Robust Growth-Optimal Portfolios

Napat Rujeerapaiboon, Daniel Kuhn
Risk Analytics and Optimization Chair, École Polytechnique Fédérale de Lausanne, Switzerland,
napat.rujeerapaiboon@epfl.ch, daniel.kuhn@epfl.ch

Wolfram Wiesemann
Imperial College Business School, Imperial College London, United Kingdom, ww@imperial.ac.uk

The growth-optimal portfolio is designed to have maximum expected log-return over the next rebalancing period. Thus, it can be computed with relative ease by solving a static optimization problem. The growth-optimal portfolio has sparked fascination among finance professionals and researchers because it can be shown to outperform any other portfolio with probability 1 in the long run. In the short run, however, it is notoriously volatile. Moreover, its computation requires precise knowledge of the asset return distribution, which is not directly observable but must be inferred from sparse data. By using methods from distributionally robust optimization, we design fixed-mix strategies that offer similar performance guarantees as the growth-optimal portfolio but for a finite investment horizon and for a whole family of distributions that share the same first and second-order moments. We demonstrate that the resulting robust growth-optimal portfolios can be computed efficiently by solving a tractable conic program whose size is independent of the length of the investment horizon. Simulated and empirical backtests show that the robust growth-optimal portfolios are competitive with the classical growth-optimal portfolio across most realistic investment horizons and for an overwhelming majority of contaminated return distributions.

Key words: Portfolio optimization, growth-optimal portfolio, distributionally robust optimization, value-at-risk, second-order cone programming, semidefinite programming

1. Introduction

Consider a portfolio invested in various risky assets and assume that this portfolio is self-financing in the sense that there are no cash withdrawals or injections after the initial endowment. Loosely speaking, the primary management objective is to design an investment strategy that ensures steady portfolio growth while controlling the fund's risk exposure. Modern portfolio theory based on the pioneering work by Markowitz (1952) suggests that in this situation investors should seek an optimal trade-off between the mean and variance of portfolio returns. The Markowitz approach has gained enormous popularity as it is intuitively appealing and lays the foundations for the celebrated capital asset pricing model due to Sharpe (1964), Mossin (1966) and Lintner (1965). Another benefit is that any mean-variance efficient portfolio can be computed rapidly even for a large asset universe by solving a tractable quadratic program.

Unfortunately, however, the Markowitz approach is static. It plans only for the next rebalancing period and ignores that the end-of-period wealth will be reinvested. This is troubling because of Roll's insight that a number of mean-variance efficient portfolios lead to almost sure ruin if the available capital is infinitely often reinvested and returns are serially independent (Roll 1973, p. 551). The Markowitz approach also burdens investors with specifying their utility functions, which are needed to find the portfolios on the efficient frontier that are best aligned with their individual risk preferences. In this context Roy (1952) aptly noted that *'a man who seeks advice about his actions will not be grateful for the suggestion that he maximize expected utility.'*

Some of the shortcomings of the Markowitz approach are alleviated by the Kelly strategy, which maximizes the expected portfolio growth rate, that is, the logarithm of the total portfolio returns' geometric mean over a sequence of consecutive rebalancing intervals. If the asset returns are serially independent and identically distributed, the strong law of large numbers implies that the portfolio growth rate over an infinite investment horizon equals the expected logarithm (i.e., the expected log-utility) of the total portfolio return over any single rebalancing period; see e.g. Cover and Thomas (1991) or Luenberger (1998) for a textbook treatment of the Kelly strategy. Kelly (1956) invented his strategy to determine the optimal wagers in repeated betting games. The strategy was then extended to the realm of portfolio management by Latané (1959). Adopting standard terminology, we refer to the portfolio managed under the Kelly strategy as the *growth-optimal portfolio*. This portfolio displays several intriguing properties that continue to fascinate finance professionals and academics alike. First and foremost, in the long run the growth-optimal portfolio can be shown to accumulate more wealth than *any* other portfolio with probability 1. This powerful result was first proved by Kelly (1956) in a binomial setting and then generalized by Breiman (1961) to situations where returns are stationary and serially independent. Algoet and Cover (1988) later showed that Breiman's result remains valid even if the independence assumption is relaxed. The growth-optimal portfolio also minimizes the expected time to reach a preassigned monetary target V asymptotically as V tends to infinity, see Breiman (1961) and Algoet and Cover (1988), and it maximizes the median of the investor's fortune, see Ethier (2004). Hakansson and Miller (1975) further established that a Kelly investor never risks ruin. Maybe surprisingly, Dempster et al. (2008) could construct examples where the growth-optimal portfolio creates value even though every tradable asset becomes almost surely worthless in the long run. Hakansson (1971b) pointed out that the growth-optimal portfolio is *myopic*, meaning that the current portfolio composition only depends on the distribution of returns over the next rebalancing period. This property has computational significance as it enables investors to compute the Kelly strategy, which is optimal across a multi-period investment horizon, by solving a single-period convex optimization problem. A comprehensive list of properties of the growth-optimal portfolio has recently been compiled by

MacLean et al. (2010). Moreover, Poundstone (2005) narrated the colorful history of the Kelly strategy in gambling and speculation, while Christensen (2012) provided a detailed review of the academic literature. Remarkably, some of the most successful investors like Warren Buffet, Bill Gross and John Maynard Keynes are reported to have used Kelly-type strategies to manage their funds; see e.g. Ziemba (2005).

The almost sure asymptotic optimality of the Kelly strategy has prompted a heated debate about its role as a normative investment rule. Latané (1959), Hakansson (1971a) and Thorp (1971) attributed the Kelly strategy an objective superiority over other strategies and argued that *every* investor with a sufficiently long planning horizon should hold the growth-optimal portfolio. Samuelson (1963, 1971) and Merton and Samuelson (1974) contested this view on the grounds that the growth-optimal portfolio can be strictly dominated under non-logarithmic preferences, irrespective of the length of the planning horizon. Nowadays there seems to be a consensus that whether or not the growth-optimal portfolio can claim a special status depends largely on one's definition of rationality. In this context Luenberger (1993) has shown that Kelly-type strategies enjoy a universal optimality property under a natural preference relation for deterministic wealth sequences.

Even though the growth-optimal portfolio is guaranteed to dominate any other portfolio with probability 1 *in the long run*, it tends to be very risky in the short term. Judicious investors might therefore ask how long it will take until the growth-optimal portfolio outperforms a given benchmark with high confidence. Unfortunately, there is evidence that the long run may be long indeed. Rubinstein (1991) demonstrates, for instance, that in a Black Scholes economy it may take 208 years to be 95% sure that the Kelly strategy beats an all-cash strategy and even 4,700 years to be 95% sure that it beats an all-stock strategy. Investors with a *finite* lifetime may thus be better off pursuing a strategy that is tailored to their individual planning horizon.

The Kelly strategy also suffers from another shortcoming that is maybe less well recognized: the computation of the optimal portfolio weights requires perfect knowledge of the joint asset return distribution. In the academic literature, this distribution is often assumed to be known. In practice, however, it is already difficult to estimate the mean returns to within workable precision, let alone the complete distribution function; see e.g. § 8.5 of Luenberger (1998). As estimation errors are unavoidable, the asset return distribution is ambiguous. Real investors have only limited prior information on this distribution, e.g. in the form of confidence intervals for its first and second-order moments. As the Kelly strategy is tailored to a single distribution, it is ignorant of ambiguity. Michaud (1989), Best and Grauer (1991) and Chopra and Ziemba (1993) have shown that portfolios optimized in view of a single nominal distribution often perform poorly in out-of-sample experiments, that is, when the data-generating distribution differs from the one used in the optimization. Therefore, ambiguity-averse investors may be better off pursuing a strategy that is optimized against *all* distributions consistent with the given prior information. We emphasize that ambiguity-aversion enjoys strong justification from decision theory, see Gilboa and Schmeidler (1989).

The family of all return distributions consistent with the available prior information is referred to as the *ambiguity set*. In this paper we will assume that the asset returns follow a *weak sense white noise process*, which means that the ambiguity set contains all distributions under which the asset returns are serially uncorrelated and have period-wise identical first and second-order moments. No other distributional information is assumed to be available. To enhance realism, we will later generalize this basic ambiguity set to allow for moment ambiguity.

The goal of this paper is to design *robust growth-optimal portfolios* that offer similar guarantees as the classical growth-optimal portfolio—but for a *finite* investment horizon and for *all* return distributions in the ambiguity set. The classical growth-optimal portfolio maximizes the return level one can guarantee to achieve *with probability 1* over an *infinite* investment horizon and for a *single known* return distribution. As it is impossible to establish almost sure guarantees for finite time periods, we strive to construct a portfolio that maximizes the return level one can guarantee to achieve *with probability* $1-\epsilon$ over a given *finite* investment horizon and for *every* return distribution in the ambiguity set. The tolerance $\epsilon \in (0,1)$ is chosen by the investor and reflects the acceptable violation probability of the guarantee. While the guaranteed return level for short periods of time and small violation probabilities ϵ is likely to be negative, we hope that attractive return guarantees will emerge for longer investment horizons even if ϵ remains small.

The overwhelming popularity of the classical Markowitz approach is owed, at least partly, to its favorable computational properties. A similar statement holds true for the classical growth-optimal portfolio, which can be computed with relative ease due to its myopic nature, see, e.g., Estrada (2010) and § 2.1 of Christensen (2012). As computational tractability is critical for the practical usefulness of an investment rule, we will not attempt to optimize over all causal portfolio strategies in this paper. Indeed, this would be a hopeless undertaking as general causal policies cannot even be represented in a computer. Instead, we will restrict attention to memoryless *fixed-mix strategies* that keep the portfolio composition constant across all rebalancing dates and observation histories. This choice is motivated by the observation that fixed-mix strategies are optimal for infinite investment horizons. Thus, we expect that the best fixed-mix strategy will achieve a similar performance as the best causal strategy even for finite (but sufficiently long) investment horizons.

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

- We introduce robust growth-optimal portfolios that offer similar performance guarantees as the classical growth-optimal portfolio but for finite investment horizons and ambiguous return distributions. Robust growth-optimal portfolios maximize a quadratic approximation of the growth rate one can guarantee to achieve with probability $1-\epsilon$ by using fixed-mix strategies.

This guarantee holds for a *finite* investment horizon and for *all* asset return distributions in the ambiguity set. Equivalently, the robust growth-optimal portfolios maximize the worst-case value-at-risk at level ϵ of a quadratic approximation of the portfolio growth rate over the given investment horizon, where the worst case is taken across all distributions in the ambiguity set.

- Using recent results from distributionally robust chance constrained programming by Zymler et al. (2013a), we show that the worst-case value-at-risk of the quadratic approximation of the portfolio growth rate can be expressed as the optimal value of a tractable semidefinite program (SDP) whose size scales with the number of assets and the length of the investment horizon. We then exploit temporal symmetries to solve this SDP analytically. This allows us to show that any robust growth-optimal portfolio can be computed efficiently as the solution of a tractable second-order cone program (SOCP) whose size scales with the number of assets but is *independent* of the length of the investment horizon.
- We show that the robust growth-optimal portfolios are near-optimal for isoelastic utility functions with relative risk aversion parameters $\varrho \gtrsim 1$. Thus, they can be viewed as *fractional Kelly strategies*, which have been suggested as heuristic remedies for over-betting in the presence of model risk, see, e.g., Christensen (2012). Our analysis provides a theoretical justification for using fractional Kelly strategies and offers a systematic method to select the fractional Kelly strategy that is most appropriate for a given investment horizon and violation probability ϵ.
- In simulated and empirical backtests we show that the robust growth-optimal portfolios are competitive with the classical growth-optimal portfolio across most realistic investment horizons and for most return distributions in the ambiguity set.

Robust growth-optimal portfolio theory is conceptually related to the *safety first principle* introduced by Roy (1952), which postulates that investors aim to minimize the ruin probability, that is, the probability that their portfolio return falls below a prescribed safety level. Roy studied portfolio choice problems in a single-period setting and assumed—as we do—that only the first and second-order moments of the asset return distribution are known. By using a Chebyshev inequality, he obtained an analytical expression for the worst-case ruin probability, which closely resembles the portfolio's Sharpe ratio. This work was influential for many later developments in behavioral finance and risk management. Our work can be seen as an extension of Roy's model to a multi-period setting, which is facilitated by recent results on distributionally robust chance constrained programming by Zymler et al. (2013a). For a general introduction to distributionally robust optimization we refer to Delage and Ye (2010), Goh and Sim (2010) or Wiesemann et al. (2014). Portfolio selection models based on the worst-case value-at-risk with moment-based ambiguity sets have previously been studied by El Ghaoui et al. (2003), Natarajan et al. (2008), Natarajan et al. (2010) and Zymler et al. (2013b). Moreover, Doan et al. (2013) investigate distributionally robust portfolio optimization models using an ambiguity set in which some marginal distributions are known, while the global dependency structure is uncertain, and Meskarian and Xu (2013) study a distributionally robust formulation of a reward-risk ratio optimization problem. However, none of these papers explicitly accounts for the dynamic effects of portfolio selection.

The universal portfolio algorithm by Cover (1991) offers an alternative way to generate a dynamic portfolio strategy without knowledge of the data-generating distribution. In its basic form, the algorithm distributes the available capital across all fixed-mix strategies. Initially each fixed-mix strategy is given the same weight, but the weights are gradually adjusted according to the empirical performance of the different strategies. The resulting universal portfolio strategy can be shown to perform at least as well as the best fixed-mix strategy selected in hindsight. As for the classical growth-optimal portfolio, however, any performance guarantees are asymptotic, and in the short run universal portfolios are susceptible to error maximization phenomena. A comprehensive survey of more sophisticated universal portfolio algorithms is provided by Györfi et al. (2012).

The rest of the paper develops as follows. In Section 2 we review the asymptotic properties of classical growth-optimal portfolios, and in Section 3 we introduce the *robust* growth-optimal portfolios and discuss their performance guarantees. An analytical formula for the worst-case value-at-risk of the portfolio growth rate is derived in Section 4, and extensions of the underlying probabilistic model are presented in Section 5. Numerical results are reported in Section 6, and Section 7 concludes.

Notation. The space of symmetric (symmetric positive semidefinite) matrices in $\mathbb{R}^{n\times n}$ is denoted by $\mathbb{S}^n$ ($\mathbb{S}^n_+$). For any $\mathbf{X},\mathbf{Y}\in\mathbb{S}^n$ we let $\langle \mathbf{X},\mathbf{Y}\rangle = \mathrm{Tr}(\mathbf{XY})$ be the trace scalar product, while the relation $\mathbf{X}\succeq\mathbf{Y}$ ($\mathbf{X}\succ\mathbf{Y}$) implies that $\mathbf{X}-\mathbf{Y}$ is positive semidefinite (positive definite). The set of eigenvalues of $\mathbf{X}\in\mathbb{S}^n$ is denoted by $\mathrm{eig}\,(\mathbf{X})$. We also define $\mathbf{1}$ as the vector of ones and $\mathbb{I}$ as the identity matrix. Their dimensions will usually be clear from the context. Random variables are represented by symbols with tildes, while their realizations are denoted by the same symbols without tildes. The set of all probability distributions on $\mathbb{R}^n$ is denoted by $\mathcal{P}^n_0$. Moreover, we define $\log(x)$ as the natural logarithm of x if $x>0$; $=-\infty$ otherwise. Finally, we define the Kronecker delta through $\delta_{ij}=1$ if $i=j$; $=0$ otherwise.

2. Growth-Optimal Portfolios

Assume that there is a fixed pool of n assets available for investment and that the portfolio composition may only be adjusted at prescribed rebalancing dates indexed by $t=1,\ldots,T$, where T represents the length of the investment horizon. By convention, period t is the interval between the rebalancing dates t and $t+1$, while the relative price change of asset i over period t, that is, the asset's rate of return, is denoted by $\tilde{r}_{t,i}\geq -1$. Based on the common belief that markets are information efficient, it is often argued that the asset returns $\tilde{\boldsymbol{r}}_t = (\tilde{r}_{t,1}, \ldots, \tilde{r}_{t,n})^\intercal$ for $t = 1, \ldots, T$ are governed by a white noise process in the sense of the following definition.

Definition 1 (Strong Sense White Noise). The random vectors $(\tilde{\boldsymbol{r}}_t)_{t=1}^T$ form a strong sense white noise process if they are mutually independent and identically distributed.

A portfolio strategy $(\boldsymbol{w}_t)_{t=1}^T$ is a rule for distributing the available capital across the given pool of assets at all rebalancing dates within the investment horizon. Formally, $w_{t,i}$ denotes the proportion of capital allocated to asset i at time t, while $\boldsymbol{w}_t = (w_{t,1}, \ldots, w_{t,n})^\intercal$ encodes the portfolio held at time t. As all available capital must be invested, we impose the budget constraint $\mathbf{1}^\intercal \boldsymbol{w}_t = 1$. Moreover, we require $\boldsymbol{w}_t \geq \mathbf{0}$ to preclude short sales. For notational simplicity, the budget and short sales constraints as well as any other regulatory or institutional portfolio constraints are captured by the abstract requirement $\boldsymbol{w}_t \in \mathcal{W}$, where $\mathcal{W}$ represents a convex polyhedral subset of the probability simplex in $\mathbb{R}^n$. We emphasize that the portfolio composition is allowed to change over time and may also depend on the asset returns observed in the past, but not on those to be revealed in the future. In general, the portfolio at time t thus constitutes a causal function $\boldsymbol{w}_t = \boldsymbol{w}_t(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{t-1})$ of the asset returns observed up to time t. Due to their simplicity and attractive theoretical properties, fixed-mix strategies represent an important and popular subclass of all causal portfolio strategies.

Definition 2 (Fixed-Mix Strategy). A portfolio strategy $(\boldsymbol{w}_t)_{t=1}^T$ is a fixed-mix strategy if there is a $\boldsymbol{w} \in \mathcal{W}$ with $\boldsymbol{w}_t(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{t-1}) = \boldsymbol{w}$ for all $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{t-1}) \in \mathbb{R}^{n \times (t-1)}$ and $t = 1, \ldots, T$.

Fixed-mix strategies are also known as constant proportions strategies. They are memoryless and keep the portfolio composition fixed across all rebalancing dates and observation histories. We emphasize, however, that fixed-mix strategies are nonetheless dynamic as they necessitate periodic trades at the rebalancing dates. Indeed, the proportions of capital invested in the different assets change randomly over any rebalancing period. Assets experiencing above average returns will have larger weights at the end of the period and must undergo a divestment to revert back to the weights prescribed by the fixed-mix strategy, while assets with a below average return require a recapitalization. This trading pattern is often condensed into the maxim '*buy low, sell high.*' By slight abuse of notation, we will henceforth use the same symbol $\boldsymbol{w} \in \mathcal{W}$ to denote individual portfolios as well as the fixed-mix strategies that they induce.

The end-of-horizon value of a portfolio with initial capital 1 that is managed under a generic causal strategy $(\boldsymbol{w}_t)_{t=1}^T$ can be expressed as

$$\tilde{V}_T = \prod_{t=1}^{T} \left[1 + \boldsymbol{w}_t(\tilde{\boldsymbol{r}}_1, \ldots, \tilde{\boldsymbol{r}}_{t-1})^\intercal \tilde{\boldsymbol{r}}_t\right],$$

where the factors in square brackets represent the total portfolio returns over the rebalancing periods. The portfolio growth rate is then defined as the natural logarithm of the geometric mean of the absolute returns, which is equivalent to the arithmetic mean of the log-returns.

$$\tilde{\gamma}_T = \log \sqrt[T]{\prod_{t=1}^{T} [1 + \boldsymbol{w}_t(\tilde{\boldsymbol{r}}_1, \ldots, \tilde{\boldsymbol{r}}_{t-1})^\intercal \tilde{\boldsymbol{r}}_t]} = \frac{1}{T} \sum_{t=1}^{T} \log [1 + \boldsymbol{w}_t(\tilde{\boldsymbol{r}}_1, \ldots, \tilde{\boldsymbol{r}}_{t-1})^\intercal \tilde{\boldsymbol{r}}_t] \tag{1}$$

The reverse formula $\tilde{V}_T = e^{\tilde{\gamma}_T T}$ highlights that there is a strictly monotonic relation between the terminal value and the growth rate of the portfolio. Thus, our informal management objective of maximizing terminal wealth is equivalent to maximizing the growth rate. Unfortunately, this maximization is generally ill-defined as $\tilde{\gamma}_T$ is uncertain. However, when the portfolio is managed under a fixed-mix strategy $\boldsymbol{w} \in \mathcal{W}$ and the asset returns $\tilde{\boldsymbol{r}}_t$, $t = 1, \ldots, T$, follow a strong sense white noise process, then $\tilde{\gamma}_T$ is asymptotically deterministic for large T.

Proposition 1 **(Asymptotic Growth Rate)**. *If $\boldsymbol{w} \in \mathcal{W}$ is a fixed-mix strategy, while the asset returns $(\tilde{\boldsymbol{r}}_t)_{t=1}^T$ follow a strong sense white noise process, then*

$$\lim_{T \to \infty} \tilde{\gamma}_T = \mathbb{E}\left(\log\left(1 + \boldsymbol{w}^\intercal \tilde{\boldsymbol{r}}_1\right)\right) \quad \textit{with probability } 1. \tag{2}$$

Proof. The claim follows immediately from (1) and the strong law of large numbers. ∎

Proposition 1 asserts that the asymptotic growth rate of a fixed-mix strategy $\boldsymbol{w} \in \mathcal{W}$ coincides almost surely with the expected log-return of portfolio $\boldsymbol{w}$ over a single (without loss of generality, the first) rebalancing period. A particular fixed-mix strategy of great conceptual and intuitive appeal is the *Kelly strategy*, which is induced by the *growth-optimal portfolio* $\boldsymbol{w}^*$ that maximizes the right hand side of (2). We henceforth assume that there are no redundant assets, that is, the second-order moment matrix of $\tilde{\boldsymbol{r}}_t$ is strictly positive definite for all t.

Definition 3 (Kelly Strategy). The Kelly strategy is the fixed-mix strategy induced by the unique growth-optimal portfolio $\boldsymbol{w}^* = \arg\max_{\boldsymbol{w} \in \mathcal{W}} \mathbb{E}\left(\log\left(1 + \boldsymbol{w}^\intercal \tilde{\boldsymbol{r}}_1\right)\right)$.

