\right) \right\|_2 \\
& p_0 = (T-1)\gamma_1\gamma^{\frac{2}{T-1}} + (T-1)^2\gamma_2\gamma^{\frac{2}{T-1}}, \quad p_1 + q_0 = (T-1)\beta\gamma^{\frac{1}{T-1}} \\
& p_2 + q_1 = \alpha - 1, \quad p_T + q_{T-1} = 2(T-1)\gamma_2\gamma^{\frac{1}{T-1}}, \quad p_{T+1} + q_T = \beta \\
& p_{2T} = \gamma_1 + \gamma_2, \quad p_t + q_{t-1} = 0 \quad \forall t = 3, \ldots, T-1, T+2, \ldots, 2T-1 \\
& p_t = \sum_{i+j=t} P_{i,j} \quad \forall t = 0, \ldots, 2T, \quad q_t = \sum_{i+j=t} Q_{i,j} \quad \forall t = 0, \ldots, 2T-2,
\end{aligned}
\tag{44}
$$

where we use the convention that the entries of $\boldsymbol{p}$, $\boldsymbol{P}$, $\boldsymbol{q}$ and $\boldsymbol{Q}$ are numbered starting from 0. For $T = 2$, $\mathrm{R}(\gamma)$ is given by a variant of (44) where the constraints $p_2 + q_1 = \alpha - 1$ and $p_T + q_{T-1} = 2(T-1)\gamma_2\gamma^{\frac{1}{T-1}}$ are combined to $p_2 + q_1 = \alpha - 1 + 2(T-1)\gamma_2\gamma^{\frac{1}{T-1}}$.

Proof: Using similar arguments as in the proof of Theorem 3.1, one first shows that the worst-case probability problem $\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}(\prod_{t=1}^{T} \tilde{\xi}_t \geq \gamma)$ admits a strong dual which constitutes a semi-infinite optimization problem. Exploiting this problem's permutation symmetry, one can further show that its optimal value amounts to

$$
\begin{aligned}
\mathrm{R}(\gamma) = \inf \quad & \alpha + T\mu\beta + T(\mu^2 + \sigma^2)\gamma_1 + T\left[T\mu^2 + \sigma^2 + (T-1)\rho\sigma^2\right]\gamma_2 \\
\text{s.t.} \quad & \alpha, \beta, \gamma_1, \gamma_2 \in \mathbb{R} \\
& \alpha + \beta\|\boldsymbol{\xi}\|_1 + \gamma_1\|\boldsymbol{\xi}\|_2^2 + \gamma_2\|\boldsymbol{\xi}\|_1^2 \geq 0 \quad \forall \boldsymbol{\xi} \geq \boldsymbol{0} \\
& \alpha + \beta\|\boldsymbol{\xi}\|_1 + \gamma_1\|\boldsymbol{\xi}\|_2^2 + \gamma_2\|\boldsymbol{\xi}\|_1^2 \geq 1 \quad \forall \boldsymbol{\xi} \geq \boldsymbol{0} : \ \prod_{t=1}^{T} \xi_t \geq \gamma.
\end{aligned}
\tag{45}
$$

Details are omitted for brevity of exposition. Lemma 3.2 then implies that (45) reduces to

$$
\begin{aligned}
\mathrm{R}(\gamma) = \ \inf \quad & \alpha + T\mu\beta + T(\mu^2 + \sigma^2)\gamma_1 + T(T\mu^2 + \sigma^2 + (T-1)\rho\sigma^2)\gamma_2 \\
\mathrm{s.\,t.} \quad & \alpha, \beta, \gamma_1, \gamma_2 \in \mathbb{R} \\
& \inf_{s \geq 0} \ \alpha + \beta s + \gamma_2 s^2 + \frac{\gamma_1}{T}s^2 \geq 0 \\
& \inf_{s \geq 0} \ \alpha + \beta s + \gamma_2 s^2 + \gamma_1 s^2 \geq 0 \\
& \inf_{s \geq T\gamma^{1/T}} \ \alpha + \beta s + \gamma_2 s^2 + \frac{\gamma_1}{T}s^2 \geq 1 \\
& \inf_{s \geq T\gamma^{1/T}} \ \alpha + \beta s + \gamma_2 s^2 + \gamma_1 s^2 g_T\left(\frac{\gamma}{s^T}, \infty\right) \geq 1.
\end{aligned}
\tag{46}
$$

By leveraging the $\mathcal{S}$-lemma and a well-known reformulation of hyperbolic constraints as second-order cone constraints, one can use similar arguments as in the proof of Theorem 3.1 to show that the first three constraints in (46) hold iff there exist $\lambda_1, \lambda_2, \lambda_3 \geq 0$ satisfying

$$
\alpha \geq 0, \quad \alpha \geq 1 - \lambda_3 T\gamma^{1/T}, \quad \gamma_1 + T\gamma_2 \geq 0, \quad \gamma_1 + \gamma_2 \geq 0
$$
$$
\gamma_2 + \frac{\gamma_1}{T} + \alpha \geq \left\|\left(\beta - \lambda_1, \gamma_2 + \frac{\gamma_1}{T} - \alpha\right)\right\|_2
$$
$$
\gamma_2 + \gamma_1 + \alpha \geq \|(\beta - \lambda_2, \gamma_2 + \gamma_1 - \alpha)\|_2
$$
$$
\gamma_2 + \frac{\gamma_1}{T} + \lambda_3 T\gamma^{1/T} + \alpha - 1 \geq \left\|\left(\beta - \lambda_3, \gamma_2 + \frac{\gamma_1}{T} - \lambda_3 T\gamma^{1/T} - \alpha + 1\right)\right\|_2.
$$

By Lemma 4.1 below, the last semi-infinite constraint in (46) can be re-expressed as

$$
\inf_{s \geq T\gamma^{1/T}, \underline{\xi}, \overline{\xi} \geq 0} \left\{\alpha + \beta s + \gamma_2 s^2 + \gamma_1 \left[\underline{\xi}^2 + (T-1)\overline{\xi}^2\right] : \underline{\xi} + (T-1)\overline{\xi} = s, \ \underline{\xi}\,\overline{\xi}^{T-1} = \gamma\right\} \geq 1,
$$

which is identical to (19). The claim then follows by replacing this constraint with its explicit semidefinite reformulation familiar from Theorem 3.1. ∎

The proof of Theorem 4.1 relies on 2 auxiliary lemmas, which we prove next.

Lemma 4.1 *For $T \geq 2$, $\overline{\gamma} = \infty$ and $\underline{\gamma} \geq 0$, the optimal value $g_T(\underline{\gamma}, \infty)$ of (25b) equals*

$$
g_T(\underline{\gamma}, \infty) = \begin{cases} \displaystyle\max_{\underline{\xi} \geq 0, \overline{\xi} \geq 0} \left\{\underline{\xi}^2 + (T-1)\overline{\xi}^2 : \underline{\xi} + (T-1)\overline{\xi} = 1, \ \underline{\xi}\,\overline{\xi}^{T-1} = \underline{\gamma}\right\} & \text{if } 0 \leq \underline{\gamma} \leq \underline{\gamma}T^{-T} \\ -\infty & \text{if } \underline{\gamma} > T^{-T}. \end{cases}
$$

Proof: If $\underline{\gamma} > T^{-T}$, then the maximization problem (25b) is infeasible due to the inequality of arithmetic and geometric means, and thus we have $g_T(\underline{\gamma}, \infty) = -\infty$. For $\underline{\gamma} = T^{-T}$, the unique feasible solution of (25b) is $\boldsymbol{\xi} = \frac{1}{T}\mathbf{1}$, which implies that $g_T(\underline{\gamma}, \infty) = \frac{1}{T}$. Moreover, for $\underline{\gamma} = 0$, the last constraint in (25b) becomes redundant. In this case $g_T(\underline{\gamma}, \infty)$ is optimized by $\boldsymbol{\xi} = \mathbf{e}_i$, and thus we find $g_T(\underline{\gamma}, \infty) = 1$. Lastly, for $0 < \underline{\gamma} < T^{-T}$, the maximization problem (25b) is feasible, and every feasible solution has strictly positive components. In addition, the product constraint $\prod_{t=1}^T \xi_t \geq \underline{\gamma}$ is binding at optimality for otherwise convex combinations of the optimal solution $\boldsymbol{\xi}$ with $\mathbf{e}_i$, where $i \in \arg\max\{\boldsymbol{\xi}_j : j = 1, \ldots, T\}$, would improve the objective function of (25b),

27

which is a contradiction. We thus conclude that

$$
g_T(\underline{\gamma}, \infty) = \begin{cases} 1 & \text{if } \underline{\gamma} = 0, \\ \max_{\boldsymbol{\xi} > \mathbf{0}} \left\{ \|\boldsymbol{\xi}\|_2^2 : \|\boldsymbol{\xi}\|_1 = 1, \ \prod_{t=1}^T \xi_t = \underline{\gamma} \right\} & \text{if } 0 < \underline{\gamma} < T^{-T}, \\ \frac{1}{T} & \text{if } \underline{\gamma} = T^{-T}, \\ -\infty & \text{if } \underline{\gamma} > T^{-T}. \end{cases}
$$

As in the proof of Lemma 3.3, for $0 < \underline{\gamma} < T^{-T}$ one can use the optimality conditions of (25b) to show that

$$
g_T(\underline{\gamma}, \infty) = \max_{k \in \{1, \ldots, \lfloor \frac{T}{2} \rfloor\}} g_{T,k}(\underline{\gamma}), \tag{47}
$$

where the functions $g_{T,k} : (0, T^{-T}) \to \mathbb{R}$, $k = 1, 2, \ldots, \lfloor \frac{T}{2} \rfloor$, are defined through

$$
g_{T,k}(\underline{\gamma}) = \max_{\underline{\xi} > 0, \overline{\xi} > 0} \left\{ k\underline{\xi}^2 + (T-k)\overline{\xi}^2 : k\underline{\xi} + (T-k)\overline{\xi} = 1, \ \underline{\xi}^k \overline{\xi}^{T-k} = \underline{\gamma} \right\}. \tag{48}
$$

Lemma 4.2 below asserts that the maximum in (47) is attained at $k = 1$. We thus obtain

$$
g_T(\underline{\gamma}, \infty) = \begin{cases} 1 & \text{if } \underline{\gamma} = 0, \\ \max_{\underline{\xi} \geq 0, \overline{\xi} \geq 0} \left\{ \underline{\xi}^2 + (T-1)\overline{\xi}^2 : \underline{\xi} + (T-1)\overline{\xi} = 1, \ \underline{\xi}\,\overline{\xi}^{T-1} = \underline{\gamma} \right\} & \text{if } 0 < \underline{\gamma} < T^{-T}, \\ \frac{1}{T} & \text{if } \underline{\gamma} = T^{-T}, \\ -\infty & \text{if } \underline{\gamma} > T^{-T}. \end{cases}
$$

The statement of the lemma now follows since the maximization problem in the equation above evaluates to 1 at $\underline{\gamma} = 0$ and to $1/T$ at $\underline{\gamma} = T^{-T}$. Indeed, the maximization problem is bounded above by $\max_{\|\boldsymbol{\xi}\|_1 = 1} \|\boldsymbol{\xi}\|_2^2$, and the optimal value 1 of this bound is achieved by the feasible solution $(\underline{\xi}, \overline{\xi}) = (1, 0)$ of the maximization problem at $\underline{\gamma} = 0$. Likewise, $g_T(T^{-T}, \infty)$ is bounded above by $\max_{\boldsymbol{\xi} \geq \mathbf{0}} \{ \|\boldsymbol{\xi}\|_2^2 : \|\boldsymbol{\xi}\|_1 = 1, \prod_t \xi_t = T^{-T} \}$, and the optimal value $1/T$ of this bound is achieved by the feasible solution $(\underline{\xi}, \overline{\xi}) = (\frac{1}{T}, \frac{1}{T})$ of the maximization problem. ∎

Lemma 4.2 For $T \geq 2$ and $0 < \underline{\gamma} < T^{-T}$, the optimal value of (47) is given by $g_{T,1}(\underline{\gamma})$.

