# Dimensionality Reduction in Dynamic Optimization under Uncertainty

THÈSE N$^{O}$ 7235 (2016)

PRÉSENTÉE LE 5 DÉCEMBRE 2016

AU COLLÈGE DU MANAGEMENT DE LA TECHNOLOGIE

CHAIRE ANALYSE DE RISQUE ET OPTIMISATION

PROGRAMME DOCTORAL EN MANAGEMENT DE LA TECHNOLOGIE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Napat RUJEERAPAIBOON

acceptée sur proposition du jury:

Prof. C. Tucci, président du jury
Prof. D. Kuhn, Prof. W. Wiesemann, directeurs de thèse
Prof. D. den Hertog, rapporteur
Prof. P. Goulart, rapporteur
Prof. D. Filipovic, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2016

*Dedicated to my parents*

*whom I never thank enough*

# Acknowledgements

First and foremost, I would like to express my deepest gratitude and highest appreciation towards my supervisor and my thesis director, Dr. Daniel Kuhn. The guidance that he has given me since the beginning of my master studies until the very last moments of my doctoral studies is truly irreplaceable. He has given me not only intellectual support but also provided me with invaluable opportunities to develop as a scholar.

I am also immensely indebted to Dr. Wolfram Wiesemann whom I have personally known and admired since my days in London, United Kingdom. His research ideas, his attention to details, and his accessibility enabled me to perform research efficiently, and I have greatly improved through working alongside him in terms of both technical skills and analytical thinking. Dr. Daniel Kuhn and Dr. Wolfram Wiesemann tremendously enhanced the quality of my work, especially this dissertation. For the success of my studies, I am mostly grateful to the two of them.

My sincere thanks go to Dr. Ruiwei Jiang, Dr. Georg Ostermaier, and Byung-Geun Choi for agreeing to collaborate with me on different research topics. I appreciate their intellectual comments as well as their industrial input and insights which contribute to the betterment of our projects as a whole.

My heartfelt thanks also go to my office mates and my other colleagues, to name a few, Dr. Grani Hanasusanto, Dr. Peyman Mohajerin Esfahani, Dr. Vladimir Roitch, Çağil Koçyiğit, Soroosh Shafieezadeh Abadeh, Mervegül Kirci, and Viet Anh Nguyen. I may have been quiet at times, but I truly enjoyed our lunchtime discussions, and I want to personally thank them for the laughters we shared and for the adventures we went through together.

In addition to those mentioned above, I have been constantly in touch with a few friends from my native country, Thailand. They brought me plenty of joy, and assisted me whenever I encountered any kind of hardships. Therefore, I would like to thank them, in particular, Awapa Prombandankul, Waranya Yongwongsoontorn, Varunya Thavornun, Jirawan Siramanont, and Pakeerat Chuprayoon.

I am also grateful to Alexia Godefroid, Stéphane Yong Wai Man and Pattara Rujeerapaiboon for their linguistic comments and suggestions.

Last but not least, I would like to also extend my gratitude to my precious friend, Sanprapha Vudhivorn, who has been my emotional and moral mainstay from afar for the last four years. Above all, however, I would like to thank my parents, Wiboon & Valapa Rujeerapaiboon, and my brother who have been with me throughout my life. If it were not for the belief that they have in me, I would not succeed in this highest level of education; in any case, I would not have had the courage to pursue my dream as an academic to begin with. For this and for their unconditional love, I will never thank them enough.

*Lausanne, 9 June 2016* Napat Rujeerapaiboon

# Abstract

Dynamic optimization problems affected by uncertainty are ubiquitous in many application domains. Decision makers typically model the uncertainty through random variables governed by a probability distribution. If the distribution is precisely known, then the emerging optimization problems constitute stochastic programs or chance constrained programs. On the other hand, if the distribution is at least partially unknown, then the emanating optimization problems represent robust or distributionally robust optimization problems. In this thesis, we leverage techniques from stochastic and distributionally robust optimization to address complex problems in finance, energy systems management and, more abstractly, applied probability. In particular, we seek to solve uncertain optimization problems where the prior distributional information includes only the first and the second moments (and, sometimes, the support).

The main objective of the thesis is to solve large instances of practical optimization problems. For this purpose, we develop complexity reduction and decomposition schemes, which exploit structural symmetries or multiscale properties of the problems at hand in order to break them down into smaller and more tractable components.

In the first part of the thesis we study the growth-optimal portfolio, which maximizes the expected log-utility over a single investment period. In a classical stochastic setting, this portfolio is known to outperform any other portfolio with probability 1 in the long run. In the short run, however, it is notoriously volatile. Moreover, its performance suffers in the presence of distributional ambiguity. We design fixed-mix strategies that offer similar performance guarantees as the classical growth-optimal portfolio but for a finite investment horizon. Moreover, the proposed performance guarantee remains valid for any asset return distribution with the same mean and covariance matrix. These results rely on a Taylor approximation of the terminal logarithmic wealth that becomes more accurate as the rebalancing frequency is increased.

In the second part of the thesis, we demonstrate that such a Taylor approximation is in fact not necessary. Specifically, we derive sharp probability bounds on the tails of a product of non-negative random variables. These generalized Chebyshev bounds can be computed numerically using semidefinite programming—in some cases even analytically. Similar techniques can also be used to derive multivariate Chebyshev bounds for sums, maxima, and minima of random variables.

In the final part of the thesis, we consider a multi-market reservoir management problem. The eroding peak/off-peak spreads on European electricity spot markets imply reduced profitability for the hydropower producers and force them to participate in the balancing markets. This motivates us to propose a two-layer stochastic programming model for the optimal operation of a cascade of hydropower plants selling energy on both spot and balancing markets. The planning problem optimizes the reservoir management over a yearly horizon with weekly granularity, and the trading subproblems optimize the market transactions over a weekly horizon with hourly granularity. We solve both the planning and trading problems in linear decision rules, and we exploit the inherent parallelizability of the trading subproblems to achieve computational tractability.

**Keywords.** Convex optimization, conic programming, distributionally robust optimization, stochastic programming, linear decision rules, portfolio optimization, growth-optimal portfolio, value-at-risk, Chebyshev inequality, electricity market

# Résumé


Les problèmes d'optimisation dynamiques affectés par des incertitudes sont omniprésents dans de nombreux domaines d'application. Les preneurs de décisions modélisent l'incertitude au travers de variables aléatoires gouvernées par une distribution de probabilité. Si la distribution est définie de manière précise, les problèmes d'optimisation émergents constituent des programmes stochastiques. D'autre part, si la distribution est partiellement indéfinie, les problèmes d'optimisation émergents représentent une optimisation robuste ou des problèmes d'optimisation distributionnellement robuste. Cette thèse est essentiellement constituée de travaux basés sur l'optimisation stochastique et sur l'optimisation distributionnellement robuste afin d'apporter une solution aux problèmes complexes dans le domaine financier, aux systèmes de gestion de l'énergie et, sur un plan abstrait, aux probabilités appliquées.

L'objectif principal de la thèse est de résoudre de grande instance d'optimisation de problème pratique. A cet effet, nous développons des schémas pour réduire et décomposer la complexité, qui exploitent des symétries structurelles ou des propriétés multi-échelles de problèmes existants, dans le but de les diviser en composants plus petits et traitables.

Dans la première partie de cette thèse, nous étudions la croissance optimale du portefeuille. Dans une configuration stochastique classique, ce portefeuille est connu pour être supérieur à tout autre portefeuille avec une probabilité 1 lors d'une exécution à long terme. Toutefois, lors d'une exécution à court terme, il est connu pour sa volatilité. De plus, sa performance diminue s'il y a du bruit dans la probabilité de distribution du rendement de l'actif. Nous concevons des stratégies de portefeuille rééquilibre constant qui offrent des garanties de performance similaire au classique croissance-optimale du portefeuille, mais pour un horizon d'investissement fini. De plus, la garantie de performance proposée reste valide pour toute distribution du rendement de l'actif avec la même valeur moyenne et même matrice de covariance. Ces résultats reposent sur une approximation Taylor de la richesse logarithmique finale qui devient plus précise au fur et à mesure que la fréquence de rééquilibrage est augmentée.

Dans la seconde partie, nous démontrons qu'une telle approximation Taylor n'est pas nécessaire. De manière spécifique, nous dérivons une probabilité précise sur la queue d'un produit de variables aléatoires non-négatives. Ces liens Chebyshev généralisés peuvent être traités numériquement en utilisant une programmation semi-définie—voire même dans certains cas de manière analytique.

En dernière partie de la thèse, nous considérons un problème de gestion de réservoir multi-marché. L'érosion du pic haut/bas propagée sur le marché au comptant d'électricité Européenne implique une rentabilité réduite aux producteurs d'énergie hydraulique et de les forcer à participer aux marchés d'équilibrage. Cela nous motive à proposer un modèle de programmation a deux-couches stochastiques pour l'optimisation opérationnelle d'une cascade d'une centrale hydraulique qui vend de l'énergie aussi bien sur un marché au comptant que sur un marché d'équilibrage. Le problème de planification optimise la gestion du réservoir avec une granularité hebdomadaire et le projet sous-commercial optimise le marché des transactions avec une granularité horaire. Nous résolvons aussi bien les problèmes de planification que de trading en règle de décision linéaire afin de parvenir à une traçabilité informatique.

**Mots clefs.** Optimisation convexe, programmation conique, optimisation distributionnellement robuste, programmation stochastique, règle de décision linéaire, optimisation de portefeuille, croissance optimale du portefeuille, valeur à risque, inégalité de Tchebychev, marché de l'électricité

# Contents

**Acknowledgements** **i**

**Abstract (English/Français)** **iii**

**List of figures** **ix**

**List of tables** **xi**

**1 Introduction** **1**

1.1 Contributions and Structure of the Thesis . . . . . . . . . . . . . . . . . . 4

1.2 Statement of Originality . . . . . . . . . . . . . . . . . . . . . . . . . . . 6

**2 Robust Growth-Optimal Portfolios** **9**

2.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9

2.2 Growth-Optimal Portfolios . . . . . . . . . . . . . . . . . . . . . . . . . 16

2.3 Robust Growth-Optimal Portfolios . . . . . . . . . . . . . . . . . . . . . 20

2.4 Worst-Case Value-at-Risk of the Growth Rate . . . . . . . . . . . . . . . 22

2.5 Extensions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 45

2.5.1 Support Information . . . . . . . . . . . . . . . . . . . . . . . . . 45

2.5.2 Moment Ambiguity . . . . . . . . . . . . . . . . . . . . . . . . . 50

2.6 Numerical Experiments . . . . . . . . . . . . . . . . . . . . . . . . . . . 52

2.6.1 Synthetic Experiments . . . . . . . . . . . . . . . . . . . . . . . . 53

2.6.2 Empirical Backtests . . . . . . . . . . . . . . . . . . . . . . . . . 58

2.6.3 Discussion of the Results . . . . . . . . . . . . . . . . . . . . . . 62

**3 Chebyshev Inequalities for Products of Random Variables** **65**

3.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 65

3.2 Optimization Perspective on Chebyshev Inequalities . . . . . . . . . . 72

3.3 Left-Sided Chebyshev Bounds . . . . . . . . . . . . . . . . . . . . . . . . . 80
3.4 Right-Sided Chebyshev Bounds . . . . . . . . . . . . . . . . . . . . . . . . 98
3.5 Covariance Bounds . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 102
3.6 Extensions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 111
3.7 Numerical Experiments . . . . . . . . . . . . . . . . . . . . . . . . . . 114
3.7.1 Comparison of Chebyshev Bounds . . . . . . . . . . . . . . . . . 114
3.7.2 Case Study: Financial Risk Management . . . . . . . . . . . . . 118

**4 Multi-Market Multi-Reservoir Management 121**
4.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 121
4.2 Hydropower Scheduling Model . . . . . . . . . . . . . . . . . . . . . . . 127
4.3 Revenue Maximization . . . . . . . . . . . . . . . . . . . . . . . . . . 132
4.4 Planner-Trader Decomposition . . . . . . . . . . . . . . . . . . . . . . 135
4.4.1 Trading Subproblems . . . . . . . . . . . . . . . . . . . . . . . 137
4.4.2 Planning Problem . . . . . . . . . . . . . . . . . . . . . . . . . 138
4.5 Multiscale Approximation . . . . . . . . . . . . . . . . . . . . . . . . . 140
4.6 Bang-Bang Strategy . . . . . . . . . . . . . . . . . . . . . . . . . . . . 142
4.6.1 Determination of Water Values . . . . . . . . . . . . . . . . . . . 144
4.6.2 Extension of the Bang-Bang Strategy . . . . . . . . . . . . . . . 145
4.7 Case Study: Austrian Electricity Market . . . . . . . . . . . . . . . . . 148
4.7.1 Experimental Setup . . . . . . . . . . . . . . . . . . . . . . . . 149
4.7.2 Numerical Results . . . . . . . . . . . . . . . . . . . . . . . . . 152

**5 Conclusions 157**
5.1 Future Research Avenues . . . . . . . . . . . . . . . . . . . . . . . . . 158

**A Appendices 161**
A.1 Distributionally Robust Chance Constraints . . . . . . . . . . . . . . . . 161
A.2 Mean and Variance of Mean-Reverting Processes . . . . . . . . . . . . . 162

**Bibliography 165**

**Curriculum Vitae 175**

# List of Figures


2.1 Comparison of the classical and robust growth-optimal portfolios under the lognormal distribution. . . . . . . . . . . . . . . . . . . . . . . . . 57

3.1 Companion figure for Lemma 3.2. . . . . . . . . . . . . . . . . . . . . . . . 87
3.2 Comparison of the left-sided and right-sided Chebyshev bounds for the products of random variables. . . . . . . . . . . . . . . . . . . . . . . . 116
3.3 Comparison of the right-sided Chebyshev bounds $\mathrm{R}(\gamma)$ and $\mathrm{R}'(\gamma)$. . . . . 117
3.4 Worst-case value-at-risk of the growth rates of the minimum-variance and maximum-expectation portfolios. . . . . . . . . . . . . . . . . . . . 119

4.1 Illustration of downward regulation (↓) and upward regulation (↑). . . 124
4.2 An example cascade of three reservoirs. . . . . . . . . . . . . . . . . . . . 128
4.3 Bang-bang strategy for reservoir management. . . . . . . . . . . . . . . . 145
4.4 Companion figure for the extended bang-bang strategy. . . . . . . . . . 147
4.5 Weekly inflows in the catchment area of the cascade. . . . . . . . . . . 149
4.6 Spot price simulation using the price forward curve. . . . . . . . . . . . 151
4.7 Comparison of the shadow prices when (i) trading in the spot market only and (ii) trading in both the spot and reserve markets. . . . . . . . 153
4.8 Comparison of the storage levels when (i) trading in the spot market only and (ii) trading in both the spot and reserve markets. . . . . . . . 154
4.9 Comparison of the cumulative revenues when (i) trading in the spot market only and (ii) trading in both the spot and reserve markets. . . . 155

# List of Tables


2.1 Markowitz risk-aversion parameter $\varrho(T,\epsilon)$ implied by the robust growth-optimal portfolio. . . . . 54

2.2 Fractional Kelly risk-aversion parameter $\kappa(T,\epsilon)$ implied by the robust growth-optimal portfolio. . . . . 55

2.3 Relative advantage (in %) of the robust growth-optimal portfolios in terms of 5% VaR ($T = 120$ months). . . . . 58

2.4 Relative advantage (in %) of the robust growth-optimal portfolios in terms of 5% VaR ($T = 360$ months). . . . . 59

2.5 Relative advantage (in %) of the robust growth-optimal portfolios in terms of 5% VaR ($T = 1{,}200$ months). . . . . 60

2.6 Out-of-sample performance of different investment strategies. . . . . 63

3.1 Chebyshev bounds equivalent to $\sup_{\mathbb{P}\in\mathscr{P}} \mathbb{P}(h(\tilde{\boldsymbol{\xi}}) \leq 0)$ for some permutation symmetric functional $h(\boldsymbol{\xi})$. . . . . 113

3.2 Runtimes (secs) required to calculate the Chebyshev bounds. . . . . 117

4.1 Description of the bang-bang strategy. . . . . 143

4.2 Description of the extended bang-bang strategy. . . . . 147

4.3 Node parameters of the cascade. . . . . 148

4.4 Arc parameters of the cascade. . . . . 148

4.5 Parameters for the stochastic differential equations (4.9). . . . . 150

4.6 Distributional parameters of the electricity reserve prices. . . . . 152

# 1 Introduction

Classical results in optimization theory usually deal with static optimization, where a decision maker implements an optimal action at a single point of time. Many static optimization problems are well understood and are supported by efficient numerical procedures. In practice, however, many optimization problems are dynamic. In this case, a decision maker seeks a sequence of decisions which optimizes some objective function. Two examples of dynamic optimization problems are discussed extensively in the remaining chapters, namely portfolio optimization problems and hydropower scheduling problems.

In a portfolio optimization problem, an investor wishes to distribute his/her current wealth over a set of available assets in order to maximize his/her terminal wealth. In a hydropower scheduling problem, on the other hand, a generation company operating a cascade of reservoirs wants to determine a generation and pumping schedule to reach a certain goal, which can be, for example, to meet an electricity demand uninterruptedly or to maximize its net revenue from trading hydroelectricity.

The main difference between static and dynamic optimization is that, in the latter case, the decision maker has to account for the effects of current decisions carried to the future. Dynamic optimization problems are further complicated by the presence of uncertainty because current decisions must be hedged against future uncertainty whereas future decisions can exploit the knowledge of the past. In this sense, these future decisions should be expressed as policies, i.e., functions of the observable information. For the two problems discussed above, the investors and the generation companies are exposed to considerable uncertainty in asset returns, electricity prices and water inflows.

Searching for optimal policies in a dynamic setting is usually much harder than searching for optimal decisions in a static setting because exact solution methods suffer from the *curse of dimensionality*, rendering them intractable for larger problems. In order to facilitate the incorporation of uncertainty and to gain tractability, one typically resorts to approximation schemes; see Powell (2014). For large instances of practical optimization problems though, approximation methods alone may not guarantee a satisfactory solving time, especially for applications where efficient decision making is not an option, but a necessity. To summarize, most dynamic optimization problems under uncertainty share in common the following hurdles.

- *Propagation effect.* Current decisions allow to hedge against future uncertainty; on the other hand, they have ramifications on the future. Hence, solving dynamic optimization problems is not equivalent to solving static optimization problems sequentially.
- *Uncertainty.* Uncertainty in optimization problems implies that decision makers have to take actions when some of the information is not yet known precisely.
- *Problem size.* The problem size (e.g. in terms of the numbers of decision stages, decisions and constraints) can be very large in real-world problems.

It is therefore important for decision makers to have additional domain-specific knowledge which allows them to solve their problems more efficiently. In some cases, the repetitive nature of dynamic models leads to an exploitable structure of the optimal policies. One interesting example advocating this notion is the growth-optimal portfolio. This portfolio is designed to have a maximum expected log-utility over a *single* rebalancing period. Kelly (1956) and Breiman (1961) independently show that the growth-optimal portfolio will eventually accumulate more wealth than any other causal portfolio with probability 1 in the long run. To many, this discovery is counterintuitive because myopic policies are rarely optimal. Nonetheless, despite its theoretical appeal, the practical relevance of the growth-optimal portfolio remains limited because of the following reasons.

- *Ambiguous asset return distribution.* The computation of the growth-optimal portfolio requires full and precise knowledge of the asset return distribution. This requirement is however rarely met in practice because the asset return distribution is typically inferred from sparse empirical observations. Hence, the growth-optimal portfolio is prone to estimation errors.
- *Asymptotic guarantees.* Many properties of the growth-optimal portfolio, including its superior wealth accumulation discussed above, hold asymptotically when the investment horizon tends to infinity. Little is known, however, about its finite-time guarantee. Indeed, empirically the growth-optimal portfolio has been shown to be unstable and amply volatile in the short run.

On the other hand, some of the main challenges for generation companies solving a hydropower scheduling problem arise, among others, from the significant uncertainty and the multiscale nature of the problem. Indeed, in European electricity markets, trading frequencies are usually high, and new information materializes every hour, be it electricity prices or water inflows. Moreover for the generation companies to fully capture the seasonality of such information, planning horizons should span at least a year. Hence, hydropower scheduling models typically consist of large numbers of decision stages and random variables. Consequently, if the generation company's objective is to maximize expected revenue, then the corresponding stochastic program is intractable.