By construction, the Kelly strategy achieves the highest asymptotic growth rate among all fixed-mix strategies. Maybe surprisingly, it also outperforms all other causal portfolio strategies in a sense made precise in the following theorem.

Theorem 1 **(Asymptotic Optimality of the Kelly Strategy)**. *Let $\tilde{\gamma}_T^*$ and $\tilde{\gamma}_T$ represent the growth rates of the Kelly strategy and any other causal portfolio strategy, respectively. If $(\tilde{\boldsymbol{r}}_t)_{t=1}^T$ is a strong sense white noise process, then $\limsup_{T \to \infty} \tilde{\gamma}_T - \tilde{\gamma}_T^* \leq 0$ with probability 1.*

Proof. See e.g. Theorem 15.3.1 of Cover and Thomas (1991). ∎

Even though the Kelly strategy has several other intriguing properties, which are discussed at length by MacLean et al. (2010), Theorem 1 lies at the root of its popularity. The theorem asserts that, in the long run, the Kelly strategy accumulates more wealth than *any other causal strategy* to first order in the exponent, that is, $e^{\tilde{\gamma}_T T} \leq e^{\tilde{\gamma}^*_T T + o(T)}$, with probability 1. However, the Kelly strategy has also a number of shortcomings that limit its practical usefulness. First, Rubinstein (1991) shows that it may take hundreds of years until the Kelly strategy starts to dominate other investment strategies with high confidence. Moreover, the computation of the growth-optimal portfolio $\boldsymbol{w}^*$ requires precise knowledge of the asset return distribution $\mathbb{P}$, which is never available in reality due to estimation errors (Luenberger 1998, § 8.5). This is problematic because Michaud (1989) showed that the growth-optimal portfolio corresponding to an inaccurate estimated distribution $\hat{\mathbb{P}}$ may perform poorly under the true data-generating distribution $\mathbb{P}$. Finally, even if $\mathbb{P}$ was known, Theorem 1 would require the asset returns to follow a strong sense white noise process under $\mathbb{P}$. This is an unrealistic requirement as there is ample empirical evidence that stock returns are serially dependent; see e.g. Jegadeesh and Titman (1993). Even though the definition of the Kelly strategy as well as Theorem 1 have been generalized by Algoet and Cover (1988) to situations where the asset returns are serially dependent, the Kelly strategy ceases to belong to the class of fixed-mix strategies in this setting and may thus no longer be easy to compute.

3. Robust Growth-Optimal Portfolios

In this section we extend the growth guarantees of Theorem 1 to *finite* investment horizons and *ambiguous* asset return distributions. In order to maintain tractability, we restrict attention to the class of fixed-mix strategies. As this class contains the Kelly strategy, which is optimal for an infinite investment horizon, we conjecture that it also contains policies that are near-optimal for finite horizons. We first observe that the portfolio growth rate $\tilde{\gamma}_T(\boldsymbol{w}) = \frac{1}{T}\sum_{t=1}^T \log(1 + \boldsymbol{w}^\intercal \tilde{\boldsymbol{r}}_t)$ of any given fixed-mix strategy $\boldsymbol{w}$ constitutes a (non-degenerate) random variable whenever the investment horizon T is finite. As $\tilde{\gamma}_T(\boldsymbol{w})$ may have a broad spectrum of very different possible outcomes, it cannot be maximized *per se*. However, one can maximize its value-at-risk (VaR) at level $\epsilon \in (0, 1)$, which is defined in terms of the chance-constrained program

$$\mathbb{P}\text{-VaR}_\epsilon(\tilde{\gamma}_T(\boldsymbol{w})) \;=\; \max_{\gamma\in\mathbb{R}} \left\{ \gamma \;:\; \mathbb{P}\left(\frac{1}{T}\sum_{t=1}^T \log(1 + \boldsymbol{w}^\intercal \tilde{\boldsymbol{r}}_t) \geq \gamma \right) \geq 1 - \epsilon \right\}.$$

The violation probability ϵ of the chance constraint reflects the investor's risk aversion and is typically chosen as a small number $\lesssim 10\%$. If γ^* denotes the optimal solution to the above chance-constrained program, then, with probability $1 - \epsilon$, the value of a portfolio managed under the fixed-mix strategy $\boldsymbol{w}$ will grow at least by a factor $e^{T\gamma^*}$ over the next T rebalancing periods. Of course, the VaR of the portfolio growth rate $\tilde{\gamma}_T(\boldsymbol{w})$ can only be computed if the distribution $\mathbb{P}$ of the asset returns is precisely known. In practice, however, $\mathbb{P}$ may only be known to belong to an ambiguity set $\mathcal{P}$, which contains all asset return distributions that are consistent with the investor's prior information. In this situation, an ambiguity-averse investor will seek protection against all distributions in $\mathcal{P}$. This is achieved by using the worst-case VaR (WVaR) of $\tilde{\gamma}_T(\boldsymbol{w})$ to assess the performance of the fixed-mix strategy $\boldsymbol{w}$.

$$\begin{aligned} \mathrm{WVaR}_\epsilon(\tilde{\gamma}_T(\boldsymbol{w})) &= \min_{\mathbb{P}\in\mathcal{P}} \ \mathbb{P}\text{-}\mathrm{VaR}_\epsilon(\tilde{\gamma}_T(\boldsymbol{w})) \\ &= \max_{\gamma\in\mathbb{R}} \left\{ \gamma \ : \ \mathbb{P}\left(\frac{1}{T}\sum_{t=1}^{T} \log\left(1+\boldsymbol{w}^\intercal \tilde{\boldsymbol{r}}_t\right) \geq \gamma \right) \geq 1-\epsilon \quad \forall \mathbb{P}\in\mathcal{P} \right\} \end{aligned} \tag{3}$$

In the remainder of this paper, we refer to the portfolios that maximize WVaR as *robust growth-optimal portfolios.* They offer the following performance guarantees.

Observation 1 (Performance Guarantees). Let $\boldsymbol{w}^*$ be the robust growth-optimal portfolio that maximizes $\mathrm{WVaR}_\epsilon(\tilde{\gamma}_T(\boldsymbol{w}))$ over $\mathcal{W}$ and denote by γ^* its objective value. Then, with probability $1-\epsilon$, the value of a portfolio managed under the fixed-mix strategy $\boldsymbol{w}^*$ will grow at least by $e^{T\gamma^*}$ over T periods. This guarantee holds for all distributions in the ambiguity set $\mathcal{P}$.

Proof. This is an immediate consequence of the definition of WVaR. ∎

We emphasize that the portfolio return in any given rebalancing period displays significant variability. Thus, the guaranteed return level γ^* corresponding to a short investment horizon is typically negative. However, positive growth rates can be guaranteed over longer investment horizons even for $\epsilon \leq 5\%$.

In the following we will assume that the asset returns are only known to follow a weak sense white noise process.

Definition 4 (Weak Sense White Noise). The random vectors $(\tilde{\boldsymbol{r}}_t)_{t=1}^T$ form a weak sense white noise process if they are mutually uncorrelated and share the same mean values $\mathbb{E}_\mathbb{P}\left(\tilde{\boldsymbol{r}}_t\right) = \boldsymbol{\mu}$ and second-order moments $\mathbb{E}_\mathbb{P}\left(\tilde{\boldsymbol{r}}_t\tilde{\boldsymbol{r}}_t^\intercal\right) = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal$ for all $1 \leq t \leq T$.

Note that every strong sense white noise process in the sense of Definition 1 is also a weak sense white noise process, while the converse implication is generally false. By modeling the asset returns as a weak sense white noise process we concede that they could be serially dependent (as long as they remain serially uncorrelated). Moreover, we deny to have any information about the return distribution except for its first and second-order moments. In particular, we also accept the possibility that the marginal return distributions corresponding to two different rebalancing periods may differ (as long as they have the *same* means and covariance matrices). In his celebrated article on the safety first principle for single-period portfolio selection, Roy (1952) provides some implicit justification for the weak sense white noise assumption. Indeed, he postulates that the first and second-order moments of the asset return distribution *'are the only quantities that can be* *distilled out of our knowledge of the past.'* Moreover, he asserts that *'the slightest acquaintance with problems of analysing economic time series will suggest that this assumption is optimistic rather than unnecessarily restrictive.'* It is thus natural to define

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0^{nT} \ : \begin{array}{l} \mathbb{E}_{\mathbb{P}}\left(\tilde{\boldsymbol{r}}_t\right) = \boldsymbol{\mu} \quad \forall t \ : \ 1 \leq t \leq T \\ \mathbb{E}_{\mathbb{P}}\left(\tilde{\boldsymbol{r}}_s \tilde{\boldsymbol{r}}_t^\intercal\right) = \delta_{st}\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal \quad \forall s,t \ : \ 1 \leq s \leq t \leq T \end{array} \right\},$$

where the mean value $\boldsymbol{\mu} \in \mathbb{R}^n$ and the covariance matrix $\boldsymbol{\Sigma} \in \mathbb{S}^n_+$ are given parameters. Note that the asset returns follow a weak sense white noise process under *any* distribution from within $\mathcal{P}$. We remark that, besides its conceptual appeal, the moment-based ambiguity set $\mathcal{P}$ has distinct computational benefits that will become apparent in Section 4. More general ambiguity sets where the moments $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are also subject to uncertainty or where the asset return distribution is supported on a prescribed subset of $\mathbb{R}^{nT}$ will be studied in Section 5.

In a single-period setting, worst-case VaR optimization problems with moment-based ambiguity sets have previously been studied by El Ghaoui et al. (2003), Natarajan et al. (2008, 2010) and Zymler et al. (2013b).

Remark 1 (Support Constraints). The ambiguity set $\mathcal{P}$ could safely be reduced by including the support constraints $\mathbb{P}\left(\tilde{\boldsymbol{r}}_t \geq -\boldsymbol{1}\right) = 1 \ \forall t : 1 \leq t \leq T$, which ensure that the stock prices remain nonnegative. Certainly, these constraints are satisfied by the unknown true asset return distribution, and ignoring them renders the worst-case VaR in (3) more conservative. In order to obtain a clean model, we first suppress these constraints but emphasize that problem (3) remains well-defined even without them. Recall that, by convention, the logarithm is defined as an extended real-valued function on all of $\mathbb{R}$. Support constraints will be studied in Section 5.1.

4. Worst-Case Value-at-Risk of the Growth Rate

Weak sense white noise ambiguity sets of the form (3) are not only physically meaningful but also computationally attractive. We will now demonstrate that useful approximations of the corresponding robust growth-optimal portfolios can be computed in polynomial time. More precisely, we will show that the worst-case VaR of a quadratic approximation of the portfolio growth rate admits an explicit analytical formula. In the remainder we will thus assume that the growth rate $\tilde{\gamma}_T(\boldsymbol{w})$ of the fixed-mix strategy $\boldsymbol{w}$ can be approximated by

$$\tilde{\gamma}_T'(\boldsymbol{w}) = \frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{w}^\intercal\tilde{\boldsymbol{r}}_t - \frac{1}{2}\left(\boldsymbol{w}^\intercal\tilde{\boldsymbol{r}}_t\right)^2\right),$$

which is obtained from (1) by expanding the logarithm to second order in $\boldsymbol{w}^\intercal\tilde{\boldsymbol{r}}_t$. This Taylor approximation has found wide application in portfolio analysis (Samuelson (1970)) and is accurate for short rebalancing periods, in which case the probability mass of $\boldsymbol{w}^\intercal\tilde{\boldsymbol{r}}_t$ accumulates around 0.

Additional theoretical justification in the context of growth-optimal portfolio selection is provided by Kuhn and Luenberger (2010). To assess the approximation quality one can expect in practice, we have computed the relative difference between $\tilde{\gamma}_T(\boldsymbol{w})$ and $\tilde{\gamma}'_T(\boldsymbol{w})$ for each individual asset and for 100,000 randomly generated fixed-mix strategies based on the *10 Industry Portfolios* and the *12 Industry Portfolios* from the Fama French online data library.[1] For a ten year investment horizon the approximation error was uniformly bounded by 1% under monthly and by 5% under yearly rebalancing, respectively, and in most cases the errors were much smaller than these upper bounds.

From now on we will also impose two non-restrictive assumptions on the moments of the asset returns.

(A1) The covariance matrix $\boldsymbol{\Sigma}$ is strictly positive definite.

(A2) For all $\boldsymbol{w} \in \mathcal{W}$, we have $1 - \boldsymbol{w}^\intercal \boldsymbol{\mu} > \sqrt{\frac{\epsilon}{(1-\epsilon)T}} \left\| \boldsymbol{\Sigma}^{1/2} \boldsymbol{w} \right\|$.

Assumption (A1) ensures that the robust growth-optimal portfolio for a particular T and ϵ is unique, and Assumption (A2) delineates the set of moments for which the quadratic approximation of the portfolio growth-rate is sensible. As the *exact* growth rate $\tilde{\gamma}_T(\boldsymbol{w})$ is increasing and concave in $\boldsymbol{w}^\intercal \tilde{\boldsymbol{r}}_t$, its worst-case VaR must be increasing in $\boldsymbol{w}^\intercal \boldsymbol{\mu}$ and decreasing in $\boldsymbol{w}^\intercal \boldsymbol{\Sigma} \boldsymbol{w}$. Assumption (A2) ensures that the worst-case VaR of the *approximate* growth rate $\tilde{\gamma}'_T(\boldsymbol{w})$ inherits these monotonicity properties and is also increasing in $\boldsymbol{w}^\intercal \boldsymbol{\mu}$ and decreasing in $\boldsymbol{w}^\intercal \boldsymbol{\Sigma} \boldsymbol{w}$. Note that the Assumptions (A1) and (A2) are readily satisfied in most situations of practical interest, even if $T = 1$. Assumption (A1) holds whenever there is no risk-free asset or portfolio, while (A2) is automatically satisfied when $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are small enough, which can always be enforced by shortening the rebalancing intervals. In fact, (A2) holds even for yearly rebalancing intervals if the means and standard deviations of the asset returns fall within their typical ranges reported in § 8 of Luenberger (1998).

In the rest of this section we compute the worst-case VaR of the approximate growth rate

$$\mathrm{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w})) \;=\; \max_{\gamma \in \mathbb{R}} \left\{ \gamma \;:\; \mathbb{P}\left(\tilde{\gamma}'_T(\boldsymbol{w}) \geq \gamma \right) \geq 1 - \epsilon \quad \forall \mathbb{P} \in \mathcal{P} \right\} \tag{4}$$

for some fixed $\boldsymbol{w} \in \mathcal{W}$, $T \in \mathbb{N}$ and $\epsilon \in (0,1)$. By exploiting a known tractable reformulation of distributionally robust quadratic chance constraints with mean and covariance information (see Theorem 6 in Appendix A), we can re-express problem (4), which involves *infinitely* many constraints parameterized by $\mathbb{P} \in \mathcal{P}$, as a *finite* semidefinite program (SDP). Thus, we obtain

$$\begin{aligned} \mathrm{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w})) \;=\; \max \quad & \gamma \\ \text{s.t.} \quad & \mathbf{M} \in \mathbb{S}^{nT+1},\; \beta \in \mathbb{R},\; \gamma \in \mathbb{R} \\ & \beta + \tfrac{1}{\epsilon} \langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq 0, \quad \mathbf{M} \succeq \mathbf{0} \\ & \mathbf{M} - \begin{bmatrix} \frac{1}{2} \sum_{t=1}^T \mathbf{P}_t^\intercal \boldsymbol{w} \boldsymbol{w}^\intercal \mathbf{P}_t & -\frac{1}{2} \sum_{t=1}^T \mathbf{P}_t^\intercal \boldsymbol{w} \\ -\frac{1}{2} \left(\sum_{t=1}^T \mathbf{P}_t^\intercal \boldsymbol{w} \right)^\intercal & \gamma T - \beta \end{bmatrix} \succeq \mathbf{0}, \end{aligned} \tag{5}$$

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

where

$$\boldsymbol{\Omega} = \left[\begin{array}{cccc|c} \boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^\intercal & \boldsymbol{\mu}\boldsymbol{\mu}^\intercal & \cdots & \boldsymbol{\mu}\boldsymbol{\mu}^\intercal & \boldsymbol{\mu} \\ \boldsymbol{\mu}\boldsymbol{\mu}^\intercal & \boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^\intercal & \cdots & \boldsymbol{\mu}\boldsymbol{\mu}^\intercal & \boldsymbol{\mu} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{\mu}\boldsymbol{\mu}^\intercal & \boldsymbol{\mu}\boldsymbol{\mu}^\intercal & \cdots & \boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^\intercal & \boldsymbol{\mu} \\ \hline \boldsymbol{\mu}^\intercal & \boldsymbol{\mu}^\intercal & \cdots & \boldsymbol{\mu}^\intercal & 1 \end{array}\right] \in \mathbb{S}^{nT+1}$$

denotes the matrix of first and second-order moments of $(\tilde{\boldsymbol{r}}_1^\intercal, \ldots, \tilde{\boldsymbol{r}}_T^\intercal)^\intercal$, while the truncation operators $\mathbf{P}_t \in \mathbb{R}^{n\times nT}$ are defined via $\mathbf{P}_t(\boldsymbol{r}_1^\intercal, \ldots, \boldsymbol{r}_T^\intercal)^\intercal = \boldsymbol{r}_t$, $t = 1, \ldots, T$. As (5) constitutes a tractable SDP, the worst-case VaR of any fixed-mix strategy's approximate growth rate can be evaluated in time polynomial in the number of assets n and the investment horizon T, see e.g. Ye (1997).

Remark 2 (Maximizing the Worst-Case VaR). In practice, we are not only interested in evaluating the worst-case VaR of a fixed portfolio, but we also aim to identify portfolios that offer attractive growth guarantees. Such portfolios can be found by treating $\boldsymbol{w} \in \mathcal{W}$ as a decision variable in (5). In this case, the last matrix inequality in (5) becomes quadratic in the decision variables, and (5) ceases to be an SDP. Fortunately, however, one can convert (5) back to an SDP by rewriting the quadratic matrix inequality as

$$\begin{aligned} 2\mathbf{M} - \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 2\gamma T - T - 2\beta \end{bmatrix} &\succeq \sum_{t=1}^{T} \begin{bmatrix} \mathbf{P}_t^\intercal \boldsymbol{w} \\ -1 \end{bmatrix} \begin{bmatrix} \mathbf{P}_t^\intercal \boldsymbol{w} \\ -1 \end{bmatrix}^\intercal \\ &= \begin{bmatrix} \mathbf{P}_1^\intercal \boldsymbol{w} & \mathbf{P}_2^\intercal \boldsymbol{w} & \cdots & \mathbf{P}_T^\intercal \boldsymbol{w} \\ -1 & -1 & \cdots & -1 \end{bmatrix} \begin{bmatrix} \mathbf{P}_1^\intercal \boldsymbol{w} & \mathbf{P}_2^\intercal \boldsymbol{w} & \cdots & \mathbf{P}_T^\intercal \boldsymbol{w} \\ -1 & -1 & \cdots & -1 \end{bmatrix}^\intercal, \end{aligned}$$

which is satisfied whenever there are $\mathbf{V} \in \mathbb{S}^{nT}$, $\boldsymbol{v} \in \mathbb{R}^{nT}$ and $v_0 \in \mathbb{R}$ with

$$\mathbf{M} = \begin{bmatrix} \mathbf{V} & \boldsymbol{v} \\ \boldsymbol{v}^\intercal & v_0 \end{bmatrix}, \quad \begin{bmatrix} 2\mathbf{V} & 2\boldsymbol{v} & \mathbf{P}_1^\intercal \boldsymbol{w} & \cdots & \mathbf{P}_T^\intercal \boldsymbol{w} \\ 2\boldsymbol{v}^\intercal & 2v_0 - 2\gamma T + T + 2\beta & -1 & \cdots & -1 \\ \boldsymbol{w}^\intercal \mathbf{P}_1 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{w}^\intercal \mathbf{P}_T & -1 & 0 & \cdots & 1 \end{bmatrix} \succeq \mathbf{0}$$

by virtue of a Schur complement argument.

Even though SDPs are polynomial-time solvable in theory, problem (5) will quickly exhaust the capabilities of state-of-the-art SDP solvers when the asset universe and the investment horizon become large. Indeed, the dimension of the underlying matrix inequalities scales with n and T, and many investors will envisage a planning horizon of several decades with monthly or weekly granularity and an asset universe comprising several hundred titles. However, we will now demonstrate that the approximate worst-case VaR problem (4) admits in fact an *analytical* solution.

Theorem 2 **(Worst-Case Value-at-Risk)**. *Under Assumptions (A1) and (A2) we have*

$$\mathrm{WVaR}_\epsilon(\tilde{\gamma}_T'(\boldsymbol{w})) = \tfrac{1}{2}\left(1 - \left(1 - \boldsymbol{w}^\intercal\boldsymbol{\mu} + \sqrt{\tfrac{1-\epsilon}{\epsilon T}}\left\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{w}\right\|\right)^2 - \tfrac{T-1}{\epsilon T}\boldsymbol{w}^\intercal\boldsymbol{\Sigma}\boldsymbol{w}\right). \tag{6}$$

The proof of Theorem 2 relies on a dimensionality reduction in three steps. We first exploit a projection property for moment-based ambiguity sets to reformulate the approximate worst-case VaR problem (4) in terms the portfolio returns $\tilde{\eta}_t = \boldsymbol{w}^\intercal \tilde{\boldsymbol{r}}_t$ as the fundamental random variables. The resulting worst-case VaR problem is equivalent to an SDP whose size scales with T but not with n. Next, we exploit the temporal permutation symmetry of this SDP to demonstrate that the matrix variable appearing in the SDP constraints has only four degrees of freedom. This insight can be used to show that the SDP involves essentially only six different decision variables. However, the matrices in the matrix inequalities still have a dimension of the order of T. Finally, by explicitly diagonalizing these matrices one can show that the SDP constraints admit an equivalent reformulation in terms of only nine traditional scalar constraints. In summary, the worst-case VaR problem (4) can be shown to be equivalent to a convex nonlinear program with six variables and nine constraints. Theorem 2 then follows by analytically solving the Karush-Kuhn-Tucker optimality conditions of this nonlinear program. A detailed formal proof of Theorem 2 is relegated to Appendix B.