Proof: The proof widely parallels that of Lemma 3.4 and is therefore omitted. ∎

We now show that in the extreme case, the weak-sense geometric random walk $\tilde{\boldsymbol{\pi}} = \{\tilde{\pi}_T\}_{T \in \mathbb{N}}$ defined through $\tilde{\pi}_T = \prod_{t=1}^T \tilde{\xi}_t$ weakly exceeds the deterministic growth process $\{\mu^T\}_{T \in \mathbb{N}}$ with certainty for any time horizon T, assuming that $\rho \geq 0$. The result can be viewed as the right-sided analogue of Theorem 3.2.

Proposition 4.1 If $\rho \geq 0$, then $\mathrm{R}(\gamma) = 1$ for all $\gamma \leq \mu^T$.

Proof: The objective function of problem (46) can be reformulated as

$$
(\alpha + T\mu\beta + T\mu^2\gamma_1 + T^2\mu^2\gamma_2) + T\sigma^2(\gamma_1 + (1 + (T-1)\rho)\gamma_2).
$$

28

For $\gamma \leq \mu^T$, the first term equals the left hand side of the third semi-infinite constraint in (46) if we set $s = T\mu$, and it must therefore be greater than or equal to 1. In the second term, the factor $(\gamma_1 + (1 + (T-1)\rho)\gamma_2)$ can be expressed as the linear combination $\rho \cdot (\gamma_1 + T\gamma_2) + (1-\rho) \cdot (\gamma_1 + \gamma_2)$. For $\rho \geq 0$, this linear combination becomes a convex combination, and the claim follows since $\gamma_1 + T\gamma_2 \geq 0$ and $\gamma_1 + \gamma_2 \geq 0$ are explicit constraints in the equivalent reformulation (44). ∎

We highlight that Proposition 4.1 breaks down for $\rho < 0$.

5 Covariance Bounds

The ambiguity set $\mathcal{P}$ reflects the assumption that the covariance matrix $\boldsymbol{\Sigma}$ is known *precisely* and that the (co-)variances of the components of $\tilde{\boldsymbol{\xi}}$ are permutation symmetric. Either assumption may prove overly restrictive in practice. In this section, we therefore assume that only an upper bound on the covariance matrix is available. More precisely, we consider the ambiguity set

$$\mathcal{P}' = \left\{ \mathbb{P} \in \mathcal{M}_+(\mathbb{R}_+^T) \ : \ \mathbb{P}\left(\tilde{\boldsymbol{\xi}} \geq \boldsymbol{0}\right) = 1, \ \mathbb{E}_{\mathbb{P}}\left(\tilde{\boldsymbol{\xi}}\right) = \boldsymbol{\mu}, \ \mathbb{E}_{\mathbb{P}}\left(\tilde{\boldsymbol{\xi}}\tilde{\boldsymbol{\xi}}^{\mathsf{T}}\right) \preceq \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^{\mathsf{T}} \right\},$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are defined as in Section 1. For $\gamma > 0$, we are then interested in quantifying *relaxed left-sided and right-sided Chebyshev bounds* of the form

$$\mathrm{L}'(\gamma) = \sup_{\mathbb{P} \in \mathcal{P}'} \mathbb{P}\left(\prod_{t=1}^{T} \tilde{\xi}_t \leq \gamma\right) \quad \text{and} \quad \mathrm{R}'(\gamma) = \sup_{\mathbb{P} \in \mathcal{P}'} \mathbb{P}\left(\prod_{t=1}^{T} \tilde{\xi}_t \geq \gamma\right).$$

In the following, we analyze each of these relaxed bounds in turn.

Theorem 5.1 (Relaxed Left-Sided Chebyshev Bound) *The relaxed left-sided Chebyshev bound satisfies* $\mathrm{L}'(\gamma) = \mathrm{L}(\gamma)$ *for all* $\gamma > 0$.

Proof: By repeating the first few steps of the proof of Theorem 3.1, one can show that $\mathrm{L}'(\gamma)$ coincides with the optimal value of (15) with the extra constraint $\boldsymbol{\Gamma} \succeq \boldsymbol{0}$. In this case Lemma 3.1 remains valid and implies that we can restrict attention to permutation-symmetric solutions of the form $\boldsymbol{\Gamma} = \gamma_1 \mathbb{I} + \gamma_2 \boldsymbol{1}\boldsymbol{1}^{\mathsf{T}}$ for some $\gamma_1, \gamma_2 \in \mathbb{R}$. As $\boldsymbol{\Gamma} = \gamma_1 \mathbb{I} + \gamma_2 \boldsymbol{1}\boldsymbol{1}^{\mathsf{T}} \succeq \boldsymbol{0}$ iff $\gamma_1 + T\gamma_2 \geq 0$ and $\gamma_1 \geq 0$ by virtue of [26, Proposition 4], we may then conclude that $\mathrm{L}'(\gamma)$ coincides with the optimal value of (16) with the extra constraints $\gamma_1 + T\gamma_2 \geq 0$ and $\gamma_1 \geq 0$. Note that (16) is equivalent to (13) and (17). As $\gamma_1 + T\gamma_2 \geq 0$ is an explicit constraint of problem (13), it is necessarily an implicit constraint of the problems (16) and (17). Thus, $\mathrm{L}'(\gamma)$ coincides with the optimal value of (17) with the extra constraint $\gamma_1 \geq 0$. To prove the identity $\mathrm{L}(\gamma) = \mathrm{L}'(\gamma)$, it is therefore sufficient to show that appending the extra constraint $\gamma_1 \geq 0$ has no impact on the optimal value of (17).

To this end, fix any feasible solution of problem (17) with $\gamma_1 < 0$. As this solution must satisfy the constraint $\alpha + s\beta + s^2\gamma_2 + s^2\gamma_1 \geq 1$ for every $s \geq 0$ and as $s = T\mu > 0$, we have

$$\alpha + T\mu\beta + T^2\mu^2\gamma_2 + T^2\mu^2\gamma_1 \geq 1. \tag{49}$$

Moreover, the objective function of (17) can be reformulated as

$$\alpha + T\mu\beta + T(\mu^2 + \sigma^2)\gamma_1 + T\left[T\mu^2 + \sigma^2 + (T-1)\rho\sigma^2\right]\gamma_2$$
$$= \left(\alpha + T\mu\beta + T^2\mu^2\gamma_2 + T^2\mu^2\gamma_1\right) + T(1-T)\left(\mu^2 + \rho\sigma^2\right)\gamma_1 + T\sigma^2(1 + (T-1)\rho)(\gamma_1 + \gamma_2),$$

which constitutes a sum of three terms. The first term in the sum is greater than or equal to 1 because of (49), and the second term is strictly positive because $T \geq 2$, $\gamma_1 < 0$ and $\mu^2 + \rho\sigma^2 > 0$. The third term is non-negative because $\rho > -1/(T-1)$ and $\gamma_1 + \gamma_2 \geq 0$ is an explicit constraint of (13) and thus an implicit constraint of (17). In summary, we have shown that the objective value of any feasible solution of (17) with $\gamma_1 < 0$ is strictly greater than 1. As the optimal value $\mathrm{L}(\gamma)$ of (17) represents a probability, however, we conclude that no feasible solution with $\gamma_1 < 0$ can optimize (17). Thus, the extra constraint $\gamma_1 \geq 0$ does not change the optimal value of (17), and the claim follows. ∎

Theorem 5.2 (Relaxed Right-Sided Chebyshev Bound) *The relaxed right-sided Chebyshev bound admits the analytical solution*

$$\mathrm{R}'(\gamma) = \begin{cases} 1 & \text{if } 0 < \gamma \leq \mu^T, \\ \mu\gamma^{-1/T} & \text{if } \mu^T < \gamma < \left(\mu + \frac{\sigma^2\theta}{T\mu}\right)^T, \\ \frac{\sigma^2\theta}{\sigma^2\theta + T(\mu - \gamma^{1/T})^2} & \text{if } \gamma \geq \left(\mu + \frac{\sigma^2\theta}{T\mu}\right)^T, \end{cases}$$

where $\theta = 1 + (T-1)\rho > 0$.