- *Intractability of multistage stochastic programs.* Solutions of stochastic programs are hard to obtain and require significant solving time. For the energy problem studied in this thesis, the numbers of decision stages and random variables further compound the complexity of the stochastic program and make it virtually impossible for the generation companies to obtain (near-) optimal solutions.

The main objective of this thesis is to formulate and solve dynamic investment problems that are not only theoretically sound but also practically relevant. To achieve this, we develop complexity reduction techniques and decomposition schemes for efficiently solving industrial size instances of the considered problems. Specifically, we aim to address the following issues related to the growth-optimal portfolios.

- For each portfolio, we aim to establish performance guarantees (similar to those offered by the growth-optimal portfolios) for finite investment horizons when the asset return distribution is ambiguous. Put differently, we wish to construct horizon-dependent guarantees that hold for any probability distribution of the underlying assets within a prescribed ambiguity set.
- An investor adopting our model would be interested in identifying a portfolio with the most attractive performance guarantee. Thus, it is important that (i) the performance guarantee of any given portfolio and (ii) the optimal portfolio in the view of the proposed performance measure can be computed efficiently.

In addition, we address the following concerns about hydropower scheduling.

- Peak/off-peak spreads on European electricity spot markets are eroding, reducing the profitability of engaged generation companies who utilize the price arbitrages. Hence, in order to recover or to outperform their original profitability, we propose an optimal strategy for trading hydroelectricity additionally in the balancing markets.
- Engaging in multiple markets simultaneously complicates the revenue maximizing stochastic program because the number of pertinent random variables increases. We therefore propose a systematic way to reduce the complexity of the problem in order to solve it within a reasonable time frame.

## 1.1 Contributions and Structure of the Thesis

In this thesis, we investigate how techniques from stochastic and distributionally robust optimization can be utilized in formulating dynamic investment problems and how these problems can be reduced or simplified so that they can be efficiently solved. Specifically, the problems considered are portfolio optimization problems and hydropower scheduling problems. Both of these problems share the same objective function which is to maximize total earnings, from their respective investment, however with respect to different risk measures, i.e., quantiles and expected values.

The main contributions of this thesis are divided into three self-contained chapters.

Chapter 2 and Chapter 4 investigate the aforementioned dynamic investment problems. Chapter 3 has a more mathematical focus but its contents nonetheless apply to an investment problem reminiscent of the one studied in Chapter 2.

In Chapter 2 we revisit the growth-optimal portfolio and alleviate some of its shortcomings. In particular, we design a robust portfolio that offers similar performance guarantees as the classical growth-optimal portfolio. This guarantee is not distribution-specific and in fact it applies for any asset return distribution sharing the same mean and covariance matrix. Relying on a conic reformulation of distributionally robust chance constraints, we show that this robust portfolio can be computed efficiently by solving a tractable second-order cone program whose size is independent of the length of the investment horizon. The contents of this chapter are published in the following paper.

(i) N. Rujeerapaiboon, D. Kuhn and W. Wiesemann. *Robust Growth-Optimal Portfolios.* Management Science **62**(7) 2090-2109, 2016.

The results in Chapter 2 rely on a second-order Taylor approximation of logarithmic terminal wealth which becomes more accurate as the rebalancing frequency increases. In Chapter 3, however, we show that the portfolio optimization problem in Chapter 2 can be solved *exactly* in polynomial time without such a Taylor approximation. To achieve this, we derive sharp probability bounds on the left tail of a product of non-negative random variables. The material of Chapter 2 can then be viewed as an application of Chapter 3 where each random variable describes the growth of the portfolio over a single rebalancing period. We prove that these generalized Chebyshev bounds (for both left and right tails) can be computed numerically using semidefinite programming. Furthermore, we demonstrate that similar techniques from duality for moment problems, polynomial optimization and semidefinite programming can be used to derive multivariate Chebyshev bounds for sums, maxima, and minima of non-negative random variables. The contents of this chapter can be found in the following paper.

(ii) N. Rujeerapaiboon, D. Kuhn and W. Wiesemann. *Chebyshev Inequalities for Products of Random Variables.* Under Review for Mathematics of Operations

Research, 2016.

In Chapter 4 we propose a stochastic program for the optimal operation of generation companies trading hydroelectricity in both spot and balancing markets simultaneously. Despite its size, this stochastic program needs to be solved efficiently because of the trading frequencies. We therefore devote Chapter 4 to the development of a scheme for decomposing the stochastic program into a two-layer stochastic program, where each layer can be solved efficiently with a linear decision rule approximation. The outer stochastic program (the planning problem) optimizes the reservoir management over a yearly horizon with weekly granularity, whereas the inner stochastic programs (the trading subproblems) optimize the market transactions over a weekly horizon with hourly granularity. Numerical experiments indicate a considerable potential of trading hydroelectricity in the balancing markets. The contents of this chapter are based on the following working paper.

(iii) N. Rujeerapaiboon, D. Kuhn and W. Wiesemann. *A Multi-Scale Decision Rule Approach for Multi-Market Multi-Reservoir Management.* Working Paper, 2016.

Finally, the main results in Chapter 2 have been extended and used as a basis for analyzing portfolio risks incurred by autocorrelations (also known as serial correlations) of asset returns. These correlations are important in many financial studies because they help explain various phenomena, for example, seasonality in asset returns and investors' beliefs about market movements. The contents of this work are not included in the thesis but can be found in the following paper.

(iv) B. Choi, N. Rujeerapaiboon and R. Jiang. *Multi-Period Portfolio Optimization: Translation of Autocorrelation Risk to Excess Variance.* Under Review for Operations Research Letters, 2016.

## 1.2 Statement of Originality

I hereby certify that this thesis is the result of my own work, where some parts are the result of collaborations with my thesis supervisors Dr. Daniel Kuhn and Dr. Wolfram Wiesemann, as well as my industrial partners, Dr. Georg Ostermaier and his colleagues from Decision Trees GmbH. No other person's work has been used without due acknowledgement.

# 2 Robust Growth-Optimal Portfolios

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.

## 2.1 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 maximise 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 (1975) 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 chapter 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 chapter 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 $\epsilon$ is likely to be negative, we hope that attractive return guarantees will emerge for longer investment horizons even if $\epsilon$ 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 chapter. 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 chapter can be summarized as follows.

(i) 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 $\epsilon$ 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.

(ii) 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.

(iii) We show that the robust growth-optimal portfolios are near-optimal for isoelastic utility functions with relative risk aversion parameters $\kappa \gtrsim 1$. 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 $\epsilon$.

(iv) 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 princi-*

*ple* 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. (2015) 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 it is 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 chapter develops as follows. In Section 2.2 we review the asymptotic properties of classical growth-optimal portfolios, and in Section 2.3 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 2.4, and extensions of the underlying probabilistic model are presented in Section 2.5. Finally, Section 2.6 reports numerical results and 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 $\mathscr{P}^n_0$. 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.

## 2.2 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 2.1** (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 2.2** (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.

$$\tilde{\gamma}_T = \log \sqrt[T]{\prod_{t=1}^{T} [1 + \boldsymbol{w}_t(\tilde{\boldsymbol{r}}_1, \ldots, \tilde{\boldsymbol{r}}_{t-1})^\intercal \tilde{\boldsymbol{r}}_t]} = \frac{1}{T} \sum_{t=1}^{T} \log \left[1 + \boldsymbol{w}_t(\tilde{\boldsymbol{r}}_1, \ldots, \tilde{\boldsymbol{r}}_{t-1})^\intercal \tilde{\boldsymbol{r}}_t\right] \tag{2.1}$$

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 2.1** (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*

$$\lim_{T \to \infty} \tilde{\gamma}_T = \mathbb{E}\left(\log\left(1 + \boldsymbol{w}^\intercal \tilde{\boldsymbol{r}}_1\right)\right) \quad \textit{with probability } 1. \tag{2.2}$$

*Proof.* The claim follows immediately from (2.1) and the strong law of large numbers. ∎

Proposition 2.1 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 (2.2). 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 2.3** (Kelly Strategy)**.** *The Kelly strategy is the fixed-mix strategy induced by the unique growth-optimal portfolio* $\boldsymbol{w}^* = \arg\max_{\boldsymbol{w} \in \mathcal{W}} \mathbb{E}\left(\log(1 + \boldsymbol{w}^\intercal \tilde{\boldsymbol{r}}_1)\right)$.

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 2.1** (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 2.1 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 2.1 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 2.1 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.

## 2.3 Robust Growth-Optimal Portfolios

In this section we extend the growth guarantees of Theorem 2.1 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

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

The violation probability $\epsilon$ of the chance constraint reflects the investor's risk aversion and is typically chosen as a small number $\lesssim 10\%$. If $\gamma^*$ 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 $\mathscr{P}$, 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 $\mathscr{P}$. 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 $\boldsymbol{w}$.

$$\begin{aligned}\text{WVaR}_\epsilon(\tilde{\gamma}_T(\boldsymbol{w})) \;&=\; \min_{\mathbb{P}\in\mathscr{P}}\; \mathbb{P}\text{-VaR}_\epsilon(\tilde{\gamma}_T(\boldsymbol{w}))\\ &=\; \max_{\gamma\in\mathbb{R}}\left\{\gamma \,:\, \mathbb{P}\left(\frac{1}{T}\sum_{t=1}^{T}\log\left(1+\boldsymbol{w}^\intercal\tilde{\boldsymbol{r}}_t\right)\geq\gamma\right)\geq 1-\epsilon \quad \forall\mathbb{P}\in\mathscr{P}\right\}\end{aligned} \tag{2.3}$$

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

**Observation 2.1** (Performance Guarantees). *Let $\boldsymbol{w}^*$ be the robust growth-optimal portfolio that maximizes WVaR$_\epsilon(\tilde{\gamma}_T(\boldsymbol{w}))$ over $\mathscr{W}$ and denote by $\gamma^*$ 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 $\mathscr{P}$.*

*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 $\gamma^*$ 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 2.4** (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 $\mathbb{E}_{\mathbb{P}}(\tilde{\boldsymbol{r}}_t) = \boldsymbol{\mu}$ 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 2.1 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

$$\mathscr{P} = \left\{ \mathbb{P} \in \mathscr{P}_0^{nT} \;:\; \begin{array}{l} \mathbb{E}_{\mathbb{P}}(\tilde{\boldsymbol{r}}_t) = \boldsymbol{\mu} \quad \forall t \,:\, 1 \le t \le T \\ \mathbb{E}_{\mathbb{P}}\left(\tilde{\boldsymbol{r}}_s \tilde{\boldsymbol{r}}_t^\intercal\right) = \delta_{st}\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal \quad \forall s,t \,:\, 1 \le s \le t \le T \end{array} \right\}, \tag{2.4}$$

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 $\mathscr{P}$. We remark that, besides its conceptual appeal, the moment-based ambiguity set $\mathscr{P}$ has distinct computational benefits that will become apparent in Section 2.4. 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 2.5.

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 2.1** (Support Constraints)**.** *The ambiguity set $\mathscr{P}$ could safely be reduced by including the support constraints $\mathbb{P}(\tilde{\boldsymbol{r}}_t \ge -\mathbf{1}) = 1 \;\forall t : 1 \le t \le T$, 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* (2.3) *more conservative. In order to obtain a clean model, we first suppress these constraints but emphasize that problem* (2.3) *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 2.5.1.*

## 2.4 Worst-Case Value-at-Risk of the Growth Rate

Weak sense white noise ambiguity sets of the form (2.4) 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 (2.1) 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 $\epsilon$ 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)

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

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

$$\text{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w})) = \max_{\gamma\in\mathbb{R}} \left\{\gamma \,:\, \mathbb{P}\left(\tilde{\gamma}'_T(\boldsymbol{w}) \geq \gamma\right) \geq 1-\epsilon \quad \forall \mathbb{P} \in \mathscr{P}\right\} \tag{2.5}$$

for some fixed $\boldsymbol{w} \in \mathscr{W}$, $T \in \mathbb{N}$ and $\epsilon \in (0,1)$. By exploiting a known tractable reformulation of distributionally robust quadratic chance constraints with mean and covariance information (see Theorem A.1 in Appendix A.1), we can re-express problem (2.5), which involves *infinitely* many constraints parameterized by $\mathbb{P} \in \mathscr{P}$, as a *finite* semidefinite program (SDP). Thus, we obtain

$$\begin{aligned} \text{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w})) = \max\ & \gamma \\ \text{s.t.}\ & \mathbf{M} \in \mathbb{S}^{nT+1},\ \beta \in \mathbb{R},\ \gamma \in \mathbb{R} \\ & \beta + \tfrac{1}{\epsilon}\langle \boldsymbol{\Omega}, \mathbf{M}\rangle \leq 0, \quad \mathbf{M} \succeq \mathbf{0} \\ & \mathbf{M} - \begin{bmatrix} \frac{1}{2}\sum_{t=1}^T \mathbf{P}_t^\top \boldsymbol{w}\boldsymbol{w}^\top \mathbf{P}_t & -\frac{1}{2}\sum_{t=1}^T \mathbf{P}_t^\top \boldsymbol{w} \\ -\frac{1}{2}\left(\sum_{t=1}^T \mathbf{P}_t^\top \boldsymbol{w}\right)^\top & \gamma T - \beta \end{bmatrix} \succeq \mathbf{0}, \end{aligned} \tag{2.6}$$

where

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

denotes the matrix of first and second-order moments of $(\tilde{\boldsymbol{r}}_1^\top, \dots, \tilde{\boldsymbol{r}}_T^\top)^\top$, while the truncation operators $\mathbf{P}_t \in \mathbb{R}^{n\times nT}$ are defined via $\mathbf{P}_t(\boldsymbol{r}_1^\top, \dots, \boldsymbol{r}_T^\top)^\top = \boldsymbol{r}_t$, $t = 1, \dots, T$. As (2.6) 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 2.2** (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* (2.6)*. In this case, the last matrix inequality in* (2.6) *becomes quadratic in the decision variables, and* (2.6) *ceases to be an SDP. Fortunately, however, one can convert* (2.6) *back to an SDP by rewriting the quadratic matrix inequality as*

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

*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*

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

*by virtue of a Schur complement argument.*

Even though SDPs are polynomial-time solvable in theory, problem (2.6) 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 (2.5) admits in fact an *analytical* solution.

We first notice that the random asset returns $\tilde{\boldsymbol{r}}_t$ enter problem (2.5) only in the form of the portfolio return $\boldsymbol{w}^\top \tilde{\boldsymbol{r}}_t$. We can thus use a well-known projection property of moment-based ambiguity sets to perform a dimensionality reduction.

**Proposition 2.2** (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 $\mathcal{P}_{\tilde{\xi}}$ and $\mathcal{P}_{\tilde{\zeta}}$ as*

$$\mathcal{P}_{\tilde{\xi}} = \left\{\mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^p\right) : \mathbb{E}_{\mathbb{P}}\left(\left[\tilde{\boldsymbol{\xi}}^\top \ 1\right]^\top \left[\tilde{\boldsymbol{\xi}}^\top \ 1\right]\right) = \boldsymbol{\Omega}_{\tilde{\xi}}\right\}$$

*and*

$$\mathcal{P}_{\tilde{\zeta}} = \left\{\mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^q\right) : \mathbb{E}_{\mathbb{P}}\left(\left[\tilde{\boldsymbol{\zeta}}^\top \ 1\right]^\top \left[\tilde{\boldsymbol{\zeta}}^\top \ 1\right]\right) = \boldsymbol{\Omega}_{\tilde{\zeta}}\right\},$$

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

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

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

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

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

Applying Proposition 2.2 to problem (2.5) yields

$$\begin{aligned} \mathrm{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w})) \ = \ \sup_{\gamma} \quad & \gamma \\ \text{s.t.} \quad & \mathbb{P}\left(\frac{1}{T}\sum_{t=1}^{T}\left(\tilde{\eta}_t - \frac{1}{2}\tilde{\eta}_t^2\right) \ge \gamma\right) \ge 1-\epsilon \quad \forall \mathbb{P} \in \mathcal{P}_{\tilde{\boldsymbol{\eta}}}(\boldsymbol{w}), \end{aligned} \tag{2.7}$$

where the projected ambiguity set

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

contains all distributions on $\mathbb{R}^T$ under which the portfolio returns $(\tilde{\eta}_1,\ldots,\tilde{\eta}_T)^\top$ follow a weak sense white noise process with (period-wise) mean $\boldsymbol{w}^\top\boldsymbol{\mu}$ and variance $\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}$.

Problem (2.7) has the same structure as the original problem (2.5), but the underlying probability space has only dimension $T$ instead of $nT$. Thus, it can be converted to a tractable SDP by using Theorem A.1 to reformulate the underlying distributionally robust chance constraint. We then obtain

$$\begin{aligned}
\mathrm{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w})) \;=\; \max \quad & \gamma \\
\text{s.t.} \quad & \mathbf{M} \in \mathbb{S}^{T+1},\ \beta \in \mathbb{R},\ \gamma \in \mathbb{R} \\
& \beta + \tfrac{1}{\epsilon}\langle \boldsymbol{\Omega}(\boldsymbol{w}), \mathbf{M} \rangle \le 0, \quad \mathbf{M} \succeq \mathbf{0} \\
& \mathbf{M} - \begin{bmatrix} \frac{1}{2}\mathbb{I} & -\frac{1}{2}\mathbf{1} \\ -\frac{1}{2}\mathbf{1}^\top & \gamma T - \beta \end{bmatrix} \succeq \mathbf{0},
\end{aligned} \tag{2.8}$$

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

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

The projected problem (2.7) can be further simplified by exploiting its compound symmetry.

**Definition 2.5** (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*

$$\mathbf{M} = \left[\begin{array}{cccc|c}
\tau_1 & \tau_2 & \cdots & \tau_2 & \tau_3 \\
\tau_2 & \tau_1 & \cdots & \tau_2 & \tau_3 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\tau_2 & \tau_2 & \cdots & \tau_1 & \tau_3 \\
\hline
\tau_3 & \tau_3 & \ldots & \tau_3 & \tau_4
\end{array}\right]. \tag{2.9}$$

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 (2.8) as well as compound symmetric.