If the set $\mathcal{W}$ of admissible portfolios is characterized by a finite number of linear constraints, then the portfolio optimization problem $\max_{\boldsymbol{w}\in\mathcal{W}} \text{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w}))$ reduces to a tractable SOCP whose size is independent of the investment horizon. In order to avoid verbose terminology, we will henceforth refer to the unique optimizer of this SOCP as the *robust growth-optimal portfolio*, even though it maximizes only an approximation of the true growth rate. Maybe surprisingly, computing the robust growth-optimal portfolio is almost as easy as computing a Markowitz portfolio. However, the robust growth-optimal portfolios offer precise performance guarantees over finite investment horizons and for a wide spectrum of different asset return distributions.

Note that the worst-case VaR (6) is increasing in the portfolio mean return $\boldsymbol{w}^\intercal\boldsymbol{\mu}$ and decreasing in the portfolio variance $\boldsymbol{w}^\intercal\boldsymbol{\Sigma}\boldsymbol{w}$ as long as $\boldsymbol{w}$ satisfies Assumption (A2). Thus, any portfolio that maximizes the worst-case VaR is mean-variance efficient. This is not surprising as the worst-case VaR is calculated solely on the basis of mean and covariance information. Markowitz investors choose freely among all mean-variance efficient portfolios based on their risk preferences, that is, they solve the Markowitz problem $\max_{\boldsymbol{w}\in\mathcal{W}} \boldsymbol{w}^\intercal\boldsymbol{\mu} - \frac{\varrho}{2}\boldsymbol{w}^\intercal\boldsymbol{\Sigma}\boldsymbol{w}$ corresponding to their idiosyncratic risk aversion parameter $\varrho \geq 0$. In contrast, a robust growth-optimal investor chooses the unique mean-variance efficient portfolio tailored to her investment horizon T and violation probability ϵ. We can thus define a function $\varrho(T,\epsilon)$ with the property that the robust growth-optimal portfolio tailored to T and ϵ coincides with the solution of the Markowitz problem with risk aversion parameter $\varrho(T,\epsilon)$. By comparing the optimality conditions of the Markowitz and robust growth-optimal portfolio problems, one can show that

$$\varrho(T,\epsilon) = \sqrt{\frac{1-\epsilon}{\epsilon T}} \cdot \frac{1}{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{w}\|} + \frac{T-1}{\epsilon T\left(1 - \boldsymbol{w}^\intercal\boldsymbol{\mu} + \sqrt{\frac{1-\epsilon}{\epsilon T}}\,\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{w}\|\right)}\,, \tag{7}$$

where $\boldsymbol{w}$ denotes the robust growth-optimal portfolio, which depends on both T and ϵ and can only be computed numerically. The function $\varrho(T,\epsilon)$ will be investigated further in Section 6.1. From the point of view of mean-variance analysis, a robust growth-optimal investor becomes less risk-averse as ϵ or T increases. Indeed, one can use Assumption (A2) to prove that $\mathrm{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w}))$ is increasing (indicating that $\varrho(T,\epsilon)$ is decreasing) in ϵ and T. We emphasize that the robust growth-optimal portfolios may lose mean-variance efficiency when the ambiguity set of the asset return distribution is no longer described in terms of exact first and second-order moments. Such situations will be studied in Section 5.

The classical growth-optimal portfolio is perceived as highly risky. Indeed, if the rebalancing intervals are short enough to justify a quadratic expansion of the logarithmic utility function, then the classical growth-optimal portfolio can be identified with the Markowitz portfolio corresponding to the aggressive risk aversion parameter $\varrho = 1$. In fact, the two portfolios are identical in the continuous-time limit if the asset prices follow a multivariate geometric Brownian motion (Luenberger 1998, § 15.5). The Markowitz portfolios associated with more moderate levels of risk aversion $\varrho \gtrsim 1$ are often viewed as *ad hoc* alternatives to the classical growth-optimal portfolio that preserve some of its attractive growth properties but mitigate its short-term variability.

According to standard convention, a *fractional Kelly strategy* with risk-aversion parameter $\kappa \geq 1$ blends the classical Kelly strategy and a risk-free asset in constant proportions of $1/\kappa$ and $(\kappa-1)/\kappa$, respectively. Fractional Kelly strategies have been suggested as heuristic remedies for over-betting in the presence of model risk, see e.g. Christensen (2012). As pointed out by MacLean et al. (2005), the fractional Kelly strategy corresponding to κ emerges as a maximizer of the Merton problem with constant relative risk aversion κ if the prices of the risky assets follow a multivariate geometric Brownian motion in continuous time. In a discrete-time market *without* a risk-free asset it is therefore natural to define the fractional Kelly strategy corresponding to κ through the portfolio that maximizes the expected isoelastic utility function $\frac{1}{1-\kappa}\mathbb{E}[(1+\boldsymbol{w}^\intercal\tilde{\boldsymbol{r}}_1)^{1-\kappa}]$. By expanding the utility function around 1, this portfolio can be closely approximated by $\boldsymbol{w}_\kappa = \arg\max_{\boldsymbol{w}\in\mathcal{W}} \boldsymbol{w}^\intercal\boldsymbol{\mu} - \frac{\kappa}{2}\boldsymbol{w}^\intercal(\boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^\intercal)\boldsymbol{w}$, which is mean-variance efficient with risk-aversion parameter $\varrho = \kappa/(1+\kappa\boldsymbol{w}_\kappa^\intercal\boldsymbol{\mu})$ whenever $\boldsymbol{w}_\kappa^\intercal\boldsymbol{\mu} \leq 1/\kappa$. Note that the last condition is reminiscent of Assumption (A2) and is satisfied for typical choices of ϵ, T $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. It then follows from (7) that the robust growth-optimal portfolio tailored to the investment horizon T and violation probability ϵ coincides with the (approximate) fractional Kelly strategy corresponding to the risk-aversion parameter

$$\kappa(T,\epsilon) = \frac{\left(1-\boldsymbol{w}^\intercal\boldsymbol{\mu}+\sqrt{\frac{1-\epsilon}{\epsilon T}}\left\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{w}\right\|\right)\sqrt{(1-\epsilon)\epsilon T}+(T-1)\left\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{w}\right\|}{\left(1-\boldsymbol{w}^\intercal\boldsymbol{\mu}+\sqrt{\frac{1-\epsilon}{\epsilon T}}\left\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{w}\right\|\right)\left(\epsilon T\left\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{w}\right\|-\sqrt{(1-\epsilon)\epsilon T}\boldsymbol{w}^\intercal\boldsymbol{\mu}\right)-(T-1)\boldsymbol{w}^\intercal\boldsymbol{\mu}\left\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{w}\right\|},$$

where the robust growth-optimal portfolio $\boldsymbol{w}$ must be computed numerically. Our work thus offers evidence for the near-optimality of fractional Kelly strategies under distributional ambiguity and provides systematic guidelines for tailoring fractional Kelly strategies to specific investment horizons and violation probabilities. The function $\kappa(T,\epsilon)$ will be studied further in Section 6.1.

REMARK 3. **(Relation to Worst-Case VaR by El Ghaoui et al. (2003))** Theorem 2 generalizes a result by El Ghaoui et al. (2003) for worst-case VaR problems in a single-period investment setting. Indeed, for $T=1$ the portfolio optimization problem $\max_{\boldsymbol{w}\in\mathcal{W}} \mathrm{WVaR}_\epsilon(\tilde{\gamma}_1'(\boldsymbol{w}))$ reduces to

$$\max_{\boldsymbol{w}\in\mathcal{W}} \tfrac{1}{2}\left(1-\left(1-\boldsymbol{w}^\intercal\boldsymbol{\mu}+\sqrt{\tfrac{1-\epsilon}{\epsilon}}\left\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{w}\right\|\right)^2\right) \iff \max_{\boldsymbol{w}\in\mathcal{W}} \boldsymbol{w}^\intercal\boldsymbol{\mu}-\sqrt{\frac{1-\epsilon}{\epsilon}}\left\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{w}\right\|.$$

Under Assumption (A2), the objective functions of the above problems are related through a strictly monotonic transformation. Thus, both problems share the same optimal solution (but have different optimal values). The second problem is readily recognized as the SOCP equivalent to the static worst-case VaR optimization problem by El Ghaoui et al. (2003).

REMARK 4. **(Long-Term Investors)** In the limit of very long investment horizons, the worst-case VaR (6) reduces to

$$\lim_{T\to\infty} \mathrm{WVaR}_\epsilon(\tilde{\gamma}_T'(\boldsymbol{w})) = \tfrac{1}{2}-\tfrac{1}{2}\left(1-\boldsymbol{w}^\intercal\boldsymbol{\mu}\right)^2-\tfrac{1}{2\epsilon}\boldsymbol{w}^\intercal\boldsymbol{\Sigma}\boldsymbol{w},$$

which can be viewed as the difference between the second-order Taylor approximation of the portfolio growth rate in the nominal scenario, $\log(1+\boldsymbol{w}^\intercal\boldsymbol{\mu})$, and a risk premium, which is inversely proportional to the violation probability ϵ.

REMARK 5. **(Worst-Case Conditional VaR)** We could use the worst-case *conditional* VaR (CVaR) instead of the worst-case VaR in (4) to quantify the desirability of the fixed-mix strategy $\boldsymbol{w}$. The CVaR at level $\epsilon\in(0,1)$ of a random reward is defined as the conditional expectation of the $\epsilon\times 100\%$ least favorable reward realizations below the VaR. CVaR is sometimes considered to be superior to the VaR because it constitutes a coherent risk measure in the sense of Artzner et al. (1999). However, it has been shown in Theorem 2.2 of Zymler et al. (2013a) that the worst-case VaR and the worst-case CVaR under mean and covariance information are actually equal on the space of reward functions that are quadratic in the uncertain parameters. As the approximate portfolio growth rate $\tilde{\gamma}_T'(\boldsymbol{w})$ is quadratic in the uncertain asset returns, its worst-case VaR thus coincides with its worst-case CVaR.

In order to perform systematic contamination or stress test experiments, it is essential to know the extremal distributions from within $\mathcal{P}$ under which the actual VaR of the approximate portfolio growth rate $\tilde{\gamma}_T'(\boldsymbol{w})$ coincides with $\mathrm{WVaR}_\epsilon(\tilde{\gamma}_T'(\boldsymbol{w}))$. We will now demonstrate that the worst case is not attained by a single distribution. However, we can explicitly construct a sequence of asset return distributions that attain the worst-case VaR asymptotically. A general computational approach to construct extremal distributions for distributionally robust optimization problems is described by Bertsimas et al. (2010). In contrast, the construction presented here is completely analytical.

By Proposition 2 in Appendix B, we have

$$\begin{aligned} \mathrm{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w})) \;=\; \sup_{\gamma} \quad & \gamma \\ \text{s.t.} \quad & \mathbb{P}\left(\frac{1}{T}\sum_{t=1}^{T}\left(\tilde{\eta}_t - \frac{1}{2}\tilde{\eta}_t^2\right) \geq \gamma\right) \geq 1-\epsilon \quad \forall \mathbb{P} \in \mathcal{P}_{\tilde{\boldsymbol{\eta}}}(\boldsymbol{w}), \end{aligned} \tag{8}$$

where the projected ambiguity set

$$\mathcal{P}_{\tilde{\boldsymbol{\eta}}}(\boldsymbol{w}) = \left\{ \mathbb{P} \in \mathcal{P}_0^T \;:\; \begin{array}{l} \mathbb{E}_{\mathbb{P}}\left(\tilde{\eta}_t\right) = \boldsymbol{w}^\intercal \boldsymbol{\mu} \quad \forall t:\; 1 \leq t \leq T \\ \mathbb{E}_{\mathbb{P}}\left(\tilde{\eta}_s \tilde{\eta}_t\right) = \delta_{st} \boldsymbol{w}^\intercal \boldsymbol{\Sigma} \boldsymbol{w} + \left(\boldsymbol{w}^\intercal \boldsymbol{\mu}\right)^2 \quad \forall s,t:\; 1 \leq s \leq t \leq T \end{array} \right\}$$

contains all distributions on $\mathbb{R}^T$ under which the portfolio returns $\tilde{\boldsymbol{\eta}} = (\tilde{\eta}_1, \ldots, \tilde{\eta}_T)^\intercal$ follow a weak sense white noise process with (period-wise) mean $\boldsymbol{w}^\intercal\boldsymbol{\mu}$ and variance $\boldsymbol{w}^\intercal\boldsymbol{\Sigma}\boldsymbol{w}$. For a fixed $\boldsymbol{w} \in \mathcal{W}$, we first construct a sequence of *portfolio* return distributions $\mathbb{P}^{\epsilon'} \in \mathcal{P}_{\tilde{\boldsymbol{\eta}}}(\boldsymbol{w})$, $\epsilon' \in (\epsilon, 1)$, that attains the worst case in problem (8) as ϵ' approaches ϵ.

To construct the distribution $\mathbb{P}^{\epsilon'}$ for a fixed $\epsilon' \in (\epsilon, 1)$, we set

$$\Delta = \sigma_{\mathrm{p}}\sqrt{\frac{T}{\epsilon'}}, \quad b = \mu_{\mathrm{p}} + \sqrt{\frac{\epsilon'}{(1-\epsilon')T}}\sigma_{\mathrm{p}}, \quad u = \mu_{\mathrm{p}} - \frac{\Delta}{T} - \sqrt{\frac{1-\epsilon'}{\epsilon' T}}\sigma_{\mathrm{p}}, \quad d = u + \frac{2\Delta}{T}$$

and introduce $2T+1$ portfolio return scenarios $\boldsymbol{\eta}^b$, $\{\boldsymbol{\eta}_t^u\}_{t=1}^T$ and $\{\boldsymbol{\eta}_t^d\}_{t=1}^T$, defined through

$$\begin{aligned} \boldsymbol{\eta}^b &= (\eta_1^b, \ldots, \eta_T^b)^\intercal && \text{where} \quad \eta_s^b = b && \forall s = 1, \ldots, T, \\ \boldsymbol{\eta}_t^u &= (\eta_{t,1}^u, \ldots, \eta_{t,T}^u)^\intercal && \text{where} \quad \eta_{t,s}^u = u + \Delta\delta_{ts} && \forall t, s = 1, \ldots, T, \\ \boldsymbol{\eta}_t^d &= (\eta_{t,1}^d, \ldots, \eta_{t,T}^d)^\intercal && \text{where} \quad \eta_{t,s}^d = d - \Delta\delta_{ts} && \forall t, s = 1, \ldots, T. \end{aligned}$$

We then define $\mathbb{P}^{\epsilon'}$ as the discrete distribution on $\mathbb{R}^T$ with

$$\begin{aligned} &\mathbb{P}^{\epsilon'}\left(\tilde{\boldsymbol{\eta}} = \boldsymbol{\eta}^b\right) = 1 - \epsilon' \\ &\mathbb{P}^{\epsilon'}\left(\tilde{\boldsymbol{\eta}} = \boldsymbol{\eta}_t^u\right) = \frac{\epsilon'}{2T} \quad \forall t = 1, \ldots, T, \\ &\mathbb{P}^{\epsilon'}\left(\tilde{\boldsymbol{\eta}} = \boldsymbol{\eta}_t^d\right) = \frac{\epsilon'}{2T} \quad \forall t = 1, \ldots, T. \end{aligned}$$

Theorem 3 below asserts that the distributions $\mathbb{P}^{\epsilon'}$ attain the worst case in (8) as $\epsilon' \downarrow \epsilon$. Before embarking on the proof of this result, we examine the properties of $\mathbb{P}^{\epsilon'}$. Note that in scenario $\boldsymbol{\eta}^b$ the portfolio returns are constant over time. Moreover, in scenarios $\boldsymbol{\eta}_t^u$ and $\boldsymbol{\eta}_t^d$ the portfolio returns are constant except for period t, in which they spike up and down, respectively. If $T \gg 1$ and/or $\epsilon' \ll 1$, then we have $\Delta \gg 1$. In this case the $\boldsymbol{\eta}_t^d$ scenarios can spike below -1. The distributions $\mathbb{P}^{\epsilon'}$

may therefore be unduly pessimistic. However, this pessimism is manifestation that we err on the side of caution. It is the price we must pay for computational (and analytical) tractability. Less pessimistic worst-case distributions can be obtained by enforcing more restrictive distributional properties in the definition of the ambiguity set $\mathcal{P}$. Examples include support constraints as studied in Section 5.

THEOREM 3. *The portfolio return distributions $\mathbb{P}^{\epsilon'}$, $\epsilon' \in (\epsilon, 1)$, have the following properties.*

(i) $\mathbb{P}^{\epsilon'} \in \mathcal{P}_{\tilde{\boldsymbol{\eta}}}(\boldsymbol{w}) \quad \forall \epsilon' \in (\epsilon, 1)$.

(ii) If $\tilde{\gamma}_T^{\eta} = \frac{1}{T}\sum_{t=1}^T (\tilde{\eta}_t - \frac{1}{2}\tilde{\eta}_t^2)$, *then* $\lim_{\epsilon' \downarrow \epsilon} \mathbb{P}^{\epsilon'}\text{-VaR}_\epsilon(\tilde{\gamma}_T^{\eta}) = \text{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w}))$.

The proof is relegated to Appendix C. As ϵ' tends to ϵ, the distributions $\mathbb{P}^{\epsilon'}$ converge weakly to $\mathbb{P}^{\epsilon}$, where $\mathbb{P}^{\epsilon}$ is defined in the same way as $\mathbb{P}^{\epsilon'}$ for $\epsilon' \in (\epsilon, 1)$. We emphasize, however, that $\mathbb{P}^{\epsilon}$ fails to be a worst-case distribution. As the scenarios $\boldsymbol{\eta}_t^u$ and $\boldsymbol{\eta}_t^d$ for $t = 1, \ldots, T$ have total weight ϵ under $\mathbb{P}^{\epsilon}$, the VaR at level ϵ of $\tilde{\gamma}_T^{\eta}$, which adopts its largest value in scenario $\tilde{\boldsymbol{\eta}} = \boldsymbol{\eta}^b$, is equal to $b - \frac{1}{2}b^2$, which implies that $\mathbb{P}^{\epsilon}\text{-VaR}_\epsilon(\tilde{\gamma}_T^{\eta}) > \text{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w}))$ (see Appendix C for more details). Hence, $\mathbb{P}\text{-VaR}_\epsilon(\tilde{\gamma}_T^{\eta})$ is discontinuous in $\mathbb{P}$ at $\mathbb{P} = \mathbb{P}^{\epsilon}$.

So far we have constructed a sequence of portfolio return distributions that asymptotically attain the worst-case VaR in (8). Next, we construct a sequence of asset return distributions that asymptotically attain the worst-case VaR in (4). To this end, we assume that the portfolio return process $(\tilde{\eta}_1, \ldots, \tilde{\eta}_T)^\intercal \in \mathbb{R}^T$ is governed by a distribution $\mathbb{P}^{\epsilon'}$ of the type constructed above, where $\epsilon' \in (\epsilon, 1)$. Moreover, we denote by $(\tilde{\boldsymbol{m}}_1^\intercal, \ldots, \tilde{\boldsymbol{m}}_T^\intercal)^\intercal \in \mathbb{R}^{nT}$ an auxiliary stochastic process that obeys any distribution under which the $\tilde{\boldsymbol{m}}_t$ are serially independent and each have the same mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, respectively. Then, we denote by $\mathbb{Q}^{\epsilon'}$ the distribution of the asset return process $(\tilde{\boldsymbol{r}}_1^\intercal, \ldots, \tilde{\boldsymbol{r}}_T^\intercal)^\intercal \in \mathbb{R}^{nT}$ defined through

$$\tilde{\boldsymbol{r}}_t = \frac{\boldsymbol{\Sigma w}}{\boldsymbol{w}^\intercal \boldsymbol{\Sigma w}} \tilde{\eta}_t + \left(\mathbb{I} - \frac{\boldsymbol{\Sigma w w}^\intercal}{\boldsymbol{w}^\intercal \boldsymbol{\Sigma w}} \right) \tilde{\boldsymbol{m}}_t \quad \forall t = 1, \ldots, T.$$

COROLLARY 1. *The asset return distributions $\mathbb{Q}^{\epsilon'}$, $\epsilon' \in (\epsilon, 1)$, have the following properties.*

(i) $\mathbb{Q}^{\epsilon'} \in \mathcal{P} \quad \forall \epsilon' \in (\epsilon, 1)$.

(ii) $\lim_{\epsilon' \downarrow \epsilon} \mathbb{Q}^{\epsilon'}\text{-VaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w})) = \text{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w}))$.

Proof. This is an immediate consequence of Theorem 3, as well as Theorem 1 of Yu et al. (2009). ∎

5. Extensions

The basic model of Section 3 can be generalized to account for support information or moment ambiguity. The inclusion of support information shrinks the ambiguity set and thus mitigates the conservatism of the basic model. In contrast, accounting for moment ambiguity enlarges the ambiguity set and enhances the realism of the basic model in situations when there is not enough raw data to obtain high-quality estimates of the means and covariances.