Proof: Using similar arguments as in the proof of the previous theorem, one can show that $\mathrm{R}'(\gamma)$ coincides with the optimal value of the following semi-infinite optimization problem:

$$\begin{aligned} \mathrm{R}'(\gamma) = \ \inf \quad & \alpha + \mu\mathbf{1}^\mathsf{T}\boldsymbol{\beta} + \left\langle (1-\rho)\sigma^2\mathbb{I} + \left(\mu^2 + \rho\sigma^2\right)\mathbf{1}\mathbf{1}^\mathsf{T}, \boldsymbol{\Gamma}\right\rangle \\ \text{s.t.} \quad & \alpha \in \mathbb{R}, \ \boldsymbol{\beta} \in \mathbb{R}^T, \ \boldsymbol{\Gamma} \in \mathbb{S}_+^T \\ & \alpha + \overline{\boldsymbol{\xi}}^\mathsf{T}\boldsymbol{\beta} + \overline{\boldsymbol{\xi}}^\mathsf{T}\boldsymbol{\Gamma}\overline{\boldsymbol{\xi}} \geq 0 \quad \forall \overline{\boldsymbol{\xi}} \geq \mathbf{0} \\ & \alpha + \underline{\boldsymbol{\xi}}^\mathsf{T}\boldsymbol{\beta} + \underline{\boldsymbol{\xi}}^\mathsf{T}\boldsymbol{\Gamma}\underline{\boldsymbol{\xi}} \geq 1 \quad \forall \underline{\boldsymbol{\xi}} \geq \mathbf{0} : \ \prod_{t=1}^T \underline{\xi}_t \geq \gamma \end{aligned} \tag{50}$$

Without loss of generality, we use different symbols $\overline{\boldsymbol{\xi}}$ and $\underline{\boldsymbol{\xi}}$ to denote the uncertain parameters in the two semi-infinite constraints, respectively. Note that (50) can be viewed as the robust counterpart of an uncertain convex program with constraint-wise uncertainty sets [3]. As the left hand sides of the robust constraints are convex in the respective uncertainties, the '*primal worst equals dual best*' duality scheme portrayed in [2, Theorem 4.1] implies that (50) is equivalent to

$$\begin{aligned} \mathrm{R}'(\gamma) = \ \sup \quad & q \\ \text{s.t.} \quad & p, q \in \mathbb{R}_+, \ \overline{\boldsymbol{\xi}}, \underline{\boldsymbol{\xi}} \in \mathbb{R}_T^+, \ \prod_{t=1}^T \underline{\xi}_t \geq \gamma \\ & p + q = 1 \\ & p\overline{\boldsymbol{\xi}} + q\underline{\boldsymbol{\xi}} = \mu\mathbf{1} \\ & p\overline{\boldsymbol{\xi}}\,\overline{\boldsymbol{\xi}}^\mathsf{T} + q\underline{\boldsymbol{\xi}}\,\underline{\boldsymbol{\xi}}^\mathsf{T} \preceq (1-\rho)\sigma^2\mathbb{I} + \left(\mu^2 + \rho\sigma^2\right)\mathbf{1}\mathbf{1}^\mathsf{T}, \end{aligned} \tag{51}$$

where p and q represent dual variables assigned to the two robust constraints in (50). Thus, the primal uncertain convex program (50) is solved under the worst possible realizations of $\overline{\boldsymbol{\xi}}$ and $\underline{\boldsymbol{\xi}}$, while the dual uncertain convex program (51) is solved under the best possible realizations, in which case $\overline{\boldsymbol{\xi}}$ and $\underline{\boldsymbol{\xi}}$ become decision variables. Problem (51) has intuitive appeal as it can be interpreted as a restriction of the original worst-case probability problem that minimizes over all two-point distributions in the ambiguity set $\mathcal{P}'$ with scenarios $\overline{\boldsymbol{\xi}}$ and $\underline{\boldsymbol{\xi}}$ and corresponding probabilities p and q, respectively. Note that (51) constitutes a non-convex program because it involves multilinear terms in the decisions. Using the variable transformations $\boldsymbol{u} \leftarrow p\overline{\boldsymbol{\xi}}$ and $\boldsymbol{v} \leftarrow q\underline{\boldsymbol{\xi}}$ we can reformulate (51) as

$$
\begin{aligned}
\mathrm{R}'(\gamma) = \ \sup \ \ &q \\
\mathrm{s.\,t.} \ \ &p, q \in \mathbb{R}_+, \ \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}_+^T \\
&\textstyle\prod_{t=1}^T v_t \geq q^T \gamma \\
&p + q = 1 \\
&\boldsymbol{u} + \boldsymbol{v} = \mu \mathbf{1} \\
&\tfrac{1}{p}\boldsymbol{u}\boldsymbol{u}^\mathsf{T} + \tfrac{1}{q}\boldsymbol{v}\boldsymbol{v}^\mathsf{T} \preceq (1 - \rho)\sigma^2 \mathbb{I} + \left(\mu^2 + \rho\sigma^2\right) \mathbf{1}\mathbf{1}^\mathsf{T}.
\end{aligned}
\tag{52}
$$

Note that if $p = 0$ ($q = 0$), then $\boldsymbol{u} = \mathbf{0}$ ($\boldsymbol{v} = \mathbf{0}$) for otherwise the matrix inequality is not satisfiable. In (52) and below we adhere to the convention that $0/0 = 0$, which reflects the idea that a scenario with zero probability mass should have zero weight in the covariance matrix. Observe that problem (52) is a convex program. In particular, the first constraint is convex because of the concavity of geometric means, and the last constraint is convex due to a standard Schur complement argument. Exploiting the problem's permutation symmetry and convexity, one can proceed as in Lemma 3.1 to show that (52) has a permutation symmetric minimizer of the form $\boldsymbol{u} = u\mathbf{1}$ and $\boldsymbol{v} = v\mathbf{1}$ for some scalar decision variables $u, v \in \mathbb{R}_+$. Restricting the search to permutation symmetric solutions, problem (52) can therefore be reformulated as

$$
\begin{aligned}
\mathrm{R}'(\gamma) = \ \sup \ \ &q \\
\mathrm{s.\,t.} \ \ &p, q, u, v \in \mathbb{R}_+ \\
&v \geq q\gamma^{1/T} \\
&p + q = 1 \\
&u + v = \mu \\
&(1 - \rho)\sigma^2 \mathbb{I} + \left(\mu^2 + \rho\sigma^2 - \tfrac{u^2}{p} - \tfrac{v^2}{q}\right) \mathbf{1}\mathbf{1}^\mathsf{T} \succeq \mathbf{0}.
\end{aligned}
\tag{53}
$$

It can be shown that the eigenvalues of the matrix $(1 - \rho)\sigma^2 \mathbb{I} + (\mu^2 + \rho\sigma^2 - \tfrac{u^2}{p} - \tfrac{v^2}{q})\mathbf{1}\mathbf{1}^\mathsf{T}$ are given by $(1 - \rho)\sigma^2$ and $(1 - \rho)\sigma^2 + T(\mu^2 + \rho\sigma^2 - \tfrac{u^2}{p} - \tfrac{v^2}{q})$; see *e.g.* [26, Proposition 4]. Since $(1 - \rho)\sigma^2 > 0$ by assumption, the matrix inequality in (53) is equivalent to the scalar constraint

$$
(1 - \rho)\sigma^2 + T\left(\mu^2 + \rho\sigma^2 - \tfrac{u^2}{p} - \tfrac{v^2}{q}\right) \geq 0.
\tag{54}
$$

Any feasible solution of (53) satisfies $q\gamma^{1/T} \leq v \leq \mu$, implying that the optimal value of (53) is bounded above by $\min\{1, \mu\gamma^{-1/T}\}$. For $0 < \gamma^{1/T} \leq \mu$, an optimal solution of (53) is then given

by $(p, q, u, v) = (0, 1, 0, \mu)$, and the optimal value is equal to 1. For $\mu < \gamma^{1/T} < \mu + \frac{(1+(T-1)\rho)\sigma^2}{T\mu}$, on the other hand, an optimal solution is given by $(p, q, u, v) = (1 - \mu\gamma^{-1/T}, \mu\gamma^{-1/T}, 0, \mu)$ with corresponding optimal value $\mu\gamma^{-1/T}$. Indeed, any larger value of q would require a larger value of v, which in turn would violate the non-negativity of u as $u + v = \mu$. One can show that the constraint (54) is always *inactive* at this solution. For $\gamma^{1/T} \geq \mu + \frac{(1+(T-1)\rho)\sigma^2}{T\mu}$, finally, the constraint (54) implies that q must not exceed $\mu\gamma^{-1/T}$, which in turn implies that the constraint must be binding. Furthermore, q has to be strictly positive for otherwise (53) would be solved by $(p, q, u, v) = (1, 0, \mu, 0)$, which contradicts our earlier finding that the constraint (54) is binding. Substituting $p = 1 - q$ and $u = \mu - v$, the left hand side of (54) becomes a quadratic function of v parametric in q. We denote the two roots of this function by v^+ and v^- and define $u^+ = \mu - v^+$ and $u^- = \mu - v^-$. A direct calculation yields

$$u^\pm = (1 - q)\mu \pm \sigma\sqrt{1 + (T-1)\rho}\sqrt{\frac{q(1-q)}{T}} \quad \text{and} \quad v^\pm = q\mu \mp \sigma\sqrt{1 + (T-1)\rho}\sqrt{\frac{q(1-q)}{T}}.$$

By construction, both (u^+, v^+) and (u^-, v^-) satisfy (54) as an equality. However, there is no $q \in (0, 1]$ for which (u^+, v^+) is feasible in (53). Indeed, a direct calculation reveals that the constraint $v^+ \geq q\gamma^{1/T}$ from (53) can hold only if

$$q(\mu - \gamma^{1/T}) \geq \sigma\sqrt{1 + (T-1)\rho}\sqrt{\frac{q(1-q)}{T}}. \tag{55}$$

However, (55) is not satisfiable as its left hand side is strictly negative by assumption, whereas its right hand side is non-negative. Therefore, (u^+, v^+) is infeasible in (53).

In contrast, the second solution (u^-, v^-) is feasible in (53) if we select $q \in (0, 1]$ with

$$u^- \geq 0 \quad \iff \quad q \leq \frac{T\mu^2}{T\mu^2 + \sigma^2(1 + (T-1)\rho)}$$

and

$$v^- \geq q\gamma^{1/T} \quad \iff \quad q \leq \frac{\sigma^2(1 + (T-1)\rho)}{\sigma^2(1 + (T-1)\rho) + T(\mu - \gamma^{1/T})^2}.$$

Problem (53) aims to maximize q, which is tantamount to setting

$$q = \min\left\{ \frac{T\mu^2}{T\mu^2 + \sigma^2(1 + (T-1)\rho)}, \frac{\sigma^2(1 + (T-1)\rho)}{\sigma^2(1 + (T-1)\rho) + T(\mu - \gamma^{1/T})^2} \right\}$$
$$= \frac{\sigma^2(1 + (T-1)\rho)}{\sigma^2(1 + (T-1)\rho) + T(\mu - \gamma^{1/T})^2},$$

where the second equality follows from $\gamma^{1/T} \geq \mu + \frac{(1+(T-1)\rho)\sigma^2}{T\mu}$. Thus, the claim follows. ∎

In addition to admitting an analytical solution, the relaxed right-sided Chebyshev bounds also allow us to determine a distribution $\mathbb{P}^\star \in \mathcal{P}'$ that attains the probability bound.