**Proposition 2.3.** *There exists a maximizer* $(\mathbf{M}, \beta, \gamma)$ *of (2.8) with* $\mathbf{M}$ *compound symmetric.*

*Proof.* Denote by $\Pi^{T+1}$ the set of all permutations $\pi$ 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 $\mathbf{P}_\pi^\top$ represents the permutation matrix corresponding to the inverse of $\pi$. A matrix $\mathbf{K} \in \mathbb{S}^{T+1}$ is compound symmetric if and only if $\mathbf{K} = \mathbf{P}_\pi \mathbf{K} \mathbf{P}_\pi^\top$ for all $\pi \in \Pi^{T+1}$. Suppose that $(\mathbf{M}, \beta, \gamma)$ is a maximizer of (2.8). Since the input matrices in (2.8) are compound symmetric and $\mathbf{P}_\pi$ is non-singular, we have

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

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

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

Hence, $\left(\mathbf{P}_\pi \mathbf{M} \mathbf{P}_\pi^\top, \beta, \gamma\right)$ is feasible in (2.8) and has the same objective value as $(\mathbf{M}, \beta, \gamma)$. It is therefore a maximizer of (2.8). As the set of maximizers is convex, the convex combination

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

is also a maximizer of (2.8). Moreover, $\mathbf{M}'$ is compound symmetric because $\rho(\Pi^{T+1}) = \Pi^{T+1}$ and, *a fortiori*, $\mathbf{P}_\rho \mathbf{M}' \mathbf{P}_\rho^\top = \mathbf{M}'$ for any $\rho \in \Pi^{T+1}$. Thus, the claim follows. ∎

By Proposition 2.3, we may assume without loss of generality that $\mathbf{M}$ in (2.8) is compound symmetric. Thus, each matrix inequality in (2.8) 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 2.4.** *For any compound symmetric matrix* $\mathbf{M} \in \mathbb{S}^{T+1}$ *of the form (2.9), the following equivalence holds.*

$$\mathbf{M} \succeq \mathbf{0} \iff \begin{cases} \tau_1 \geq \tau_2 & (2.10a) \\ \tau_4 \geq 0 & (2.10b) \\ \tau_1 + (T-1)\tau_2 \geq 0 & (2.10c) \\ \tau_4\left(\tau_1 + (T-1)\tau_2\right) \geq T\tau_3^2 & (2.10d) \end{cases}$$

*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

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

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 $\tau_4$, respectively. Thus $\mathbf{M} \succeq \mathbf{0}$ if and only if (2.10a), (2.10b), and (2.10c) hold. Moreover, (2.10d) is trivially implied by (2.10b) and (2.10c) 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 $\lambda$ 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 $\lambda$ yields the two eigenvalues

$$\lambda = \frac{1}{2}\left(\tau_1 + (T-1)\tau_2 + \tau_4 \pm \sqrt{(\tau_1 + (T-1)\tau_2 + \tau_4)^2 + 4\left(T\tau_3^2 - \tau_4(\tau_1 + (T-1)\tau_2)\right)}\right) \tag{2.11a}$$

$$= \frac{1}{2}\left(\tau_1 + (T-1)\tau_2 + \tau_4 \pm \sqrt{(\tau_1 + (T-1)\tau_2 - \tau_4)^2 + 4T\tau_3^2}\right). \tag{2.11b}$$

From equation (2.11b) it is evident that the square root term constitutes a strictly positive real number. The two eigenvalues are thus nonnegative if and only if

$$\tau_1 + (T-1)\tau_2 + \tau_4 \geq 0, \quad (\tau_1 + (T-1)\tau_2)\,\tau_4 \geq T\tau_3^2. \tag{2.12}$$

The second inequality in (2.12) ensures that the square root term in (2.11a) does not exceed $\tau_1 + (T-1)\tau_2 + \tau_4$, which implies (2.10d). By (2.12), both the product and the sum of $\tau_1 + (T-1)\tau_2$ and $\tau_4$ are nonnegative, which implies that each of them must be individually nonnegative, i.e., (2.10b) and (2.10c) hold. The claim now follows from the fact that (2.10a)–(2.10d) also imply (2.12). ∎

**Corollary 2.1.** *For any compound symmetric matrix* $\mathbf{M} \in \mathbb{S}^{T+1}$ *of the form (2.9), the semidefinite constraint* $\mathbf{M} \succeq \mathbf{0}$ *is equivalent to a system of second-order cone constraints.*

$$\mathbf{M} \succeq \mathbf{0} \iff \begin{cases} \tau_1 \geq \tau_2 \\ \tau_1 + (T-1)\tau_2 + \tau_4 \geq \sqrt{(\tau_1 + (T-1)\tau_2 - \tau_4)^2 + 4T\tau_3^2} \end{cases}$$

*Proof.* The inequalities (2.10b), (2.10c), and (2.10d) can be viewed as a hyperbolic constraint. The claim then follows from Boyd and Vandenberghe (2004, Exercise 4.26). ∎

We now demonstrate that the semidefinite program (2.8), which involves $\mathcal{O}(T^2)$ decision variables, can be reduced to an equivalent non-linear program with only six decision variables. First, by Proposition 2.3, we may assume without any loss of generality that the decision variable $\mathbf{M}$ is of the form (2.9) for some $\boldsymbol{\tau} \in \mathbb{R}^4$. Thus, we can use Proposition 2.4 to re-express both semidefinite constraints in (2.8) 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.

$$
\begin{aligned}
\mathrm{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w})) \; = \; \max \quad & \gamma \\
\text{s.t.} \quad & \boldsymbol{\tau} \in \mathbb{R}^4,\ \beta \in \mathbb{R},\ \gamma \in \mathbb{R} \\
& \beta + \tfrac{1}{\epsilon}\left[ T\left(\sigma_\mathrm{p}^2 + \mu_\mathrm{p}^2\right)\tau_1 + T(T-1)\mu_\mathrm{p}^2\tau_2 + 2T\mu_\mathrm{p}\tau_3 + \tau_4 \right] \le 0 && \text{(2.14a)} \\
& \tau_1 \ge \tau_2 && \text{(2.14b)} \\
& \tau_4 \ge 0 && \text{(2.14c)} \\
& \tau_1 + (T-1)\tau_2 \ge 0 && \text{(2.14d)} \\
& \tau_4\left(\tau_1 + (T-1)\tau_2\right) \ge T\tau_3^2 && \text{(2.14e)} \\
& \left(\tau_1 - \tfrac{1}{2}\right) \ge \tau_2 && \text{(2.14f)} \\
& \tau_4 - \gamma T + \beta \ge 0 && \text{(2.14g)} \\
& \left(\tau_1 - \tfrac{1}{2}\right) + (T-1)\tau_2 \ge 0 && \text{(2.14h)} \\
& \left(\tau_4 - \gamma T + \beta\right)\left(\left(\tau_1 - \tfrac{1}{2}\right) + (T-1)\tau_2\right) \ge T\left(\tau_3 + \tfrac{1}{2}\right)^2 && \text{(2.14i)}
\end{aligned}
$$

Note that (2.14a) corresponds to the trace inequality, while (2.14b)–(2.14e) encodes the positive semidefiniteness of $\mathbf{M}$, and (2.14f)–(2.14i) is a reformulation of the last matrix inequality in (2.8).

We first note that (2.14a) is binding at optimality. Indeed, if (2.14a) 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, (2.14b) and (2.14d) are redundant in view of (2.14f) and (2.14h) and can thus be dropped. Finally, there exists an optimal solution for which (2.14f) is binding. Indeed, if (2.14f) is not binding at $(\boldsymbol{\tau}, \beta, \gamma)$, then $\left(\frac{\tau_1 + (T-1)\tau_2 - \frac{1}{2}}{T} + \frac{1}{2}, \frac{\tau_1 + (T-1)\tau_2 - \frac{1}{2}}{T}, \tau_3, \tau_4, \gamma, \beta\right)$ remains feasible with the same objective value but satisfies (2.14f) as an equality. Without loss of generality, we can thus eliminate the decision variable $\tau_1$ by using the substitution $\tau_1 = \tau_2 + \frac{1}{2}$. In summary, we have

$$
\begin{aligned}
\mathrm{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w})) \; = \; \max \quad & \gamma \\
\text{s.t.} \quad & \tau_2 \in \mathbb{R},\ \tau_3 \in \mathbb{R},\ \tau_4 \in \mathbb{R},\ \beta \in \mathbb{R},\ \gamma \in \mathbb{R} \\
& \beta + \frac{1}{\epsilon}\left[\frac{T}{2}\left(\sigma_\mathrm{p}^2 + \mu_\mathrm{p}^2\right) + T\left(\sigma_\mathrm{p}^2 + T\mu_\mathrm{p}^2\right)\tau_2 + 2T\mu_\mathrm{p}\tau_3 + \tau_4\right] = 0 \\
& \tau_4 \geq 0 \\
& \tau_4 - \gamma T + \beta \geq 0 \\
& \tau_2 \geq 0 \\
& \tau_4\left(\tau_2 + \tfrac{1}{2T}\right) \geq \tau_3^2 \\
& \left(\tau_4 - \gamma T + \beta\right)\tau_2 \geq \left(\tau_3 + \tfrac{1}{2}\right)^2 .
\end{aligned}
\tag{2.15}
$$

Problem (2.15) can be written more compactly as

$$
\begin{aligned}
\mathrm{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w})) \; = \; -\min \quad & aw + bx + cy + dz + e \\
\text{s.t.} \quad & w, x, y, z \in \mathbb{R} \\
& w \geq 0,\ x \geq 0,\ y \geq 1 \\
& (\tfrac{1}{2}z + 1)^2 \leq w(y+1) \\
& (\tfrac{1}{2}z - 1)^2 \leq x(y-1),
\end{aligned}
\tag{2.16}
$$

where the decision variables $w$, $x$, $y$ and $z$ in (2.16) are related to the variables $\tau_2$, $\tau_3$, $\tau_4$, $\beta$ and $\gamma$ in (2.15) 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 $a = \frac{1}{4\epsilon} - \frac{1}{4}$, $b = \frac{1}{4}$, $c = \frac{\sigma_\mathrm{p}^2 + T\mu_\mathrm{p}^2}{4\epsilon T}$, $d = -\frac{\mu_\mathrm{p}}{4\epsilon}$ and $e = \frac{\mu_\mathrm{p}^2}{4\epsilon} - \frac{\mu_\mathrm{p}}{2\epsilon} + \frac{\sigma_\mathrm{p}^2}{2\epsilon} - \frac{\sigma_\mathrm{p}^2}{4\epsilon T}$. Note that the inequality constraints in (2.15) correspond to the inequality constraints in (2.16) in the same order, while the equality constraint in (2.15) has been eliminated via the following substitution.

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

Problem (2.16) admits an explicit analytical solution as stated in the following lemma.

**Lemma 2.1.** *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*

$$\begin{aligned}
\min \quad & aw + bx + cy + dz \\
\text{s.t.} \quad & w, x, y, z \in \mathbb{R} \\
& w \geq 0,\ x \geq 0,\ y \geq 1 \\
& (\tfrac{1}{2}z + 1)^2 \leq w(y+1) \\
& (\tfrac{1}{2}z - 1)^2 \leq x(y-1)
\end{aligned} \tag{2.17}$$

*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 (2.17) is bounded, while assumption (ii) guarantees that $\Delta = \sqrt{(a+b)c - d^2}$ is real. Assumption (iii) is not strictly needed, and problem (2.17) admits a *generalized* closed-form solution even if this assumption is violated. However, this more general solution is not needed for this chapter, and therefore we will not derive it.

Note that (2.17) 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 (2.17). 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 (2.17) are binding by the definition of $w$ and $x$.

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

$$ap+bq+d = 0, \quad ap^2+bq^2 = c, \quad p = \frac{z+2}{2y+2}, \quad q = \frac{z-2}{2y-2} \tag{2.18}$$

Moreover, we denote by $\alpha_1$, $\alpha_2$ and $\alpha_3$ the Lagrange multipliers of the three linear inequalities and by $\lambda$ and $\delta$ the Lagrange multipliers of the two hyperbolic constraints in (2.17), 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 (2.18) 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 (2.17), we first use (2.18) to show that $w = p\left(\frac{1}{2}z+1\right)$ and $x = q\left(\frac{1}{2}z-1\right)$. This enables us to express the optimal objective value as

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

where the last equality follows again from (2.18). 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 state the main result of this section.

**Theorem 2.2** (Worst-Case Value-at-Risk)**.** *Under Assumptions (A1) and (A2) we have*

$$\mathrm{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w})) = \frac{1}{2}\left(1-\left(1-\boldsymbol{w}^\top\boldsymbol{\mu}+\sqrt{\tfrac{1-\epsilon}{\epsilon T}}\left\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{w}\right\|\right)^2-\tfrac{T-1}{\epsilon T}\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}\right). \tag{2.19}$$

*Proof.* We know that the worst-case VaR of $\tilde{\gamma}'_T(\boldsymbol{w})$ is given by the optimal value of problem (2.16). By construction, the objective function coefficients $a$, $b$ and $c$ in (2.16) 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

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

As all conditions of Lemma 2.1 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. ∎

If the set $\mathcal{W}$ of admissible portfolios is characterized by a finite number of linear constraints, then the portfolio optimization problem $\max_{\boldsymbol{w}\in\mathcal{W}} \mathrm{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w}))$ 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 (2.19) is increasing in the portfolio mean return $\boldsymbol{w}^\top\boldsymbol{\mu}$ and decreasing in the portfolio variance $\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}$ 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 $\max_{\boldsymbol{w}\in\mathcal{W}} \boldsymbol{w}^\top\boldsymbol{\mu} - \frac{\varrho}{2}\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}$ 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 $\epsilon$. We can thus define a function $\varrho(T,\epsilon)$ with the property that the robust growth-optimal portfolio tailored to $T$ and $\epsilon$ 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

$$\varrho(T,\epsilon) = \sqrt{\frac{1-\epsilon}{\epsilon T}} \cdot \frac{1}{\left\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{w}\right\|} + \frac{T-1}{\epsilon T\left(1 - \boldsymbol{w}^\top\boldsymbol{\mu} + \sqrt{\frac{1-\epsilon}{\epsilon T}}\left\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{w}\right\|\right)}, \tag{2.20}$$

where $\boldsymbol{w}$ denotes the robust growth-optimal portfolio, which depends on both $T$ and $\epsilon$ and can only be computed numerically. The function $\varrho(T,\epsilon)$ will be investigated further in Section 2.6.1. From the point of view of mean-variance analysis, a robust growth-optimal investor becomes less risk-averse as $\epsilon$ 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 $\epsilon$ 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 2.5.

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 $\kappa$ emerges as a maximizer of the Merton problem with constant relative risk aversion $\kappa$ 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 $\kappa$ through the portfolio that maximizes the expected isoelastic utility function $\frac{1}{1-\kappa}\mathbb{E}[(1+\boldsymbol{w}^\top \tilde{\boldsymbol{r}}_1)^{1-\kappa}]$. By expanding the utility function around 1, this portfolio can be closely approximated by $\boldsymbol{w}_\kappa = \arg\max_{\boldsymbol{w}\in\mathcal{W}} \boldsymbol{w}^\top \boldsymbol{\mu} - \frac{\kappa}{2}\boldsymbol{w}^\top\left(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\top\right)\boldsymbol{w}$, which is mean-variance efficient with risk-aversion parameter $\varrho = \kappa/(1+\kappa\,\boldsymbol{w}_\kappa^\top \boldsymbol{\mu})$ whenever $\boldsymbol{w}_\kappa^\top \boldsymbol{\mu} \leq 1/\kappa$. Note that the last condition is reminiscent of Assumption (A2) and is satisfied for typical choices of $\epsilon$, $T$, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. It then follows from (2.20) that the robust growth-optimal portfolio tailored to the investment horizon $T$ and violation probability $\epsilon$ coincides with the (approximate) fractional Kelly strategy corresponding to the risk-aversion parameter

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

where the robust growth-optimal portfolio $\boldsymbol{w}$ must be computed numerically and $\upsilon$

is defined as $1 - \boldsymbol{w}^\top \boldsymbol{\mu} + \sqrt{\frac{1-\epsilon}{\epsilon T}}\, \|\boldsymbol{\Sigma}^{1/2}\boldsymbol{w}\|$. 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 2.6.1.

**Remark 2.3.** ***(Relation to Worst-Case VaR by El Ghaoui et al. (2003))*** *Theorem 2.2 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* $\max_{\boldsymbol{w}\in\mathcal{W}} \mathit{WVaR}_\epsilon(\tilde{\gamma}'_1(\boldsymbol{w}))$ *reduces to*

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

*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 2.4.** ***(Long-Term Investors)*** *In the limit of very long investment horizons, the worst-case VaR* (2.19) *reduces to*

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

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

**Remark 2.5.** ***(Worst-Case Conditional VaR)*** *We could use the worst-case* conditional *VaR (CVaR) instead of the worst-case VaR in* (2.5) *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 $\mathscr{P}$ 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. (2010a). In contrast, the construction presented here is completely analytical.

For a fixed $\boldsymbol{w} \in \mathscr{W}$, we first construct a sequence of *portfolio* return distributions $\mathbb{P}^{\epsilon'} \in \mathscr{P}_{\tilde{\boldsymbol{\eta}}}(\boldsymbol{w})$, $\epsilon' \in (\epsilon, 1)$, that attains the worst case in problem (2.7) as $\epsilon'$ approaches $\epsilon$. Recall that $\tilde{\boldsymbol{\eta}}$ represents a weak sense white noise process with mean $\mu_{\mathrm{p}} = \boldsymbol{w}^\top \boldsymbol{\mu}$ and standard deviation $\sigma_{\mathrm{p}} = \sqrt{\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}}$.

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 $\boldsymbol{\eta}^b$, $\{\boldsymbol{\eta}^u_t\}_{t=1}^T$ and $\{\boldsymbol{\eta}^d_t\}_{t=1}^T$, defined through

$$\begin{aligned}
\boldsymbol{\eta}^b &= (\eta^b_1, \ldots, \eta^b_T)^\top && \text{where} \quad \eta^b_s = b && \forall s = 1, \ldots, T, \\
\boldsymbol{\eta}^u_t &= (\eta^u_{t,1}, \ldots, \eta^u_{t,T})^\top && \text{where} \quad \eta^u_{t,s} = u + \Delta\delta_{ts} && \forall t, s = 1, \ldots, T, \\
\boldsymbol{\eta}^d_t &= (\eta^d_{t,1}, \ldots, \eta^d_{t,T})^\top && \text{where} \quad \eta^d_{t,s} = d - \Delta\delta_{ts} && \forall t, s = 1, \ldots, T.
\end{aligned}$$

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

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

Theorem 2.3 below asserts that the distributions $\mathbb{P}^{\epsilon'}$ attain the worst case in (2.7) 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 $\mathscr{P}$. Examples include support constraints as studied in Section 2.5.

**Theorem 2.3.** *The portfolio return distributions* $\mathbb{P}^{\epsilon'}$, $\epsilon' \in (\epsilon, 1)$, *have the following properties.*

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

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

*Proof.* We first establish some identities for the parameters $\Delta$, $b$, $u$ and $d$ that will be useful for the proof of the two assertions. From the definition of $d$ we conclude that

$$(T-1)u^2 + (u+\Delta)^2 = (T-1)d^2 + (d-\Delta)^2, \tag{2.21}$$

while the definitions of $b$ and $u$ imply that

$$(1-\epsilon')b + \epsilon'\left(u + \frac{\Delta}{T}\right) = \mu_{\text{p}}. \tag{2.22}$$

For later reference we define

$$\gamma = \frac{1}{2}\left(1 - \left(1 - \mu_{\mathrm{p}} + \sqrt{\frac{1-\epsilon'}{\epsilon' T}}\sigma_{\mathrm{p}}\right)^2 - \frac{T-1}{\epsilon' T}\sigma_{\mathrm{p}}^2\right). \tag{2.23}$$

By construction, $\gamma$ is equal to the worst-case VaR of $\tilde{\gamma}'_T(\boldsymbol{w})$ at the tolerance level $\epsilon'$; see Theorem 2.2. Basic algebraic manipulations yield the following equation equivalent to (2.23).

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

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

$$b^2 - 2b + \frac{2\mu_{\mathrm{p}} - \mu_{\mathrm{p}}^2 - \sigma_{\mathrm{p}}^2 - 2\gamma\epsilon'}{1-\epsilon'} = 0. \tag{2.24}$$

Similarly, from the definition of $u$ we obtain

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

where the second equality uses (2.23) and the third equality uses the definition of $\Delta$. Therefore, $u$ can be viewed a root of the quadratic equation

$$Tu^2 + 2u(\Delta - T) + 2T\gamma - 2\Delta + \Delta^2 = 0. \tag{2.25}$$

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 $\mathscr{P}_{\tilde{\boldsymbol{\eta}}}(\boldsymbol{w})$. By construction, it is clear that $\mathbb{P}^{\epsilon'}$ is indeed a probability distribution. As for the first order moment conditions, we observe that

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

where the second equality follows from the definition of $d$, while the third equality exploits (2.22). As for the second order moment conditions, we have

$$
\begin{aligned}
\mathbb{E}_{\mathbb{P}^{\epsilon'}}\left(\tilde{\eta}_t^2\right) &= \left(1-\epsilon'\right) b^2 + \frac{\epsilon'}{2T}\left((T-1)\,u^2 + (u+\Delta)^2\right) + \frac{\epsilon'}{2T}\left((T-1)\,d^2 + (d-\Delta)^2\right) \\
&= \left(1-\epsilon'\right) b^2 + \frac{\epsilon'}{T}\left((T-1)u^2 + (u+\Delta)^2\right) \\
&= \left(1-\epsilon'\right) b^2 + \frac{\epsilon'}{T}\left(Tu^2 + 2u\Delta + \Delta^2\right) \\
&= \left(1-\epsilon'\right)\left(2b - \frac{2\mu_{\mathrm{p}} - \mu_{\mathrm{p}}^2 - \sigma_{\mathrm{p}}^2 - 2\gamma\epsilon'}{1-\epsilon'}\right) + \frac{2\epsilon'}{T}\left(uT - T\gamma + \Delta\right) \\
&= \mu_{\mathrm{p}}^2 + \sigma_{\mathrm{p}}^2 - 2\left(\mu_{\mathrm{p}} - (1-\epsilon')b - \epsilon' u - \frac{\epsilon'\Delta}{T}\right) \;=\; \mu_{\mathrm{p}}^2 + \sigma_{\mathrm{p}}^2,
\end{aligned}
\tag{2.26}
$$

where the second equality follows from (2.21), the fourth equality exploits (2.24) to re-express $b^2$ and (2.25) to re-express $Tu^2 + 2u\Delta + \Delta^2$, and the last equality holds due to (2.22). Similarly, we find

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

for $s \neq t$. A comparison with (2.26) shows that the first two terms in the second line of the above expression are equal to $\mu_{\mathrm{p}}^2 + \sigma_{\mathrm{p}}^2$. The third equality then follows from the definition of $\Delta$. Thus, $\mathbb{P}^{\epsilon'} \in \mathscr{P}_{\tilde{\boldsymbol{\eta}}}(\boldsymbol{w})$.