5.1. Support Information

Assume that, besides the usual first and second-order moment information, the asset returns $(\tilde{\boldsymbol{r}}_1^\intercal, \ldots, \tilde{\boldsymbol{r}}_T^\intercal)^\intercal$ are known to materialize within an ellipsoidal support set of the form

$$\Xi = \left\{ (\boldsymbol{r}_1^\intercal, \ldots, \boldsymbol{r}_T^\intercal)^\intercal \in \mathbb{R}^{nT} : \frac{1}{T} \sum_{t=1}^{T} (\boldsymbol{r}_t - \boldsymbol{\nu})^\intercal \boldsymbol{\Lambda}^{-1} (\boldsymbol{r}_t - \boldsymbol{\nu}) \le \delta \right\},$$

where $\boldsymbol{\nu} \in \mathbb{R}^n$ determines the center, $\boldsymbol{\Lambda} \in \mathbb{S}^n$ $(\boldsymbol{\Lambda} \succ \mathbf{0})$ the shape and $\delta \in \mathbb{R}$ $(\delta > 0)$ the size of Ξ. By construction, the ellipsoid Ξ is invariant under permutations of the rebalancing intervals $t = 1, \ldots, T$. This permutation symmetry is instrumental to ensure that any robust growth-optimal portfolio can be computed by solving a tractable conic program of size independent of T. If the usual moment information is complemented by support information, we must replace the standard ambiguity set $\mathcal{P}$ with the (smaller) ambiguity set

$$\mathcal{P}_\Xi = \left\{ \mathbb{P} \in \mathcal{P}_0^{nT} : \begin{array}{c} \mathbb{E}_\mathbb{P}(\tilde{\boldsymbol{r}}_t) = \boldsymbol{\mu} \quad \forall t \ : \ 1 \le t \le T \\ \mathbb{E}_\mathbb{P}(\tilde{\boldsymbol{r}}_s \tilde{\boldsymbol{r}}_t^\intercal) = \delta_{st}\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal \quad \forall s,t \ : \ 1 \le s \le t \le T \\ \mathbb{P}((\tilde{\boldsymbol{r}}_1^\intercal, \ldots, \tilde{\boldsymbol{r}}_T^\intercal)^\intercal \in \Xi) = 1 \end{array} \right\}$$

when computing the worst-case VaR (4). By using a tractable conservative approximation for distributionally robust chance constraints with mean, covariance and support information (see Theorem 6 in Appendix A), we can lower bound this generalized worst-case VaR by the optimal value of a tractable SDP.

$$\begin{array}{rll} \mathrm{WVaR}_\epsilon(\tilde{\gamma}_T'(\boldsymbol{w})) \ \ge & \max & \gamma \\ & \mathrm{s.t.} & \mathbf{M} \in \mathbb{S}^{nT+1}, \ \alpha \in \mathbb{R}, \ \beta \in \mathbb{R}, \ \gamma \in \mathbb{R}, \ \lambda \in \mathbb{R} \\ & & \alpha \ge 0, \quad \beta \le 0, \quad \lambda \ge 0, \quad \beta + \frac{1}{\epsilon}\langle \boldsymbol{\Omega}, \mathbf{M} \rangle \le 0 \\ & & \mathbf{M} \succeq \alpha \begin{bmatrix} -\sum_{t=1}^T \mathbf{P}_t^\intercal \boldsymbol{\Lambda}^{-1} \mathbf{P}_t & \sum_{t=1}^T \mathbf{P}_t^\intercal \boldsymbol{\Lambda}^{-1} \boldsymbol{\nu} \\ \left(\sum_{t=1}^T \mathbf{P}_t^\intercal \boldsymbol{\Lambda}^{-1} \boldsymbol{\nu}\right)^\intercal & T(\delta - \boldsymbol{\mu}^\intercal \boldsymbol{\Lambda}^{-1} \boldsymbol{\mu}) \end{bmatrix} \\ & & \mathbf{M} - \begin{bmatrix} \frac{1}{2}\sum_{t=1}^T \mathbf{P}_t^\intercal \boldsymbol{w}\boldsymbol{w}^\intercal \mathbf{P}_t & -\frac{1}{2}\sum_{t=1}^T \mathbf{P}_t^\intercal \boldsymbol{w} \\ -\frac{1}{2}\left(\sum_{t=1}^T \mathbf{P}_t^\intercal \boldsymbol{w}\right)^\intercal & \gamma T - \beta \end{bmatrix} \succeq \\ & & \qquad \lambda \begin{bmatrix} -\sum_{t=1}^T \mathbf{P}_t^\intercal \boldsymbol{\Lambda}^{-1} \mathbf{P}_t & \sum_{t=1}^T \mathbf{P}_t^\intercal \boldsymbol{\Lambda}^{-1} \boldsymbol{\nu} \\ \left(\sum_{t=1}^T \mathbf{P}_t^\intercal \boldsymbol{\Lambda}^{-1} \boldsymbol{\nu}\right)^\intercal & T(\delta - \boldsymbol{\mu}^\intercal \boldsymbol{\Lambda}^{-1} \boldsymbol{\mu}) \end{bmatrix} \end{array} \tag{9}$$

Here, the truncation operators $\mathbf{P}_t$, $t = 1, \ldots, T$, are defined as in Section 3. We emphasize that even though the SDP (9) offers only a lower bound on the true worst-case VaR *with* support information, it still provides an upper bound on the worst-case VaR of Section 3 *without* support information. This can be seen by fixing $\alpha = \lambda = 0$, in which case (9) reduces to the SDP (5). Note that the SDP (9) is polynomial-time solvable in theory but computationally burdensome in practice because the dimension of the underlying matrix inequalities scales with n and T. We will now show that (9) can be substantially simplified by exploiting its inherent temporal symmetry.

THEOREM 4 **(Support Information)**. *A lower bound on* $\mathrm{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w}))$ *under the ambiguity set* $\mathcal{P}_\Xi$ *is provided by the SDP* (9)*, which is equivalent to*

$$\begin{array}{ll} \max & \gamma \\ \text{s.t.} & \mathbf{A} \in \mathbb{S}^n,\ \mathbf{B} \in \mathbb{S}^n,\ \boldsymbol{c} \in \mathbb{R}^n,\ d \in \mathbb{R},\ \alpha \in \mathbb{R},\ \beta \in \mathbb{R},\ \gamma \in \mathbb{R},\ \lambda \in \mathbb{R} \\ & \alpha \geq 0,\ \beta \leq 0,\ \lambda \geq 0 \\ & \beta + \frac{1}{\epsilon}\left(T\left\langle \mathbf{A}, \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal \right\rangle + T(T-1)\left\langle \mathbf{B}, \boldsymbol{\mu}\boldsymbol{\mu}^\intercal \right\rangle + 2T\boldsymbol{c}^\intercal\boldsymbol{\mu} + d\right) \leq 0 \\ & \begin{bmatrix} \mathbf{A} + (T-1)\mathbf{B} & \boldsymbol{c} \\ \boldsymbol{c}^\intercal & \frac{d}{T} \end{bmatrix} \succeq \alpha \begin{bmatrix} -\boldsymbol{\Lambda}^{-1} & \boldsymbol{\Lambda}^{-1}\boldsymbol{\nu} \\ \boldsymbol{\nu}^\intercal\boldsymbol{\Lambda}^{-1} & \delta - \boldsymbol{\nu}^\intercal\boldsymbol{\Lambda}^{-1}\boldsymbol{\nu} \end{bmatrix} \\ & \mathbf{A} - \mathbf{B} \succeq -\alpha\boldsymbol{\Lambda}^{-1} \\ & \begin{bmatrix} \mathbf{A} + (T-1)\mathbf{B} + \lambda\boldsymbol{\Lambda}^{-1} & c - \lambda\boldsymbol{\Lambda}^{-1}\boldsymbol{\nu} & \boldsymbol{w} \\ c^\intercal - \lambda\boldsymbol{\nu}^\intercal\boldsymbol{\Lambda}^{-1} & \frac{1}{2} + \frac{d+\beta}{T} - \gamma - \lambda(\delta - \boldsymbol{\nu}^\intercal\boldsymbol{\Lambda}^{-1}\boldsymbol{\nu}) & -1 \\ \boldsymbol{w}^\intercal & -1 & 2 \end{bmatrix} \succeq \mathbf{0} \\ & \begin{bmatrix} \mathbf{A} - \mathbf{B} + \lambda\boldsymbol{\Lambda}^{-1} & \boldsymbol{w} \\ \boldsymbol{w}^\intercal & 2 \end{bmatrix} \succeq \mathbf{0}. \end{array} \tag{10}$$

Note that the size of the SDP (10) scales only with the number of assets n but not with T. Theorem 4 implies that one can maximize $\mathrm{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w}))$ approximately over $\boldsymbol{w} \in \mathcal{W}$ by solving a tractable SDP, and thus the robust growth-optimal portfolios can be approximated efficiently even in the presence of support information. The proof of Theorem 4 is relegated to Appendix D.

5.2. Moment Ambiguity

Assume that $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ are possibly inaccurate estimates of the true mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ of the asset returns, respectively. Assume further that $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are known to reside in a convex uncertainty set of the form

$$\mathcal{U} = \left\{ (\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \mathbb{R}^n \times \mathbb{S}^n : (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^\intercal \hat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) \leq \delta_1, \delta_3\hat{\boldsymbol{\Sigma}} \preceq \boldsymbol{\Sigma} \preceq \delta_2\hat{\boldsymbol{\Sigma}} \right\}, \tag{11}$$

where $\delta_1 \geq 0$ reflects our confidence in the estimate $\hat{\boldsymbol{\mu}}$, while the parameters δ_2 and δ_3, $\delta_2 \geq 1 \geq \delta_3 > 0$, express our confidence in the estimate $\hat{\boldsymbol{\Sigma}}$. Guidelines for selecting $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\Sigma}}$, δ_1, δ_2 and δ_3 based on historical data are provided by Delage and Ye (2010). If the asset returns follow a weak sense white noise process with ambiguous means and covariances described by the uncertainty set $\mathcal{U}$ and if the Assumptions (A1) and (A2) are satisfied for all $(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \mathcal{U}$, then, by Theorem 2, the worst-case VaR of the approximate portfolio growth rate $\tilde{\gamma}'_T(\boldsymbol{w})$ is given by

$$\mathrm{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w})) = \min_{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \mathcal{U}} \frac{1}{2}\left(1 - \left(1 - \boldsymbol{w}^\intercal\boldsymbol{\mu} + \sqrt{\frac{1-\epsilon}{\epsilon T}}\left\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{w}\right\|\right)^2 - \frac{T-1}{\epsilon T}\boldsymbol{w}^\intercal\boldsymbol{\Sigma}\boldsymbol{w}\right).$$

We will now demonstrate that the above minimization problem admits an analytical solution.

THEOREM 5. *If (A1) and (A2) hold for all* $(\boldsymbol{\mu},\boldsymbol{\Sigma})\in\mathcal{U}$*, then*

$$\mathrm{WVaR}_\epsilon(\tilde{\gamma}_T'(\boldsymbol{w})) = \frac{1}{2}\left(1-\left(1-\boldsymbol{w}^\intercal\hat{\boldsymbol{\mu}}+\left(\sqrt{\delta_1}+\sqrt{\frac{(1-\epsilon)\,\delta_2}{\epsilon T}}\right)\left\|\hat{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{w}\right\|\right)^2-\frac{\delta_2\,(T-1)}{\epsilon T}\boldsymbol{w}^\intercal\hat{\boldsymbol{\Sigma}}\boldsymbol{w}\right).$$

Proof. Recall that under Assumptions (A1) and (A2) the worst-case VaR (6) is increasing in the portfolio mean return $\boldsymbol{w}^\intercal\boldsymbol{\mu}$ and decreasing in the portfolio standard deviation $\left\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{w}\right\|$. Thus, the worst case of (6) is achieved at the minimum portfolio mean return

$$\min_{(\boldsymbol{\mu},\boldsymbol{\Sigma})\in\mathcal{U}} \boldsymbol{w}^\intercal\boldsymbol{\mu} \;=\; \boldsymbol{w}^\intercal\hat{\boldsymbol{\mu}}-\sqrt{\delta_1}\left\|\hat{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{w}\right\|$$

and at the maximum portfolio standard deviation

$$\max_{(\boldsymbol{\mu},\boldsymbol{\Sigma})\in\mathcal{U}} \left\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{w}\right\| \;=\; \sqrt{\delta_2}\left\|\hat{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{w}\right\|.$$

The claim now follows by substituting the above expressions into (6). ∎

We remark that maximizing $\mathrm{WVaR}_\epsilon(\tilde{\gamma}_T'(\boldsymbol{w}))$ over $\boldsymbol{w}\in\mathcal{W}$ gives rise to a tractable SOCP, and thus the robust growth-optimal portfolios can be computed efficiently even under moment ambiguity. We further remark that the Assumptions (A1) and (A2) hold for all $(\boldsymbol{\mu},\boldsymbol{\Sigma})\in\mathcal{U}$ if and only if

$$\delta_3\hat{\boldsymbol{\Sigma}}\succ\boldsymbol{0} \quad\text{and}\quad \boldsymbol{w}^\intercal\hat{\boldsymbol{\mu}}+\sqrt{\delta_1}\left\|\hat{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{w}\right\|+\sqrt{\frac{\epsilon\delta_2}{(1-\epsilon)T}}\left\|\hat{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{w}\right\|<1.$$

These semidefinite and second-order conic constraints can be verified efficiently.

6. Numerical Experiments

We now assess the robust growth-optimal portfolios in several synthetic and empirical backtests. The emerging second-order cone and semidefinite programs are solved with SDPT3 using the MATLAB interface Yalmip by Löfberg (2004). On a 3.4 GHz machine with 16.0 GB RAM, all portfolio optimization problems of this section are solved in less than 0.40 seconds. Thus, the runtimes are negligible for practical purposes. All experiments rely on one of the following time series with monthly resolution. The *10 Industry Portfolios* (10Ind) and *12 Industry Portfolios* (12Ind) datasets from the Fama French online data library[2] comprise U.S. stock portfolios grouped by industries. The *Dow Jones Industrial Average* (DJIA) dataset is obtained from Yahoo Finance[3] and comprises the 30 constituents of the DJIA index as of August 2013. The *iShares Exchange-Traded Funds* (iShares) dataset is also obtained from Yahoo Finance and comprises the following nine funds: EWG (Germany), EWH (Hong Kong), EWI (Italy), EWK (Belgium), EWL (Switzerland), EWN (Netherlands), EWP (Spain), EWQ (France) and EWU (United Kingdom).

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

[3] `http://finance.yahoo.com`

We will consistently use the shrinkage estimators proposed by DeMiguel et al. (2013) to estimate the mean $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ of a time series. The shrinkage estimator of $\boldsymbol{\mu}$ ($\boldsymbol{\Sigma}$) constitutes a weighted average of the sample mean $\hat{\boldsymbol{\mu}}$ (sample covariance matrix $\hat{\boldsymbol{\Sigma}}$) and the vector of ones scaled by $\frac{\mathbf{1}^\intercal \hat{\boldsymbol{\mu}}}{n}$ (the identity matrix scaled by $\frac{\mathrm{Tr}(\hat{\boldsymbol{\Sigma}})}{n}$). The underlying shrinkage intensities are obtained via the bootstrapping procedure proposed by DeMiguel et al. (2013) using 500 bootstrap samples. Shrinkage estimators have been promoted as a means to combat the impact of estimation errors in portfolio selection. We emphasize that the moments estimated in this manner satisfy the technical Assumptions (A1) and (A2) for all data sets considered in this section.

We henceforth distinguish two different Kelly investors. *Ambiguity-neutral investors* believe that the asset returns follow the unique multivariate lognormal distribution $\mathbb{P}_{\ln}$ consistent with the mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Note that in this model the asset prices follow a discrete-time geometric Brownian motion. We assume that the ambiguity-neutral investors hold the classical growth-optimal portfolio $\boldsymbol{w}_{\mathrm{go}}$, which is defined as the unique maximizer of $\mathbb{E}_{\mathbb{P}_{\ln}}(\log(1+\boldsymbol{w}^\intercal\tilde{\boldsymbol{r}}_t))$ over $\boldsymbol{w} \in \mathcal{W} = \{\boldsymbol{w} \in \mathbb{R}^n : \boldsymbol{w} \geq \mathbf{0},\ \mathbf{1}^\intercal \boldsymbol{w} = 1\}$. By using the second-order Taylor expansion of the logarithm around 1, we may approximate $\boldsymbol{w}_{\mathrm{go}}$ with $\hat{\boldsymbol{w}}_{\mathrm{go}} = \arg\max_{\boldsymbol{w}\in\mathcal{W}} \boldsymbol{w}^\intercal\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{w}^\intercal(\boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^\intercal)\,\boldsymbol{w}$. This approximation is highly accurate under a lognormal distribution if the rebalancing intervals are of the order of a few months or shorter, see Kuhn and Luenberger (2010). *Ambiguity-averse investors* hold the robust growth-optimal portfolio $\boldsymbol{w}_{\mathrm{rgo}}$, which is defined as the unique maximizer of $\mathrm{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w}))$ over $\mathcal{W}$ for $\epsilon = 5\%$. Unless otherwise stated, the worst-case VaR is evaluated with respect to the weak sense white noise ambiguity set $\mathcal{P}$ with known first and second-order moments but without support information.

6.1. Synthetic Experiments

We first illustrate the relation between the parameters T and ϵ of the robust-growth optimal portfolio and the risk-aversion parameters $\varrho(T,\epsilon)$ and $\kappa(T,\epsilon)$ of the Markowitz and fractional Kelly portfolios, respectively. Afterwards, we showcase the benefits of accounting for horizon effects and distributional ambiguity when designing portfolio strategies. In all synthetic experiments we set $n = 10$ and assume that the *true* mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ of the asset returns coincide with the respective estimates obtained from the 120 samples of the 10Ind dataset between 01/2003 and 12/2012. In this setting the growth-optimal portfolio $\boldsymbol{w}_{\mathrm{go}}$ and its approximation $\hat{\boldsymbol{w}}_{\mathrm{go}}$ are virtually indistinguishable. We may thus identify the growth-optimal portfolio with $\hat{\boldsymbol{w}}_{\mathrm{go}}$.

In Section 4 we have seen that each robust growth-optimal portfolio tailored to an investment horizon T and violation probability ϵ is identical to a Markowitz portfolio $\boldsymbol{w}_\varrho = \arg\max_{\boldsymbol{w}\in\mathcal{W}} \boldsymbol{w}^\intercal\boldsymbol{\mu} - \frac{\varrho}{2}\boldsymbol{w}^\intercal\boldsymbol{\Sigma}\boldsymbol{w}$ for some risk aversion parameter $\varrho = \varrho(T,\epsilon)$. Table 1 shows $\varrho(T,\epsilon)$ for different values of T and ϵ, based on the means and covariances obtained from the 10Ind dataset. As expected from

Table 1 Markowitz risk-aversion parameter $\varrho(T,\epsilon)$ implied by the robust growth-optimal portfolio that is tailored to the investment horizon T (in months) and the violation probability ϵ. The reported values are specific to the 10Ind dataset.

T \ ϵ	5%	10%	15%	20%	25%	T \ ϵ	5%	10%	15%	20%	25%
24	46.87	28.80	21.66	17.64	14.97	336	27.45	15.16	10.76	8.44	6.97
48	39.20	23.40	17.35	13.99	11.79	360	27.20	14.99	10.62	8.32	6.87
72	35.77	20.98	15.42	12.36	10.38	384	26.98	14.83	10.50	8.22	6.78
96	33.71	19.54	14.26	11.39	9.53	408	26.78	14.69	10.39	8.12	6.69
120	32.30	18.54	13.47	10.72	8.95	432	26.60	14.56	10.29	8.04	6.62
144	31.25	17.81	12.88	10.23	8.53	456	26.42	14.44	10.19	7.95	6.55
168	30.44	17.25	12.43	9.85	8.19	480	26.27	14.33	10.11	7.88	6.49
192	29.78	16.79	12.06	9.53	7.92	504	26.12	14.23	10.02	7.82	6.43
216	29.24	16.41	11.76	9.28	7.70	528	25.98	14.13	9.94	7.75	6.37
240	28.77	16.08	11.50	9.06	7.51	552	25.86	14.04	9.88	7.69	6.32
264	28.38	15.80	11.28	8.88	7.34	576	25.74	13.97	9.81	7.63	6.27
288	28.03	15.56	11.09	8.71	7.21	600	25.62	13.88	9.75	7.58	6.22
312	27.72	15.35	10.91	8.56	7.08						

the discussion after Theorem 2, $\varrho(T,\epsilon)$ is decreasing in T and ϵ. Note that $\varrho(T,\epsilon)$ exceeds the risk aversion parameter of the classical growth-optimal portfolio ($\varrho = 1$) uniformly for all investment horizons up to 50 years and for all violations probabilities up to 25%. We have also observed that all robust growth-optimal portfolios under consideration are distributed over the leftmost decile of the efficient frontier in the mean-standard deviation plane. Thus, even though they are significantly more conservative than the classical growth-optimal portfolio, the robust growth-optimal portfolios display a significant degree of heterogeneity across different values of T and ε.

In Section 4 we have also seen that the robust growth-optimal portfolio tailored to T and ϵ can be interpreted as a fractional Kelly strategy $\boldsymbol{w}_\kappa = \arg\max_{\boldsymbol{w}\in\mathcal{W}} \boldsymbol{w}^\intercal\boldsymbol{\mu} - \frac{\kappa}{2}\boldsymbol{w}^\intercal(\boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^\intercal)\boldsymbol{w}$ for some risk aversion parameter $\kappa = \kappa(T,\epsilon)$. Table 2 shows $\kappa(T,\epsilon)$ for different values of T and ϵ in the context of the 10Ind dataset. The fractional Kelly and Markowitz risk-aversion parameters display qualitatively similar dependencies on T and ϵ.

Horizon Effects We assume that the asset returns follow the multivariate lognormal distribution $\mathbb{P}_{\text{ln}}$, implying that the beliefs of the ambiguity-neutral investors are correct. In contrast, the ambiguity-averse investors have only limited distributional information and are therefore at a disadvantage. Figure 1(a) displays the 5% VaR of the portfolio growth rate over T months for the classical and the robust growth-optimal portfolios, where the VaR is computed on the basis of 50,000 independent samples from $\mathbb{P}_{\text{ln}}$. Recall that only the robust growth-optimal portfolios are tailored to T. Thus, under the true distribution $\mathbb{P}_{\text{ln}}$ the robust growth-optimal portfolios offer superior performance guarantees (at the desired 95% confidence level) to the classical growth-optimal portfolio across all investment horizons of less than 170 years. Note that longer investment horizons are only of limited practical interest.

Table 2 Fractional Kelly risk-aversion parameter $\kappa(T,\epsilon)$ implied by the robust growth-optimal portfolio that is tailored to the investment horizon T (in months) and the violation probability ϵ. The reported values are specific to the 10Ind dataset.