Corollary 5.1 (Extremal Distribution) *A distribution $\mathbb{P}^\star \in \mathcal{P}'$ attaining the relaxed right-*

sided Chebyshev bound $R'(\gamma)$ *is given by* $\mathbb{P}^\star = p^\star \delta_{[u^\star/p^\star]\mathbf{1}} + q^\star \delta_{[v^\star/q^\star]\mathbf{1}}$, *where*

$$q^\star = \begin{cases} 1 & \text{if } 0 < \gamma \leq \mu^T, \\ \mu\gamma^{-1/T} & \text{if } \mu^T < \gamma < \left(\mu + \frac{\sigma^2\theta}{T\mu}\right)^T, \\ \frac{\sigma^2\theta}{\sigma^2\theta + T(\mu - \gamma^{1/T})^2} & \text{if } \gamma \geq \left(\mu + \frac{\sigma^2\theta}{T\mu}\right)^T, \end{cases}$$

and $p^\star = 1 - q^\star$, *as well as*

$$v^\star = \begin{cases} \mu & \text{if } 0 < \gamma < \left(\mu + \frac{\sigma^2\theta}{T\mu}\right)^T, \\ q^\star\mu + \sigma\sqrt{\frac{\theta q^\star(1-q^\star)}{T}} & \text{if } \gamma \geq \left(\mu + \frac{\sigma^2\theta}{T\mu}\right)^T \end{cases}$$

and $u^\star = \mu - v^\star$, *where* $\theta = 1 + (T-1)\rho > 0$.

Proof: The proof follows directly from that of Theorem 5.2 and is thus omitted. ∎

The relaxed left-sided and right-sided Chebyshev bounds differ in the sense that the left-sided bound coincides with $L(\gamma)$, whereas $R'(\gamma)$ does not equal $R(\gamma)$ in general. The relaxed right-sided Chebyshev bound does coincide with $R(\gamma)$, however, when T is sufficiently large.

Proposition 5.1 *If* $\mu > \sqrt{\frac{1-\rho}{T}}\sigma$, *then* $R'(\gamma) = R(\gamma)$ *for all* $\gamma \geq \overline{\gamma}$, *where*

$$\overline{\gamma}^{1/T} = \mu + \frac{1}{2ab}\left(1 + \sqrt{4ab\sqrt{\frac{1-\rho}{T}}\sigma + 1}\right)$$

with $a = \mu - \sqrt{\frac{1-\rho}{T}}\sigma$, $b = \frac{T}{\sigma^2\theta}$ *and* $\theta = 1 + (T-1)\rho$.

Note that $ab \to \infty$ and thus $\overline{\gamma}^{1/T} \to \mu$ whenever $T \to \infty$. The rate of convergence depends on μ, σ and ρ, and the fastest convergence is observed for large μ and small σ and ρ.

Proof: We first show that $\overline{\gamma}^{1/T} > \mu + \frac{\sigma^2\theta}{T\mu}$ (Step 1), which allows us to invoke Theorem 5.2 to conclude that $R'(\gamma) = \frac{\sigma^2\theta}{\sigma^2\theta + T(\mu - \gamma^{1/T})^2}$. We then employ Corollary 5.1 to construct a distribution $\mathbb{P}^\star \in \mathcal{P}'$ that satisfies $\mathbb{P}^\star\left(\prod_{t=1}^T \tilde{\xi}_t \geq \gamma\right) = R'(\gamma)$ (Step 2), and we show that a suitable perturbation of $\mathbb{P}^\star$ results in a distribution $\mathbb{P} \in \mathcal{P}$ that satisfies $\mathbb{P}\left(\prod_{t=1}^T \tilde{\xi}_t \geq \gamma\right) = \mathbb{P}^\star\left(\prod_{t=1}^T \tilde{\xi}_t \geq \gamma\right)$ (Step 3). The statement then follows from the fact that $R(\gamma)$ is bounded above by $R'(\gamma)$.

Step 1: We show that $\overline{\gamma}^{1/T}$ is the maximum root of the convex quadratic function

$$q(x) = \sigma^2\theta\left[a\left(1 + b(\mu - x)^2\right) - x\right],$$

where a and b are defined in the statement of the theorem, and that this root satisfies $\overline{\gamma}^{1/T} > \mu + \frac{\sigma^2\theta}{T\mu}$. From the quadratic formula we know that the maximum root $x^\star$ of $q(x)$ satisfies

$$x^\star = \frac{2ab\mu + 1 + \sqrt{(2ab\mu + 1)^2 - 4a^2b(b\mu^2 + 1)}}{2ab} = \mu + \frac{1}{2ab}\left(1 + \sqrt{4ab(\mu - a) + 1}\right),$$

and replacing a and b with their definitions reveals that $x^\star = \overline{\gamma}^{1/T}$. To show that $\overline{\gamma}^{1/T} > \mu + \frac{\sigma^2\theta}{T\mu}$, we observe that

$$q\left(\mu + \frac{\sigma^2\theta}{T\mu}\right) = \left(\mu - \sqrt{\frac{1-\rho}{T}}\sigma\right)\left(\sigma^2\theta + \frac{\sigma^4\theta^2}{T\mu^2}\right) - \sigma^2\theta\left(\mu + \frac{\sigma^2\theta}{T\mu}\right)$$

$$= \sigma^2\theta(\sigma^2\theta + T\mu^2)\left(\frac{\mu - \sqrt{(1-\rho)/T}\sigma}{T\mu^2} - \frac{1}{T\mu}\right) < 0,$$

as well as $q(x) \to \infty$ for $x \to \infty$ since $\mu > \sqrt{\frac{1-\rho}{T}}\sigma$. Since $q(x)$ is quadratic, both observations imply that the maximum root $x^\star = \overline{\gamma}^{1/T}$ of $q(x)$ indeed belongs to the interval $\left(\mu + \frac{\sigma^2\theta}{T\mu}, \infty\right)$.

Step 2: The distribution $\mathbb{P}^\star$ in Corollary 5.1 satisfies $\mathbb{P}^\star\left(\prod_{t=1}^T \tilde{\xi}_t \geq \gamma\right) = \mathrm{R}'(\gamma)$. For later reference, we remark that $\mathbb{P}^\star = p^\star \delta_{[u^\star/p^\star]\mathbf{1}} + q^\star \delta_{[v^\star/q^\star]\mathbf{1}}$ satisfies the properties

$$v^\star = q^\star\gamma^{1/T}, \quad u^\star + v^\star = \mu \quad \text{and} \quad \frac{(u^\star)^2}{p^\star} + \frac{(v^\star)^2}{q^\star} = \mu^2 + \frac{1}{T}(1 + (T-1)\rho)\sigma^2. \tag{56}$$

Note that the last condition holds because (54) is binding when $\gamma^{1/T} \geq \mu + \frac{\sigma^2\theta}{T\mu}$.

Step 3: Consider the distribution $\mathbb{P}$ defined through

$$\mathbb{P}\left(\tilde{\boldsymbol{\xi}} = \left(\frac{u^\star}{p^\star} - \lambda\right)\mathbf{1} + T\lambda\mathbf{e}_i\right) = \frac{1}{T}p^\star, \quad i = 1, \ldots, T, \quad \text{and} \quad \mathbb{P}\left(\tilde{\boldsymbol{\xi}} = \frac{v^\star}{q^\star}\mathbf{1}\right) = q^\star$$

with $\lambda = \sqrt{\frac{1-\rho}{p^\star T}}\sigma$. If $\mathbb{P} \in \mathcal{P}$, then we find that

$$\mathrm{R}(\gamma) \geq \mathbb{P}\left(\prod_{t=1}^T \tilde{\xi}_t = \gamma\right) \geq \mathbb{P}\left(\tilde{\boldsymbol{\xi}} = \frac{v^\star}{q^\star}\mathbf{1}\right) = q^\star = \mathrm{R}'(\gamma),$$

which implies $R(\gamma) = R'(\gamma)$. We thus need to show that $\mathbb{P} \in \mathcal{P}$. To this end, we first observe that the first two moments of $\tilde{\boldsymbol{\xi}}$ under $\mathbb{P}$ satisfy

$$\mathbb{E}_\mathbb{P}\left(\tilde{\boldsymbol{\xi}}\right) = \frac{p^\star}{T} \sum_{i=1}^T \left(\left(\frac{u^\star}{p^\star} - \lambda\right)\mathbf{1} + T\lambda\mathbf{e}_i\right) + v^\star\mathbf{1} = (u^\star + v^\star)\mathbf{1} = \boldsymbol{\mu}$$

$$\mathbb{E}_\mathbb{P}\left(\tilde{\boldsymbol{\xi}}\tilde{\boldsymbol{\xi}}^\intercal\right) = \frac{p^\star}{T} \sum_{i=1}^T \left(\left(\frac{u^\star}{p^\star} - \lambda\right)\mathbf{1} + T\lambda\mathbf{e}_i\right)\left(\left(\frac{u^\star}{p^\star} - \lambda\right)\mathbf{1} + T\lambda\mathbf{e}_i\right)^\intercal + \frac{(v^\star)^2}{q^\star}\mathbf{1}\mathbf{1}^\intercal$$

$$= \frac{p^\star}{T}\left(\left(T\left(\frac{u^\star}{p^\star} - \lambda\right)^2 + 2\left(\frac{u^\star}{p^\star} - \lambda\right)T\lambda\right)\mathbf{1}\mathbf{1}^\intercal + T^2\lambda^2\mathbb{I}\right) + \frac{(v^\star)^2}{q^\star}\mathbf{1}\mathbf{1}^\intercal$$

$$= \left(\frac{(u^\star)^2}{p^\star} + \frac{(v^\star)^2}{q^\star} - p^\star\lambda^2\right)\mathbf{1}\mathbf{1}^\intercal + p^\star T\lambda^2\mathbb{I}$$

$$= (\mu^2 + \rho\sigma^2)\mathbf{1}\mathbf{1}^\intercal + (1-\rho)\sigma^2\mathbb{I} = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal,$$

where the last row is due to (56) and our definition of λ. It remains to be shown that $\tilde{\boldsymbol{\xi}}$ is non-negative $\mathbb{P}$-a.s. By construction of $\mathbb{P}$, this is the case iff $u^\star \geq p^\star\lambda$. We now observe that

$$u^\star = \mu - q^\star\gamma^{1/T} = \mu - \frac{\sigma^2\theta\gamma^{1/T}}{\sigma^2\theta + T(\mu - \gamma^{1/T})^2} \geq \sqrt{\frac{1-\rho}{T}}\sigma,$$

where the first identity follows from (56), the second one is due to the definition of $q^\star$ in Corollary 5.1, and the inequality holds since there is $C > 0$ such that

$$q\left(\gamma^{1/T}\right) = C\left[\mu - \frac{\sigma^2\theta\gamma^{1/T}}{\sigma^2\theta + T(\mu - \gamma^{1/T})^2} - \sqrt{\frac{1-\rho}{T}}\sigma\right],$$

and this expression is non-negative whenever $\gamma \geq \overline{\gamma}$. We thus conclude that

$$\frac{(u^\star)^2}{p^\star} \geq (u^\star)^2 \geq \frac{(1-\rho)\sigma^2}{T},$$

which in turn implies that $u^\star \geq \sqrt{\frac{(1-\rho)p^\star}{T}}\sigma = p^\star\lambda$ as desired. The claim now follows. ∎

6 Extensions

The techniques developed in this paper can also be used to construct Chebyshev bounds for sums, minima and maxima of non-negative random variables. All these Cheybshev bounds can be reduced to computing $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}(h(\tilde{\boldsymbol{\xi}}) \leq 0)$ for some permutation-symmetric functional $h(\boldsymbol{\xi})$.