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 (2.25). A similar calculation shows that $\tilde{\gamma}_T^{\eta}$ is also equal to $\gamma$ 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

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

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

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

where the first equivalence follows from basic algebraic manipulations, while the second equivalence is due to (2.24). 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 $\gamma$ as a function of $\epsilon'$, which follows from the definition of $\gamma$ in (2.23), 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. ∎

As $\epsilon'$ tends to $\epsilon$, 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 $\epsilon$ under $\mathbb{P}^{\epsilon}$, the VaR at level $\epsilon$ 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 $\mathbb{P}^{\epsilon}\text{-VaR}_{\epsilon}(\tilde{\gamma}_T^{\eta}) > \text{WVaR}_{\epsilon}(\tilde{\gamma}'_T(\boldsymbol{w}))$. 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 (2.7). Next, we construct a sequence of asset return distributions that asymptotically attain the worst-case VaR in (2.5). 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 $\left(\tilde{\boldsymbol{m}}_1^\intercal, \ldots, \tilde{\boldsymbol{m}}_T^\intercal\right)^\intercal \in \mathbb{R}^{nT}$ 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 $\left(\tilde{\boldsymbol{r}}_1^\intercal, \ldots, \tilde{\boldsymbol{r}}_T^\intercal\right)^\intercal \in \mathbb{R}^{nT}$ defined through

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

**Corollary 2.2.** *The asset return distributions* $\mathbb{Q}^{\epsilon'}$, $\epsilon' \in (\epsilon, 1)$, *have the following properties.*

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

*(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 2.3, as well as Theorem 1 of Yu et al. (2009). ∎

## 2.5 Extensions

The basic model of Section 2.3 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.

### 2.5.1 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

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

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 $\Xi$. By construction, the ellipsoid $\Xi$ 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 $\mathscr{P}$ with the (smaller) ambiguity set

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

when computing the worst-case VaR (2.5). By using a tractable conservative approximation for distributionally robust chance constraints with mean, covariance and support information (see Theorem A.1 in Appendix A.1), we can lower bound this generalized worst-case VaR by the optimal value of a tractable SDP.

$$
\begin{aligned}
\mathrm{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w})) \;\geq\; \max \quad & \gamma \\
\text{s.t.} \quad & \mathbf{M} \in \mathbb{S}^{nT+1},\ \alpha \in \mathbb{R},\ \beta \in \mathbb{R},\ \gamma \in \mathbb{R},\ \lambda \in \mathbb{R} \\
& \alpha \geq 0, \quad \beta \leq 0, \quad \lambda \geq 0, \quad \beta + \tfrac{1}{\epsilon}\langle \boldsymbol{\Omega}, \mathbf{M} \rangle \leq 0 \\
& \mathbf{M} \succeq \alpha \begin{bmatrix} -\sum_{t=1}^T \mathbf{P}_t^\top \boldsymbol{\Lambda}^{-1} \mathbf{P}_t & \sum_{t=1}^T \mathbf{P}_t^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{\nu} \\ \left(\sum_{t=1}^T \mathbf{P}_t^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{\nu}\right)^\top & T\left(\delta - \boldsymbol{\nu}^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{\nu}\right) \end{bmatrix} \\
& \mathbf{M} - \begin{bmatrix} \frac{1}{2}\sum_{t=1}^T \mathbf{P}_t^\top \boldsymbol{w}\boldsymbol{w}^\top \mathbf{P}_t & -\frac{1}{2}\sum_{t=1}^T \mathbf{P}_t^\top \boldsymbol{w} \\ -\frac{1}{2}\left(\sum_{t=1}^T \mathbf{P}_t^\top \boldsymbol{w}\right)^\top & \gamma T - \beta \end{bmatrix} \succeq \\
& \qquad \lambda \begin{bmatrix} -\sum_{t=1}^T \mathbf{P}_t^\top \boldsymbol{\Lambda}^{-1} \mathbf{P}_t & \sum_{t=1}^T \mathbf{P}_t^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{\nu} \\ \left(\sum_{t=1}^T \mathbf{P}_t^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{\nu}\right)^\top & T\left(\delta - \boldsymbol{\nu}^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{\nu}\right) \end{bmatrix}
\end{aligned}
\tag{2.27}
$$

Here, the truncation operators $\mathbf{P}_t$, $t = 1, \ldots, T$, are defined as in Section 2.3. We emphasize that even though the SDP (2.27) 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 2.3 *without* support information. This can be seen by fixing $\alpha = \lambda = 0$, in which case (2.27) reduces to the SDP (2.6). Note that the SDP (2.27) 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 (2.27) can be substantially simplified by exploiting its inherent temporal symmetry. To this end, we first introduce the notion of a block compound symmetric matrix.

**Definition 2.6** (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*

$$
\mathbf{M} = \left[\begin{array}{cccc|c} \mathbf{A} & \mathbf{B} & \cdots & \mathbf{B} & \boldsymbol{c} \\ \mathbf{B} & \mathbf{A} & \cdots & \mathbf{B} & \boldsymbol{c} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{B} & \mathbf{B} & \ldots & \mathbf{A} & \boldsymbol{c} \\ \hline \boldsymbol{c}^\top & \boldsymbol{c}^\top & \cdots & \boldsymbol{c}^\top & d \end{array}\right]
\tag{2.28}
$$

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

Next, we argue that without any loss of generality the decision matrix $\mathbf{M}$ in (2.27) may be assumed to be block compound symmetric.

**Proposition 2.5.** *There exists a maximizer* $(\mathbf{M}, \alpha, \beta, \gamma, \lambda)$ *of (2.27) where* $\mathbf{M}$ *is block compound symmetric with blocks of size* $n \times n$.

*Proof.* The proof widely parallels that of Proposition 2.3 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 2.6.** *For any matrix* $\mathbf{K} \in \mathbb{S}^{nT}$ *of the form*

$$\mathbf{K} = \begin{bmatrix} \mathbf{A} & \mathbf{B} & \cdots & \mathbf{B} \\ \mathbf{B} & \mathbf{A} & \cdots & \mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B} & \mathbf{B} & \cdots & \mathbf{A} \end{bmatrix} \tag{2.29}$$

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

*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 $[\boldsymbol{v}_i^\top, \boldsymbol{v}_i^\top, \ldots, \boldsymbol{v}_i^\top]^\top$ is an eigenvector of $\mathbf{K}$ with eigenvalue $\lambda_i$. 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 $\mathbf{1}^\top \boldsymbol{k} = 0$ 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, $\left[k_1 \boldsymbol{u}_i^\top, k_2 \boldsymbol{u}_i^\top, \ldots, k_T \boldsymbol{u}_i^\top\right]^\top$ is an eigenvector of $\mathbf{K}$ with eigenvalue $\theta_i$. Hence, there are $T-1$ linearly independent eigenvectors that share the same eigenvalue $\theta_i$. 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 $\mathrm{eig}(\mathbf{K}) = \mathrm{eig}(\mathbf{A} - \mathbf{B}) \cup \mathrm{eig}(\mathbf{A} + (T-1)\mathbf{B})$. ∎

**Proposition 2.7.** *For any block compound symmetric matrix $\mathbf{M} \in \mathbb{S}^{nT+1}$ of the form (2.28), the following equivalence holds.*

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

*Proof.* For ease of exposition, we set

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

where $\boldsymbol{c}' = [\boldsymbol{c}^\top, \boldsymbol{c}^\top, \ldots, \boldsymbol{c}^\top]^\top$, and $\mathbf{K}$ is the block matrix defined in (2.29). 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 2.6. Thus, the claim follows. Assume next that $d \neq 0$. Then,

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

where the first and the last equivalences follow from standard Schur complement arguments, while the third equivalence holds due to Proposition 2.6. Thus, the claim follows again. ∎

By Proposition 2.5, we may assume without any loss of generality that the decision variable $\mathbf{M}$ is of the form (2.28), 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 2.7 then allows us to re-express each $(nT+1)$-dimensional SDP constraint in (2.27) 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 can thus reformulate the SDP (2.27) as

$$
\begin{aligned}
\max \quad & \gamma \\
\text{s.t.} \quad & \mathbf{A} \in \mathbb{S}^n,\ \mathbf{B} \in \mathbb{S}^n,\ \boldsymbol{c} \in \mathbb{R}^n,\ d \in \mathbb{R},\ \alpha \in \mathbb{R},\ \beta \in \mathbb{R},\ \gamma \in \mathbb{R},\ \lambda \in \mathbb{R} \\
& \alpha \geq 0,\ \beta \leq 0,\ \lambda \geq 0 \\
& \beta + \tfrac{1}{\epsilon}\left(T\langle \mathbf{A}, \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\top\rangle + T(T-1)\langle \mathbf{B}, \boldsymbol{\mu}\boldsymbol{\mu}^\top\rangle + 2T\boldsymbol{c}^\top\boldsymbol{\mu} + d\right) \leq 0 \\
& \begin{bmatrix} \mathbf{A} + (T-1)\mathbf{B} & \boldsymbol{c} \\ \boldsymbol{c}^\top & \frac{d}{T} \end{bmatrix} \succeq \alpha \begin{bmatrix} -\boldsymbol{\Lambda}^{-1} & \boldsymbol{\Lambda}^{-1}\boldsymbol{\nu} \\ \boldsymbol{\nu}^\top\boldsymbol{\Lambda}^{-1} & \delta - \boldsymbol{\nu}^\top\boldsymbol{\Lambda}^{-1}\boldsymbol{\nu} \end{bmatrix} \\
& \mathbf{A} - \mathbf{B} \succeq -\alpha\boldsymbol{\Lambda}^{-1} \\
& \begin{bmatrix} \mathbf{A} + (T-1)\mathbf{B} + \lambda\boldsymbol{\Lambda}^{-1} & c - \lambda\boldsymbol{\Lambda}^{-1}\boldsymbol{\nu} & \boldsymbol{w} \\ c^\top - \lambda\boldsymbol{\nu}^\top\boldsymbol{\Lambda}^{-1} & \frac{1}{2} + \frac{d+\beta}{T} - \gamma - \lambda(\delta - \boldsymbol{\nu}^\top\boldsymbol{\Lambda}^{-1}\boldsymbol{\nu}) & -1 \\ \boldsymbol{w}^\top & -1 & 2 \end{bmatrix} \succeq \mathbf{0} \\
& \begin{bmatrix} \mathbf{A} - \mathbf{B} + \lambda\boldsymbol{\Lambda}^{-1} & \boldsymbol{w} \\ \boldsymbol{w}^\top & 2 \end{bmatrix} \succeq \mathbf{0}.
\end{aligned}
\tag{2.31}
$$

This simplified SDP provides a lower bound on $\mathrm{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w}))$ in the presence of support information. Note that the size of the SDP (2.31) scales only with the number of assets $n$ but not with $T$. This observation 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.

### 2.5.2 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

$$
\mathcal{U} = \left\{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \mathbb{R}^n \times \mathbb{S}^n : (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^\top \hat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) \leq \delta_1, \delta_3\hat{\boldsymbol{\Sigma}} \preceq \boldsymbol{\Sigma} \preceq \delta_2\hat{\boldsymbol{\Sigma}}\right\}, \tag{2.32}
$$

where $\delta_1 \geq 0$ reflects our confidence in the estimate $\hat{\boldsymbol{\mu}}$, while the parameters $\delta_2$ and $\delta_3$, $\delta_2 \geq 1 \geq \delta_3 > 0$, express our confidence in the estimate $\hat{\boldsymbol{\Sigma}}$. Guidelines for selecting $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\Sigma}}$, $\delta_1$, $\delta_2$ and $\delta_3$ 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 $\mathscr{U}$ and if the Assumptions (A1) and (A2) are satisfied for all $\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right) \in \mathscr{U}$, then, by Theorem 2.2, the worst-case VaR of the approximate portfolio growth rate $\tilde{\gamma}'_T(\boldsymbol{w})$ is given by

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

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

**Theorem 2.4.** *If (A1) and (A2) hold for all* $\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right) \in \mathscr{U}$, *then* $\mathrm{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w}))$ *is equal to*

$$\frac{1}{2}\left(1 - \left(1 - \boldsymbol{w}^\top \hat{\boldsymbol{\mu}} + \left(\sqrt{\delta_1} + \sqrt{\frac{(1-\epsilon)\,\delta_2}{\epsilon T}}\right)\left\|\hat{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{w}\right\|\right)^2 - \frac{\delta_2\,(T-1)}{\epsilon T}\boldsymbol{w}^\top \hat{\boldsymbol{\Sigma}}\boldsymbol{w}\right).$$

*Proof.* Recall that under Assumptions (A1) and (A2) the worst-case VaR (2.19) is increasing in the portfolio mean return $\boldsymbol{w}^\top\boldsymbol{\mu}$ and decreasing in the portfolio standard deviation $\left\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{w}\right\|$. Thus, the worst case of (2.19) is achieved at the minimum portfolio mean return

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

and at the maximum portfolio standard deviation

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

The claim now follows by substituting the above expressions into (2.19). ∎

We remark that maximizing $\mathrm{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w}))$ over $\boldsymbol{w}\in\mathscr{W}$ 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 $\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right) \in \mathscr{U}$ if and only if

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

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

## 2.6 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).

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

[2] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
[3] http://finance.yahoo.com

$\mathbb{P}_{\ln}$ 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 $\boldsymbol{w}_{\text{go}}$, which is defined as the unique maximizer of $\mathbb{E}_{\mathbb{P}_{\ln}}\left(\log(1+\boldsymbol{w}^\top\tilde{\boldsymbol{r}}_t)\right)$ over $\boldsymbol{w}\in\mathscr{W}=\{\boldsymbol{w}\in\mathbb{R}^n:\boldsymbol{w}\geq\mathbf{0},\ \mathbf{1}^\top\boldsymbol{w}=1\}$. By using the second-order Taylor expansion of the logarithm around 1, we may approximate $\boldsymbol{w}_{\text{go}}$ with $\hat{\boldsymbol{w}}_{\text{go}}=\arg\max_{\boldsymbol{w}\in\mathscr{W}}\boldsymbol{w}^\top\boldsymbol{\mu}-\frac{1}{2}\boldsymbol{w}^\top\left(\boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^\top\right)\boldsymbol{w}$. 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 $\boldsymbol{w}_{\text{rgo}}$, which is defined as the unique maximizer of $\text{WVaR}_\epsilon(\tilde{\gamma}'_T(\boldsymbol{w}))$ over $\mathscr{W}$ for $\epsilon=5\%$. Unless otherwise stated, the worst-case VaR is evaluated with respect to the weak sense white noise ambiguity set $\mathscr{P}$ with known first and second-order moments but without support information.

### 2.6.1 Synthetic Experiments

We first illustrate the relation between the parameters $T$ and $\epsilon$ 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 $\boldsymbol{w}_{\text{go}}$ and its approximation $\hat{\boldsymbol{w}}_{\text{go}}$ are virtually indistinguishable. We may thus identify the growth-optimal portfolio with $\hat{\boldsymbol{w}}_{\text{go}}$.

In Section 2.4 we have seen that each robust growth-optimal portfolio tailored to an investment horizon $T$ and violation probability $\epsilon$ is identical to a Markowitz portfolio $\boldsymbol{w}_\varrho=\arg\max_{\boldsymbol{w}\in\mathscr{W}}\boldsymbol{w}^\top\boldsymbol{\mu}-\frac{\varrho}{2}\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}$ for some risk aversion parameter $\varrho=\varrho(T,\epsilon)$. Table 2.1 shows $\varrho(T,\epsilon)$ for different values of $T$ and $\epsilon$, based on the means and covariances obtained from the 10Ind dataset. As expected from the discussion after Theorem 2.2, $\varrho(T,\epsilon)$ is decreasing in $T$ and $\epsilon$. 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 violation probabilities up to 25%. We

Table 2.1: 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 $\epsilon$. The reported values are specific to the 10Ind dataset.

| $T$ \ $\epsilon$ | 5% | 10% | 15% | 20% | 25% | $T$ \ $\epsilon$ | 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 | | | | | | |

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 $\varepsilon$.

In Section 2.4 we have also seen that the robust growth-optimal portfolio tailored to $T$ and $\epsilon$ can be interpreted as a fractional Kelly strategy $\boldsymbol{w}_\kappa = \arg\max_{\boldsymbol{w}\in\mathcal{W}} \boldsymbol{w}^\top\boldsymbol{\mu} - \frac{\kappa}{2}\boldsymbol{w}^\top\left(\boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^\top\right)\boldsymbol{w}$ for some risk aversion parameter $\kappa=\kappa(T,\epsilon)$. Table 2.2 shows $\kappa(T,\epsilon)$ for different values of $T$ and $\epsilon$ in the context of the 10Ind dataset. The fractional Kelly and Markowitz risk-aversion parameters display qualitatively similar dependencies on $T$ and $\epsilon$.

**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 2.1(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

Table 2.2: 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 $\epsilon$. The reported values are specific to the 10Ind dataset.

| $T$ \ $\epsilon$ | 5% | 10% | 15% | 20% | 25% | $T$ \ $\epsilon$ | 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 | | | | | | |

$T$. Thus, under the true distribution $\mathbb{P}_{\ln}$ 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.

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 $\mathbb{P}_{\ln}$. 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 $\boldsymbol{\mu}^\top \boldsymbol{w}_{\text{rgo}} / \sqrt{\boldsymbol{w}_{\text{rgo}}^\top \boldsymbol{\Sigma} \boldsymbol{w}_{\text{rgo}}}$ and $\boldsymbol{\mu}^\top \boldsymbol{w}_{\text{go}} / \sqrt{\boldsymbol{w}_{\text{go}}^\top \boldsymbol{\Sigma} \boldsymbol{w}_{\text{go}}}$, respectively, when $T$ tends to infinity. The boxplot in Figure 2.1(b) visualizes the distribution of

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

where $\widehat{SR}_{\text{rgo}}$ and $\widehat{SR}_{\text{go}}$ 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 $\mathbb{P}_{\text{ln}}$.

**Ambiguity Effects** We now perform a stress test inspired by Bertsimas et al. (2010a), where we contamine the lognormal distribution $\mathbb{P}_{\text{ln}}$ with the worst-case distributions for the classical and robust growth-optimal portfolios, respectively. More precisely, by Corollary 2.3 we can construct two near-worst-case distributions $\mathbb{P}_{\text{go}}$ and $\mathbb{P}_{\text{rgo}}$ satisfying

$$\begin{aligned}\mathbb{P}_{\text{go}}\text{-VaR}_{5\%}(\tilde{\gamma}_T(\boldsymbol{w}_{\text{go}})) &\leq \text{WVaR}_{5\%}(\tilde{\gamma}_T(\boldsymbol{w}_{\text{go}})) + \delta,\\ \mathbb{P}_{\text{rgo}}\text{-VaR}_{5\%}(\tilde{\gamma}_T(\boldsymbol{w}_{\text{rgo}})) &\leq \text{WVaR}_{5\%}(\tilde{\gamma}_T(\boldsymbol{w}_{\text{rgo}})) + \delta,\end{aligned}$$

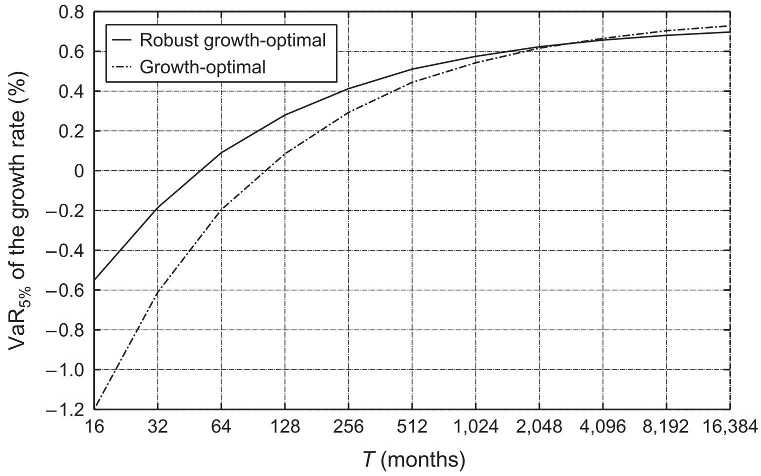
where $\delta$ is a small constant such as $10^{-6}$. We can then construct a contaminated distribution

$$\mathbb{P} = \psi_{\text{go}}\mathbb{P}_{\text{go}} + \psi_{\text{rgo}}\mathbb{P}_{\text{rgo}} + \left(1 - \psi_{\text{go}} - \psi_{\text{rgo}}\right)\mathbb{P}_{\text{ln}} \tag{2.33}$$

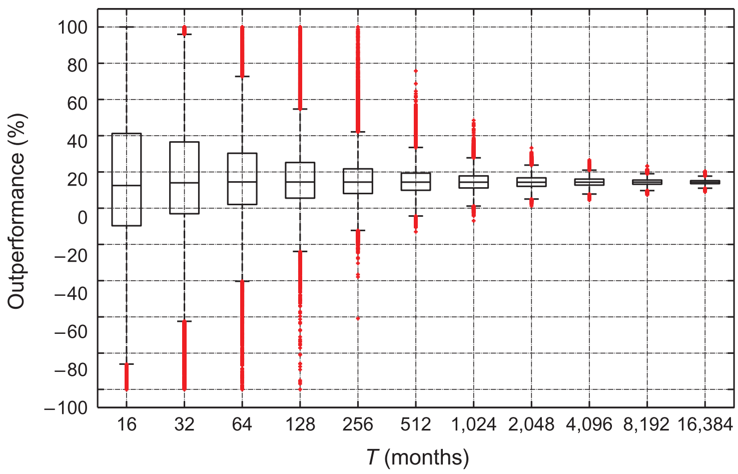
using the contamination weights $\psi_{\text{go}}, \psi_{\text{rgo}} \geq 0$ with $\psi_{\text{go}} + \psi_{\text{rgo}} \leq 1$. Note that $\mathbb{P} \in \mathscr{P}$ because $\mathbb{P}_{\text{go}}, \mathbb{P}_{\text{rgo}}, \mathbb{P}_{\text{ln}} \in \mathscr{P}$, which implies that the ambiguity-averse investors hedge against all distributions of the form (2.33). In contrast, the ambiguity-neutral investors exclusively account for the distribution with $\psi_{\text{go}} = \psi_{\text{rgo}} = 0$. 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

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

for all distributions of the form (2.33), where each VaR is evaluated using 250,000 samples from $\mathbb{P}$. The resulting percentage values are reported in Tables 2.3, 2.4, and 2.5 for investment horizons of 120 months, 360 months and 1,200 months, respectively.