T \ ϵ	5%	10%	15%	20%	25%	T \ ϵ	5%	10%	15%	20%	25%
24	74.35	37.26	26.14	20.50	16.98	336	35.05	17.22	11.77	9.04	7.38
48	56.76	28.70	20.10	15.73	13.01	360	34.65	17.00	11.60	8.91	7.26
72	49.83	25.15	17.55	13.70	11.31	384	34.29	16.80	11.46	8.79	7.16
96	45.92	23.10	16.07	12.51	10.31	408	33.96	16.62	11.32	8.68	7.07
120	43.34	21.73	15.07	11.71	9.64	432	33.67	16.46	11.20	8.58	6.99
144	41.48	20.73	14.35	11.13	9.14	456	33.39	16.30	11.09	8.49	6.91
168	40.06	19.97	13.78	10.68	8.76	480	33.14	16.16	10.98	8.41	6.84
192	38.93	19.35	13.34	10.32	8.45	504	32.91	16.03	10.89	8.33	6.77
216	38.01	18.85	12.96	10.02	8.20	528	32.69	15.91	10.80	8.26	6.71
240	37.23	18.42	12.65	9.76	7.99	552	32.49	15.80	10.72	8.19	6.65
264	36.57	18.06	12.39	9.55	7.80	576	32.30	15.70	10.64	8.13	6.60
288	36.00	17.75	12.15	9.36	7.64	600	32.13	15.60	10.56	8.07	6.55
312	35.49	17.47	11.95	9.19	7.50						

We also compare the realized Sharpe ratios of the classical and robust growth-optimal portfolios along 50,000 sample paths of length T drawn from $\mathbb{P}_{\rm ln}$. The Sharpe ratio along a given path is defined as the ratio of the sample mean and the sample standard deviation of the monthly portfolio returns on that path. It can be viewed as a signal-to-noise ratio of the portfolio return process and therefore constitutes a popular performance measure for investment strategies. The random *ex post* Sharpe ratios display a high variability for small T but converge almost surely to the deterministic *a priori* Sharpe ratios $\boldsymbol{\mu}^\intercal \boldsymbol{w}_{\rm rgo}/\sqrt{\boldsymbol{w}_{\rm rgo}^\intercal \boldsymbol{\Sigma} \boldsymbol{w}_{\rm rgo}}$ and $\boldsymbol{\mu}^\intercal \boldsymbol{w}_{\rm go}/\sqrt{\boldsymbol{w}_{\rm go}^\intercal \boldsymbol{\Sigma} \boldsymbol{w}_{\rm go}}$, respectively, when T tends to infinity. The boxplot in Figure 1(b) visualizes the distribution of

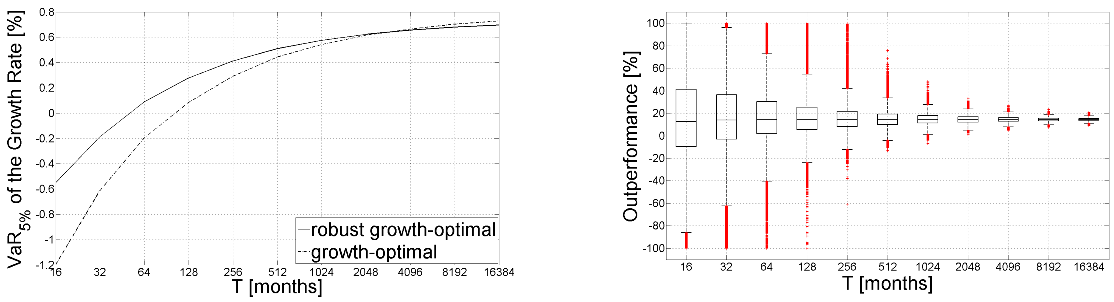
$$\frac{\widehat{SR}_{\rm rgo} - \widehat{SR}_{\rm go}}{\left|\widehat{SR}_{\rm rgo}\right| + \left|\widehat{SR}_{\rm go}\right|},$$

where $\widehat{SR}_{\rm rgo}$ and $\widehat{SR}_{\rm go}$ denote the *ex post* Sharpe ratios of the robust and the classical growth-optimal portfolios, respectively. We observe that the Sharpe ratio of the robust growth-optimal portfolio exceeds that of the classical growth-optimal portfolio by 14.24% on average.

As they are tailored to the investment horizon T, the robust growth-optimal portfolios can offer higher performance guarantees and *ex post* Sharpe ratios than the classical growth-optimal portfolios even though they are ignorant of the exact data-generating distribution $\mathbb{P}_{\rm ln}$.

Ambiguity Effects We now perform a stress test inspired by Bertsimas et al. (2010), where we contamine the lognormal distribution $\mathbb{P}_{\rm ln}$ with the worst-case distributions for the classical and robust growth-optimal portfolios, respectively. More precisely, by Corollary 3 we can construct two near-worst-case distributions $\mathbb{P}_{\rm go}$ and $\mathbb{P}_{\rm rgo}$ satisfying

$$\mathbb{P}_{\rm go}\text{-VaR}_{5\%}(\tilde{\gamma}_T(\boldsymbol{w}_{\rm go})) \;\leq\; \text{WVaR}_{5\%}(\tilde{\gamma}_T(\boldsymbol{w}_{\rm go})) + \delta,$$



(a) 5% VaR of the monthly portfolio growth rates for the classical and robust growth-optimal portfolios.

(b) Relative difference of realized Sharpe-ratios (shown are the 10%, 25%, 50%, 75% and 90% quantiles and outliers)

Figure 1 Comparison of the classical and robust growth-optimal portfolios under the lognormal distribution $\mathbb{P}_{\rm ln}$.

$$\mathbb{P}_{\rm rgo}\text{-VaR}_{5\%}(\tilde{\gamma}_T(\boldsymbol{w}_{\rm rgo})) \;\leq\; \text{WVaR}_{5\%}(\tilde{\gamma}_T(\boldsymbol{w}_{\rm rgo})) + \delta,$$

where δ is a small constant such as 10^{-6}. We can then construct a contaminated distribution

$$\mathbb{P} = \psi_{\rm go}\,\mathbb{P}_{\rm go} + \psi_{\rm rgo}\,\mathbb{P}_{\rm rgo} + (1 - \psi_{\rm go} - \psi_{\rm rgo})\,\mathbb{P}_{\rm ln} \tag{12}$$

using the contamination weights $\psi_{\rm go}, \psi_{\rm rgo} \geq 0$ with $\psi_{\rm go} + \psi_{\rm rgo} \leq 1$. Note that $\mathbb{P} \in \mathcal{P}$ because $\mathbb{P}_{\rm go}, \mathbb{P}_{\rm rgo}, \mathbb{P}_{\rm ln} \in \mathcal{P}$, which implies that the ambiguity-averse investors hedge against all distributions of the form (12). In contrast, the ambiguity-neutral investors exclusively account for the distribution with $\psi_{\rm go} = \psi_{\rm rgo} = 0$. In order to assess the benefits of an ambiguity-averse investment strategy, we evaluate the relative advantage of the robust growth-optimal portfolios over their classical counterparts in terms of their performance guarantees. Thus, we compute

$$\frac{\mathbb{P}\text{-VaR}_{5\%}(\tilde{\gamma}_T(\boldsymbol{w}_{\rm rgo})) - \mathbb{P}\text{-VaR}_{5\%}(\tilde{\gamma}_T(\boldsymbol{w}_{\rm go}))}{|\mathbb{P}\text{-VaR}_{5\%}(\tilde{\gamma}_T(\boldsymbol{w}_{\rm rgo}))| + |\mathbb{P}\text{-VaR}_{5\%}(\tilde{\gamma}_T(\boldsymbol{w}_{\rm go}))|}$$

for all distributions of the form (12), where each VaR is evaluated using 250,000 samples from $\mathbb{P}$. The resulting percentage values are reported in Tables 3, 4, and 5 for investment horizons of 120 months, 360 months and 1,200 months, respectively.

We observe that the robust growth-optimal portfolios outperform their classical counterparts under *all* contaminated probability distributions of the form (12). Even for $\psi_{\rm go} = \psi_{\rm rgo} = 0$ the robust portfolios are at an advantage because they are tailored to the investment horizon. As expected, their advantage increases with the contamination level and is more pronounced for short investment horizons. Only for unrealistically long horizons of more than 100 years and for low contamination levels the classical growth-optimal portfolio becomes competitive.

Table 3 **Relative advantage (in %) of the robust growth-optimal portfolios in terms of 5% VaR ($T = 120$ months)**

ψ_{rgo} \ ψ_{go}	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.0	63.87	62.41	65.83	63.28	65.61	68.20	71.18	67.20	98.19	100.0	72.65
0.1	62.89	63.52	67.78	63.96	71.03	64.81	77.36	78.47	100.0	58.41	
0.2	64.16	64.70	66.56	69.35	68.64	72.99	83.81	100.0	53.81		
0.3	68.97	67.98	68.42	71.10	74.59	90.50	100.0	52.44			
0.4	64.84	66.94	69.93	76.00	88.32	100.0	50.13				
0.5	68.43	72.95	73.92	79.92	100.0	49.85					
0.6	70.19	73.53	89.97	100.0	47.97						
0.7	74.08	71.88	100.0	43.69							
0.8	100.0	100.0	41.11								
0.9	100.0	34.76									
1.0	8.52										

Table 4 **Relative advantage (in %) of the robust growth-optimal portfolios in terms of 5% VaR ($T = 360$ months)**

ψ_{rgo} \ ψ_{go}	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.0	11.60	11.65	11.53	11.50	11.58	11.80	11.90	12.09	13.27	16.30	81.01
0.1	11.61	11.54	11.60	11.87	12.00	12.09	12.41	12.21	16.71	68.62	
0.2	11.42	11.61	11.49	11.46	12.12	12.08	12.05	18.26	65.81		
0.3	11.42	12.05	12.20	12.56	12.87	13.12	16.20	63.49			
0.4	11.37	12.12	11.76	12.00	13.15	18.69	65.79				
0.5	11.58	12.06	12.75	13.17	15.96	61.86					
0.6	12.07	11.90	12.61	17.24	60.05						
0.7	12.19	13.62	16.27	59.91							
0.8	12.78	16.01	58.52								
0.9	15.40	48.91									
1.0	14.75										

Table 5 **Relative advantage (in %) of the robust growth-optimal portfolios in terms of 5% VaR ($T = 1{,}200$ months)**

ψ_{rgo} \ ψ_{go}	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.0	2.49	2.53	2.48	2.50	2.41	2.59	2.58	2.72	2.96	3.03	91.56
0.1	2.52	2.44	2.46	2.47	2.50	2.46	2.68	3.00	3.87	84.23	
0.2	2.48	2.48	2.41	2.57	2.54	2.78	2.74	3.34	82.83		
0.3	2.51	2.54	2.67	2.71	2.66	2.84	3.74	82.98			
0.4	2.56	2.61	2.61	2.64	2.55	3.56	80.90				
0.5	2.59	2.53	2.64	2.70	2.82	80.70					
0.6	2.52	2.69	2.87	3.90	78.64						
0.7	2.46	2.61	4.48	79.01							
0.8	2.93	3.26	75.02								
0.9	3.48	69.17									
1.0	23.57										

6.2. Empirical Backtests

We now assess the performance of the robust growth-optimal portfolio without (RGOP) and with (RGOP$^+$) moment uncertainty on different empirical datasets. RGOP$^+$ optimizes the worst-case VaR over all means and covariance matrices in the uncertainty set (11). We compare the robust growth-optimal portfolios against the equally weighted portfolio ($1/n$), the classical growth-optimal portfolio (GOP), the fractional Kelly strategy corresponding to the risk-aversion parameter $\kappa = 2$ (1/2-Kelly), two mean-variance efficient portfolios corresponding to the risk-aversion parameters $\varrho = 1$ and $\varrho = 3$ (MV) and Cover's universal portfolio (UNIV). The equally weighted portfolio contains all assets in equal proportions. This seemingly naïve investment strategy is immune to estimation errors and surprisingly difficult to outperform with optimization-based portfolio strategies, see DeMiguel et al. (2009). In the presence of a risk-free instrument, the fractional Kelly strategy corresponding to $\kappa = 2$ invests approximately half of the capital in the classical growth-optimal portfolio and the other half in cash. This so-called '*half-Kelly*' strategy enjoys wide popularity among investors wishing to trade off growth versus security, see e.g. MacLean et al. (2005). The Markowitz portfolio corresponding to $\varrho = 1$ closely approximates the classical growth-optimal portfolio, while the Markowitz portfolio corresponding to $\varrho = 3$ provides a more conservative alternative. Moreover, the universal portfolio by Cover (1991) learns adaptively the best fixed-mix strategy from the history of observed asset returns. We compute the universal portfolio using a weighted average of 10^6 portfolios chosen uniformly at random from $\mathcal{W}$ where the weights are proportional to their empirical performance; see Blum and Kalai (1999).

To increase the practical relevance of our experiments, we evaluate all investment strategies under proportional transaction costs of $c = 50$ basis points per dollar traded. Note that the RGOP, RGOP$^+$, GOP, MV and 1/2-Kelly strategies all depend on estimates $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ of the (unknown) true mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ of the asset returns, respectively. The RGOP$^+$ strategy further depends on estimates of the confidence parameters δ_1, δ_2 and δ_3 characterizing the uncertainty set (11). All moments and confidence parameters are re-estimated every 12 months using the most recent 120 observations. Accordingly, the portfolio weights of all fixed-mix strategies are recalculated every 12 months based on the new estimates and (in the case of RGOP and RGOP$^+$) a shrunk investment horizon. Strictly speaking, the resulting investment strategies are thus no longer of fixed-mix type. Instead, the portfolio weights are periodically updated in a greedy fashion. We stress that our numerical results do not change qualitatively if we use a shorter re-estimation interval of 6 months or a longer interval of 24 months. For the sake of brevity, we only report the results for a re-estimation window of 12 months.

We choose δ_1 and δ_2 such that the moment uncertainty set (11) constructed from the estimates $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ contains the true mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ with confidence 95%. This is achieved via the bootstrapping procedure proposed by Delage and Ye (2010), implemented with 500 iterations and two bootstrap datasets of size 120 per iteration. The parameter δ_3 defines a lower bound on the covariance matrix that is never binding; see also Remark 1 of Delage and Ye (2010). Thus, we can set $\delta_3 = 0$ without loss of generality.

We evaluate the performance of the different investment strategies on the 10Ind, 12Ind, iShares and DJIA datasets. We denote by $\boldsymbol{w}_t^-$ and $\boldsymbol{w}_t$ the portfolio weights before and after rebalancing at the beginning of interval t, respectively. Thus, $\boldsymbol{w}_t$ represents the target portfolio prescribed by the underlying strategy. The following performance measures are recorded for every strategy:

1. *Mean return*:
$$\hat{r}_p = \frac{1}{T}\sum_{t=1}^{T}\left((1+\boldsymbol{w}_t^\intercal \boldsymbol{r}_t)\left(1-c\sum_{i=1}^{n}\left|w_{t,i}-w_{t,i}^-\right|\right)-1\right).$$
2. *Standard deviation*:
$$\hat{\sigma}_p = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}\left((1+\boldsymbol{w}_t^\intercal \boldsymbol{r}_t)\left(1-c\sum_{i=1}^{n}\left|w_{t,i}-w_{t,i}^-\right|\right)-1-\hat{r}_p\right)^2}.$$
3. *Sharpe ratio*:
$$\widehat{SR} = \frac{\hat{r}_p}{\hat{\sigma}_p}.$$
4. *Turnover rate*:
$$\widehat{TR} = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{n}\left|w_{t,i}-w_{t,i}^-\right|.$$
5. *Net Aggregate Return*:
$$\widehat{NR} = \widehat{V}_T, \quad \widehat{V}_t = \prod_{s=1}^{t}(1+\boldsymbol{w}_s^\intercal \boldsymbol{r}_s)\left(1-c\sum_{i=1}^{n}\left|w_{s,i}-w_{s,i}^-\right|\right).$$
6. *Maximum drawdown*:
$$\widehat{MDD} = \max_{1\le s<t\le T}\frac{\widehat{V}_s-\widehat{V}_t}{\widehat{V}_s}.$$

The results of the empirical backtests are reported in Table 6. We observe that the robust growth-optimal portfolios with and without moment uncertainty consistently outperform the other strategies in terms of out-of-sample Sharpe ratios and thus generate the smoothest wealth dynamics. Moreover, the robust growth-optimal portfolios achieve the lowest standard deviation and the lowest maximum drawdown (maximum percentage loss over any subinterval of the backtest period) across all datasets. These results suggest that the robust growth-optimal portfolios are only moderately risky. The universal portfolio as well as the equally weighted portfolio achieve the lowest turnover rate, which determines the total amount of transaction costs incurred by an investment strategy. This is not surprising as these two portfolios are independent of the investors' changing beliefs about the future asset returns. Nonetheless, the robust growth-optimal portfolios achieve higher terminal wealth than the equally weighted portfolio in the majority of backtests. Maybe surprisingly, despite its theoretical appeal, the classical growth-optimal portfolio is strictly dominated by most other strategies. In fact, it is highly susceptible to error maximization phenomena as it aggressively invests in assets whose estimated mean returns are high.

We also tested whether the Sharpe ratio of the RGOP$^+$ strategy statistically exceeds those of the other strategies by using a significance test proposed in Jobson and Korkie (1981) and Memmel (2003). The corresponding one-sided p-values are reported in Table 6 (in parenthesis). Star symbols (*) identify p-values that are significant at the 5% level. We observe that the RGOP$^+$ strategy achieves a significantly higher Sharpe ratio than all other benchmarks in the majority of experiments.

Table 6 Out-of-sample performance of different investment strategies. The first column specifies the dataset used in the respective experiment as well as the underlying backtest period (excluding the ten-year estimation window prior to the first rebalancing interval). The best performance measures found in each experiment are highlighted by gray shading.

Dataset	Portfolio	$\hat{r}_p$	$\hat{\sigma}_p$	$\widehat{SR}$	$\widehat{TR}$	$\widehat{NR}$	$\widehat{MDD}$
10Ind (01/2000–12/2012)	RGOP	0.0062	0.0360	0.1718	0.0437	2.3628	0.3555
	RGOP$^+$	0.0064	0.0361	0.1775	0.0433	2.4465	0.3544
	$1/n$	0.0050	0.0444	0.1130* (0.0311)	0.0325	1.8714	0.4818
	GOP	0.0008	0.0583	0.0143* (0.0105)	0.0812	0.8681	0.6738
	1/2-Kelly	0.0015	0.0502	0.0301* (0.0113)	0.0760	1.0352	0.6319
	MV ($\varrho = 1$)	0.0009	0.0583	0.0150* (0.0107)	0.0805	0.8740	0.6731
	MV ($\varrho = 3$)	0.0025	0.0442	0.0555* (0.0133)	0.0706	1.2550	0.5617
	UNIV	0.0050	0.0441	0.1139* (0.0310)	0.0323	1.8763	0.4796
12Ind (01/2000–12/2012)	RGOP	0.0063	0.0359	0.1744	0.0445	2.3925	0.3605
	RGOP$^+$	0.0065	0.0361	0.1805	0.0444	2.4875	0.3606
	$1/n$	0.0049	0.0449	0.1097* (0.0207)	0.0320	1.8374	0.4966
	GOP	0.0013	0.0585	0.0225* (0.0134)	0.0763	0.9338	0.6457
	1/2-Kelly	0.0018	0.0499	0.0368* (0.0134)	0.0763	1.0920	0.6146
	MV ($\varrho = 1$)	0.0013	0.0584	0.0231* (0.0136)	0.0761	0.9395	0.6451
	MV ($\varrho = 3$)	0.0026	0.0437	0.0596* (0.0135)	0.0712	1.2897	0.5489
	UNIV	0.0049	0.0446	0.1103* (0.0206)	0.0318	1.8402	0.4951
iShares (04/2006–07/2013)	RGOP	0.0033	0.0573	0.0575	0.0388	1.1548	0.5867
	RGOP$^+$	0.0033	0.0573	0.0576	0.0388	1.1555	0.5865
	$1/n$	0.0029	0.0689	0.0425 (0.3086)	0.0321	1.0466	0.6045
	GOP	0.0032	0.0628	0.0503 (0.3723)	0.0649	1.1042	0.6165
	1/2-Kelly	0.0030	0.0592	0.0505 (0.2482)	0.0514	1.1114	0.6022
	MV ($\varrho = 1$)	0.0032	0.0627	0.0505 (0.3753)	0.0646	1.1054	0.6154
	MV ($\varrho = 3$)	0.0030	0.0582	0.0516 (0.1695)	0.0457	1.1195	0.5964
	UNIV	0.0030	0.0687	0.0431 (0.3141)	0.0321	1.0509	0.6045
DJIA (04/2000–07/2013)	RGOP	0.0049	0.0381	0.1296	0.0668	1.9569	0.3966
	RGOP$^+$	0.0057	0.0381	0.1498	0.0651	2.2118	0.4000
	$1/n$	0.0066	0.0460	0.1424 (0.4230)	0.0527	2.4017	0.4824
	GOP	−0.0025	0.0801	−0.0313* (0.0107)	0.1034	0.3831	0.8389
	1/2-Kelly	−0.0029	0.0685	−0.0430* (0.0055)	0.1002	0.4105	0.8269
	MV ($\varrho = 1$)	−0.0025	0.0805	−0.0308* (0.0109)	0.1024	0.3828	0.8398
	MV ($\varrho = 3$)	−0.0009	0.0563	−0.0160* (0.0068)	0.0931	0.6605	0.7325
	UNIV	0.0066	0.0459	0.1434 (0.4333)	0.0527	2.4153	0.4804

7. Conclusions

The classical growth-optimal portfolio maximizes the growth-rate of wealth that can be guaranteed with certainty over an infinite planning horizon if the asset return distribution is precisely known.