Theorem 6.1 *For any permutation-symmetric continuous functional $h : \mathbb{R}_+^T \to \mathbb{R}$, we have*

$$
\begin{aligned}
\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}(h(\tilde{\boldsymbol{\xi}}) \leq 0) = \quad \inf \quad & \alpha + T\mu\beta + T(\mu^2 + \sigma^2)\gamma_1 + T\left[T\mu^2 + \sigma^2 + (T-1)\rho\sigma^2\right]\gamma_2 \\
\text{s.t.} \quad & \alpha, \lambda_1, \lambda_2 \in \mathbb{R}_+, \ \ \beta, \gamma_1, \gamma_2 \in \mathbb{R} \\
& \gamma_1 + \gamma_2 \geq 0, \ \ \gamma_2 + \gamma_1 + \alpha \geq \left\|(\beta - \lambda_1, \gamma_2 + \gamma_1 - \alpha)\right\|_2 \\
& \tfrac{\gamma_1}{T} + \gamma_2 \geq 0, \ \ \gamma_2 + \tfrac{\gamma_1}{T} + \alpha \geq \left\|\left(\beta - \lambda_2, \gamma_2 + \tfrac{\gamma_1}{T} - \alpha\right)\right\|_2 \\
& \alpha + \beta s + \gamma_2 s^2 + \gamma_1 \underline{\phi}(s) \geq 1 \quad \forall s \in \mathcal{S} \\
& \alpha + \beta s + \gamma_2 s^2 + \gamma_1 \overline{\phi}(s) \geq 1 \quad \forall s \in \mathcal{S},
\end{aligned}
\tag{57}
$$

where the optimal value functions $\underline{\phi}(s)$ and $\overline{\phi}(s)$ are defined as

$$
\underline{\phi}(s) = \inf_{\boldsymbol{\xi} \geq \mathbf{0}} \left\{ \|\boldsymbol{\xi}\|_2^2 : \|\boldsymbol{\xi}\|_1 = s, \ h(\boldsymbol{\xi}) \leq 0 \right\} \quad and \quad \overline{\phi}(s) = \sup_{\boldsymbol{\xi} \geq \mathbf{0}} \left\{ \|\boldsymbol{\xi}\|_2^2 : \|\boldsymbol{\xi}\|_1 = s, \ h(\boldsymbol{\xi}) \leq 0 \right\}
$$

for all $s \geq 0$, while $\mathcal{S} = \left\{ s \in \mathbb{R}_+ : \underline{\phi}(s) < +\infty \right\}$ denotes the effective domain of $\underline{\phi}(s)$ and $\overline{\phi}(s)$.

Proof: The proof is largely based on arguments familiar from Theorems 3.1 and 4.1. Details are omitted for brevity of exposition. ∎

The significance of Theorem 6.1 is that it enables us to compute $\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}(h(\tilde{\boldsymbol{\xi}}) \leq 0)$ by solving a semidefinite program whenever $\underline{\phi}(s)$ and $\overline{\phi}(s)$ are piecewise polynomials. In this case the last two constraints in (57) reduce to the requirement that a univariate piecewise polynomial, whose coefficients depend affinely on the decision variables, must be non-negative uniformly on $\mathcal{S}$. Such conditions can systematically be reformulated as linear matrix inequalities [23].

Table 1 lists examples of permutation-symmetric functionals $h(\boldsymbol{\xi})$ that lead to piecewise polynomial mappings $\underline{\phi}(s)$ and $\overline{\phi}(s)$ and thus to computable Chebyshev bounds. Theorems 6.2 and 6.3 below present two special cases in which these bounds can be evaluated analytically.

Theorem 6.2 (Left-Sided Chebyshev Bound for Sums) *For any $\gamma > 0$ we have*

$$
\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\left(\sum_{t=1}^T \tilde{\xi}_t \geq \gamma\right) = \begin{cases} \frac{T\sigma^2\theta}{T\sigma^2\theta + (\gamma - T\mu)^2} & \text{if } \gamma \geq T\mu + \sigma^2\theta/\mu, \\ \frac{T\mu}{\gamma} & \text{if } T\mu \leq \gamma < T\mu + \sigma^2\theta/\mu, \\ 1 & \text{if } \gamma < T\mu, \end{cases}
$$

where $\theta = 1 + (T-1)\rho > 0$.

Proof: By Theorem 6.1 the Chebyshev bound $\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}(\sum_{t=1}^T \tilde{\xi}_t \geq \gamma)$ can be reformulated as the semi-infinite program (57), where the functions $\underline{\phi}(s)$ and $\overline{\phi}(s)$ are specified in Table 1. Distinguishing the cases $\gamma_1 \geq 0$ and $\gamma_1 < 0$, this semi-infinite program can be reduced to a robust optimization problem with a scalar uncertain parameter by using the '*primal worst equals dual best*' duality scheme from robust optimization [2]. One can further show that the optimal value of this problem coincides with the univariate Chebyshev bound $\sup_{\mathbb{P}_1 \in \mathcal{P}} \mathbb{P}(\tilde{\xi} \geq \gamma)$, where $\mathcal{P}_1$ contains all distributions of $\tilde{\xi}$ supported on $\mathbb{R}_+$ with mean $T\mu$ and variance $\sigma^2 T(1 + (T-1)\rho)$. The latter Chebyshev bound has an analytical formula, which can be obtained from (3). ∎

Chebyshev bound	$h(\boldsymbol{\xi})$	$\underline{\phi}(s)$	$\overline{\phi}(s)$
$\displaystyle\sup_{\mathbb{P}\in\mathcal{P}}\mathbb{P}\left(\min_{t=1,\ldots,T}\tilde{\xi}_t\leq\gamma\right)$	$\displaystyle\min_{t=1,\ldots,T}\xi_t-\gamma$	$\begin{cases}\dfrac{s^2}{T} & \text{if } s\geq\gamma T\\[4pt]\gamma^2+\dfrac{1}{T-1}(s-\gamma)^2 & \text{if } 0\leq s<\gamma T\end{cases}$	s^2
$\displaystyle\sup_{\mathbb{P}\in\mathcal{P}}\mathbb{P}\left(\min_{t=1,\ldots,T}\tilde{\xi}_t\geq\gamma\right)$	$\displaystyle\gamma-\min_{t=1,\ldots,T}\xi_t$	$\begin{cases}\dfrac{s^2}{T} & \text{if } s\geq\gamma T\\[4pt]+\infty & \text{if } 0\leq s<\gamma T\end{cases}$	$\begin{cases}(s-\gamma T)^2+2\gamma(s-\gamma T)+T\gamma^2 & \text{if } s\geq\gamma T\\[4pt]-\infty & \text{if } 0\leq s<\gamma T\end{cases}$
$\displaystyle\sup_{\mathbb{P}\in\mathcal{P}}\mathbb{P}\left(\max_{t=1,\ldots,T}\tilde{\xi}_t\leq\gamma\right)$	$\displaystyle\max_{t=1,\ldots,T}\xi_t-\gamma$	$\begin{cases}\dfrac{s^2}{T} & \text{if } 0\leq s\leq\gamma T\\[4pt]+\infty & \text{if } s>\gamma T\end{cases}$	$\begin{cases}s^2 & \text{if } 0\leq s\leq\gamma\\[2pt]\gamma^2+(s-\gamma)^2 & \text{if } \gamma<s\leq 2\gamma\\[2pt]\cdots & \vdots\\[2pt](T-1)\gamma^2+(s-(T-1)\gamma)^2 & \text{if }(T-1)\gamma<s\leq T\gamma\\[2pt]-\infty & \text{if } s>T\gamma\end{cases}$
$\displaystyle\sup_{\mathbb{P}\in\mathcal{P}}\mathbb{P}\left(\max_{t=1,\ldots,T}\tilde{\xi}_t\geq\gamma\right)$	$\displaystyle\gamma-\max_{t=1,\ldots,T}\xi_t$	$\begin{cases}\dfrac{s^2}{T} & \text{if } s\geq\gamma T\\[4pt]\gamma^2+\dfrac{1}{T-1}(s-\gamma)^2 & \text{if } \gamma\leq s<\gamma T\\[4pt]+\infty & \text{if } 0\leq s<\gamma\end{cases}$	$\begin{cases}s^2 & \text{if } s\geq\gamma\\[4pt]-\infty & \text{if } 0\leq s<\gamma\end{cases}$
$\displaystyle\sup_{\mathbb{P}\in\mathcal{P}}\mathbb{P}\left(\sum_{t=1}^{T}\tilde{\xi}_t\leq\gamma\right)$	$\displaystyle\sum_{t=1}^{T}\tilde{\xi}_t-\gamma$	$\begin{cases}\dfrac{s^2}{T} & \text{if } 0\leq s\leq\gamma\\[4pt]+\infty & \text{if } s>\gamma\end{cases}$	$\begin{cases}s^2 & \text{if } 0\leq s\leq\gamma\\[4pt]-\infty & \text{if } s>\gamma\end{cases}$
$\displaystyle\sup_{\mathbb{P}\in\mathcal{P}}\mathbb{P}\left(\sum_{t=1}^{T}\tilde{\xi}_t\geq\gamma\right)$	$\displaystyle\gamma-\sum_{t=1}^{T}\tilde{\xi}_t$	$\begin{cases}\dfrac{s^2}{T} & \text{if } s\geq\gamma\\[4pt]+\infty & \text{if } 0\leq s<\gamma\end{cases}$	$\begin{cases}s^2 & \text{if } s\geq\gamma\\[4pt]-\infty & \text{if } 0\leq s<\gamma\end{cases}$

Table 1: Chebyshev bounds equivalent to $\sup_{\mathbb{P}\in\mathcal{P}}\mathbb{P}(h(\tilde{\boldsymbol{\xi}})\leq 0)$ for some permutation symmetric functional $h(\boldsymbol{\xi})$. These bounds coincide with the optimal value of (57), instantiated with the respective piecewise polynomials $\underline{\phi}(s)$ and $\overline{\phi}(s)$.