(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 2.1: Comparison of the classical and robust growth-optimal portfolios under the lognormal distribution $\mathbb{P}_{\ln}$.

Table 2.3: Relative advantage (in %) of the robust growth-optimal portfolios in terms of 5% VaR ($T = 120$ months).

| $\psi_{\rm rgo}$ \ $\psi_{\rm go}$ | 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 | | | | | | | | | | |

We observe that the robust growth-optimal portfolios outperform their classical counterparts under *all* contaminated probability distributions of the form (2.33). 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.

### 2.6.2 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 (2.32). 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

Table 2.4: Relative advantage (in %) of the robust growth-optimal portfolios in terms of 5% VaR ($T = 360$ months).

| $\psi_{\text{rgo}}$ \ $\psi_{\text{go}}$ | 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 | | | | | | | | | | |

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 $\mathscr{W}$ 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 $\delta_1$, $\delta_2$ and $\delta_3$ characterizing the uncertainty set (2.32). 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

Table 2.5: Relative advantage (in %) of the robust growth-optimal portfolios in terms of 5% VaR ($T$ = 1,200 months).

| $\psi_{rgo}$ \ $\psi_{go}$ | 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 | | | | | | | | | | |

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 $\delta_1$ and $\delta_2$ such that the moment uncertainty set (2.32) 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 $\delta_3$ 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*:

$$\hat{r}_p = \frac{1}{T}\sum_{t=1}^{T}\left(\left(1+\boldsymbol{w}_t^\top \boldsymbol{r}_t\right)\left(1-c\sum_{i=1}^{n}\left|w_{t,i}-w_{t,i}^{-}\right|\right)-1\right).$$

2. *Standard deviation*:

$$\hat{\sigma}_p = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}\left(\left(1+\boldsymbol{w}_t^\top \boldsymbol{r}_t\right)\left(1-c\sum_{i=1}^{n}\left|w_{t,i}-w_{t,i}^{-}\right|\right)-1-\hat{r}_p\right)^2}.$$

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*:

$$\widehat{NR} = \widehat{V}_T, \quad \widehat{V}_t = \prod_{s=1}^{t}\left(1+\boldsymbol{w}_s^\top \boldsymbol{r}_s\right)\left(1-c\sum_{i=1}^{n}\left|w_{s,i}-w_{s,i}^{-}\right|\right).$$

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 2.6. 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 $\text{RGOP}^+$ 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 2.6 (in parenthesis). Star symbols (*) identify $p$-values that are significant at the 5% level. We observe that the $\text{RGOP}^+$ strategy achieves a significantly higher Sharpe ratio than all other benchmarks in the majority of experiments.

### 2.6.3 Discussion of the Results

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 chapter 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 $\epsilon$, 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

Table 2.6: 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 |

growth guarantees for all investment horizons up to $\sim 170$ years even if the classical growth-optimal portfolio has access to the true data-generating distribution. The out-performance 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.

# 3 Chebyshev Inequalities for Products of Random Variables

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

## 3.1 Introduction

The classical one-sided Chebyshev inequality (Bienaymeé 1853; Chebyshev 1867) for a random variable $\tilde{\xi}$ with mean $\mu$ and variance $\sigma^2$ can be represented as

$$\mathbb{P}(\tilde{\xi} \geq \gamma) \leq \begin{cases} \frac{\sigma^2}{\sigma^2+(\gamma-\mu)^2} & \text{if } \gamma \geq \mu, \\ 1 & \text{if } \gamma < \mu. \end{cases} \tag{3.1}$$

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

$$\mathbb{P}^\star = \begin{cases} \frac{\sigma^2}{\sigma^2+(\gamma-\mu)^2}\delta_\gamma + \frac{(\gamma-\mu)^2}{\sigma^2+(\gamma-\mu)^2}\delta_{\mu-\sigma^2/(\gamma-\mu)} & \text{if } \gamma > \mu, \\ \frac{\sigma^2}{\sigma^2+(\mu-\gamma)^2}\delta_\gamma + \frac{(\mu-\gamma)^2}{\sigma^2+(\mu-\gamma)^2}\delta_{\mu+\sigma^2/(\mu-\gamma)} & \text{if } \gamma < \mu. \end{cases} \tag{3.2}$$

In the degenerate case $\gamma = \mu$, the inequality (3.1) is still sharp because the distributions

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

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

If we have the extra information that the random variable $\tilde{\xi}$ is non-negative (and without much loss of generality that $\mu > 0$), then one can strengthen the Chebyshev inequality (3.1) to

$$\mathbb{P}(\tilde{\xi} \geq \gamma) \leq \begin{cases} \frac{\sigma^2}{\sigma^2+(\gamma-\mu)^2} & \text{if } \gamma \geq \mu + \sigma^2/\mu, \\ \frac{\mu}{\gamma} & \text{if } \mu \leq \gamma < \mu + \sigma^2/\mu, \\ 1 & \text{if } \gamma < \mu, \end{cases} \tag{3.3}$$

see, e.g., Godwin (1955); Shohat and Tamarkin (1943). The extremal distributions (3.2) are supported on the non-negative real line if either $\gamma \geq \mu + \sigma^2/\mu > \mu$ or if $\gamma < \mu$. Thus, they certify the sharpness of (3.3) in the respective parameter domains. For $\mu \leq \gamma < \mu + \sigma^2/\mu$ the Chebyshev inequality (3.3) for non-negative random variables reduces in fact to the classical Markov inequality $\mathbb{P}(\tilde{\xi} \geq \gamma) \leq \mu/\gamma$. In this Markov regime, the Chebyshev inequality (3.3) remains sharp because the distributions

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

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

In the rest of the chapter we consider a sequence of $T$ random variables $\tilde{\xi}_1, \tilde{\xi}_2, \ldots, \tilde{\xi}_T$ and assume that the first two moments of these random variables are known and permutation symmetric. Specifically, assume that all random variables share the same mean $\mu$ and variance $\sigma^2$, respectively, while all pairs of mutually distinct random variables share the same correlation coefficient $\rho$. Thus, the mean vector and the covariance matrix of $\tilde{\boldsymbol{\xi}} = (\tilde{\xi}_1, \ldots, \tilde{\xi}_T)^\intercal$ are given by

$$\boldsymbol{\mu} = \begin{bmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{bmatrix} \in \mathbb{R}^T \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \cdots & \rho\sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho\sigma^2 & \rho\sigma^2 & \cdots & \sigma^2 \end{bmatrix} \in \mathbb{S}^T, \tag{3.4}$$

respectively. Throughout the chapter we assume that $\sigma > 0$ and $-\frac{1}{T-1} < \rho < 1$. These conditions are necessary and sufficient for the covariance matrix $\boldsymbol{\Sigma}$ to be strictly positive definite. Note that $\tilde{\boldsymbol{\xi}}$ constitutes a weak-sense stationary stochastic process in the sense of Lindgren (2012).

An elementary calculation reveals that the sum $\sum_{t=1}^T \tilde{\xi}_t$ has mean value $T\mu$ and variance $T\sigma^2(1+(T-1)\rho)$. The classical Chebyshev inequality (3.1) applied to $\sum_{t=1}^T \tilde{\xi}_t$ thus implies

$$\mathbb{P}\left(\sum_{t=1}^T \tilde{\xi}_t \ge \gamma\right) \le \begin{cases} \frac{T\sigma^2(1+(T-1)\rho)}{T\sigma^2(1+(T-1)\rho)+(\gamma-T\mu)^2} & \text{if } \gamma \ge T\mu, \\ 1 & \text{if } \gamma < T\mu. \end{cases} \tag{3.5}$$

This inequality is still sharp due to a projection property of distribution families with compatible first and second moments. Indeed, for any distribution $\mathbb{P}_\zeta$ of a random variable $\tilde{\zeta}$ with mean value $T\mu$ and variance $T\sigma^2(1+(T-1)\rho)$ there exists a distribution $\mathbb{P}$ of the random vector $\tilde{\boldsymbol{\xi}}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ such that $\mathbb{P}_\zeta$ coincides with the marginal distribution of $\sum_{t=1}^T \tilde{\xi}_t$ under $\mathbb{P}$, that is, $\mathbb{P}_\zeta(\tilde{\zeta} \in B) = \mathbb{P}(\sum_{t=1}^T \tilde{\xi}_t \in B)$ for every Borel set $B \subseteq \mathbb{R}$ (Yu et al. 2009). The extremal distributions (3.2) certifying the sharpness of (3.1) can therefore be used to construct multivariate extremal distributions of $\tilde{\boldsymbol{\xi}}$ certifying the sharpness of (3.5). This result may be unexpected. Indeed, if $\tilde{\xi}_1, \ldots, \tilde{\xi}_T$ are independent and identically distributed, then, by the central limit theorem, their sum is approximately normally distributed with mean $T\mu$ and variance $T\sigma^2$. In contrast, if $\tilde{\xi}_1, \ldots, \tilde{\xi}_T$ are only known to be uncorrelated with a common mean and variance (but not necessarily independent and identically distributed), then, by the projection theorem, their sum may follow *any* distribution with mean $T\mu$ and variance $T\sigma^2$.

Assume now that $\tilde{\xi}_t$ is non-negative for every $t = 1, \ldots, T$ (and without much loss of generality that $\mu > 0$). As we will prove in Proposition 3.1 below, a distribution $\mathbb{P}$ supported on $\mathbb{R}_+^T$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ as given in (3.4) exists iff $\mu^2 + \rho\sigma^2 \geq 0$. We will assume that this condition holds throughout the rest of the chapter. In this setting, the generalized Chebyshev inequality (3.3) applied to the non-negative random variable $\sum_{t=1}^T \tilde{\xi}_t$ implies

$$\mathbb{P}(\textstyle\sum_{t=1}^T \tilde{\xi}_t \geq \gamma) \leq \begin{cases} \frac{T\sigma^2(1+(T-1)\rho)}{T\sigma^2(1+(T-1)\rho)+(\gamma-T\mu)^2} & \text{if } \gamma \geq T\mu + \sigma^2(1+(T-1)\rho)/\mu, \\ \frac{T\mu}{\gamma} & \text{if } T\mu \leq \gamma < T\mu + \sigma^2(1+(T-1)\rho)/\mu, \\ 1 & \text{if } \gamma < T\mu. \end{cases} \tag{3.6}$$

Even though the multivariate extension (3.6) of the univariate Chebyshev inequality (3.3) can still be shown to be sharp, we are not aware of an elementary proof; see Theorem 3.9 below.

In this chapter we aim to derive Chebyshev inequalities for *products* of non-negative random variables. Specifically, we will derive sharp upper bounds on the left and right tail probabilities $\mathbb{P}(\prod_{t=1}^T \tilde{\xi}_t \leq \gamma)$ and $\mathbb{P}(\prod_{t=1}^T \tilde{\xi}_t \geq \gamma)$, respectively. Products of random variables frequently arise in physics, statistics, finance, number theory and many other branches of science (Galambos and Simonelli 2004). Indeed, they are at the heart of stochastic models of many complex phenomena. When rocks are crushed, for example, the size of a fragment is multiplied by a random factor (that is smaller than 1) in every single breakup event (Frisch and Sornette 1997). Similar multiplicative phenomena explain the distribution of body weights, stock prices, the sizes of biological populations, income, rainfall etc. (Aitchison and Brown 1957).

Note that the stochastic process $\tilde{\boldsymbol{\pi}} = \{\tilde{\pi}_T\}_{T\in\mathbb{N}}$ defined through $\tilde{\pi}_T = \prod_{t=1}^T \tilde{\xi}_t$ can be interpreted as a geometric random walk driven by the weak-sense stationary process

$\tilde{\boldsymbol{\xi}} = \{\tilde{\xi}_t\}_{t\in\mathbb{N}}$. Chebyshev inequalities for the products of the $\tilde{\xi}_t$ thus provide tight bounds on the quantiles of a geometric random walk when there is limited distributional information. Consequently, they are potentially relevant for the many applications in economics and operations research, where geometric Brownian motions are traditionally used to model the prices of assets (Karatzas and Shreve 1991). An improved understanding of weak-sense geometric random walks may also stimulate new research directions in distributionally robust optimziation (Delage and Ye 2010; Goh and Sim 2010; Wiesemann et al. 2014) and optimal uncertainty quantification (Hanasusanto et al. 2015c; Owhadi et al. 2013).

**Remark 3.1** (Chebyshev in Log-Space)**.** *It seems natural to reduce Chebyshev inequalities for products of non-negative random variables to Chebyshev inequalities for their logarithms. Assume thus that the first two moments of the logarithmic random variables* $\tilde{\eta}_t = \log(\tilde{\xi}_t)$, $1,\ldots,T$*, are known and permutation symmetric. Specifically, denote by* $\mu_\eta$, $\sigma_\eta^2$ *and* $\rho_\eta$ *the mean, variance and correlation coefficient in log-space. Then, the Chebyshev inequality* (3.5) *for sums implies*

$$\mathbb{P}(\textstyle\prod_{t=1}^T \tilde{\xi}_t \geq \gamma) = \mathbb{P}(\sum_{t=1}^T \tilde{\eta}_t \geq \log\gamma) \leq \begin{cases} \frac{T\sigma_\eta^2(1+(T-1)\rho_\eta)}{T\sigma_\eta^2(1+(T-1)\rho_\eta)+(\log\gamma - T\mu_\eta)^2} & \textit{if } \log\gamma \geq T\mu_\eta, \\ 1 & \textit{if } \log\gamma < T\mu_\eta. \end{cases} \tag{3.7}$$

*Note that* (3.7) *is sharp because* (3.5) *is sharp. However, there is no one-to-one correspondence between the moments of the original and the logarithmic random variables. Even worse, it is possible that* $\mu$ *is finite while* $\mu_\eta = -\infty$ *(e.g., if* $\xi_t = 0$ *with positive probability), or that* $\mu_\eta$ *is finite while* $\mu = +\infty$ *(e.g., if* $\tilde{\xi}_t$ *follows a Pareto distribution with unit shape parameter). In this work we focus on the case where the* $\tilde{\xi}_t$ *have known finite first and second moments, and we explicitly allow the event* $\tilde{\xi}_t = 0$ *to have positive probability. This assumption can be crucial for truthfully capturing the bankruptcy risks in financial applications, for instance.*

The starting point of this chapter is the intriguing observation that modern optimization theory provides powerful tools for constructing and analyzing probability inequalities (Bertsimas and Popescu 2005). Assume for instance that we aim to find a sharp probability inequality for a target event characterized through finitely many polynomial inequalities on a random vector $\tilde{\boldsymbol{\xi}}$. Assume further that the desired inequality should hold for all distributions of $\tilde{\boldsymbol{\xi}}$ satisfying finitely many polynomial support and moment constraints. In the special case of the Chebyshev inequality (3.1), the target event corresponds to the set $\{\xi \in \mathbb{R} : \xi \geq \gamma\}$, while the relevant distribution family corresponds to the class of all distributions on $\mathbb{R}$ with mean $\mu$ and variance $\sigma^2$. Constructing the desired probability inequality is thus tantamount to maximizing the probability of the target event over the given distribution family. This leads to a generalized moment problem over probability measures. Under a mild regularity condition, this moment problem admits a strong dual linear program subject to polynomially parameterized semi-infinite constraints; see e.g. Isii (1960, 1962); Karlin and Studden (1966). A key insight of Bertsimas and Popescu (2005) is that this dual problem can be approximated systematically by tractable semidefinite programs. The resulting approximations are safe (i.e., they are guaranteed to provide *upper* bounds on the probability of the semialgebraic event). Moreover, these approximations are always tight in the univariate case but generically loose in the multivariate setting.

Stronger statements are available for probability inequalities that rely exclusively on first- and second-order moments. Specifically, if the support of the random vector $\tilde{\boldsymbol{\xi}}$ is unrestricted, the best upper bound on the probability of a *convex* target event is given by $1/(1+d^2)$, where $d$ represents the distance of the target event from the mean vector of $\tilde{\boldsymbol{\xi}}$ under the Mahalanobis norm induced by the covariance matrix of $\tilde{\boldsymbol{\xi}}$ (Marshall and Olkin 1960). More generally, if the target event constitutes a *union* of finitely many convex sets, over each of which convex quadratic optimization problems can be solved in polynomial time, then the best Chebyshev bound can be computed by an efficient algorithm reminiscent of the ellipsoid method of convex optimization (Bertsimas and Popescu 2005). Recently it has been observed that if the target event is defined by quadratic inequalities, the best Chebyshev bound coincides exactly with the optimal value of a single tractable semidefinite program (Vandenberghe et al. 2007). In spite of these encouraging results, the computation of Chebyshev bounds becomes hard in the presence of support constraints. Specifically, if $\tilde{\boldsymbol{\xi}}$ is supported on the non-negative orthant, it is already NP-hard to find sharp Chebyshev bounds for convex polyhedral target events (Bertsimas and Popescu 2005).

For a random vector $\tilde{\boldsymbol{\xi}}$ with zero mean and unrestricted support, the above methods have been used to derive a sharp Chebyshev bound on $\mathbb{P}(\prod_{t=1}^{T}\tilde{\xi}_t \geq 1, \tilde{\xi}_t > 0\ \forall t)$, which is expressed in terms of the solution of a tractable convex program (Marshall and Olkin 1960). As the $\tilde{\xi}_t$ are allowed to adopt negative values, however, we believe that the practical relevance of this bound is limited. In this chapter we aim to derive sharp Chebyshev bounds on $\mathbb{P}(\prod_{t=1}^{T}\tilde{\xi}_t \geq \gamma)$ and $\mathbb{P}(\prod_{t=1}^{T}\tilde{\xi}_t \leq \gamma)$ under the explicit assumption that $\tilde{\boldsymbol{\xi}}$ is supported on the non-negative orthant. Note that the second target event $\{\boldsymbol{\xi} \in \mathbb{R}_+^T : \prod_{t=1}^{T}\xi_t \leq \gamma\}$ is neither convex nor representable as a finite union of convex sets, nor representable through finitely many quadratic constraints in $\boldsymbol{\xi}$. Thus, none of the existing techniques could be used to bound its probability even if there were no support constraints. As support constraints generically lead to intractability (Bertsimas and Popescu 2005), we focus here on the special case where the first- and second-order moments are permutation-symmetric.