The robust growth-optimal portfolios introduced in this paper maximize the growth-rate of wealth that can be guaranteed with probability $1-\epsilon$ over a finite investment horizon of T periods if the asset return distribution is ambiguous. We show that any robust growth-optimal portfolio can be computed almost as efficiently as a Markowitz portfolio by solving a convex optimization problem whose size is constant in T. If the distributional uncertainty is captured by a weak-sense white noise ambiguity set, then the robust growth-optimal portfolios can naturally be identified with classical Markowitz portfolios or fractional Kelly strategies. However, in contrast to Markowitz and fractional Kelly investors, robust growth-optimal investors are absolved from the burden of determining their risk-aversion parameter. Instead, they only have to specify their investment horizon T and violation tolerance ϵ, both of which admit a simple physical interpretation. Simulated backtests indicate that the robust growth-optimal portfolio tailored to a finite investment horizon T can outperform the classical growth-optimal portfolio in terms of Sharpe ratio and growth guarantees for all investment horizons up to ~ 170 years even if the classical growth-optimal portfolio has access to the true data-generating distribution. The outperformance becomes more dramatic if the out-of-sample distribution deviates from the in-sample distribution used to compute the classical growth-optimal portfolio. Empirical backtests suggest that robust growth-optimal portfolios compare favorably against popular benchmark strategies such as the $1/n$ portfolio, various Markowitz portfolios, the classical growth-optimal portfolio, the half-Kelly strategy and Cover's universal portfolio. The $1/n$ portfolio achieves a lower turnover rate but is dominated by the robust growth-optimal portfolio in terms of the realized Sharpe ratio, realized net return and several other indicators even in the presence of high proportional transaction costs of 50 basis points. Our backtest experiments further showcase the benefits of accounting for moment ambiguity.

References

Algoet, P., T. Cover. 1988. Asymptotic optimality and asymptotic equipartition properties of log-optimum investment. *Annals of Probability* **16**(2) 876–898.

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

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

Best, M., R. Grauer. 1991. On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results. *Review of Financial Studies* **4**(2) 315–342.

Blum, A., A. Kalai. 1999. Universal portfolios with and without transaction costs. *Machine Learning* **35**(3) 193–205.

Boyd, S., L. Vandenberghe. 2004. *Convex Optimization*. Cambridge University Press.

Breiman, L. 1961. Optimal gambling systems for favourable games. *Fourth Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, 65–78.

Chopra, V., W. Ziemba. 1993. The effect of errors in means, variances, and covariances on optimal portfolio choice. *Journal of Portfolio Management* **19**(2) 6–11.

Christensen, M. 2012. On the history of the growth optimal portfolio. L. Györfi, G. Ottucsák, H. Walk, eds., *Machine Learning for Financial Engineering*. Imperial College Press, London, 1–80.

Cover, T. 1991. Universal portfolios. *Mathematical Finance* **1**(1) 1–29.

Cover, T., J. Thomas. 1991. *Elements of Information Theory*. Wiley Series in Telecommunications, Wiley.

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

DeMiguel, V., L. Garlappi, R. Uppal. 2009. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies* **22**(5) 1915–1953.

DeMiguel, V., A. Martin-Utrera, F. Nogales. 2013. Size matters: Optimal calibration of shrinkage estimators for portfolio selection. *Journal of Banking & Finance* **37**(8) 3018–3034.

Dempster, M., I. Evstigneev, K. Schenk-Hoppé. 2008. Financial markets. The joy of volatility. *Quantitative Finance* **8**(1) 1–3.

Doan, X., X. Li, K. Natarajan. 2013. Robustness to dependency in portfolio optimization using overlapping marginals. *Available on Optimization Online* .

El Ghaoui, L., M. Oks, F. Oustry. 2003. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research* **51**(4) 543–556.

Estrada, J. 2010. Geometric mean maximization: An overlooked portfolio approach? *Journal of Investing* **19**(4) 134–147.

Ethier, S. 2004. The Kelly system maximizes median fortune. *Journal of Applied Probability* **41**(4) 1230–1236.

Gilboa, I., D. Schmeidler. 1989. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics* **18**(2) 141–153.

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

Györfi, L., G. Ottucsák, A. Urbán. 2012. Empirical log-optimal portfolio selections: A survey. L. Györfi, G. Ottucsák, H. Walk, eds., *Machine Learning for Financial Engineering*. Imperial College Press, London, 81–118.

Hakansson, N. 1971a. Multi-period mean-variance analysis: Toward a general theory of portfolio choice. *Journal of Finance* **26**(4) 857–884.

Hakansson, N. 1971b. On optimal myopic portfolio policies, with and without serial correlation of yields. *The Journal of Business* **44**(3) 324–334.

Hakansson, N., B. Miller. 1975. Compound-return mean-variance efficient portfolios never risk ruin. *Management Science* **22**(4) 391–400.

Jegadeesh, N., S. Titman. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance* **48**(1) 65–91.

Jobson, J., B. Korkie. 1981. Performance hypothesis testing with the Sharpe and Treynor measures. *Journal of Finance* **36**(4) 889–908.

Kelly, J. 1956. A new interpretation of information rate. *Bell System Technical Journal* **35**(4) 917–926.

Kuhn, D., D. Luenberger. 2010. Analysis of the rebalancing frequency in log-optimal portfolio selection. *Quantitative Finance* **10**(2) 221–234.

Latané, H. 1959. Criteria for choice among risky ventures. *Journal of Political Economy* **67**(2) 144–155.

Lintner, J. 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics Statistics* **47**(1) 13–37.

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

Luenberger, D. 1993. A preference foundation for log mean-variance criteria in portfolio choice problems. *Journal of Economic Dynamics and Control* **17**(5-6) 887–906.

Luenberger, D. 1998. *Investment Science*. Oxford University Press.

MacLean, L., E. Thorp, W. Ziemba. 2010. Good and bad properties of the Kelly criterion. *The Kelly Capital Growth Investment Criterion: Theory and Practice*. World Scientific, 563–574.

MacLean, L. C., W. T. Ziemba, Y. Li. 2005. Time to wealth goals in capital accumulation and the optimal trade-off of growth versus security. *Quantitative Finance* **5**(4) 343–357.

Markowitz, H. 1952. Portfolio selection. *Journal of Finance* **7**(1) 77–91.

Memmel, C. 2003. Performance hypothesis testing with the Sharpe ratio. *Finance Letters* **1**(1) 21–23.

Merton, R., P. Samuelson. 1974. Fallacy of the log-normal approximation to optimal portfolio decision-making over many periods. *Journal of Financial Economics* **1**(1) 67–94.

Meskarian, R., H. Xu. 2013. Robust reward-risk ratio optimization: Modeling and numerical methods. Technical report, Singapore University of Technology and Design.

Michaud, R. 1989. The Markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal* **45**(1) 31–42.

Mossin, J. 1966. Equilibrium in a capital asset market. *Econometrica* **34**(4) 768–783.

Natarajan, K., D. Pachamanova, M. Sim. 2008. Incorporating asymmetric distributional information in robust value-at-risk optimization. *Management Science* **54**(3) 573–585.

Natarajan, K., M. Sim, J. Uichanco. 2010. Tractable robust expected utility and risk models for portfolio optimization. *Mathematical Finance* **20**(4) 695–731.

Poundstone, W. 2005. *Fortune's Formula: The Untold Story of the Scientific Betting System that Beat the Casinos and Wall Street*. Hill & Wang.

Roll, R. 1973. Evidence on the "growth-optimum" model. *Journal of Finance* **28**(3) 551–566.

Roy, A. 1952. Safety first and the holding of assets. *Econometrica* **20**(3) 431–449.

Rubinstein, M. 1991. Continuously rebalanced investment strategies. *Journal of Portfolio Management* **18**(1) 78–81.

Samuelson, P. 1963. Risk and uncertainty: A fallacy of large numbers. *Scientia* **57**(6) 50–55.

Samuelson, P. 1970. The fundamental approximation theorem of portfolio analysis in terms of means, variances and higher moments. *Review of Economic Studies* **37**(4) 537–542.

Samuelson, P. 1971. The fallacy of maximizing the geometric mean in long sequences of investing or gambling. *Proceedings of the National Academy of Sciences* **68**(10) 2493–2496.

Sharpe, W. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* **19**(3) 425–442.

Thorp, E. 1971. Portfolio choice and the Kelly criterion. *Business and Economics Statistics Section of Proceedings of the American Statistical Association* 215–224.

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

Votaw, D. 1948. Testing compound symmetry in a normal multivariate distribution. *Annals of Mathematical Statistics* **19**(4) 447–473.

Wiesemann, W., D. Kuhn, M. Sim. 2014. Distributionally robust convex optimization. *To appear in Operations Research* .

Ye, Y. 1997. *Interior Point Algorithms: Theory and Analysis*. Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley.

Yu, Y., Y. Li, D. Schuurmans, C. Szepesvári. 2009. A general projection property for distribution families. *Advances in Neural Information Processing Systems* **22** 2232–2240.

Ziemba, W. 2005. The symmetric downside-risk Sharpe ratio. *Journal of Portfolio Management* **32**(1) 108–122.

Zymler, S., D. Kuhn, B. Rustem. 2013a. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming A* **137**(1-2) 167–198.

Zymler, S., D. Kuhn, B. Rustem. 2013b. Worst-case value at risk of nonlinear portfolios. *Management Science* **59**(1) 172–188.

Appendix A: Distributionally Robust Quadratic Chance Constraints

THEOREM 6. *Let $\mathcal{P}$ be the set of all probability distributions of $\tilde{\boldsymbol{\xi}} \in \mathbb{R}^n$ that share the same mean $\boldsymbol{\mu} \in \mathbb{R}^n$ and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{S}_+^n$, $\boldsymbol{\Sigma} \succ \mathbf{0}$. Moreover, let $\mathcal{P}(\Xi)$ be the subset of $\mathcal{P}$ that contains only distributions supported on the ellipsoid $\Xi = \{\boldsymbol{\xi} \in \mathbb{R}^n : (\boldsymbol{\xi} - \boldsymbol{\nu})^\intercal \boldsymbol{\Lambda}^{-1}(\boldsymbol{\xi} - \boldsymbol{\nu}) \le \delta\}$, where $\boldsymbol{\Lambda} \in \mathbb{S}^n$, $\boldsymbol{\Lambda} \succ \mathbf{0}$, $\boldsymbol{\nu} \in \mathbb{R}^n$ and $\delta \in \mathbb{R}$, $\delta > 0$. Then, for $\mathbf{Q} \in \mathbb{S}^n$, $\boldsymbol{q} \in \mathbb{R}^n$ and $q^0 \in \mathbb{R}$ we find:*

(i) the distributionally robust chance constraint $\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(\tilde{\boldsymbol{\xi}}^\intercal \mathbf{Q} \tilde{\boldsymbol{\xi}} + \tilde{\boldsymbol{\xi}}^\intercal \boldsymbol{q} + q^0 \le 0\right) \ge 1 - \epsilon$ with moment information is equivalent to

$$\exists \mathbf{M} \in \mathbb{S}^{n+1},\ \beta \in \mathbb{R}: \quad \beta + \frac{1}{\epsilon}\langle \boldsymbol{\Omega}, \mathbf{M}\rangle \le 0, \quad \mathbf{M} \succeq \mathbf{0}, \quad \mathbf{M} \succeq \begin{bmatrix} \mathbf{Q} & \frac{1}{2}\boldsymbol{q} \\ \frac{1}{2}\boldsymbol{q}^\intercal & q^0 - \beta \end{bmatrix}; \tag{13}$$

(ii) the distributionally robust chance constraint $\inf_{\mathbb{P}\in\mathcal{P}(\Xi)} \mathbb{P}\left(\tilde{\boldsymbol{\xi}}^\intercal \mathbf{Q} \tilde{\boldsymbol{\xi}} + \tilde{\boldsymbol{\xi}}^\intercal \boldsymbol{q} + q^0 \le 0\right) \ge 1 - \epsilon$ with moment and support information is implied (conservatively approximated) by

$$\begin{aligned} &\exists \mathbf{M} \in \mathbb{S}^{n+1},\ \alpha \in \mathbb{R},\ \beta \in \mathbb{R},\ \lambda \in \mathbb{R}: \\ &\qquad \beta + \frac{1}{\epsilon}\langle \boldsymbol{\Omega}, \mathbf{M}\rangle \le 0, \quad \mathbf{M} \succeq \alpha \begin{bmatrix} -\boldsymbol{\Lambda}^{-1} & \boldsymbol{\Lambda}^{-1}\boldsymbol{\nu} \\ (\boldsymbol{\Lambda}^{-1}\boldsymbol{\nu})^\intercal & \delta - \boldsymbol{\mu}^\intercal \boldsymbol{\Lambda}^{-1}\boldsymbol{\mu} \end{bmatrix}, \\ &\qquad \mathbf{M} \succeq \begin{bmatrix} \mathbf{Q} & \frac{1}{2}\boldsymbol{q} \\ \frac{1}{2}\boldsymbol{q}^\intercal & q^0 - \beta \end{bmatrix} + \lambda \begin{bmatrix} -\boldsymbol{\Lambda}^{-1} & \boldsymbol{\Lambda}^{-1}\boldsymbol{\nu} \\ (\boldsymbol{\Lambda}^{-1}\boldsymbol{\nu})^\intercal & \delta - \boldsymbol{\mu}^\intercal \boldsymbol{\Lambda}^{-1}\boldsymbol{\mu} \end{bmatrix}. \end{aligned} \tag{14}$$

In the above expressions, $\boldsymbol{\Omega}$ is a notational shorthand for the second-order moment matrix of $\tilde{\boldsymbol{\xi}}$,

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal & \boldsymbol{\mu} \\ \boldsymbol{\mu}^\intercal & 1 \end{bmatrix}.$$

Proof. Assertion (i) follows from Vandenberghe et al. (2007) or Theorem 2.3 of Zymler et al. (2013b). Assertion (ii) is an immediate consequence of Theorem 3.7 of Zymler et al. (2013b). Note that (14) reduces to (13) if we set $\alpha = \lambda = 0$. ∎

Appendix B: Proof of Theorem 2

In order to prove Theorm 2, we simplify the semidefinite program (4) in several steps. We first notice that the random asset returns $\tilde{\boldsymbol{r}}_t$ enter problem (4) only in the form of the portfolio return $\boldsymbol{w}^\intercal \tilde{\boldsymbol{r}}_t$. We can thus use a well-known projection property of moment-based ambiguity sets to perform a dimensionality reduction.

PROPOSITION 2 **(General Projection Property)**. *Let $\tilde{\boldsymbol{\xi}}$ and $\tilde{\boldsymbol{\zeta}}$ be random vectors valued in $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively, and define the ambiguity sets $\mathcal{P}_{\tilde{\boldsymbol{\xi}}}$ and $\mathcal{P}_{\tilde{\boldsymbol{\zeta}}}$ as*

$$\mathcal{P}_{\tilde{\boldsymbol{\xi}}} = \left\{ \mathbb{P} \in \mathcal{P}_0^p :\ \mathbb{E}_{\mathbb{P}}\left(\begin{bmatrix} \tilde{\boldsymbol{\xi}}^\intercal & 1 \end{bmatrix}^\intercal \begin{bmatrix} \tilde{\boldsymbol{\xi}}^\intercal & 1 \end{bmatrix} \right) = \boldsymbol{\Omega}_{\tilde{\boldsymbol{\xi}}} \right\}$$

and

$$\mathcal{P}_{\tilde{\boldsymbol{\zeta}}} = \left\{ \mathbb{P} \in \mathcal{P}_0^q :\ \mathbb{E}_{\mathbb{P}}\left(\begin{bmatrix} \tilde{\boldsymbol{\zeta}}^\intercal & 1 \end{bmatrix}^\intercal \begin{bmatrix} \tilde{\boldsymbol{\zeta}}^\intercal & 1 \end{bmatrix} \right) = \boldsymbol{\Omega}_{\tilde{\boldsymbol{\zeta}}} \right\},$$

where the moment matrices $\boldsymbol{\Omega}_{\tilde{\boldsymbol{\xi}}} \in \mathbb{S}_+^{p+1}$ and $\boldsymbol{\Omega}_{\tilde{\boldsymbol{\zeta}}} \in \mathbb{S}_+^{q+1}$ are related through

$$\boldsymbol{\Omega}_{\tilde{\boldsymbol{\zeta}}} = \begin{bmatrix} \boldsymbol{\Lambda} & \mathbf{0} \\ \mathbf{0}^\intercal & 1 \end{bmatrix} \boldsymbol{\Omega}_{\tilde{\boldsymbol{\xi}}} \begin{bmatrix} \boldsymbol{\Lambda} & \mathbf{0} \\ \mathbf{0}^\intercal & 1 \end{bmatrix}^\intercal$$

for some matrix $\boldsymbol{\Lambda} \in \mathbb{R}^{q \times p}$. Then, for any Borel measurable function $f : \mathbb{R}^q \to \mathbb{R}$, we have

$$\inf_{\mathbb{P}\in\mathcal{P}_{\tilde{\boldsymbol{\zeta}}}} \mathbb{P}\left(f(\tilde{\boldsymbol{\zeta}}) \le 0 \right) = \inf_{\mathbb{P}\in\mathcal{P}_{\tilde{\boldsymbol{\xi}}}} \mathbb{P}\left(f(\boldsymbol{\Lambda}\tilde{\boldsymbol{\xi}}) \le 0 \right).$$

Proof. This is an immediate consequence of Theorem 1 of Yu et al. (2009). ∎

Applying Proposition 2 to problem (4) yields

$$\begin{aligned} \mathrm{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w})) \;=\; & \sup_{\gamma} \;\; \gamma \\ & \text{s.t.} \;\; \mathbb{P}\left(\frac{1}{T}\sum_{t=1}^{T}\left(\tilde{\eta}_t - \frac{1}{2}\tilde{\eta}_t^2\right) \geq \gamma\right) \geq 1-\epsilon \quad \forall \mathbb{P} \in \mathcal{P}_{\tilde{\boldsymbol{\eta}}}(\boldsymbol{w}), \end{aligned} \tag{15}$$

where the projected ambiguity set

$$\mathcal{P}_{\tilde{\boldsymbol{\eta}}}(\boldsymbol{w}) = \left\{ \mathbb{P} \in \mathcal{P}_0^T \;:\; \begin{array}{ll} \mathbb{E}_{\mathbb{P}}(\tilde{\eta}_t) = \boldsymbol{w}^\intercal \boldsymbol{\mu} & \forall t:\; 1 \leq t \leq T \\ \mathbb{E}_{\mathbb{P}}(\tilde{\eta}_s \tilde{\eta}_t) = \delta_{st} \boldsymbol{w}^\intercal \boldsymbol{\Sigma} \boldsymbol{w} + (\boldsymbol{w}^\intercal \boldsymbol{\mu})^2 & \forall s,t:\; 1 \leq s \leq t \leq T \end{array} \right\}$$

contains all distributions on $\mathbb{R}^T$ under which the portfolio returns $(\tilde{\eta}_1, \ldots, \tilde{\eta}_T)^\intercal$ follow a weak sense white noise process with (period-wise) mean $\boldsymbol{w}^\intercal\boldsymbol{\mu}$ and variance $\boldsymbol{w}^\intercal\boldsymbol{\Sigma}\boldsymbol{w}$. Problem (15) has the same structure as the original problem (4), but the underlying probability space has only dimension T instead of nT. Thus, it can be converted to a tractable SDP by using Theorem 6 to reformulate the underlying distributionally robust chance constraint. We then obtain

$$\begin{aligned} \mathrm{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w})) \;=\; & \max \;\; \gamma \\ & \text{s.t.} \;\; \mathbf{M} \in \mathbb{S}^{T+1},\; \beta \in \mathbb{R},\; \gamma \in \mathbb{R} \\ & \qquad \beta + \tfrac{1}{\epsilon}\langle \boldsymbol{\Omega}(\boldsymbol{w}), \mathbf{M} \rangle \leq 0, \quad \mathbf{M} \succeq \mathbf{0} \\ & \qquad \mathbf{M} - \begin{bmatrix} \frac{1}{2}\mathbb{I} & -\frac{1}{2}\mathbf{1} \\ -\frac{1}{2}\mathbf{1}^\intercal & \gamma T - \beta \end{bmatrix} \succeq \mathbf{0}, \end{aligned} \tag{16}$$

where $\boldsymbol{\Omega}(\boldsymbol{w}) \in \mathbb{S}^{T+1}$ denotes the matrix of first and second-order moments of $(\tilde{\eta}_1, \ldots, \tilde{\eta}_T)^\intercal$.

$$\boldsymbol{\Omega}(\boldsymbol{w}) = \left[\begin{array}{cccc|c} \boldsymbol{w}^\intercal\boldsymbol{\Sigma}\boldsymbol{w} + (\boldsymbol{w}^\intercal\boldsymbol{\mu})^2 & (\boldsymbol{w}^\intercal\boldsymbol{\mu})^2 & \cdots & (\boldsymbol{w}^\intercal\boldsymbol{\mu})^2 & \boldsymbol{w}^\intercal\boldsymbol{\mu} \\ (\boldsymbol{w}^\intercal\boldsymbol{\mu})^2 & \boldsymbol{w}^\intercal\boldsymbol{\Sigma}\boldsymbol{w} + (\boldsymbol{w}^\intercal\boldsymbol{\mu})^2 & \cdots & (\boldsymbol{w}^\intercal\boldsymbol{\mu})^2 & \boldsymbol{w}^\intercal\boldsymbol{\mu} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ (\boldsymbol{w}^\intercal\boldsymbol{\mu})^2 & (\boldsymbol{w}^\intercal\boldsymbol{\mu})^2 & \cdots & \boldsymbol{w}^\intercal\boldsymbol{\Sigma}\boldsymbol{w} + (\boldsymbol{w}^\intercal\boldsymbol{\mu})^2 & \boldsymbol{w}^\intercal\boldsymbol{\mu} \\ \hline \boldsymbol{w}^\intercal\boldsymbol{\mu} & \boldsymbol{w}^\intercal\boldsymbol{\mu} & \cdots & \boldsymbol{w}^\intercal\boldsymbol{\mu} & 1 \end{array}\right]$$

The projected problem (15) can be further simplified by exploiting its compound symmetry.