Theorem 6.3 (Right-Sided Chebyshev Bound for Sums) *For any $\gamma > 0$ we have*

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\left(\sum_{t=1}^{T} \tilde{\xi}_t \leq \gamma\right) = \begin{cases} 1 & \text{if } \gamma \geq T\mu, \\ \frac{T\sigma^2\theta}{T\sigma^2\theta + (\gamma - T\mu)^2} & \text{otherwise,} \end{cases}$$

where $\theta = 1 + (T-1)\rho > 0$.

Proof: The proof is widely parallel to that of Theorem 6.2 and is thus omitted for brevity.

7 Numerical Experiments

We first compare our Chebyshev bounds $\mathrm{R}(\gamma)$ and $\mathrm{L}(\gamma)$ with alternative bounds proposed in the literature, as well as the relaxed Chebyshev bound $\mathrm{R}'(\gamma)$ from Section 5. We then present a case study that employs our left-sided Chebyshev bound $\mathrm{L}(\gamma)$ to select financial portfolios under imprecise knowledge of the asset return distributions. All optimization problems are solved with the SDPT3 optimization software using the YALMIP interface [21, 29].

7.1 Comparison of Chebyshev Bounds

Instead of employing the bounds $\mathrm{R}(\gamma)$ and $\mathrm{L}(\gamma)$ from Sections 3 and 4, which are exact but may result in computationally challenging optimization problems, one can employ existing results to derive approximate bounds on the tail probabilities of a product of non-negative, permutation-symmetric random variables. In the following, we compare our bounds with two such approximations based on earlier results of Marshall and Olkin [22] and Vandenberghe et al. [30]. Both approximations rely on the larger ambiguity set

$$\mathcal{P}^0 = \left\{ \mathbb{P} \in \mathcal{M}_+(\mathbb{R}^T) \; : \; \mathbb{E}_{\mathbb{P}}\left(\tilde{\boldsymbol{\xi}}\right) = \boldsymbol{\mu}, \; \mathbb{E}_{\mathbb{P}}\left(\tilde{\boldsymbol{\xi}}\tilde{\boldsymbol{\xi}}^\intercal\right) = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal \right\}$$

with support $\mathbb{R}^T$, where $\boldsymbol{\mu} \in \mathbb{R}^T$ and $\boldsymbol{\Sigma} \in \mathbb{S}_+^T$, $\boldsymbol{\Sigma} \succ \mathbf{0}$, need not be permutation-symmetric.

Marshall and Olkin [22] derive a convex optimization problem that provides a tight upper bound on the probability that the random vector $\tilde{\boldsymbol{\xi}}$ is contained in a closed convex set $\mathcal{C}$, assuming that $\tilde{\boldsymbol{\xi}}$ can be governed by any distribution from the ambiguity set $\mathcal{P}^0$. The choice $\mathcal{C} = \left\{\boldsymbol{\xi} \in \mathbb{R}^T \; : \; \prod_{t=1}^{T} \xi_t \geq \gamma\right\}$ allows us to approximate the right-sided Chebyshev bound $\mathrm{R}(\gamma)$. For this special case, the bound of Marshall and Olkin has the analytical solution

$$\mathrm{R}^{\mathrm{MO}}(\gamma) = \begin{cases} 1 & \text{if } 0 < \gamma \leq \mu^T, \\ \frac{\sigma^2(1+(T-1)\rho)}{\sigma^2(1+(T-1)\rho) + T(\mu - \gamma^{1/T})^2} & \text{if } \gamma > \mu^T, \end{cases}$$

which follows from [4, Theorem 6.1]. By construction, $\mathrm{R}^{\mathrm{MO}}(\gamma) \geq \mathrm{R}(\gamma)$ since $\mathcal{P} \subset \mathcal{P}^0$. Note that $\mathrm{R}^{\mathrm{MO}}(\gamma)$ coincides with our relaxed Chebyshev bound $\mathrm{R}'(\gamma)$ for $\gamma \geq \left(\mu + \frac{\sigma^2\theta}{T\mu}\right)^T$, see Theorem 5.2. Thus, $\mathrm{R}^{\mathrm{MO}}(\gamma)$ also coincides with our right-sided Chebyshev bound $\mathrm{R}(\gamma)$ for large values of γ, see Proposition 5.1. Note that the bound of Marshall and Olkin cannot be used to approximate our left-sided Chebyshev bound $\mathrm{L}(\gamma)$ since the complement of $\mathcal{C}$ fails to be convex.

	Number of random variables T									
	4	8	12	16	20	24	28	32	36	40
VBC bounds	1.02	1.01	1.06	1.07	1.11	1.23	1.29	1.48	1.72	2.02
Our bounds	1.63	1.81	2.19	2.64	3.43	4.71	6.38	9.34	13.37	18.35

Table 2: Runtimes (secs) required to calculate the Chebyshev bounds. Each runtime is averaged over 10 instances with randomly selected μ, σ and γ, and it includes the calculation of both the left-sided and the right-sided bounds.

Vandenberghe et al. [30] derive a semidefinite program that provides a tight upper bound on the probability that $\tilde{\boldsymbol{\xi}} \in \mathcal{C}$ for a (not necessarily convex) set $\mathcal{C} = \{\boldsymbol{\xi} \in \mathbb{R}^T \; : \; \boldsymbol{\xi}^\intercal \boldsymbol{A}_i \boldsymbol{\xi} + 2\boldsymbol{b}_i^\intercal \boldsymbol{\xi} + c_i < 0 \;\; \forall i = 1, \ldots, m\}$, assuming that the random vector $\tilde{\boldsymbol{\xi}}$ can be governed by any distribution from the ambiguity set $\mathcal{P}^0$. Employing a second-order Taylor approximation of $\prod_{t=1}^T \xi_t$ around $\mu \mathbf{1}$,

$$\prod_{t=1}^T \xi_t \approx \mu^{T-2} \left(\mu^2 + \mu(\boldsymbol{\xi} - \mu\mathbf{1})^\intercal \mathbf{1} + \frac{1}{2}(\boldsymbol{\xi} - \mu\mathbf{1})^\intercal(\mathbf{1}\mathbf{1}^\intercal - \mathbb{I})(\boldsymbol{\xi} - \mu\mathbf{1}) \right)$$

$$= \mu^{T-2} \left((1 - T)\mu^2 + \mu\boldsymbol{\xi}^\intercal \mathbf{1} + \frac{1}{2}\boldsymbol{\xi}^\intercal(\mathbf{1}\mathbf{1}^\intercal - \mathbb{I})\boldsymbol{\xi} + \frac{1}{2}\mu^2 T(T-1) - (T-1)\mu\boldsymbol{\xi}^\intercal \mathbf{1} \right)$$

$$= \frac{1}{2}\mu^{T-2} \left((T-1)(T-2)\mu^2 - 2(T-2)\mu\boldsymbol{\xi}^\intercal \mathbf{1} + \boldsymbol{\xi}^\intercal(\mathbf{1}\mathbf{1}^\intercal - \mathbb{I})\boldsymbol{\xi} \right),$$

we can derive an approximate right-sided Chebyshev bound $\mathrm{R}^{\mathrm{VBC}}(\gamma) = \sup_{\mathbb{P} \in \mathcal{P}^0} \mathbb{P}\left(\tilde{\boldsymbol{\xi}} \in \mathcal{C}\right)$ by replacing the product $\prod_{t=1}^T \xi_t$ with its Taylor approximation in the definition of the set $\mathcal{C}$:

$$\mathcal{C} = \left\{ \boldsymbol{\xi} \in \mathbb{R}^T \; : \; \frac{1}{2}\mu^{T-2}\left((T-1)(T-2)\mu^2 - 2(T-2)\mu\boldsymbol{\xi}^\intercal \mathbf{1} + \boldsymbol{\xi}^\intercal(\mathbf{1}\mathbf{1}^\intercal - \mathbb{I})\boldsymbol{\xi} \right) > \gamma \right\}$$

A similar approximation $\mathrm{L}^{\mathrm{VBC}}(\gamma)$ can be derived for our left-sided Chebyshev bound $\mathrm{L}(\gamma)$ by considering the strict complement of $\mathcal{C}$. Note that $\mathrm{R}^{\mathrm{VBC}}(\gamma)$ and $\mathrm{L}^{\mathrm{VBC}}(\gamma)$ can over- or underestimate our bounds $\mathrm{R}(\gamma)$ and $\mathrm{L}(\gamma)$ due to the use of the Taylor approximation.

Figure 3 compares our Chebyshev bounds $\mathrm{L}(\gamma)$ and $\mathrm{R}(\gamma)$ with the approximate bounds $\mathrm{L}^{\mathrm{VBC}}(\gamma)$ and $\mathrm{R}^{\mathrm{VBC}}(\gamma)$ ('VBC bounds') as well as $\mathrm{R}^{\mathrm{MO}}(\gamma)$ ('MO bound'). As expected, the VBC bounds can over- and underestimate our bounds $\mathrm{L}(\gamma)$ and $\mathrm{R}(\gamma)$, whereas the MO bound consistently overestimates $\mathrm{R}(\gamma)$. Moreover, the MO bound coincides with our right-sided Chebyshev bound for large values of γ. The quality of both approximations deteriorates with increasing σ and decreasing γ. Interestingly, the VBC bound deterioates with increasing numbers of random variables, whereas the MO bound improves with increasing T. The figure shows that both approximate bounds can misestimate the bounds $\mathrm{L}(\gamma)$ and $\mathrm{R}(\gamma)$ substantially.