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

(i) If the distribution $\mathbb{P}$ of the non-negative random variables has mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ as given in (3.4), then the sharp upper Chebyshev bounds on $\mathbb{P}(\prod_{t=1}^{T}\tilde{\xi}_t \geq \gamma)$ and $\mathbb{P}(\prod_{t=1}^{T}\tilde{\xi}_t \leq \gamma)$ can both be expressed as the optimal values of explicit semidefinite programs, which are amenable to efficient numerical solution via interior point algorithms.

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

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

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

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

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

The rest of the chapter is structured as follows. In Section 3.2 we formalize the connection between probability inequalities and convex optimization. Left- and right-sided Chebyshev inequalities for products of random variables are then derived in Sections 3.3 and 3.4, respectively, while generalized Chebyshev inequalities that account for imprecise knowledge of the covariances are discussed in Section 3.5. Chebyshev inequalities for other permutation-symmetric functionals of the random variables are presented in Section 3.6, and examples are given in Section 3.7.

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

## 3.2 Optimization Perspective on Chebyshev Inequalities

To analyze probability bounds using tools from optimization, we first introduce an *ambiguity set* $\mathscr{P}$, that is, a family of distributions for which the desired probability bound should hold. In this chapter we mainly focus on the ambiguity set of all distributions supported on $\mathbb{R}_+^T$ that share the permutation-symmetric mean and covariance matrix defined in (3.4), that is, we set

$$\mathscr{P} = \left\{ \mathbb{P} \in \mathscr{M}_+(\mathbb{R}_+^T) \; : \; \mathbb{P}\left(\tilde{\boldsymbol{\xi}} \geq \mathbf{0}\right) = 1, \, \mathbb{E}_{\mathbb{P}}\left(\tilde{\boldsymbol{\xi}}\right) = \boldsymbol{\mu}, \, \mathbb{E}_{\mathbb{P}}\left(\tilde{\boldsymbol{\xi}}\tilde{\boldsymbol{\xi}}^\intercal\right) = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal \right\}. \tag{3.8}$$

We highlight that $\mathscr{P}$ is characterized by only four parameters: $T, \mu, \sigma, \rho$. Without much loss of generality, we assume henceforth that $\mu > 0$, $\sigma > 0$ and $-\frac{1}{T-1} < \rho < 1$. The last two conditions are equivalent to $\boldsymbol{\Sigma} \succ \mathbf{0}$. To rule out trivial special cases, we further restrict attention to $T \geq 2$. However, all of these conditions do not yet guarantee that $\mathscr{P}$ is non-empty. Proposition 3.1 below provides a necessary and sufficient condition for the non-emptiness of $\mathscr{P}$.

**Proposition 3.1** (Non-emptiness of $\mathscr{P}$)**.** *The ambiguity set $\mathscr{P}$ is non-empty iff $\mu^2 + \rho\sigma^2 \geq 0$.*

*Proof.* If $\mathscr{P}$ is non-empty, then any $\mathbb{P} \in \mathscr{P}$ satisfies

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

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

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

$$\mathbb{P}\left(\tilde{\boldsymbol{\xi}} = y\mathbf{1} + (x-y)\mathbf{e}_i\right) = \frac{p}{T}, \quad i = 1, \ldots, T, \quad \text{and} \quad \mathbb{P}\left(\tilde{\boldsymbol{\xi}} = z\mathbf{1}\right) = 1 - p \tag{3.9}$$

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

$$\begin{aligned}
&\text{(i)} \;\; \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\xi}}] = \boldsymbol{\mu} && \iff && \frac{p}{T}(x + (T-1)y) + (1-p)z = \mu; \\
&\text{(ii)} \;\; \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\xi}}\tilde{\boldsymbol{\xi}}^\intercal] = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal && \iff && \frac{p}{T}(x^2 + (T-1)y^2) + (1-p)z^2 = \mu^2 + \sigma^2, \\
& && && \frac{p}{T}(2xy + (T-2)y^2) + (1-p)z^2 = \mu^2 + \rho\sigma^2.
\end{aligned}$$

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

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

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

$$\begin{aligned}
&\text{(i')} \ \ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\xi}}] = \boldsymbol{\mu} && \iff && p m_1 + (1-p) z = \mu; \\
&\text{(ii')} \ \ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\xi}}\tilde{\boldsymbol{\xi}}^\top] = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\top && \iff && p m_2 + (1-p) z^2 = \mu^2 + \sigma^2, \\
& && && \frac{p}{T-1}(T m_1^2 - m_2) + (1-p) z^2 = \mu^2 + \rho\sigma^2.
\end{aligned}$$

In the remainder of the proof, we thus need to show that there exist $m_1, m_2, z \geq 0$, $m_1^2 \leq m_2 \leq T m_1^2$, and $p \in [0, 1]$ satisfying (i') and (ii'). To this end, consider the choice

$$p = \begin{cases} \min\left\{\frac{T\mu^2}{T\mu^2 + (1+(T-1)\rho)\sigma^2}, \frac{\rho T}{1+(T-1)\rho}\right\} & \text{if } \rho > 0, \\ \frac{T\mu^2}{T\mu^2+\sigma^2} & \text{if } \rho = 0, \\ \frac{-\rho T}{1-\rho} & \text{if } \rho < 0, \end{cases} \tag{3.10}$$

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

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

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

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

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

$$\begin{aligned} m_2 \leq Tm_1^2 \quad &\Longleftrightarrow \quad \sqrt{\frac{1-\rho}{pT}}\sigma \leq m_1 \\ &\Longleftrightarrow \quad \mu\sqrt{pT} + \left(\sqrt{(1+(T-1)\rho)(1-p)} - \sqrt{1-\rho}\right)\sigma \geq 0, \end{aligned} \tag{3.11}$$

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

For $\rho > 0$, we observe that the expression $\sqrt{(1+(T-1)\rho)(1-p)} - \sqrt{1-\rho}$ in (3.11) evaluates to 0 for $p = \frac{T\rho}{1+(T-1)\rho}$ and that it is decreasing in $p$. Since $\mu\sqrt{pT} \geq 0$ by construction, we thus conclude that the last inequality in (3.11) holds, and hence $m_2 \leq Tm_1^2$ when $\rho \geq 0$. In combination with (3.10) and (3.11), the above inequality ensures that $m_2 \leq Tm_1^2$.

For $\rho = 0$, equation (3.11) simplifies to

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

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

For $\rho < 0$, substituting $p$ in (3.11) with its definition from (3.10) yields

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

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

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

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

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

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

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

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

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

**Theorem 3.1** (Slater Condition)**.** *If* $\mu^2 + \rho\sigma^2 > 0$*, then the moment vector* $(1, \boldsymbol{\mu}, \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal)$ *is contained in the interior of the moment cone* $\mathcal{K}$ *defined through*

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

*Proof.* We first show that $\mathscr{P}$ contains a distribution of the form (3.9) where the inequalities $x \geq y \geq 0$, $z \geq 0$ and $p \in [0,1]$ hold *strictly*, as well as $x + (T-1)y > Tz$ (Step 1). This distribution allows us to show that $(1, \boldsymbol{\mu}, \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal)$ is in the relative interior of $\mathcal{K}_1 = \mathcal{K} \cap (\{1\} \times \mathbb{R}_+^T \times \mathbb{S}_+^T)$ (Step 2), from which the result follows directly by re-scaling the measures in $\mathcal{K}_1$ (Step 3).

**Step 1:** We distinguish the cases $\rho < 0$ and $\rho \geq 0$. For $\rho < 0$, one readily verifies that the choice of $p$, $x$, $y$ and $z$ in the proof of Proposition 3.1 satisfies $x > y > 0$, $z > 0$, $p \in (0,1)$ and $x + (T-1)y > Tz$ by construction. Moreover, these inequalities are also satisfied strictly for $\rho \geq 0$ if we replace $p$ in (3.10) with any value from the open interval $(0, p)$.

**Step 2:** To prove that $(1, \boldsymbol{\mu}, \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal) \in \operatorname{relint} \mathcal{K}_1$, we show that all perturbed ambiguity sets

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

with $\boldsymbol{\mu}^\epsilon \in \mathscr{B}_\epsilon(\boldsymbol{\mu})$ and $\boldsymbol{\Omega}^\epsilon \in \mathscr{B}_\epsilon(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal)$ are non-empty for sufficiently small $\epsilon$, where $\mathscr{B}_\epsilon(\boldsymbol{x})$ denotes the $\epsilon$-ball around $\boldsymbol{x}$ in the respective space. Note that the covariance matrix of any distribution in $\mathscr{P}(\boldsymbol{\mu}^\epsilon, \boldsymbol{\Omega}^\epsilon)$ is positive definite for small $\epsilon$ since $\boldsymbol{\Sigma} \succ \mathbf{0}$ and the eigenvalues are continuous functions of the second-order moment matrix. In the following, we construct a discrete distribution $\mathbb{P}^\epsilon \in \mathscr{P}(\boldsymbol{\mu}^\epsilon, \boldsymbol{\Omega}^\epsilon)$ with

$$\mathbb{P}^\epsilon\left(\tilde{\boldsymbol{\xi}} = \boldsymbol{\xi}^{\epsilon,i}\right) = \frac{p}{T} \quad i = 1, \ldots, T \quad \text{and} \quad \mathbb{P}^\epsilon\left(\tilde{\boldsymbol{\xi}} = \boldsymbol{\xi}^{\epsilon,T+1}\right) = 1 - p, \tag{3.12}$$

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

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

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

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

These moment conditions represent a system of nonlinear equations

$$\boldsymbol{F}(\boldsymbol{\mu}^\epsilon, \boldsymbol{\Omega}^\epsilon; \{\boldsymbol{\xi}^{\epsilon,i}\}_{i=1}^{T+1}) = \mathbf{0}$$

in the moments $\boldsymbol{\mu}^\epsilon$ and $\boldsymbol{\Omega}^\epsilon$ as well as the atoms $\boldsymbol{\xi}^{\epsilon,i}$, $i = 1,\dots,T+1$, of the distribution $\mathbb{P}^\epsilon$. From Step 1 we know that $\boldsymbol{F}(\boldsymbol{\mu}, \boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^\intercal; \{\boldsymbol{\xi}^i\}_{i=1}^{T+1}) = \mathbf{0}$ for $\boldsymbol{\xi}^i = y\mathbf{1} + (x-y)\mathbf{e}_i$, $i = 1,\dots,T$, $\boldsymbol{\xi}^{T+1} = z\mathbf{1}$ and for some $x, y, z \in \mathbb{R}_+$ satisfying $x > y > 0$, $z > 0$ and $x + (T-1)y > Tz$. Moreover, the implicit function theorem proves the existence of continuously differentiable functions $\boldsymbol{g}^i : \mathbb{R}_+^T \times \mathbb{S}_+^T \to \mathbb{R}^T$, $i = 1,\dots,T+1$, such that $\boldsymbol{F}(\boldsymbol{\mu}^\epsilon, \boldsymbol{\Omega}^\epsilon; \{\boldsymbol{g}^i(\boldsymbol{\mu}^\epsilon, \boldsymbol{\Omega}^\epsilon)\}_{i=1}^{T+1}) = \mathbf{0}$ for all $\boldsymbol{\mu}^\epsilon \in \mathscr{B}_\epsilon(\boldsymbol{\mu})$ and $\boldsymbol{\Omega}^\epsilon \in \mathscr{B}_\epsilon(\boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^\intercal)$, provided that $\epsilon$ is sufficiently small, $\boldsymbol{F}$ is continuously differentiable, and the Jacobian of $\boldsymbol{F}$ with respect to $\boldsymbol{\xi}^{\epsilon,i}$ has full row rank at $(\boldsymbol{\mu}^\epsilon, \boldsymbol{\Omega}^\epsilon, \{\boldsymbol{\xi}^{\epsilon,i}\}_{i=1}^{T+1}) = (\boldsymbol{\mu}, \boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^\intercal, \{\boldsymbol{\xi}^i\}_{i=1}^{T+1})$. Thus, the functions $\boldsymbol{g}^i$ allow us to construct distributions of the form (3.12) that satisfy the moment conditions of the perturbed ambiguity sets $\mathscr{P}(\boldsymbol{\mu}^\epsilon, \boldsymbol{\Omega}^\epsilon)$ for all $\boldsymbol{\mu}^\epsilon \in \mathscr{B}_\epsilon(\boldsymbol{\mu})$ and $\boldsymbol{\Omega}^\epsilon \in \mathscr{B}_\epsilon(\boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^\intercal)$. Since each $\boldsymbol{g}^i$ is continuous, we have $\boldsymbol{g}^i(\boldsymbol{\mu}^\epsilon, \boldsymbol{\Omega}^\epsilon) > \mathbf{0}$ for all $\boldsymbol{\mu}^\epsilon \in \mathscr{B}_\epsilon(\boldsymbol{\mu})$ and $\boldsymbol{\Omega}^\epsilon \in \mathscr{B}_\epsilon(\boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^\intercal)$ when $\epsilon$ is sufficiently small, that is, the support of $\mathbb{P}^\epsilon$ is contained in $\mathbb{R}_+^T$, and thus $\mathbb{P}^\epsilon$ is indeed contained in $\mathscr{P}(\boldsymbol{\mu}^\epsilon, \boldsymbol{\Omega}^\epsilon)$.

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

We then obtain

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

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

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

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

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

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

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

$$m_1 + zv_1 + z\sum_{t=2}^{T} c_{1t} = 0 \quad \text{and} \quad m_2 + zv_2 + z\left(c_{21} + \sum_{t=3}^{T} c_{2t}\right) = 0.$$

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

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

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

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

Theorem 3.1 will allow us to use the strong duality theorem of Shapiro (2001, Proposition 3.4), which states that a linear optimization problem over the distributions in $\mathscr{P}$ has the same optimal value as its associated dual problem. In the remainder of the chapter, we will make extensive use of this insight, and we therefore assume from now on that $\mu^2 + \rho\sigma^2 > 0$.

## 3.3 Left-Sided Chebyshev Bounds

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

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

where the ambiguity set $\mathscr{P}$ is defined in (3.8). We begin with the main result of this section.

**Theorem 3.2** (Left-Sided Chebyshev Bound). *Let* $\gamma > 0$. *For all* $T \geq 3$, *the left-sided Chebyshev bound* $\mathrm{L}(\gamma)$ *coincides with the optimal objective value of the semidefinite program*

$$
\begin{array}{ll}
\inf & \alpha + T\mu\beta + T(\mu^2+\sigma^2)\gamma_1 + T\left[T\mu^2+\sigma^2+(T-1)\rho\sigma^2\right]\gamma_2 \\
\text{s.t.} & \alpha,\beta,\gamma_1,\gamma_2 \in \mathbb{R},\ \lambda_1,\lambda_2,\lambda_3 \geq 0,\ \boldsymbol{p} \in \mathbb{R}^{2T+1},\ \boldsymbol{P} \in \mathbb{S}_+^{T+1},\ \boldsymbol{q} \in \mathbb{R}^{2T-1},\ \boldsymbol{Q} \in \mathbb{S}_+^{T} \\
& \alpha \geq 1, \quad \gamma_1+\gamma_2 \geq 0, \quad \gamma_1 + T\gamma_2 \geq 0 \\
& \gamma_2 + \frac{\gamma_1}{T} + \alpha \geq \left\| \left(\beta-\lambda_1, \gamma_2 + \frac{\gamma_1}{T} - \alpha\right) \right\|_2 \\
& \gamma_2+\gamma_1+\alpha-1 \geq \left\| \left(\beta-\lambda_2, \gamma_2+\gamma_1-\alpha+1\right) \right\|_2 \\
& \gamma_2 + \frac{\gamma_1}{T} + \lambda_3 + \alpha - 1 \geq \left\| \left(\beta - \lambda_3 T\gamma^{1/T}, \gamma_2 + \frac{\gamma_1}{T} + \lambda_3 - \alpha + 1\right) \right\|_2 \\
& p_0 = (T-1)\gamma_1\gamma^{\frac{2}{T-1}} + (T-1)^2\gamma_2\gamma^{\frac{2}{T-1}}, \quad p_1 + q_0 = (T-1)\beta\gamma^{\frac{1}{T-1}} \\
& p_2 + q_1 = \alpha - 1, \quad p_T + q_{T-1} = 2(T-1)\gamma_2\gamma^{\frac{1}{T-1}}, \quad p_{T+1} + q_T = \beta \\
& p_{2T} = \gamma_1 + \gamma_2, \quad p_t + q_{t-1} = 0 \quad \forall t = 3,\ldots,T-1,T+2,\ldots,2T-1 \\
& p_t = \sum_{i+j=t} P_{i,j} \quad \forall t = 0,\ldots,2T, \quad q_t = \sum_{i+j=t} Q_{i,j} \quad \forall t = 0,\ldots,2T-2,
\end{array}
\tag{3.13}
$$

*where we use the convention that the entries of* $\boldsymbol{p}$, $\boldsymbol{P}$, $\boldsymbol{q}$ *and* $\boldsymbol{Q}$ *are numbered starting from* 0. *For* $T = 2$, $\mathrm{L}(\gamma)$ *is given by a variant of* (3.13) *where the constraints* $p_2 + q_1 = \alpha - 1$ *and* $p_T + q_{T-1} = 2(T-1)\gamma_2\gamma^{\frac{1}{T-1}}$ *are combined to* $p_2 + q_1 = \alpha - 1 + 2(T-1)\gamma_2\gamma^{\frac{1}{T-1}}$.

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

$$
\begin{array}{rll}
\mathrm{L}(\gamma) = & \sup & \int_{\mathbb{R}_+^T} 1_{\{\prod_{t=1}^T \xi_t \leq \gamma\}}\, \mathbb{P}(\mathrm{d}\boldsymbol{\xi}) \\
& \text{s.t.} & \mathbb{P} \in \mathcal{M}_+(\mathbb{R}_+^T) \\
& & \int_{\mathbb{R}_+^T} \mathbb{P}(\mathrm{d}\boldsymbol{\xi}) = 1 \\
& & \int_{\mathbb{R}_+^T} \boldsymbol{\xi}\, \mathbb{P}(\mathrm{d}\boldsymbol{\xi}) = \boldsymbol{\mu} \\
& & \int_{\mathbb{R}_+^T} \boldsymbol{\xi}\boldsymbol{\xi}^\intercal\, \mathbb{P}(\mathrm{d}\boldsymbol{\xi}) = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal.
\end{array}
\tag{3.14}
$$

This moment problem admits a strong conic dual in the Lagrange multipliers $\alpha \in \mathbb{R}$, $\boldsymbol{\beta} \in \mathbb{R}^T$ and $\Gamma \in \mathbb{S}^T$ corresponding to the normalization, mean and covariance constraints in (3.14), respectively, see Theorem 3.1 and Shapiro (2001, Proposition 3.4).