Definition 5 (Compound Symmetry, Votaw (1948)). A matrix $\mathbf{M} \in \mathbb{S}^{T+1}$ is compound symmetric if there exist $\tau_1, \tau_2, \tau_3, \tau_4 \in \mathbb{R}$ with

$$\mathbf{M} = \left[\begin{array}{cccc|c} \tau_1 & \tau_2 & \cdots & \tau_2 & \tau_3 \\ \tau_2 & \tau_1 & \cdots & \tau_2 & \tau_3 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \tau_2 & \tau_2 & \cdots & \tau_1 & \tau_3 \\ \hline \tau_3 & \tau_3 & \ldots & \tau_3 & \tau_4 \end{array}\right]. \tag{17}$$

Note that the second-order moment matrix $\boldsymbol{\Omega}(\boldsymbol{w})$ is compound symmetric because of the temporal symmetry of the random returns. More generally, the second-order moment matrix of any univariate weak sense white noise process is compound symmetric. The next proposition shows that there exists a matrix $\mathbf{M}$ that is both optimal in (16) as well as compound symmetric.

Proposition 3. *There exists a maximizer* $(\mathbf{M}, \beta, \gamma)$ *of (16) with* $\mathbf{M}$ *compound symmetric.*

Proof. Denote by Π^{T+1} the set of all permutations π of the integers $\{1,2,\ldots,T+1\}$ with $\pi(T+1)=T+1$. For any $\pi\in\Pi^{T+1}$ we define the corresponding permutation matrix $\mathbf{P}_\pi\in\mathbb{R}^{(T+1)\times(T+1)}$ through $(\mathbf{P}_\pi)_{ij}=1$ if $\pi(i)=j$; $=0$ otherwise. Note that $\mathbf{P}_\pi^\intercal$ represents the permutation matrix corresponding to the inverse of π. A matrix $\mathbf{K}\in\mathbb{S}^{T+1}$ is compound symmetric if and only if $\mathbf{K}=\mathbf{P}_\pi\mathbf{K}\mathbf{P}_\pi^\intercal$ for all $\pi\in\Pi^{T+1}$. Suppose that $(\mathbf{M},\beta,\gamma)$ is a maximizer of (16). Since the input matrices in (16) are compound symmetric and $\mathbf{P}_\pi$ is non-singular, we have

$$\mathbf{M}-\begin{bmatrix}\frac{1}{2}\mathbb{I} & -\frac{\mathbf{1}}{2}\\ -\left(\frac{\mathbf{1}}{2}\right)^\intercal & \gamma T-\beta\end{bmatrix}\succeq\mathbf{0}\iff\mathbf{P}_\pi\left(\mathbf{M}-\begin{bmatrix}\frac{1}{2}\mathbb{I} & -\frac{\mathbf{1}}{2}\\ -\left(\frac{\mathbf{1}}{2}\right)^\intercal & \gamma T-\beta\end{bmatrix}\right)\mathbf{P}_\pi^\intercal\succeq\mathbf{0}$$

$$\iff\mathbf{P}_\pi\mathbf{M}\mathbf{P}_\pi^\intercal-\begin{bmatrix}\frac{1}{2}\mathbb{I} & -\frac{\mathbf{1}}{2}\\ -\left(\frac{\mathbf{1}}{2}\right)^\intercal & \gamma T-\beta\end{bmatrix}\succeq\mathbf{0}.$$

The compound symmetry of $\boldsymbol{\Omega}(\boldsymbol{w})$ and the cyclicity of the trace further imply

$$\langle\boldsymbol{\Omega}(\boldsymbol{w}),\mathbf{M}\rangle=\mathrm{Tr}\left(\mathbf{M}\boldsymbol{\Omega}(\boldsymbol{w})\right)=\mathrm{Tr}\left(\mathbf{M}\mathbf{P}_\pi^\intercal\boldsymbol{\Omega}(\boldsymbol{w})\mathbf{P}_\pi\right)=\mathrm{Tr}\left(\mathbf{P}_\pi\mathbf{M}\mathbf{P}_\pi^\intercal\boldsymbol{\Omega}(\boldsymbol{w})\right)=\langle\boldsymbol{\Omega}(\boldsymbol{w}),\mathbf{P}_\pi\mathbf{M}\mathbf{P}_\pi^\intercal\rangle.$$

Hence, $(\mathbf{P}_\pi\mathbf{M}\mathbf{P}_\pi^\intercal,\beta,\gamma)$ is feasible in (16) and has the same objective value as $(\mathbf{M},\beta,\gamma)$. It is therefore a maximizer of (16). As the set of maximizers is convex, the convex combination

$$\mathbf{M}'=\frac{1}{T!}\sum_{\pi\in\Pi^{T+1}}\mathbf{P}_\pi\mathbf{M}\mathbf{P}_\pi^\intercal$$

is also a maximizer of (16). Moreover, $\mathbf{M}'$ is compound symmetric because $\rho(\Pi^{T+1})=\Pi^{T+1}$ and, *a fortiori*, $\mathbf{P}_\rho\mathbf{M}'\mathbf{P}_\rho^\intercal=\mathbf{M}'$ for any $\rho\in\Pi^{T+1}$. Thus, the claim follows. ∎

By Proposition 3, we may assume without loss of generality that $\mathbf{M}$ in (16) is compound symmetric. Thus, each matrix inequality in (16) requires a compound symmetric matrix to be positive semidefinite. The next proposition shows that semidefinite constraints involving compound symmetric matrices of any dimension can be reduced to four simple scalar constraints.

PROPOSITION 4. *For any compound symmetric matrix $\mathbf{M}\in\mathbb{S}^{T+1}$ of the form (17), the following equivalence holds.*

$$\mathbf{M}\succeq\mathbf{0}\iff\begin{cases}\tau_1\geq\tau_2 & \text{(18a)}\\ \tau_4\geq 0 & \text{(18b)}\\ \tau_1+(T-1)\tau_2\geq 0 & \text{(18c)}\\ \tau_4\left(\tau_1+(T-1)\tau_2\right)\geq T\tau_3^2 & \text{(18d)}\end{cases}$$

Proof. We use the well-known fact that a symmetric matrix is positive semidefinite if and only if all of its eigenvalues are nonnegative. First, it is easy to verify that any vector of the form $\boldsymbol{v}=[v_1,v_2,\ldots,v_T,0]^\intercal$ with $\sum_{i=1}^T v_i=0$ constitutes an eigenvector of $\mathbf{M}$ with eigenvalue $\tau_1-\tau_2$. Indeed, we have

$$\mathbf{M}\boldsymbol{v}=\begin{bmatrix}\tau_1v_1+\tau_2(v_2+v_3+\cdots+v_T)\\ \tau_1v_2+\tau_2(v_1+v_3+\cdots+v_T)\\ \vdots\\ \tau_1v_T+\tau_2(v_2+v_3+\cdots+v_{T-1})\\ \tau_3(v_1+v_2+\cdots+v_T)\end{bmatrix}=\begin{bmatrix}(\tau_1-\tau_2)v_1\\ (\tau_1-\tau_2)v_2\\ \vdots\\ (\tau_1-\tau_2)v_T\\ 0\end{bmatrix}=(\tau_1-\tau_2)\boldsymbol{v}.$$

There are $T-1$ linearly independent eigenvectors of the above type. Next, we assume first that $\tau_3=0$. In this case, the two remaining eigenvectors can be chosen as $[1,1,\ldots,1,0]^\intercal$ and $[0,0,\ldots,0,1]^\intercal$ with eigenvalues

$\tau_1 + (T-1)\tau_2$ and τ_4, respectively. Thus $\mathbf{M} \succeq \mathbf{0}$ if and only if (18a), (18b), and (18c) hold. Moreover, (18d) is trivially implied by (18b) and (18c) whenever $\tau_3 = 0$. Assume now that $\tau_3 \neq 0$. In this case, the two remaining eigenvectors are representable as $\boldsymbol{v} = [1, 1, \ldots, 1, v]^\intercal$ for some $v \in \mathbb{R}$. Observe that λ is a corresponding eigenvalue if and only if $\mathbf{M}\boldsymbol{v} = \lambda\boldsymbol{v}$, which is equivalent to

$$\tau_1 + (T-1)\tau_2 + v\tau_3 = \lambda, \quad T\tau_3 + v\tau_4 = \lambda v.$$

The second equation above thus implies that $v(\lambda - \tau_4) = T\tau_3 \neq 0$, and thus $v = \frac{T\tau_3}{\lambda - \tau_4}$. Substituting this expression for v into the first equation above, we obtain

$$\tau_1 + (T-1)\tau_2 + \frac{T\tau_3^2}{\lambda - \tau_4} = \lambda.$$

Solving this equation for λ yields the two eigenvalues

$$\lambda = \frac{1}{2}\left(\tau_1 + (T-1)\tau_2 + \tau_4 \pm \sqrt{(\tau_1 + (T-1)\tau_2 + \tau_4)^2 + 4\left(T\tau_3^2 - \tau_4(\tau_1 + (T-1)\tau_2)\right)}\right) \tag{19a}$$

$$= \frac{1}{2}\left(\tau_1 + (T-1)\tau_2 + \tau_4 \pm \sqrt{(\tau_1 + (T-1)\tau_2 - \tau_4)^2 + 4T\tau_3^2}\right). \tag{19b}$$

From equation (19b) it is evident that the square root term constitutes a strictly positive real number. The two eigenvalues are thus nonnegative if and only if

$$\tau_1 + (T-1)\tau_2 + \tau_4 \geq 0, \quad (\tau_1 + (T-1)\tau_2)\,\tau_4 \geq T\tau_3^2. \tag{20}$$

The second inequality in (20) ensures that the square root term in (19a) does not exceed $\tau_1 + (T-1)\tau_2 + \tau_4$, which implies (18d). By (20), both the product and the sum of $\tau_1 + (T-1)\tau_2$ and τ_4 are nonnegative, which implies that each of them must be individually nonnegative, i.e., (18b) and (18c) hold. The claim now follows from the fact that (18b)–(18d) also imply (20). ■

Note that the inequalities (18b), (18c) and (18d) represent a hyperbolic constraint and are therefore second-order cone representable; see Exercise 4.26 of Boyd and Vandenberghe (2004).

We now demonstrate that the semidefinite program (16), which involves $\mathcal{O}(T)$ decision variables, can be reduced to an equivalent non-linear program with only six decision variables. First, by Proposition 3, we may assume without any loss of generality that the decision variable $\mathbf{M}$ in (16) is of the form (17) for some $\boldsymbol{\tau} \in \mathbb{R}^4$. Thus, we can use Proposition 4 to re-express both semidefinite constraints in (16) in terms of one non-linear and three linear constraints, respectively. Using the notational shorthands $\mu_\mathrm{p} = \boldsymbol{w}^\intercal\boldsymbol{\mu}$ and $\sigma_\mathrm{p} = \sqrt{\boldsymbol{w}^\intercal\boldsymbol{\Sigma}\boldsymbol{w}}$ for the mean and the standard deviation of the portfolio return, we obtain the following non-linear program.

$$\begin{aligned}
\mathrm{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w})) \;=\; \max \;& \gamma \\
\text{s.t.}\;& \boldsymbol{\tau} \in \mathbb{R}^4,\ \beta \in \mathbb{R},\ \gamma \in \mathbb{R} \\
& \beta + \tfrac{1}{\epsilon}\Big[T\left(\sigma_\mathrm{p}^2 + \mu_\mathrm{p}^2\right)\tau_1 + T(T-1)\mu_\mathrm{p}^2\tau_2 + 2T\mu_\mathrm{p}\tau_3 + \tau_4\Big] \leq 0 && \text{(21a)} \\
& \tau_1 \geq \tau_2 && \text{(21b)} \\
& \tau_4 \geq 0 && \text{(21c)} \\
& \tau_1 + (T-1)\tau_2 \geq 0 && \text{(21d)}
\end{aligned}$$

$$\tau_4\left(\tau_1 + (T-1)\tau_2\right) \geq T\tau_3^2 \tag{21e}$$

$$\left(\tau_1 - \tfrac{1}{2}\right) \geq \tau_2 \tag{21f}$$

$$\tau_4 - \gamma T + \beta \geq 0 \tag{21g}$$

$$\left(\tau_1 - \tfrac{1}{2}\right) + (T-1)\tau_2 \geq 0 \tag{21h}$$

$$\left(\tau_4 - \gamma T + \beta\right)\left(\left(\tau_1 - \tfrac{1}{2}\right) + (T-1)\tau_2\right) \geq T\left(\tau_3 + \tfrac{1}{2}\right)^2 \tag{21i}$$

Note that (21a) corresponds to the trace inequality, while (21b)–(21e) encodes the positive semidefiniteness of $\mathbf{M}$, and (21f)–(21i) is a reformulation of the last matrix inequality in (16).

We first note that (21a) is binding at optimality. Indeed, if (21a) is not binding at $(\boldsymbol{\tau}, \beta, \gamma)$, then $(\boldsymbol{\tau}, \gamma + \frac{\Delta}{T}, \beta + \Delta)$ remains feasible but has a higher objective value for a sufficiently small $\Delta > 0$. Moreover, (21b) and (21d) are redundant in view of (21f) and (21h) and can thus be dropped. Finally, there exists an optimal solution for which (21f) is binding. Indeed, if (21f) is not binding at $(\boldsymbol{\tau}, \beta, \gamma)$, then $(\frac{\tau_1 + (T-1)\tau_2 - \frac{1}{2}}{T} + \frac{1}{2}, \frac{\tau_1 + (T-1)\tau_2 - \frac{1}{2}}{T}, \tau_3, \tau_4, \gamma, \beta)$ remains feasible with the same objective value but satisfies (21f) as an equality. Without loss of generality, we can thus eliminate the decision variable τ_1 by using the substitution $\tau_1 = \tau_2 + \frac{1}{2}$. In summary, we have

$$\begin{array}{rcll}
\mathrm{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w})) & = & \max & \gamma \\
& & \text{s.t.} & \tau_2 \in \mathbb{R},\ \tau_3 \in \mathbb{R},\ \tau_4 \in \mathbb{R},\ \beta \in \mathbb{R},\ \gamma \in \mathbb{R} \\
& & & \beta + \frac{1}{\epsilon}\left[\frac{T}{2}\left(\sigma_{\mathrm{p}}^2 + \mu_{\mathrm{p}}^2\right) + T\left(\sigma_{\mathrm{p}}^2 + T\mu_{\mathrm{p}}^2\right)\tau_2 + 2T\mu_{\mathrm{p}}\tau_3 + \tau_4\right] = 0 \\
& & & \tau_4 \geq 0 \\
& & & \tau_4 - \gamma T + \beta \geq 0 \\
& & & \tau_2 \geq 0 \\
& & & \tau_4\left(\tau_2 + \frac{1}{2T}\right) \geq \tau_3^2 \\
& & & \left(\tau_4 - \gamma T + \beta\right)\tau_2 \geq \left(\tau_3 + \frac{1}{2}\right)^2 .
\end{array} \tag{22}$$

Problem (22) can be written more compactly as

$$\begin{array}{rcll}
\mathrm{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w})) & = & -\ \min & aw + bx + cy + dz + e \\
& & \text{s.t.} & w, x, y, z \in \mathbb{R} \\
& & & w \geq 0,\ x \geq 0,\ y \geq 1 \\
& & & (\frac{1}{2}z + 1)^2 \leq w(y+1) \\
& & & (\frac{1}{2}z - 1)^2 \leq x(y-1),
\end{array} \tag{23}$$

where the decision variables w, x, y and z in (23) are related to the variables τ_2, τ_3, τ_4, β and γ in (22) through the transformations

$$w = \frac{4\tau_4}{T}, \quad x = \frac{4\left(\tau_4 - \gamma T + \beta\right)}{T}, \quad y = 4T\tau_2 + 1, \quad z = -8\tau_3 - 2,$$

while the objective function coefficients are given by $a = \frac{1}{4\epsilon} - \frac{1}{4}$, $b = \frac{1}{4}$, $c = \frac{\sigma_\text{p}^2 + T\mu_\text{p}^2}{4\epsilon T}$, $d = -\frac{\mu_\text{p}}{4\epsilon}$ and $e = \frac{\mu_\text{p}^2}{4\epsilon} - \frac{\mu_\text{p}}{2\epsilon} + \frac{\sigma_\text{p}^2}{2\epsilon} - \frac{\sigma_\text{p}^2}{4\epsilon T}$. Note that the inequality constraints in (22) correspond to the inequality constraints in (23) in the same order, while the equality constraint in (22) has been eliminated via the following substitution.

$$\begin{aligned}
\gamma &= \tfrac{1}{4}(w-x) + \tfrac{\beta}{T} \\
&= \tfrac{1}{4}(w-x) - \tfrac{1}{\epsilon T}\left(\tfrac{T}{2}\left(\sigma_\text{p}^2 + \mu_\text{p}^2\right) + T\left(\sigma_\text{p}^2 + T\mu_\text{p}^2\right)\tau_2 + 2T\mu_\text{p}\tau_3 + \tau_4\right) \\
&= \tfrac{1}{4}(w-x) - \tfrac{1}{\epsilon T}\left(\tfrac{T}{2}\left(\sigma_\text{p}^2 + \mu_\text{p}^2\right) + \tfrac{1}{4}\left(\sigma_\text{p}^2 + T\mu_\text{p}^2\right)(y-1) - \tfrac{1}{4}T\mu_\text{p}(z+2) + \tfrac{1}{4}Tw\right) \\
&= -\left(\left(\tfrac{1}{4\epsilon} - \tfrac{1}{4}\right)w + \left(\tfrac{1}{4}\right)x + \left(\tfrac{\sigma_\text{p}^2 + T\mu_\text{p}^2}{4\epsilon T}\right)y + \left(-\tfrac{\mu_\text{p}}{4\epsilon}\right)z + \tfrac{\mu_\text{p}^2}{4\epsilon} - \tfrac{\mu_\text{p}}{2\epsilon} + \tfrac{\sigma_\text{p}^2}{2\epsilon} - \tfrac{\sigma_\text{p}^2}{4\epsilon T}\right) \\
&= -(aw + bx + cy + dz + e)
\end{aligned}$$

Problem (23) admits an explicit analytical solution as stated in the following lemma.

LEMMA 1. *For any given real numbers a, b, c and d that satisfy the conditions*

(i) $a, b, c > 0$,

(ii) $(a+b)c > d^2$ *and*

(iii) $a + b + d > \Delta\sqrt{b/a}$, *where* $\Delta = \sqrt{(a+b)c - d^2} > 0$,

the optimal value of the optimization problem

$$\begin{array}{ll}
\min & aw + bx + cy + dz \\
\text{s. t.} & w, x, y, z \in \mathbb{R} \\
& w \geq 0,\ x \geq 0,\ y \geq 1 \\
& (\tfrac{1}{2}z + 1)^2 \leq w(y+1) \\
& (\tfrac{1}{2}z - 1)^2 \leq x(y-1)
\end{array} \tag{24}$$

is given by

$$\frac{2bd + d^2 + \Delta^2 + 2\Delta\sqrt{ab}}{a+b} + \frac{2\sqrt{ab}}{(a+b)^2}\left(\Delta - d\sqrt{a/b}\right)\left(a + b + d - \Delta\sqrt{b/a}\right).$$

Proof. Assumption (i) ensures that problem (24) is bounded, while assumption (ii) guarantees that $\Delta = \sqrt{(a+b)c - d^2}$ is real. Assumption (iii) is not strictly needed, and problem (24) admits a *generalized* closed-form solution even if this assumption is violated. However, this more general solution is not needed for this paper, and therefore we will not derive it.

Note that (24) constitutes a (convex) SOCP with two hyperbolic constraints, and the Karush-Kuhn-Tucker (KKT) optimality conditions are necessary *and* sufficient. We will now prove the lemma constructively by showing that the candidate solution

$$y = \frac{2-p-q}{p-q}, \quad z = \frac{2(p+q-2pq)}{p-q}, \quad w = \frac{\left(\frac{1}{2}z+1\right)^2}{y+1}, \quad x = \frac{\left(\frac{1}{2}z-1\right)^2}{y-1}$$

with

$$p = \frac{-d + \Delta\sqrt{b/a}}{a+b}, \quad q = \frac{-d - \Delta\sqrt{a/b}}{a+b}$$

satisfies the KKT conditions and is thus optimal in (24). Note first that this solution is feasible. Indeed, by the assumptions (i)–(iii) we have $q < p < 1$. We conclude that

$$y = \frac{2-p-q}{p-q} = 1 + \frac{2(1-p)}{p-q} > 1,$$

which in turn implies that $w \geq 0$ and $x \geq 0$. The two hyperbolic constraints in (24) are binding by the definition of w and x.

For later reference we state the following identities, which are easy to verify.

$$ap + bq + d = 0, \quad ap^2 + bq^2 = c, \quad p = \frac{z+2}{2y+2}, \quad q = \frac{z-2}{2y-2} \tag{25}$$

Moreover, we denote by α_1, α_2 and α_3 the Lagrange multipliers of the three linear inequalities and by λ and δ the Lagrange multipliers of the two hyperbolic constraints in (24), respectively. To prove that the suggested candidate solution is indeed optimal, we show that it satisfies the KKT conditions with $\alpha_1 = \alpha_2 = \alpha_3 = 0$, $\lambda = \frac{a}{y+1} > 0$ and $\delta = \frac{b}{y-1} > 0$. Note that these Lagrange multipliers are dual feasible and satisfy complementary slackness. By using (25) together with the explicit formulas for the candidate solution and the Lagrange multipliers, we can further verify the stationarity conditions:

$$\begin{aligned} a - \alpha_1 - \lambda(y+1) &= 0 \\ b - \alpha_2 - \delta(y-1) &= 0 \\ c - \alpha_3 - \lambda w - \delta x = c - ap^2 - bq^2 &= 0 \\ d + \tfrac{1}{2}\lambda(z+2) + \tfrac{1}{2}\delta(z-2) = d + ap + bq &= 0 \end{aligned}$$

As all KKT conditions are met, we conclude that the proposed candidate solution is optimal. In order to evaluate the optimal objective value of problem (24), we first use (25) to show that $w = p(\frac{1}{2}z + 1)$ and $x = q(\frac{1}{2}z - 1)$. This enables us to express the optimal objective value as

$$\begin{aligned} aw + bx + cy + dz &= ap\left(\tfrac{1}{2}z + 1\right) + bq\left(\tfrac{1}{2}z - 1\right) + cy + dz \\ &= ap - bq + cy + \tfrac{1}{2}z\,(2d + ap + bq) \\ &= ap - bq + cy + \tfrac{1}{2}dz, \end{aligned}$$

where the last equality follows again from (25). As y and z are defined in terms of p and q, we can now express the optimal objective value as a function of p and q only. The claim then follows by substituting the definitions of p and q into the resulting formula. ∎

We are now ready to prove Theorem 2.