The MO bound has an analytical solution and can therefore be computed in negligible time. In contrast, the VBC bounds and our bounds require the solution of semidefinite programs with two LMIs of size $\mathcal{O}(T^2)$. Table 2 compares the computation times of both bounds for products of different size T on a computer with a 3.40GHz i7 CPU and 16GB RAM. While both bounds can be computed within seconds, the VBC bounds require significantly less runtime than our bounds. We attribute this to the LMI reformulations of the polynomial constraints in Theorems 3.1 and 4.1, which seem to lack structure that can be exploited by SDPT3.
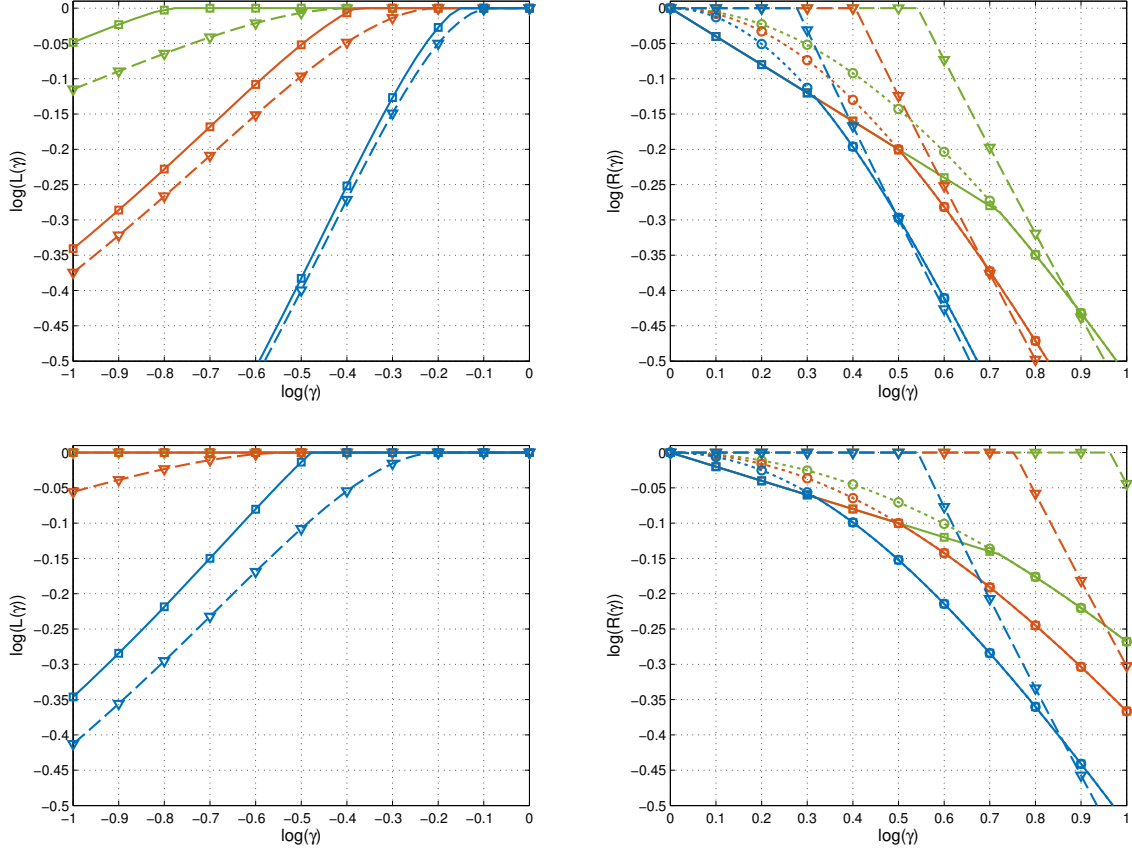
Figure 3: Comparison of the left-sided (left) and right-sided (right) Chebyshev bounds for the products of $T = 5$ (top) and $T = 10$ (bottom) random variables with $\mu = 1$ and $\rho = 0$. The solid lines with squares, the dashed lines with triangles and the dotted lines with circles represent our bounds, the VBC bounds and the MO bounds, respectively. From bottom to top, the blue, red and green lines correspond to $\sigma = 0.2$, 0.3 and 0.4 (left) and $\sigma = 0.4$, 0.5 and 0.6 (right), respectively.

Figure 4 compares the right-sided Chebyshev bound $R(\gamma)$ with the relaxed right-sided bound $R'(\gamma)$ and the MO bound $R^{MO}(\gamma)$. The figure illustrates that $R^{MO}(\gamma)$ coincides with $R'(\gamma)$ for $\gamma \geq \left(\mu + \frac{\sigma^2 \theta}{T\mu}\right)^T$, and subsequently both bounds coincide with $R(\gamma)$ for large values of γ. The gaps between the bounds increase with larger variances σ^2, and they decrease with larger numbers of random variables T.

7.2 Case Study: Financial Risk Management

Consider an investor who allocates a limited budget to a fixed pool of n assets over a time horizon of T periods. We denote by $\tilde{r}_{t,i} \geq -1$, $t = 1, \ldots, T$ and $i = 1, \ldots, n$, the relative price change of asset i between periods t and $t + 1$. We assume that the investor pursues a fixed-mix (or constant proportions) strategy which rebalances the portfolio composition to a pre-selected set of weights $\boldsymbol{w} \in \mathcal{W} = \{\boldsymbol{z} \in \mathbb{R}_+^n : \mathbf{e}^\intercal \boldsymbol{z} = 1\}$ at the beginning of each period. Note that despite being memoryless, fixed-mix strategies are dynamic since they recapitalize those assets whose returns were below average ('buy low') and divest assets whose returns were above average ('sell high'). Fixed-mix strategies generalize the well-known $1/N$-portfolio [9], and they have received significant attention among both academics and practitioners.
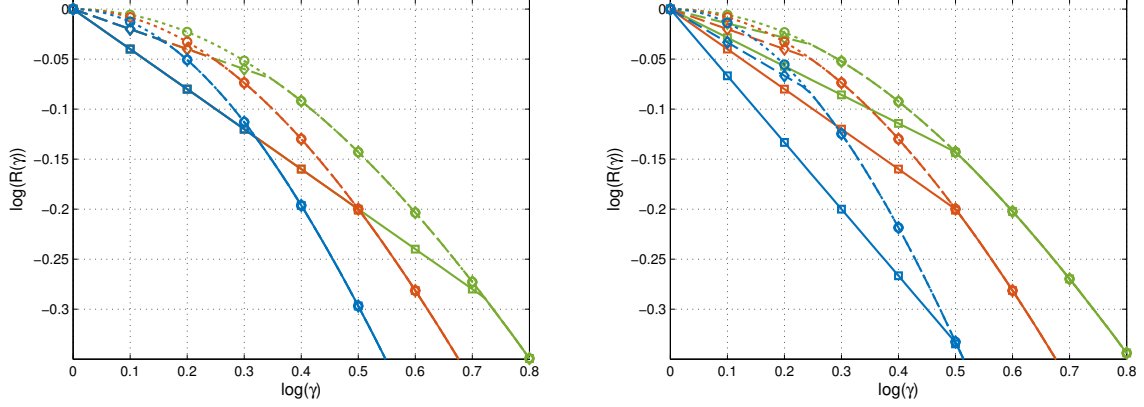
Figure 4: Comparison of the right-sided Chebyshev bounds $\mathrm{R}(\gamma)$ (solid lines with squares), $\mathrm{R}'(\gamma)$ (dashed lines with diamonds) and $\mathrm{R^{MO}}(\gamma)$ (dotted lines with circles) with $\mu = 1$ and $\rho = 0$. From bottom to top, the blue, red and green lines correspond to $\sigma = 0.4$, 0.5 and 0.6 in the left graph (with $T = 5$ fixed) and to $T = 3$, 5 and 7 in the right graph (with $\sigma = 0.5$ fixed), respectively.

We assume that the investor assesses the fixed-mix strategy $\boldsymbol{w}$ in view of the value-at-risk of the portfolio's terminal wealth, which is defined as

$$\mathrm{VaR}_\epsilon(\boldsymbol{w}) = \sup_{\gamma \in \mathbb{R}} \left\{ \gamma : \mathbb{P}\left(\prod_{t=1}^{T}(1 + \boldsymbol{w}^\intercal \tilde{\boldsymbol{r}}_t) > \gamma\right) \geq 1 - \epsilon \right\}.$$

Here, the asset returns $\tilde{\boldsymbol{r}}_t = (\tilde{r}_{t,i})_{i=1}^{n}$ are governed by the probability distribution $\mathbb{P}$, and ϵ is a pre-specified parameter that reflects the investor's risk tolerance.

Calculating the value-at-risk of a portfolio's terminal wealth requires perfect knowledge of the joint asset return distribution $\mathbb{P}$, which is unavailable in practice. Following [26], we will assume that it is only known that the asset returns $(\tilde{\boldsymbol{r}}_t)_{t=1}^{T}$ follow a weak-sense white noise process with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$, that is, the asset returns are serially uncorrelated and have period-wise identical first and second-order moments. In that case, the wealth evolution $(\tilde{\xi}_t)_{t=1}^{T} = (1 + \boldsymbol{w}^\intercal \tilde{\boldsymbol{r}}_t)_{t=1}^{T}$ also follows a weak-sense stochastic process governed by a distribution $\mathbb{P}_{\boldsymbol{w}}$ supported on $\mathbb{R}_+^T$, under which the $\tilde{\xi}_t$ have mean $\boldsymbol{w}^\intercal \boldsymbol{\mu}$ and variance $\boldsymbol{w}^\intercal \boldsymbol{\Sigma} \boldsymbol{w}$ and are serially uncorrelated. We denote the set of all these distributions by $\mathcal{P}_{\boldsymbol{w}}$. In this setting, an ambiguity-averse investor may assess the fixed-mix strategy $\boldsymbol{w}$ in view of the *worst-case* value-at-risk of the portfolio's terminal wealth over all distributions $\mathbb{P}_{\boldsymbol{w}} \in \mathcal{P}_{\boldsymbol{w}}$:

$$\mathrm{WVaR}_\epsilon(\boldsymbol{w}) = \sup_{\gamma \in \mathbb{R}} \left\{ \gamma : \inf_{\mathbb{P}_{\boldsymbol{w}} \in \mathcal{P}_{\boldsymbol{w}}} \mathbb{P}_{\boldsymbol{w}}\left(\prod_{t=1}^{T} \tilde{\xi}_t > \gamma\right) \geq 1 - \epsilon \right\}.$$