Recalling that $\boldsymbol{\mu} = \mu \mathbf{1}$ and $\boldsymbol{\Sigma} = (1-\rho)\sigma^2 \mathbb{I} + \rho\sigma^2 \mathbf{1}\mathbf{1}^\intercal$, the dual problem can be expressed as

$$\begin{aligned} \mathrm{L}(\gamma) = \ \inf \quad & \alpha + \mu \mathbf{1}^\intercal \boldsymbol{\beta} + \langle (1-\rho)\sigma^2 \mathbb{I} + \left(\mu^2 + \rho\sigma^2\right) \mathbf{1}\mathbf{1}^\intercal, \boldsymbol{\Gamma} \rangle \\ \text{s.t.} \quad & \alpha \in \mathbb{R},\ \boldsymbol{\beta} \in \mathbb{R}^T,\ \boldsymbol{\Gamma} \in \mathbb{S}^T \\ & \alpha + \boldsymbol{\xi}^\intercal \boldsymbol{\beta} + \boldsymbol{\xi}^\intercal \boldsymbol{\Gamma} \boldsymbol{\xi} \geq 0 \quad \forall \boldsymbol{\xi} \geq \mathbf{0} \\ & \alpha + \boldsymbol{\xi}^\intercal \boldsymbol{\beta} + \boldsymbol{\xi}^\intercal \boldsymbol{\Gamma} \boldsymbol{\xi} \geq 1 \quad \forall \boldsymbol{\xi} \geq \mathbf{0} :\ \textstyle\prod_{t=1}^T \xi_t \leq \gamma. \end{aligned} \tag{3.15}$$

By Lemma 3.1 below, the symmetry of problem (3.15) implies that we may restrict attention to permutation-symmetric solutions of the form $\left(\alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma}\right)$ with $\boldsymbol{\beta} = \beta \mathbf{1}$ and $\boldsymbol{\Gamma} = \gamma_1 \mathbb{I} + \gamma_2 \mathbf{1}\mathbf{1}^\intercal$ for some $\beta, \gamma_1, \gamma_2 \in \mathbb{R}$. Thus, problem (3.15) simplifies to

$$\begin{aligned} \mathrm{L}(\gamma) = \ \inf \quad & \alpha + T\mu\beta + T(\mu^2 + \sigma^2)\gamma_1 + T\left[T\mu^2 + \sigma^2 + (T-1)\rho\sigma^2\right]\gamma_2 \\ \text{s.t.} \quad & \alpha, \beta, \gamma_1, \gamma_2 \in \mathbb{R} \\ & \alpha + \beta\|\boldsymbol{\xi}\|_1 + \gamma_1 \|\boldsymbol{\xi}\|_2^2 + \gamma_2 \|\boldsymbol{\xi}\|_1^2 \geq 0 \quad \forall \boldsymbol{\xi} \geq \mathbf{0} \\ & \alpha + \beta\|\boldsymbol{\xi}\|_1 + \gamma_1 \|\boldsymbol{\xi}\|_2^2 + \gamma_2 \|\boldsymbol{\xi}\|_1^2 \geq 1 \quad \forall \boldsymbol{\xi} \geq \mathbf{0} :\ \textstyle\prod_{t=1}^T \xi_t \leq \gamma. \end{aligned} \tag{3.16}$$

Lemma 3.2 then implies that (3.16) can be reduced to

$$\begin{aligned} \mathrm{L}(\gamma) = \ \inf \quad & \alpha + T\mu\beta + T(\mu^2 + \sigma^2)\gamma_1 + T\left[T\mu^2 + \sigma^2 + (T-1)\rho\sigma^2\right]\gamma_2 \\ \text{s.t.} \quad & \alpha, \beta, \gamma_1, \gamma_2 \in \mathbb{R} \\ & \inf_{s \geq 0} \ \alpha + \beta s + \gamma_2 s^2 + \frac{\gamma_1}{T} s^2 \geq 0 \\ & \inf_{s \geq 0} \ \alpha + \beta s + \gamma_2 s^2 + \gamma_1 s^2 \geq 1 \\ & \inf_{s \geq 0} \ \alpha + \beta s + \gamma_2 s^2 + \gamma_1 s^2 f_T\left(0, \frac{\gamma}{s^T}\right) \geq 1. \end{aligned} \tag{3.17}$$

By assigning a Lagrange multiplier $\lambda_1 \geq 0$ to the constraint $s \geq 0$ and using the $\mathscr{S}$-lemma (Pólik and Terlaky 2007), the first constraint in (3.17) can be reformulated as the linear matrix inequality

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

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

where the first equivalence follows from the observation that a $2 \times 2$-matrix is positive semidefinite iff it has non-negative diagonal elements as well as a non-negative determinant, while the second equivalence uses a well-known reformulation of hyperbolic constraints as second-order cone constraints (Boyd and Vandenberghe 2004, p. 197). Similarly, the second constraint in (3.17) holds iff there exists $\lambda_2 \geq 0$ with

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

Lemma 3.3 below further allows us to decompose the third constraint in (3.17) into two simpler semi-infinite constraints.

$$\inf_{s \in [0, T\gamma^{1/T}]} \alpha + \beta s + \gamma_2 s^2 + \gamma_1 \frac{s^2}{T} \geq 1 \tag{3.18a}$$

$$\inf_{s \geq T\gamma^{1/T}} \left\{ \alpha + \beta s + \gamma_2 s^2 + \gamma_1 \min_{\underline{\xi}, \overline{\xi} \geq 0} \left\{ \underline{\xi}^2 + (T-1)\overline{\xi}^2 : \underline{\xi} + (T-1)\overline{\xi} = s,\ \underline{\xi}\,\overline{\xi}^{T-1} = \gamma \right\} \right\} \geq 1 \tag{3.18b}$$

As $s \in \left[0, T\gamma^{1/T}\right]$ iff $s(T\gamma^{1/T} - s) \geq 0$, we can once again use the $\mathscr{S}$-lemma to show that (3.18a) holds iff there exists $\lambda_3 \geq 0$ with

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

Finally, it remains to be shown that (3.18b) also admits a conic reformulation. To do so, we first argue that one can replace (3.18b) with

$$\inf_{s \geq T\gamma^{1/T}, \underline{\xi}, \overline{\xi} \geq 0} \left\{ \alpha + \beta s + \gamma_2 s^2 + \gamma_1 \left[ \underline{\xi}^2 + (T-1)\overline{\xi}^2 \right] : \underline{\xi} + (T-1)\overline{\xi} = s,\ \underline{\xi}\,\overline{\xi}^{T-1} = \gamma \right\} \geq 1 \tag{3.19}$$

without changing the optimal value of problem (3.17). If $\gamma_1 \geq 0$, then (3.19) is indeed equivalent to (3.18b). On the other hand, if $\gamma_1 < 0$, we find

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

which means that (3.18b) is implied by the second semi-infinite constraint in problem (3.17). By eliminating $s = \underline{\xi} + (T-1)\overline{\xi}$, the maximization problem on the left hand side of (3.19) reduces to

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

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

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

By setting $\kappa = \underline{\xi}^{1/(T-1)}$, we can further replace $\underline{\xi}$ and $\overline{\xi}$ with $\kappa^{T-1}$ and $\gamma^{1/(T-1)}/\kappa$, respectively. Using elementary manipulations, one can then show that (3.19) reduces to

$$
\begin{aligned}
\inf_{\kappa \geq 0} \ (T-1)\gamma_1 \gamma^{\frac{2}{T-1}} + (T-1)^2 \gamma_2 \gamma^{\frac{2}{T-1}} + (T-1)\beta\gamma^{\frac{1}{T-1}}\kappa + (\alpha-1)\kappa^2 \\
+ 2(T-1)\gamma_2\gamma^{\frac{1}{T-1}}\kappa^T + \beta\kappa^{T+1} + (\gamma_1+\gamma_2)\kappa^{2T} \ \geq \ 0. \qquad (3.20)
\end{aligned}
$$

Note that the objective of the maximization problem on the left hand side of (3.20) constitutes a polynomial of degree $2T$ in $\kappa$ and is therefore representable as $l(\kappa) =$

$\sum_{i=0}^{2T} a_i \kappa^i$, where

$$a_i = \begin{cases} (T-1)\gamma_1 \gamma^{\frac{2}{T-1}} + (T-1)^2 \gamma_2 \gamma^{\frac{2}{T-1}} & \text{if } i = 0, \\ (T-1)\beta\gamma^{\frac{1}{T-1}} & \text{if } i = 1, \\ \alpha - 1 & \text{if } i = 2, \\ 2(T-1)\gamma_2 \gamma^{\frac{1}{T-1}} & \text{if } i = T, \\ \beta & \text{if } i = T+1, \\ \gamma_1 + \gamma_2 & \text{if } i = 2T, \\ 0 & \text{otherwise.} \end{cases} \tag{3.21}$$

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

Constraint (3.19) thus requires the polynomial $l(\kappa)$ to be non-negative for all $\kappa \geq 0$. By the Markov-Lukacs Theorem (Krein and Nudelman 1977), this is equivalent to postulating that $l(\kappa)$ admits a sum-of-squares representation of the form $l(\kappa) = p(\kappa) + \kappa q(\kappa)$, where $p(\kappa) = \sum_{i=0}^{2T} p_i \kappa^i$ and $q(\kappa) = \sum_{i=0}^{2T-2} q_i \kappa^i$ are sum-of-squares polynomials of degrees $2T$ and $2T-2$, respectively. By matching the coefficients of all monomials, one verifies that the identity $l(\kappa) = p(\kappa) + \kappa q(\kappa)$ holds iff

$$p_0 = a_0, \quad p_t + q_{t-1} = a_t \quad \forall t = 1, \ldots, 2T-1 \quad \text{and} \quad p_{2T} = a_{2T}. \tag{3.22}$$

Moreover, by Nesterov (2000, Theorem 3), $p(\kappa)$ and $q(\kappa)$ are sum-of-squares polynomials iff there exist positive semidefinite matrices $\boldsymbol{P} \in \mathbb{S}_+^{T+1}$ and $\boldsymbol{Q} \in \mathbb{S}_+^{T}$ such that

$$p_t = \sum_{i+j=t} P_{i,j} \quad \forall t = 0, \ldots, 2T \quad \text{and} \quad q_t = \sum_{i+j=t} Q_{i,j} \quad \forall t = 0, \ldots, 2T-2. \tag{3.23}$$

Thus, (3.19) holds iff the conic constraints (3.22) and (3.23) are satisfied. The claim now follows by replacing the three semi-infinite constraints in (3.17) with their explicit conic reformulations. ∎

The proof of Theorem 3.2 relies on 4 auxiliary lemmas, which we prove next.

**Lemma 3.1.** *Problem (3.15) has a permutation symmetric minimizer* $(\alpha^\star, \boldsymbol{\beta}^\star, \boldsymbol{\Gamma}^\star)$ *that* *satisfies* $\boldsymbol{\beta}^\star = \beta^\star \mathbf{1}$ *and* $\boldsymbol{\Gamma}^\star = \gamma_1^\star \mathbb{I} + \gamma_2^\star \mathbf{1}\mathbf{1}^\intercal$ *for some* $\beta^\star, \gamma_1^\star, \gamma_2^\star \in \mathbb{R}$.

*Proof.* Let $\mathfrak{P}$ be the set of all permutations of the index set $\{1,\ldots,T\}$. For any $\pi \in \mathfrak{P}$ we denote by $\mathbf{P}_\pi \in \mathbb{R}^{T\times T}$ the permutation matrix defined through $(\mathbf{P}_\pi)_{ij} = 1$ if $\pi(i) = j$; $= 0$ otherwise. Let $(\alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma})$ by any optimal solution to (3.15), which exists by Shapiro (2001, Proposition 3.4). We first show that the permuted solution $(\alpha_\pi, \boldsymbol{\beta}_\pi, \boldsymbol{\Gamma}_\pi) = (\alpha, \mathbf{P}_\pi \boldsymbol{\beta}, \mathbf{P}_\pi \boldsymbol{\Gamma} \mathbf{P}_\pi^\intercal)$ is also optimal in (3.15). To this end, we observe that

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

where the first equality follows from the definition of $\alpha_\pi$, $\boldsymbol{\beta}_\pi$ and $\boldsymbol{\Gamma}_\pi$, the second equality exploits the cyclicity property of the trace scalar product, and the third equality holds due to the permutation symmetry of $\mathbf{1}$ and the fact that $\mathbf{P}_\pi^\intercal = \mathbf{P}_{\pi^{-1}} = \mathbf{P}_\pi^{-1}$. Thus, $(\alpha_\pi, \boldsymbol{\beta}_\pi, \boldsymbol{\Gamma}_\pi)$ has the same objective value as $(\alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma})$. To show that $(\alpha_\pi, \boldsymbol{\beta}_\pi, \boldsymbol{\Gamma}_\pi)$ is feasible in (3.15), we note that

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

where the first equivalence follows from the definition of $\alpha_\pi$, $\boldsymbol{\beta}_\pi$ and $\boldsymbol{\Gamma}_\pi$ and because $\mathbf{P}_\pi^\intercal = \mathbf{P}_\pi^{-1}$, the second equivalence holds because permutations are bijective, and the third equivalence relies on the permutation symmetry of the non-negative orthant. Thus, $(\alpha_\pi, \boldsymbol{\beta}_\pi, \boldsymbol{\Gamma}_\pi)$ satisfies the semi-infinite constraints in (3.15) whenever $(\alpha, \boldsymbol{\beta}, \boldsymbol{\Gamma})$ does. We conclude that $(\alpha_\pi, \boldsymbol{\beta}_\pi, \boldsymbol{\Gamma}_\pi)$ is feasible and thus optimal in (3.15) for every $\pi \in \mathfrak{P}$.

Due to the convexity of the (semi-infinite) linear program (3.15), the equally weighted average $(\alpha^\star, \boldsymbol{\beta}^\star, \boldsymbol{\Gamma}^\star) = \frac{1}{T!} \sum_{\pi \in \mathfrak{P}} (\alpha_\pi, \boldsymbol{\beta}_\pi, \boldsymbol{\Gamma}_\pi)$ constitutes another optimal solution. It is now clear that $\mathbf{P}_\pi \boldsymbol{\beta}^\star = \boldsymbol{\beta}^\star$ and $\mathbf{P}_\pi \boldsymbol{\Gamma}^\star \mathbf{P}_\pi^\intercal = \boldsymbol{\Gamma}^\star$ for any $\pi \in \mathfrak{P}$ since $\pi(\mathfrak{P}) = \mathfrak{P}$. Thus, the claim follows. ■

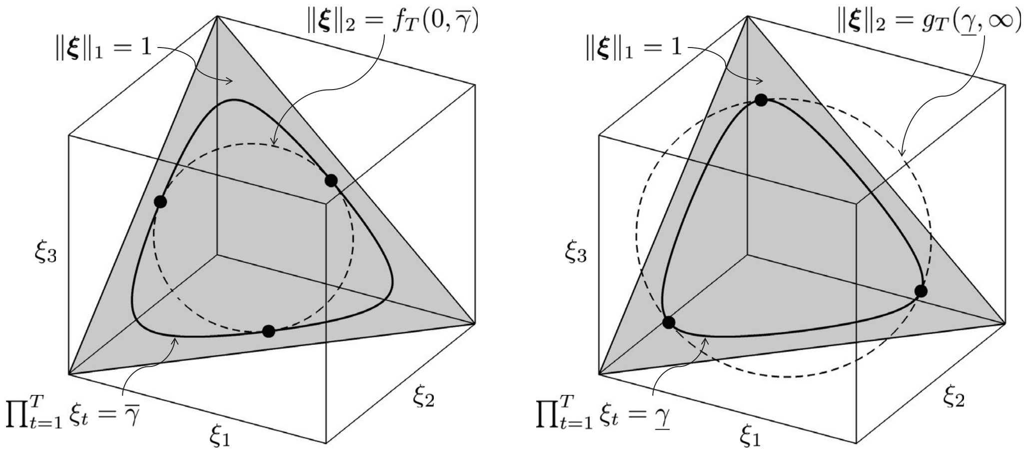



Figure 3.1: The subproblems (3.25a) (top) and (3.25b) (bottom) determine the smallest and the largest spheres centered at the origin that intersect with the hyperplane $\|\boldsymbol{\xi}\|_1 = 1$ (shaded areas) and the hyperbola $\prod_{t=1}^T \xi_t = \overline{\gamma}, \underline{\gamma}$ (solid lines). The dashed circles represent level sets of the objective function $\|\boldsymbol{\xi}\|_2^2$. Both graphs illustrate the case where $T = 3$.

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

$$\inf_{\boldsymbol{\xi} \geq \mathbf{0}} \left\{ \alpha + \beta \|\boldsymbol{\xi}\|_1 + \gamma_1 \|\boldsymbol{\xi}\|_2^2 + \gamma_2 \|\boldsymbol{\xi}\|_1^2 \,:\, \textstyle\prod_{t=1}^T \xi_t \in [\underline{\gamma}, \overline{\gamma}] \right\} \geq \Delta$$
$$\iff \begin{cases} \inf\limits_{s \geq T\underline{\gamma}^{1/T}} \; \alpha + \beta s + \gamma_2 s^2 + \gamma_1 s^2 \; f_T(\underline{\gamma}/s^T, \overline{\gamma}/s^T) \geq \Delta \\ \inf\limits_{s \geq T\underline{\gamma}^{1/T}} \; \alpha + \beta s + \gamma_2 s^2 + \gamma_1 s^2 \; g_T(\underline{\gamma}/s^T, \overline{\gamma}/s^T) \geq \Delta, \end{cases} \tag{3.24}$$

*where*

$$f_T(\underline{\gamma}, \overline{\gamma}) = \inf_{\boldsymbol{\xi} \geq \mathbf{0}} \left\{ \|\boldsymbol{\xi}\|_2^2 : \|\boldsymbol{\xi}\|_1 = 1,\ \textstyle\prod_{t=1}^T \xi_t \in [\underline{\gamma}, \overline{\gamma}] \right\} \tag{3.25a}$$

$$\textit{and} \quad g_T(\underline{\gamma}, \overline{\gamma}) = \sup_{\boldsymbol{\xi} \geq \mathbf{0}} \left\{ \|\boldsymbol{\xi}\|_2^2 : \|\boldsymbol{\xi}\|_1 = 1,\ \textstyle\prod_{t=1}^T \xi_t \in [\underline{\gamma}, \overline{\gamma}] \right\}. \tag{3.25b}$$

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

Figure 3.1 visualizes the two parametric subproblems (3.25a) and (3.25b). Note that both problems are non-convex whenever $\overline{\gamma} < \infty$ as their last constraints are equivalent to $(\prod_{t=1}^T \xi_t)^{1/T} \in [\underline{\gamma}^{1/T}, \overline{\gamma}^{1/T}]$ and because geometric means are concave (Boyd and

Vandenberghe 2004, § 3.1). Moreover, the subproblem (3.25b) remains non-convex for $\overline{\gamma} = \infty$ since it maximizes a convex objective function.

*Proof of Lemma 3.2:* The first constraint in (3.24) can be reduced to

$$\inf_{s \geq T\underline{\gamma}^{1/T}} \alpha + \beta s + \gamma_2 s^2 + \inf_{\boldsymbol{\xi} \geq \mathbf{0}} \left\{ \gamma_1 \|\boldsymbol{\xi}\|_2^2 : \|\boldsymbol{\xi}\|_1 = s,\ \prod_{t=1}^T \xi_t \in [\underline{\gamma}, \overline{\gamma}] \right\} \geq \Delta \tag{3.26}$$

by decomposing the maximization over all $\boldsymbol{\xi} \geq \mathbf{0}$ into two nested maximization problems over all $s \geq T\underline{\gamma}^{1/T}$ and over all $\boldsymbol{\xi} \geq \mathbf{0}$ with $\|\boldsymbol{\xi}\|_1 = s$, respectively. Here, the lower bound on $s$ is owed to the fact that there is $\boldsymbol{\xi} \geq \mathbf{0}$ satisfying $\|\boldsymbol{\xi}\|_1 = s$ and $\prod_{t=1}^T \xi_t \in [\underline{\gamma}, \overline{\gamma}]$ if and only if $s \geq T\underline{\gamma}^{1/T}$. A case distinction on the sign of $\gamma_1$ shows that constraint (3.26) holds if and only if

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

is satisfied. The change of variables $\boldsymbol{\xi} \to s\boldsymbol{\xi}$ shows that this constraint system is equivalent to the second constraint system in (3.24). Finally, we have $f_T(\underline{\gamma}, \infty) = 1/T$ for $\underline{\gamma} \leq T^{-T}$ and $g_T(0, \overline{\gamma}) = 1$ for $\overline{\gamma} \in \mathbb{R}_+ \cup \{\infty\}$ since the inequalities $\frac{1}{T}\|\boldsymbol{\xi}\|_1^2 \leq \|\boldsymbol{\xi}\|_2^2 \leq \|\boldsymbol{\xi}\|_1^2$ are tight for $\boldsymbol{\xi} = \frac{1}{T}\mathbf{1}$ and $\boldsymbol{\xi} = \mathbf{e}_i$, respectively. ∎

**Lemma 3.3.** *For $T \geq 2$, $\underline{\gamma} = 0$ and $\overline{\gamma} \geq 0$, the optimal value $f_T(0, \overline{\gamma})$ of* (3.25a) *equals*

$$f_T(0, \overline{\gamma}) = \begin{cases} \min_{\underline{\xi} \geq 0, \overline{\xi} \geq 0} \left\{ \underline{\xi}^2 + (T-1)\overline{\xi}^2 : \underline{\xi} + (T-1)\overline{\xi} = 1,\ \underline{\xi}\,\overline{\xi}^{T-1} = \overline{\gamma} \right\} & \text{if } 0 \leq \overline{\gamma} \leq T^{-T}, \\ \frac{1}{T} & \text{if } \overline{\gamma} > T^{-T}. \end{cases} \tag{3.27}$$