Proof of Theorem 2. We know that the worst-case VaR of $\tilde{\gamma}'_T(\boldsymbol{w})$ is given by the optimal value of problem (23). By construction, the objective function coefficients a, b and c in (23) are strictly positive. Moreover, the square root discriminant $\Delta = \sqrt{(a+b)c - d^2} = \frac{\sigma_\mathrm{p}}{4\epsilon\sqrt{T}}$ is strictly positive by Assumption (A1), while Assumption (A2) implies that

$$a + b + d = \frac{1}{4\epsilon}(1 - \mu_\mathrm{p}) > \frac{1}{4\epsilon}\sqrt{\frac{\epsilon}{(1-\epsilon)T}}\sigma_\mathrm{p} = \sqrt{b/a}\Delta.$$

As all conditions of Lemma 1 are satisfied, we may conclude that

$$\mathrm{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w})) = -\left(\tfrac{2bd + d^2 + \Delta^2 + 2\Delta\sqrt{ab}}{a+b} + \tfrac{2\sqrt{ab}}{(a+b)^2}\left(\Delta - d\sqrt{a/b}\right)\left(a + b + d - \Delta\sqrt{b/a}\right) + e\right).$$

The claim then follows by substituting the definitions of a, b, c, d and e into the above expression and rearranging terms. ∎

Appendix C: Proof of Theorem 3

We first establish some identities for the parameters Δ, b, u and d defined on page 17 that will be useful for the proof of the two assertions. From the definition of d we conclude that

$$(T-1)u^2 + (u+\Delta)^2 = (T-1)d^2 + (d-\Delta)^2, \tag{26}$$

while the definitions of b and u imply that

$$(1-\epsilon')b + \epsilon'\left(u + \frac{\Delta}{T}\right) = \mu_{\rm p}. \tag{27}$$

For later reference we define

$$\gamma = \frac{1}{2}\left(1 - \left(1 - \mu_{\rm p} + \sqrt{\frac{1-\epsilon'}{\epsilon' T}}\sigma_{\rm p}\right)^2 - \frac{T-1}{\epsilon' T}\sigma_{\rm p}^2\right). \tag{28}$$

By construction, γ is equal to the worst-case VaR of $\tilde{\gamma}'_T(\boldsymbol{w})$ at the tolerance level ϵ'; see Theorem 2. Basic algebraic manipulations yield the following equation equivalent to (28).

$$(1-\epsilon') - \left(2\mu_{\rm p} - \mu_{\rm p}^2 - \sigma_{\rm p}^2 - 2\gamma\epsilon'\right) = \left(\sqrt{(1-\epsilon')}(1-\mu_{\rm p}) - \sqrt{\frac{\epsilon'}{T}}\sigma_{\rm p}\right)^2$$

By using the definition of b, this equation can be reformulated as

$$b^2 - 2b + \frac{2\mu_{\rm p} - \mu_{\rm p}^2 - \sigma_{\rm p}^2 - 2\gamma\epsilon'}{1-\epsilon'} = 0. \tag{29}$$

Similarly, from the definition of u we obtain

$$\begin{aligned}
u &= \mu_{\rm p} - \frac{\Delta}{T} - \sqrt{\frac{1-\epsilon'}{\epsilon' T}}\sigma_{\rm p} \\
&= 1 - \frac{\Delta}{T} - \sqrt{1 - 2\gamma - \frac{T-1}{\epsilon' T}\sigma_{\rm p}^2} \\
&= 1 - \frac{\Delta}{T} - \sqrt{1 - 2\gamma + \frac{\Delta^2}{T^2} - \frac{\Delta^2}{T}} \\
&= \frac{2(T-\Delta) - \sqrt{4(T-\Delta)^2 - 4T\left(2T\gamma - 2\Delta + \Delta^2\right)}}{2T},
\end{aligned}$$

where the second equality uses (28) and the third equality the definition of Δ. Therefore, u can be viewed a root of the quadratic equation

$$Tu^2 + 2u(\Delta - T) + 2T\gamma - 2\Delta + \Delta^2 = 0. \tag{30}$$

We are now ready to prove assertion (i). We will show that the distributions $\mathbb{P}^{\epsilon'}$, $\epsilon' \in (\epsilon, 1)$, satisfy the moment conditions in the definition of $\mathcal{P}_{\tilde{\boldsymbol{\eta}}}(\boldsymbol{w})$ on page 36. By construction, it is clear that $\mathbb{P}^{\epsilon'}$ is indeed a probability distribution. As for the first order moment conditions, we observe that

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}^{\epsilon'}}(\tilde{\eta}_t) &= (1-\epsilon')\,b + \frac{\epsilon'}{2T}(Tu + \Delta) + \frac{\epsilon'}{2T}(Td - \Delta) \\
&= (1-\epsilon')\,b + \frac{\epsilon'}{T}(Tu + \Delta) \;=\; \mu_{\rm p} \quad \forall t = 1, \ldots, T,
\end{aligned}$$

where the second equality follows from the definition of d, while the third equality exploits (27). As for the second order moment conditions, we have

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}^{\epsilon'}}\left(\tilde{\eta}_t^2\right) &= (1-\epsilon')\,b^2 + \frac{\epsilon'}{2T}\left((T-1)\,u^2 + (u+\Delta)^2\right) + \frac{\epsilon'}{2T}\left((T-1)\,d^2 + (d-\Delta)^2\right) \\
&= (1-\epsilon')\,b^2 + \frac{\epsilon'}{T}\left((T-1)u^2 + (u+\Delta)^2\right) \\
&= (1-\epsilon')\,b^2 + \frac{\epsilon'}{T}\left(Tu^2 + 2u\Delta + \Delta^2\right) \\
&= (1-\epsilon')\left(2b - \frac{2\mu_{\rm p} - \mu_{\rm p}^2 - \sigma_{\rm p}^2 - 2\gamma\epsilon'}{1-\epsilon'}\right) + \frac{2\epsilon'}{T}\left(uT - T\gamma + \Delta\right) \\
&= \mu_{\rm p}^2 + \sigma_{\rm p}^2 - 2\left(\mu_{\rm p} - (1-\epsilon')b - \epsilon' u - \frac{\epsilon'\Delta}{T}\right) \;=\; \mu_{\rm p}^2 + \sigma_{\rm p}^2,
\end{aligned} \tag{31}$$

where the second equality follows from (26), the fourth equality exploits (29) to re-express b^2 and (30) to re-express $Tu^2 + 2u\Delta + \Delta^2$, and the last equality holds due to (27). Similarly, we find

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}^{\epsilon'}}\left(\tilde{\eta}_s\tilde{\eta}_t\right) &= (1-\epsilon')\,b^2 + \frac{\epsilon'}{2T}\left((T-2)\,u^2 + 2u\,(u+\Delta)\right) + \frac{\epsilon'}{2T}\left((T-2)\,d^2 + 2d\,(d-\Delta)\right) \\
&= (1-\epsilon')\,b^2 + \frac{\epsilon'}{2T}\left((T-1)\left(u^2+d^2\right) + (u+\Delta)^2 + (d-\Delta)^2\right) - \frac{\epsilon'}{T}\Delta^2 \;=\; \mu_{\rm p}^2
\end{aligned}$$

for $s \neq t$. A comparison with (31) shows that the first two terms in the second line of the above expression are equal to $\mu_{\rm p}^2 + \sigma_{\rm p}^2$. The third equality then follows from the definition of Δ. Thus, $\mathbb{P}^{\epsilon'} \in \mathcal{P}_{\tilde{\boldsymbol{\eta}}}(\boldsymbol{w})$.

To prove assertion (ii), we first evaluate the distribution of the (quadratic approximation of the) uncertain portfolio growth rate $\tilde{\gamma}_T^\eta$ under $\mathbb{P}^{\epsilon'}$. Indeed, $\tilde{\gamma}_T^\eta$ will adopt only one of two different possible values depending on whether the realization of $\tilde{\boldsymbol{\eta}}$ is equal to $\boldsymbol{\eta}^b$ or any of the other scenarios ($\boldsymbol{\eta}_t^u$ or $\boldsymbol{\eta}_t^d$ for any $t = 1,\ldots,T$), respectively. If $\tilde{\boldsymbol{\eta}} = \boldsymbol{\eta}^b$, it is easy to verify that $\tilde{\gamma}_T^\eta = b - \frac{1}{2}b^2$. On the other hand, if $\tilde{\boldsymbol{\eta}} = \boldsymbol{\eta}_t^u$ for any $t = 1,\ldots,T$, then

$$\begin{aligned}
\tilde{\gamma}_T^\eta &= \frac{1}{T}\left((T-1)\left(u - \frac{1}{2}u^2\right) + (u+\Delta) - \frac{1}{2}(u+\Delta)^2\right) \\
&= \frac{1}{2T}\left(-Tu^2 - 2u\,(\Delta - T) + 2\Delta - \Delta^2\right) = \gamma,
\end{aligned}$$

where the second equality follows from basic manipulations, and the third equality holds due to (30). A similar calculation shows that $\tilde{\gamma}_T^\eta$ is also equal to γ if $\tilde{\boldsymbol{\eta}} = \boldsymbol{\eta}_t^d$ for any $t = 1,\ldots,T$. Details are omitted for brevity. Next, we will demonstrate that $b - \frac{1}{2}b^2 > \gamma$, that is, the growth rate $\tilde{\gamma}_T^\eta$ adopts its largest value in scenario $\tilde{\boldsymbol{\eta}} = \boldsymbol{\eta}^b$. To this end, we observe that

$$\begin{aligned}
\gamma &= \frac{1}{2}\left(1 - \left(1 - \mu_{\rm p} + \sqrt{\frac{1-\epsilon'}{\epsilon' T}}\sigma_{\rm p}\right)^2 - \frac{T-1}{\epsilon' T}\sigma_{\rm p}^2\right) \\
&= \frac{1}{2}\left(1 - (1-\mu_{\rm p})^2 - 2\sqrt{\frac{1-\epsilon'}{\epsilon' T}}\,(1-\mu_{\rm p})\,\sigma_{\rm p} - \frac{T-\epsilon'}{\epsilon' T}\sigma_{\rm p}^2\right) \\
&< \frac{1}{2}\left(1 - (1-\mu_{\rm p})^2 - \frac{2}{T}\sigma_{\rm p}^2 - \frac{T-\epsilon'}{\epsilon' T}\sigma_{\rm p}^2\right) \\
&< \frac{1}{2}\left(1 - (1-\mu_{\rm p})^2 - \sigma_{\rm p}^2\right),
\end{aligned}$$

where the first inequality holds because of Assumption (A2) and because $\epsilon' \in (\epsilon, 1)$. Thus,

$$\frac{1}{2}\left(1 - (1-\mu_{\rm p})^2 - \sigma_{\rm p}^2\right) > \gamma \iff \frac{2\mu_{\rm p} - \mu_{\rm p}^2 - \sigma_{\rm p}^2 - 2\gamma\epsilon'}{2(1-\epsilon')} > \gamma \iff b - \frac{1}{2}b^2 > \gamma\,,$$

where the first equivalence follows from basic algebraic manipulations, while the second equivalence is due to (29). In summary, we have shown that

$$\mathbb{P}^{\epsilon'}(\tilde{\gamma}_T^\eta = \gamma) = \mathbb{P}^{\epsilon'}(\tilde{\boldsymbol{\eta}} \neq \boldsymbol{\eta}^b) = \epsilon' \quad \text{and} \quad \mathbb{P}^{\epsilon'}(\tilde{\gamma}_T^\eta > \gamma) = \mathbb{P}^{\epsilon'}(\tilde{\boldsymbol{\eta}} = \boldsymbol{\eta}^b) = 1 - \epsilon'$$

for all $\epsilon' \in (\epsilon, 1)$. Together with the continuity of γ as a function of ϵ', which follows from the definition of γ in (28), we may thus conclude that

$$\lim_{\epsilon' \downarrow \epsilon} \mathbb{P}^{\epsilon'}\text{-VaR}_\epsilon(\tilde{\gamma}_T^\eta) = \text{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w}))\,.$$

This observation completes the proof. ∎

Appendix D: Proof of Theorem 4

The proof of Theorem 4 parallels that of Theorem 2 and relies on three auxiliary propositions. We will first show that the matrix $\mathbf{M}$ in (9) may be assumed to be block compound symmetric in the sense of the following definition.

Definition 6 (Block Compound Symmetry). A matrix $\mathbf{M} \in \mathbb{S}^{nT+1}$ is block compound symmetric with blocks of size $n \times n$ if it is representable as

$$\mathbf{M} = \left[\begin{array}{cccc|c} \mathbf{A} & \mathbf{B} & \cdots & \mathbf{B} & \boldsymbol{c} \\ \mathbf{B} & \mathbf{A} & \cdots & \mathbf{B} & \boldsymbol{c} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{B} & \mathbf{B} & \ldots & \mathbf{A} & \boldsymbol{c} \\ \hline \boldsymbol{c}^\intercal & \boldsymbol{c}^\intercal & \cdots & \boldsymbol{c}^\intercal & d \end{array}\right] \tag{32}$$

for some $\mathbf{A} \in \mathbb{S}^n$, $\mathbf{B} \in \mathbb{S}^n$, $\boldsymbol{c} \in \mathbb{R}^n$ and $d \in \mathbb{R}$.

Formally, we can prove the following result.

Proposition 5. *There exists a maximizer* $(\mathbf{M}, \alpha, \beta, \gamma, \lambda)$ *of (9) where* $\mathbf{M}$ *is block compound symmetric with blocks of size* $n \times n$.

Proof. The proof widely parallels that of Proposition 3 in Appendix B and is thus omitted. ∎

The next two propositions demonstrate that the positive semidefiniteness of a block compound symmetric matrix $\mathbf{M} \in \mathbb{R}^{nT+1}$ can be enforced by two linear matrix inequalities of dimensions only n and $n+1$, respectively.

Proposition 6. *For any matrix* $\mathbf{K} \in \mathbb{S}^{nT}$ *of the form*

$$\mathbf{K} = \begin{bmatrix} \mathbf{A} & \mathbf{B} & \cdots & \mathbf{B} \\ \mathbf{B} & \mathbf{A} & \cdots & \mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B} & \mathbf{B} & \cdots & \mathbf{A} \end{bmatrix} \tag{33}$$

for some $\mathbf{A}, \mathbf{B} \in \mathbb{S}^n$ *we have that* $\text{eig}(\mathbf{K}) = \text{eig}\,(\mathbf{A} - \mathbf{B}) \cup \text{eig}(\mathbf{A} + (T-1)\mathbf{B})$.

Proof. We prove this proposition constructively by determining all eigenvalues as well as the corresponding eigenvectors of $\mathbf{K}$. Let $\{(\boldsymbol{v}_i, \lambda_i)\}_{i=1}^n$ denote all n pairs of eigenvectors and eigenvalues of the matrix $\mathbf{A} + (T-1)\mathbf{B}$. For any $i = 1, \ldots, n$, we have

$$\mathbf{K} \begin{bmatrix} \boldsymbol{v}_i \\ \boldsymbol{v}_i \\ \vdots \\ \boldsymbol{v}_i \end{bmatrix} = \begin{bmatrix} (\mathbf{A} + (T-1)\mathbf{B})\,\boldsymbol{v}_i \\ (\mathbf{A} + (T-1)\mathbf{B})\,\boldsymbol{v}_i \\ \vdots \\ (\mathbf{A} + (T-1)\mathbf{B})\,\boldsymbol{v}_i \end{bmatrix} = \lambda_i \begin{bmatrix} \boldsymbol{v}_i \\ \boldsymbol{v}_i \\ \vdots \\ \boldsymbol{v}_i \end{bmatrix},$$

which implies that $[\boldsymbol{v}_i^\intercal, \boldsymbol{v}_i^\intercal, \ldots, \boldsymbol{v}_i^\intercal]^\intercal$ is an eigenvector of $\mathbf{K}$ with eigenvalue λ_i. Next, denote by $\{(\boldsymbol{u}_i, \theta_i)\}_{i=1}^n$ the n pairs of eigenvectors and eigenvalues of the matrix $\mathbf{A}-\mathbf{B}$. For any $i=1,\ldots,n$ and for any $\boldsymbol{k}\in\mathbb{R}^T$ with $\mathbf{1}^\intercal\boldsymbol{k}=0$ we have

$$\mathbf{K}\begin{bmatrix} k_1\boldsymbol{u}_i \\ k_2\boldsymbol{u}_i \\ \vdots \\ k_T\boldsymbol{u}_i \end{bmatrix} = \begin{bmatrix} k_1(\mathbf{A}-\mathbf{B})\boldsymbol{u}_i \\ k_2(\mathbf{A}-\mathbf{B})\boldsymbol{u}_i \\ \vdots \\ k_T(\mathbf{A}-\mathbf{B})\boldsymbol{u}_i \end{bmatrix} = \theta_i \begin{bmatrix} k_1\boldsymbol{u}_i \\ k_2\boldsymbol{u}_i \\ \vdots \\ k_T\boldsymbol{u}_i \end{bmatrix}.$$

Thus, $[k_1\boldsymbol{u}_i^\intercal, k_2\boldsymbol{u}_i^\intercal, \ldots, k_T\boldsymbol{u}_i^\intercal]^\intercal$ is an eigenvector of $\mathbf{K}$ with eigenvalue θ_i. Hence, there are $T-1$ linearly independent eigenvectors that share the same eigenvalue θ_i. In summary, we have found all $n+n(T-1)=nT$ eigenvalues of $\mathbf{K}$ counted by their multiplicities, and we may thus conclude that $\text{eig}(\mathbf{K}) = \text{eig}(\mathbf{A}-\mathbf{B}) \cup \text{eig}(\mathbf{A}+(T-1)\mathbf{B})$. ∎

PROPOSITION 7. *For any block compound symmetric matrix $\mathbf{M}\in\mathbb{S}^{nT+1}$ of the form (32), the following equivalence holds.*

$$\mathbf{M}\succeq\mathbf{0} \iff \begin{cases} \mathbf{A}\succeq\mathbf{B} \\ \begin{bmatrix} \mathbf{A}+(T-1)\mathbf{B} & \boldsymbol{c} \\ \boldsymbol{c}^\intercal & \frac{d}{T} \end{bmatrix} \succeq \mathbf{0} \end{cases}$$

Proof. For ease of exposition, we set

$$\mathbf{M} = \begin{bmatrix} \mathbf{K} & \boldsymbol{c}' \\ \boldsymbol{c}'^\intercal & d \end{bmatrix},$$

where $\boldsymbol{c}' = [\boldsymbol{c}^\intercal, \boldsymbol{c}^\intercal, \ldots, \boldsymbol{c}^\intercal]^\intercal$, and $\mathbf{K}$ is the block matrix defined in (33). Assume first that $d=0$. Then, $\mathbf{M}\succeq\mathbf{0}$ if and only if $\boldsymbol{c}=\mathbf{0}$ and $\mathbf{K}\succeq\mathbf{0}$, which in turn is equivalent to $\boldsymbol{c}=\mathbf{0}$, $\mathbf{A}\succeq\mathbf{B}$, and $\mathbf{A}+(T-1)\mathbf{B}\succeq\mathbf{0}$; see Proposition 6. Thus, the claim follows. Assume next that $d\neq 0$. Then,

$$\begin{aligned}
\mathbf{M}\succeq\mathbf{0} &\iff d>0,\ \mathbf{K}\succeq \frac{1}{d}\boldsymbol{c}'\boldsymbol{c}'^\intercal \\
&\iff d>0,\ \begin{bmatrix} \mathbf{A}-\boldsymbol{c}\boldsymbol{c}^\intercal/d & \mathbf{B}-\boldsymbol{c}\boldsymbol{c}^\intercal/d & \ldots & \mathbf{B}-\boldsymbol{c}\boldsymbol{c}^\intercal/d \\ \mathbf{B}-\boldsymbol{c}\boldsymbol{c}^\intercal/d & \mathbf{A}-\boldsymbol{c}\boldsymbol{c}^\intercal/d & \ldots & \mathbf{B}-\boldsymbol{c}\boldsymbol{c}^\intercal/d \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B}-\boldsymbol{c}\boldsymbol{c}^\intercal/d & \mathbf{B}-\boldsymbol{c}\boldsymbol{c}^\intercal/d & \ldots & \mathbf{A}-\boldsymbol{c}\boldsymbol{c}^\intercal/d \end{bmatrix} \succeq \mathbf{0} \\
&\iff d>0,\ \mathbf{A}\succeq\mathbf{B},\ \mathbf{A}+(T-1)\mathbf{B}\succeq \frac{T}{d}\boldsymbol{c}\boldsymbol{c}^\intercal \\
&\iff \mathbf{A}\succeq\mathbf{B},\ \begin{bmatrix} \mathbf{A}+(T-1)\mathbf{B} & \boldsymbol{c} \\ \boldsymbol{c}^\intercal & \frac{d}{T} \end{bmatrix} \succeq \mathbf{0},
\end{aligned}$$

where the first and the last equivalences follow from standard Schur complement arguments, while the third equivalence holds due to Proposition 6. Thus, the claim follows again. ∎

We are now ready to prove Theorem 4.

Proof of Theorem 4. By Proposition 5, we may assume without any loss of generality that the decision variable $\mathbf{M}$ in (9) is of the form (32), where $\mathbf{A}\in\mathbb{S}^n$, $\mathbf{B}\in\mathbb{S}^n$, $\boldsymbol{c}\in\mathbb{R}^n$ and $d\in\mathbb{R}$ represent new auxiliary decision variables to be used instead of $\mathbf{M}$. Proposition 7 then allows us to re-express each $(nT+1)$-dimensional SDP constraint in (9) in terms of two linear matrix inequalities of dimensions n and $n+1$. Using a standard Schur complement argument to linearize all terms quadratic in $\boldsymbol{w}$, we may conclude that the SDPs (9) and (10) are equivalent. Thus, Theorem 4 follows. ∎