In [26], the worst-case value-at-risk of the portfolio's terminal wealth is replaced with a quadratic approximation. The Chebyshev bounds proposed in this paper allow us to calculate the worst-case value-at-risk exactly without resorting to any approximation. Indeed, one verifies that

$$\mathrm{WVaR}_\epsilon(\boldsymbol{w}) = \sup_{\gamma \in \mathbb{R}} \left\{ \gamma : \sup_{\mathbb{P}_{\boldsymbol{w}} \in \mathcal{P}_{\boldsymbol{w}}} \mathbb{P}_{\boldsymbol{w}}\left(\prod_{t=1}^{T} \tilde{\xi}_t \leq \gamma\right) \leq \epsilon \right\} = \sup_{\gamma \in \mathbb{R}} \left\{ \gamma : L(\gamma; \boldsymbol{w}^\intercal \boldsymbol{\mu}, \boldsymbol{w}^\intercal \boldsymbol{\Sigma} \boldsymbol{w}) \leq \epsilon \right\},$$
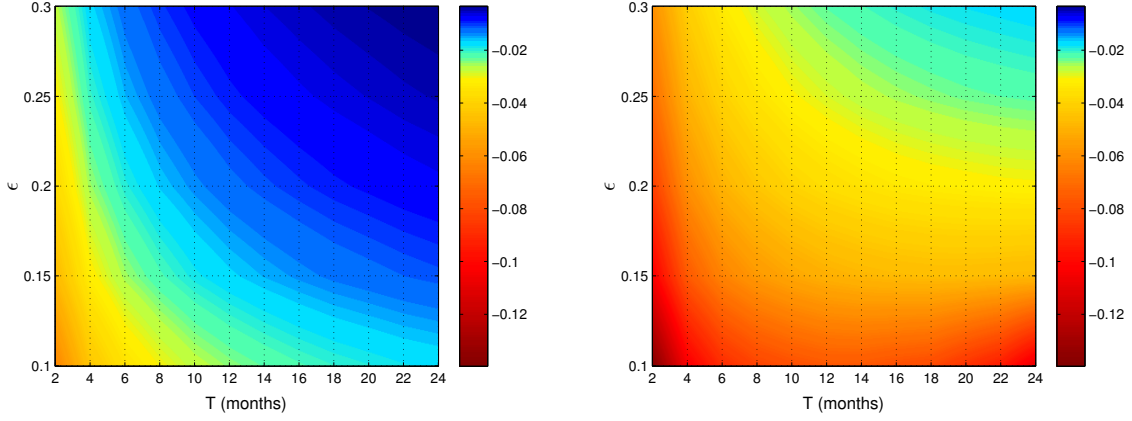
Figure 5: Wort-case value-at-risk of the growth rates of the minimum-variance (left) and maximum-expectation (right) portfolios for different investment horizons T and risk tolerances ϵ.

where we have made explicit the dependence of the left-sided Chebyshev bound L on the mean $\boldsymbol{w}^\mathsf{T}\boldsymbol{\mu}$ and the variance $\boldsymbol{w}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{w}$ of the wealth evolution $(\tilde{\xi}_t)_{t=1}^T$. Since L is monotonically non-decreasing in γ, the last expression can be evaluated efficiently through bisection on γ.

Figure 5 reports the worst-case value-at-risk of two portfolios over different time horizons T, where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are calibrated to the 2003–2012 period of Fama and French's 10 Industry Portfolios data set.[2] The minimum-variance portfolio (left graph) corresponds to the weight vector $\boldsymbol{w} \in \mathcal{W}$ that minimizes $\boldsymbol{w}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{w}$, whereas the maximum-expectation portfolio (right graph) invests all wealth into the asset i with the highest expected return μ_i. To facilitate a fair comparison among different time horizons, the graphs report the growth rates of the portfolios, that is, the logarithms of the terminal wealth, divided by the number of investment periods T. As expected, the minimum-variance portfolio is less risky than the maximum-expectation portfolio, and the risk of both portfolios tends to decrease when the investment horizon T grows. Interestingly, however, the risk of the maximum-expectation portfolio *increases* with large T for low risk tolerances $\epsilon \lesssim 0.15$. This seemingly counter-intuitive effect is explained by Theorem 3.2, which states that the wealth evolution $\prod_{t=1}^T \tilde{\xi}_t$ is absorbed at 0 for large investment horizons T.

In addition to *evaluating* the worst-case value-at-risk of a pre-selected portfolio $\boldsymbol{w}$, an investor often seeks to determine a portfolio $\boldsymbol{w}^\star$ that *optimizes* the worst-case value-at-risk. The search for optimal portfolios is greatly simplified by the observation that there is always a portfolio $\boldsymbol{w}^\star$ on the mean-variance efficient frontier that maximizes $\mathrm{WVaR}_\epsilon(\boldsymbol{w})$ over (subsets of) $\mathcal{W}$. Indeed, Theorem 5.1 implies that $\mathrm{L}(\gamma; \boldsymbol{w}^\mathsf{T}\boldsymbol{\mu}, \boldsymbol{w}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{w}) = \mathrm{L}'(\gamma; \boldsymbol{w}^\mathsf{T}\boldsymbol{\mu}, \boldsymbol{w}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{w})$, and one readily verifies that $\mathrm{L}'(\gamma; \boldsymbol{w}^\mathsf{T}\boldsymbol{\mu}, \boldsymbol{w}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{w})$ is non-decreasing in both γ and $\boldsymbol{w}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{w}$. This implies that

$$\sup_{\gamma \in \mathbb{R}} \left\{ \gamma : \mathrm{L}'(\gamma; \boldsymbol{w}^\mathsf{T}\boldsymbol{\mu}, \boldsymbol{w}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{w}) \leq \epsilon \right\} \leq \sup_{\gamma \in \mathbb{R}} \left\{ \gamma : \mathrm{L}'(\gamma; \boldsymbol{w}'^\mathsf{T}\boldsymbol{\mu}, \boldsymbol{w}'^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{w}') \leq \epsilon \right\}$$

for two portfolios $\boldsymbol{w}$ and $\boldsymbol{w}'$ that satisfy $\boldsymbol{w}^\mathsf{T}\boldsymbol{\mu} = \boldsymbol{w}'^\mathsf{T}\boldsymbol{\mu}$ and $\boldsymbol{w}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{w} \geq \boldsymbol{w}'^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{w}'$. We thus conclude that among all portfolios $\boldsymbol{w} \in \mathcal{W}$ that achieve the same mean return $\boldsymbol{w}^\mathsf{T}\boldsymbol{\mu}$, the portfolio with

---

[2]See http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

smallest variance $\boldsymbol{w}^\intercal \boldsymbol{\Sigma} \boldsymbol{w}$ provides the best worst-case value-at-risk. We can therefore identify an optimal portfolio through a one-dimensional line search over the mean-variance efficient frontier.

References

[1] AITCHISON, J., AND BROWN, J. *The Lognormal Distribution with Special Reference to Its Use in Economics*. Cambridge University Press, 1957.

[2] BECK, A., AND BEN-TAL, A. Duality in robust optimization: Primal worst equals dual best. *Operations Research Letters 37*, 1 (2009), 1–6.

[3] BEN-TAL, A., EL GHAOUI, L., AND NEMIROVSKI, A. *Robust Optimization*. Princeton University Press, 2009.

[4] BERTSIMAS, D., AND POPESCU, I. Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal on Optimization 15*, 3 (2005), 780–804.

[5] BIENAYMÉE, I. Considérations à l'appui de la découverte de laplace sur la loi de probabilité dans la méthode des moindres carrés. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences de Paris 37* (1853), 309–324.

[6] BOYD, S., AND VANDENBERGHE, L. *Convex Optimization*. Cambridge University Press, 2004.

[7] CHEBYSHEV, P. Des valeurs moyennes. *Journal de Mathématiques Pures et Appliquées 12*, 2 (1867), 177–184.

[8] DELAGE, E., AND YE, Y. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research 58*, 3 (2010), 595–612.

[9] DEMIGUEL, V., GARLAPPI, L., AND UPPAL, R. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies 22*, 5 (2009), 1915–1953.

[10] FRISCH, U., AND SORNETTE, D. Extreme deviations and applications. *Journal of Physics I France 7* (1997), 1155–1171.

[11] GALAMBOS, J., AND SIMONELLI, I. *Products of Random Variables: Applications to Problems of Physics and to Arithmetical Functions*. Taylor & Francis, 2004.

[12] GODWIN, H. On generalizations of Tchebycheff's inequality. *Journal of the American Statistical Association 50* (1955), 923–945.

[13] Goh, J., and Sim, M. Distributionally robust optimization and its tractable approximations. *Operations Research 58*, 4 (2010), 902–917.

[14] Hanasusanto, G. A., Roitch, V., Kuhn, D., and Wiesemann, W. A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Mathematical Programming 151*, 1 (2015), 35–62.

[15] Isii, K. The extrema of probability determined by generalized moments (I) bounded random variables. *Annals of the Institute of Statistical Mathematics 12*, 2 (1960), 119–134.

[16] Isii, K. On sharpness of Tchebycheff-type inequalities. *Annals of the Institute of Statistical Mathematics 14*, 1 (1962), 185–197.

[17] Karatzas, I., and Shreve, S. *Brownian Motion and Stochastic Calculus*. Springer, 1991.

[18] Karlin, S., and Studden, W. *Tchebycheff Systems: With Applications in Analysis and Statistics*. Interscience Publishers, 1966.

[19] Krein, M., and Nudelman, A. *The Markov Moment Problem and Extremal Problems*. American Mathematical Society, 1977.

[20] Lindgren, G. *Stationary Stochastic Processes: Theory and Applications*. Taylor & Francis, 2012.

[21] Löfberg, J. YALMIP: A toolbox for modeling and optimization in Matlab. In *Proceedings of the 2004 International Symposium on Computer Aided Control Systems Design* (2004), pp. 284–289.

[22] Marshall, A. W., and Olkin, I. Multivariate Chebyshev inequalities. *The Annals of Mathematical Statistics 31*, 4 (1960), 1001–1014.

[23] Nesterov, Y. Squared functional systems and optimization problems. In *High Performance Optimization*. Kluwer Academic Publishers, 2000, pp. 405–440.

[24] Owhadi, H., Scovel, C., Sullivan, T. J., McKerns, M., and Ortiz, M. Optimal uncertainty quantification. *SIAM Review 55*, 2 (2013), 271–345.

[25] Pólik, I., and Terlaky, T. A survey of the S-lemma. *SIAM Review 49*, 3 (2007), 371–481.

[26] Rujeerapaiboon, N., Kuhn, D., and Wiesemann, W. Robust growth-optimal portfolios. *To appear in Management Science*.

[27] Shapiro, A. On duality theory of conic linear problems. In *Semi-Infinite Programming: Recent Advances*. Kluwer Academic Publishers, 2001, pp. 135–165.

[28] Shohat, J., and Tamarkin, J. *The Problem of Moments*. American Mathematical Society, 1943.

[29] Toh, K. C., Todd, M. J., and Tutuncu, R. H. SDPT3–a Matlab software package for semidefinite programming. *Optimization Methods and Software 11*, 1–4 (1999), 545–581.

[30] Vandenberghe, L., Boyd, S., and Comanor, K. Generalized Chebyshev bounds via semidefinite programming. *SIAM Review 49*, 1 (2007), 52–64.

[31] Wiesemann, W., Kuhn, D., and Sim, M. Distributionally robust convex optimization. *Operations Research 62*, 6 (2014), 1358–1376.

[32] Yu, Y., Li, Y., Schuurmans, D., and Szepesvári, C. A general projection property for distribution families. In *Advances in Neural Information Processing Systems 22*. Curran Associates, Inc., 2009, pp. 2232–2240.