*Proof.* We first observe that the non-convex optimization problem (3.25a) is bounded below by its relaxation $\min_{\|\boldsymbol{\xi}\|_1 = 1} \|\boldsymbol{\xi}\|_2^2$. Note, however, that the optimal solution $\boldsymbol{\xi} = \frac{1}{T}\mathbf{1}$ of this relaxation is feasible and thus optimal in (3.25a) whenever $\overline{\gamma} \geq T^{-T}$. Thus, we have $f_T(0, \overline{\gamma}) = \frac{1}{T}$ for $\overline{\gamma} \geq T^{-T}$. For $0 \leq \overline{\gamma} < T^{-T}$, on the other hand, the product constraint $\prod_{t=1}^T \xi_t \leq \overline{\gamma}$ must be binding, for otherwise convex combinations of the optimal solution $\boldsymbol{\xi}$ with $\frac{1}{T}\mathbf{1}$ would improve the objective function of $f_T(0, \overline{\gamma})$, which is a contradiction. In summary, we thus find

$$f_T(0,\overline{\gamma}) = \begin{cases} \inf_{\boldsymbol{\xi}\geq \mathbf{0}} \left\{\|\boldsymbol{\xi}\|_2^2 : \|\boldsymbol{\xi}\|_1 = 1,\ \prod_{t=1}^T \xi_t = \overline{\gamma}\right\} & \text{if } 0 \leq \overline{\gamma} < T^{-T}, \\ \frac{1}{T} & \text{if } \overline{\gamma} \geq T^{-T}. \end{cases} \tag{3.28}$$

When $\overline{\gamma} = 0$, the product constraint in the first line of (3.28) can only be satisfied if $\xi_t = 0$ for at least one $t$. By permutation symmetry, we may assume without loss of generality that $\xi_T = 0$. Then, the product constraint is automatically satisfied and may be disregarded, implying that the minimization problem in the first line of (3.28) is solved by $\xi_1 = \xi_2 = \cdots = \xi_{T-1} = \frac{1}{T-1}$ and $\xi_T = 0$. We thus conclude that $f_T(0,0) = \frac{1}{T-1}$ and therefore

$$f_T(0,\overline{\gamma}) = \begin{cases} \frac{1}{T-1} & \text{if } \overline{\gamma} = 0, \\ \inf_{\boldsymbol{\xi}> \mathbf{0}} \left\{\|\boldsymbol{\xi}\|_2^2 : \|\boldsymbol{\xi}\|_1 = 1,\ \prod_{t=1}^T \xi_t = \overline{\gamma}\right\} & \text{if } 0 < \overline{\gamma} < T^{-T}, \\ \frac{1}{T} & \text{if } \overline{\gamma} \geq T^{-T}. \end{cases} \tag{3.29}$$

We now study the non-convex parametric optimization problem

$$\min_{\boldsymbol{\xi}>\mathbf{0}} \left\{\|\boldsymbol{\xi}\|_2^2 : \|\boldsymbol{\xi}\|_1 = 1,\ \prod_{t=1}^T \xi_t = \overline{\gamma}\right\} \tag{3.30}$$

on the domain $0 < \overline{\gamma} < T^{-T}$. Observe that (3.30) has a non-empty compact feasible set for any admissible $\overline{\gamma}$ and is therefore solvable. Assigning Lagrange multipliers $a$ and $b$ to the norm and product constraints, respectively, we find that any optimal solution to (3.30) must satisfy the stationarity conditions

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

where the equivalence follows from primal feasibility. Note that each $\xi_t$ needs to satisfy an identical quadratic equation, which must have two distinct positive real roots[1] $\underline{\xi}$ and $\overline{\xi}$. The roots depend on $a$, $b$ and $\overline{\gamma}$, but this dependence is notationally suppressed to avoid clutter. At optimality, the decision variables $\xi_1, \xi_2 \ldots, \xi_T$ can thus be partitioned into two groups, where all variables in the first group are equal to $\underline{\xi}$,

[1]The existence of at least one real root is guaranteed because (3.30) is solvable and because any optimal solution must satisfy the stationarity conditions. In fact, the stationarity conditions must admit *two* distinct positive real roots because otherwise $\boldsymbol{\xi} = \frac{1}{T}\mathbf{1}$ would be the only conceivable optimal solution, which is impossible for $\overline{\gamma} < T^{-T}$.

and all variables in the second group are equal to $\overline{\xi}$. This structural insight allows us to simplify problem (3.30). Indeed, by permutation symmetry, it is sufficient to consider only solutions that satisfy $\xi_1 = \cdots = \xi_k = \underline{\xi}$ and $\xi_{k+1} = \cdots = \xi_T = \overline{\xi}$ for some $\underline{\xi}, \overline{\xi} > 0$ and for some $k \in \{1, \ldots, \lfloor \frac{T}{2} \rfloor\}$. Thus, the optimal value of (3.30) coincides with

$$\min_{k \in \{1,\ldots,\lfloor \frac{T}{2} \rfloor\}} f_{T,k}(\overline{\gamma}), \tag{3.31}$$

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

$$f_{T,k}(\overline{\gamma}) = \min_{\underline{\xi}>0, \overline{\xi}>0} \left\{ k\underline{\xi}^2 + (T-k)\overline{\xi}^2 : k\underline{\xi} + (T-k)\overline{\xi} = 1,\ \underline{\xi}^k \overline{\xi}^{T-k} = \overline{\gamma} \right\}. \tag{3.32}$$

By Lemma 3.4 below, the optimal value of (3.31) is given by $f_{T,1}(\overline{\gamma})$. Hence, if we replace the minimization problem in (3.29) with $f_{T,1}(\overline{\gamma})$, we obtain

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

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

**Lemma 3.4.** *For $T \geq 2$ and $0 < \overline{\gamma} < T^{-T}$, the optimal value of* (3.31) *is given by $f_{T,1}(\overline{\gamma})$.*

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

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

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

where the second equality follows from the substitution $\underline{\xi} \leftarrow 2\underline{\xi}$ and $\overline{\xi} \leftarrow 2\overline{\xi}$, and the last equality holds because $f_{2,1}(\overline{\gamma}) = 1 - 2\overline{\gamma}$ for any $\overline{\gamma} \in (0, 2^{-2})$, which can be verified by direct calculation. Thus, we need to show that $f_{4,1}(\overline{\gamma}) < \frac{1}{2} - 4\sqrt{\overline{\gamma}}$, where

$$\begin{aligned} f_{4,1}(\overline{\gamma}) &= \min_{\underline{\xi}>0,\overline{\xi}>0} \left\{ \underline{\xi}^2 + 3\overline{\xi}^2 : \underline{\xi} + 3\overline{\xi} = 1,\ \underline{\xi}\overline{\xi}^3 = \overline{\gamma} \right\} \\ &= \min_{\overline{\xi}>0} \left\{ (1-3\overline{\xi})^2 + 3\overline{\xi}^2 : (1-3\overline{\xi})\overline{\xi}^3 = \overline{\gamma} \right\}. \end{aligned} \tag{3.33}$$

It is therefore sufficient to find $\overline{\xi}^\star$ feasible in (3.33) with

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

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

$$(1-3\zeta^-)(\zeta^-)^3 - \overline{\gamma} > 0 \quad \text{and} \quad (1-3\zeta^+)(\zeta^+)^3 - \overline{\gamma} < 0. \tag{3.34}$$

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

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

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

Fix now some $T \geq 5$ and assume for the sake of argument that there exist $k \in \{2, \ldots, \lfloor \frac{T}{2} \rfloor\}$ and $\overline{\gamma} \in (0, T^{-T})$ with $f_T(0, \overline{\gamma}) = f_{T,k}(\overline{\gamma}) < f_{T,1}(\overline{\gamma})$. Hence, there are some $\underline{\xi} > 0$ and $\overline{\xi} > 0$ with $\underline{\xi} \neq \overline{\xi}$ such that the minimum of $f_T(0, \overline{\gamma})$ in (3.25) is attained by the solution $\xi_1 = \cdots = \xi_k = \underline{\xi}$ and $\xi_{k+1} = \cdots = \xi_T = \overline{\xi}$. Fixing $\xi_1, \ldots, \xi_{k-2}$ and $\xi_{k+3}, \ldots, \xi_T$ at their optimal values and optimizing only over the remaining four decision variables in $f_T(0, \overline{\gamma})$ yields

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

Defining the strictly positive constant $c = 1 - (k-2)\underline{\xi} - (T-k-2)\overline{\xi} = 2\underline{\xi} + 2\overline{\xi}$ and using the substitution $y_t \leftarrow \xi_{k-2+t}/c$ for $t = 1, \ldots, 4$ further yields

$$
\begin{aligned}
f_T(0,\overline{\gamma}) \;&=\; (k-2)\underline{\xi}^2 + (T-k-2)\overline{\xi}^2 + \\
&\quad \min_{y_1,y_2,y_3,y_4 \geq 0} \left\{ \textstyle\sum_{t=1}^4 c^2 y_t^2 : \sum_{t=1}^4 y_t = 1,\ \prod_{t=1}^4 y_t \leq \frac{\overline{\gamma}}{c^4\,\underline{\xi}^{k-2}\,\overline{\xi}^{T-k-2}} \right\} \qquad (3.35)\\
&=\; (k-2)\underline{\xi}^2 + (T-k-2)\overline{\xi}^2 + c^2 f_4\left(0, \frac{\overline{\gamma}}{c^4\,\underline{\xi}^{k-2}\,\overline{\xi}^{T-k-2}}\right),
\end{aligned}
$$

where the second equality follows from the definition of $f_4(0, \overline{\gamma})$ in (3.25). By construction, the minimization problem in (3.35) must be solved by $y_1 = y_2 = \underline{\xi}$ and $y_3 = y_4 = \overline{\xi}$. However, this contradicts our previous results. In fact, we know that the solution of $f_4(0, \overline{\gamma})$ must have the following properties for $T = 4$. If $\overline{\gamma}/[c^4\,\underline{\xi}^{k-2}\,\overline{\xi}^{T-k-2}] < 4^{-4}$, then three out of the four $\xi_t$ variables must be equal at optimality. Conversely, if $\overline{\gamma}/[c^4\,\underline{\xi}^{k-2}\,\overline{\xi}^{T-k-2}] \geq 4^{-4}$, then all four $\xi_t$ variables must be equal. This contradicts our assumption that there exist $k \in \{2, \ldots, \lfloor \frac{T}{2} \rfloor\}$ and $\overline{\gamma} \in (0, T^{-T})$ with $f_T(0, \overline{\gamma}) = f_{T,k}(\overline{\gamma}) < f_{T,1}(\overline{\gamma})$. Thus, the assertion holds for all $T > 4$. ∎

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

**Theorem 3.3** (Certainty of Absorption)**.** *For* $T > \frac{\mu^2+\sigma^2}{(1-\rho)\sigma^2} + 1$ *we have* $\mathrm{L}(\gamma) = 1$ *for every* $\gamma > 0$.

*Proof.* From the proof of Proposition 3.1 we know that there exists a discrete distribution $\mathbb{P}_0 = \sum_{k\in\mathcal{K}} p_k \delta_{\boldsymbol{\xi}^k} \in \mathscr{P}$ with scenarios $\boldsymbol{\xi}^k$ and associated probabilities $p_k > 0$, where $k$ ranges over a finite index set $\mathcal{K}$ of cardinality $T+1$. By the permutation symmetry, any discrete distribution of the form $\mathbb{P}_0 \in \mathscr{P}$ can be used to construct a corresponding symmetric distribution

$$\mathbb{P} = \frac{1}{T!} \sum_{\pi\in\mathfrak{P}} \sum_{k\in\mathcal{K}} p_k \delta_{\mathbf{P}_\pi \boldsymbol{\xi}^k}, \tag{3.36}$$

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

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

where the first equality follows from the definition of $\mathbb{P}$ and because the $t$-th component of $\mathbf{P}_\pi \boldsymbol{\xi}^{(k)}$ is given by $\xi_{\pi(t)}^k$, while the third equality holds because there are $(T-2)!$ permutations that map $s$ to $r$ and $t$ to any fixed index different from $r$. Similarly, one can show that

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

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

$$\sum_{k\in\mathcal{K}} p_k = 1 \tag{3.37a}$$

$$\sum_{k\in\mathcal{K}} p_k m_1^k = \mu \tag{3.37b}$$

$$\sum_{k\in\mathcal{K}} p_k m_2^k = \mu^2 + \sigma^2 \tag{3.37c}$$

$$\sum_{k\in\mathcal{K}} \frac{p_k}{T-1}\left(T(m_1^k)^2 - m_2^k\right) = \mu^2 + \rho\sigma^2. \tag{3.37d}$$

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

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

$$\inf_{\boldsymbol{\xi}\geq \mathbf{0}} \left\{ \prod_{t=1}^T \xi_t : \frac{1}{T}\sum_{t=1}^T \xi_t = m_1^k,\ \frac{1}{T}\sum_{t=1}^T \xi_t^2 = m_2^k \right\}, \tag{3.38}$$

which depends parametrically on $m_1^k$ and $m_2^k$. By Lemma 3.5 (i) below, problem (3.38) is indeed solvable for every $k \in \mathcal{K}$. The new distribution with updated scenarios still belongs to $\mathscr{P}$ because we did not change $p_k$, $m_1^k$ and $m_2^k$, implying that the moment conditions (3.37) remain valid. To gain a better understanding of the updated distribution, we define the disjoint index sets

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

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

$$
\begin{aligned}
k \in \mathcal{K}^+ &\iff T \geq \frac{m_2^k}{(m_1^k)^2} \geq \frac{T}{T-1} \iff \frac{1}{T} \leq \frac{m_2^k - (m_1^k)^2}{m_2^k} \leq \frac{T-1}{T} \\
&\iff \prod_{t=1}^{T} \xi_t^k = 0
\end{aligned} \tag{3.39a}
$$

and

$$
\begin{aligned}
k \in \mathcal{K}^- &\iff 1 \leq \frac{m_2^k}{(m_1^k)^2} < \frac{T}{T-1} \iff 0 \leq \frac{m_2^k - (m_1^k)^2}{m_2^k} < \frac{1}{T} \\
&\iff \prod_{t=1}^{T} \xi_t^k > 0.
\end{aligned} \tag{3.39b}
$$

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

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

$$
m_2^k(\lambda) = \begin{cases} (1-\lambda) m_2^k + \lambda \frac{T}{T-1} (m_1^k)^2 & \text{for } k \in \mathcal{K}^+ \\ m_2^i + \lambda \sum_{k \in \mathcal{K}^+} \frac{p_k}{p_i} (m_2^k - \frac{T}{T-1} (m_1^k)^2) & \text{for } k = i \\ m_2^k & \text{for } k \in \mathcal{K}^- \backslash \{i\} \end{cases} \tag{3.40}
$$

By construction, $p_k$, $m_1^k$ and $m_2^k = m_2^k(\lambda)$ satisfy the moment conditions (3.37) for every $\lambda \in [0,1]$. As in Step 1, the scenario $\boldsymbol{\xi}^k(\lambda)$ is then chosen to be a minimizer of problem (3.38) with inputs $m_1^k$ and $m_2^k = m_2^k(\lambda)$. However, (3.38) could fail to be solvable for $\lambda \lesssim 1$, in which case the proposed construction would fail. Indeed, Lemma 3.5 (i) shows that (3.38) is only solvable when $1 \leq m_2^k(\lambda)/(m_1^k)^2 \leq T$. In the remainder we will demonstrate that there is $\lambda^\star \in (0,1)$ such that $\boldsymbol{\xi}^k(\lambda^\star)$ exists for every $k \in \mathcal{K}$ and such that all scenarios $k \in \mathcal{K}^+ \cup \{i\}$ are absorbing.

Subtracting (3.37d) from (3.37c) and dividing the difference by (3.37c) yields

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

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

$$\frac{\sum_{k\in\mathcal{K}^+} p_k\, m_2^k\,\frac{m_2^k-(m_1^k)^2}{m_2^k}+\sum_{k\in\mathcal{K}^-} p_k\, m_2^k\,\frac{m_2^k-(m_1^k)^2}{m_2^k}}{\sum_{k\in\mathcal{K}^+} p_k m_2^k+\sum_{k\in\mathcal{K}^-} p_k m_2^k}>\frac{1}{T}. \tag{3.41}$$

The expression on the left hand side of the above inequality represents a weighted average of the fractions $(m_2^k-(m_1^k)^2)/m_2^k$ across all $k\in\mathcal{K}$. Recall from (3.39a) and (3.39b) that the fractions indexed by $k\in\mathcal{K}^+$ are larger or equal to $1/T$, while those indexed by $k\in\mathcal{K}^-$ are strictly smaller than $1/T$. The inequality (3.41) asserts that the fractions corresponding to $k\in\mathcal{K}^+$ dominate those corresponding to $k\in\mathcal{K}^-$. Thus, (3.41) remains valid if we replace $\mathcal{K}^-$ with $\{i\}$, that is,

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

which is equivalent to

$$\frac{\sum_{k\in\mathcal{K}^+} p_k\frac{(m_1^k)^2}{T-1}+p_i\left(m_2^i+\sum_{k\in\mathcal{K}^+}\frac{p_k}{p_i}\left(m_2^k-\frac{T}{T-1}(m_1^k)^2\right)-(m_1^i)^2\right)}{\sum_{k\in\mathcal{K}^+} p_k\frac{T}{T-1}(m_1^k)^2+p_i\left(m_2^i+\sum_{k\in\mathcal{K}^+}\frac{p_k}{p_i}\left(m_2^k-\frac{T}{T-1}(m_1^k)^2\right)\right)}>\frac{1}{T}. \tag{3.42}$$

Using the notation introduced in (3.40), the inequality (3.42) can be reformulated as

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

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

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

As $i \in \mathcal{K}^-$, the relation (3.39b) further implies that

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

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

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

By construction, we thus have $1 \le m_2^k(\lambda^\star)/(m_1^k)^2 \le T$ for every $k \in \mathcal{K}$, which implies via Lemma 3.5 (i) that the corresponding scenarios $\boldsymbol{\xi}^k(\lambda^\star)$ are well-defined. Our construction also guarantees that $\frac{T}{T-1} \le m_2^k(\lambda^\star)/(m_1^k)^2 \le T$ for every $k \in \mathcal{K}^+ \cup \{i\}$, which implies via Lemma 3.5 (ii) that the corresponding scenarios $\boldsymbol{\xi}^k(\lambda^\star)$ are absorbing. Thus, by replacing $\boldsymbol{\xi}^k$ with $\boldsymbol{\xi}^k(\lambda^\star)$ in (3.36) we obtain a new distribution $\mathbb{P} \in \mathcal{P}$ with more absorbing scenarios. As the total number of scenarios is finite, we can repeat Step 2 finitely many times to construct a distribution $\mathbb{P} \in \mathcal{P}$ that has only absorbing scenarios. Thus, the claim follows. ∎

The proof of Theorem 3.3 relies on the following auxiliary result.

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

$$\inf_{\boldsymbol{\xi} \ge \mathbf{0}} \left\{ \prod_{t=1}^T \xi_t : \frac{1}{T} \sum_{t=1}^T \xi_t = m_1,\ \frac{1}{T} \sum_{t=1}^T \xi_t^2 = m_2 \right\}. \tag{3.43}$$

*Then, the following statements hold:*

*(i) Problem* (3.43) *is feasible and solvable iff $T \ge \frac{m_2}{m_1^2} \ge 1$.*

*(ii) The optimal value of* (3.43) *is zero iff $T \ge \frac{m_2}{m_1^2} \ge \frac{T}{T-1}$.*

*Proof.* As for assertion (i), assume that there is $\boldsymbol{\xi}$ feasible in (3.43). We then have

$\frac{1}{T}\sum_{t=1}^{T}\xi_t = m_1$, which implies that $Tm_1^2 \geq m_2 \geq m_1^2$ since $\|\boldsymbol{\xi}\|_1 \geq \|\boldsymbol{\xi}\|_2 \geq \frac{1}{\sqrt{T}}\|\boldsymbol{\xi}\|_1$. Conversely, if $T \geq \frac{m_2}{m_1^2} \geq 1$, we may define $\boldsymbol{\xi} = (z, \frac{m_1 T - z}{T-1}, \ldots, \frac{m_1 T - z}{T-1})$ for some $z \in [m_1, Tm_1]$ to be chosen later. By construction, we have $\frac{1}{T}\sum_{t=1}^{T}\xi_t = m_1$ irrespective of $z$, while

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

changes continuously from $m_1^2$ to $Tm_1^2$ when $z$ is swept from $m_1$ to $Tm_1$. Thus, by the intermediate value theorem, we may assume that $\frac{1}{T}\sum_{t=1}^{T}\xi_t^2 = m_2 \in [m_1^2, Tm_1^2]$ for some suitably chosen $z \in [m_1, Tm_1]$. We conclude that (3.43) is feasible whenever $T \geq \frac{m_2}{m_1^2} \geq 1$. In that case, however, (3.43) is also solvable as the objective function is continuous and the feasible set is compact.

To prove assertion (ii), we observe that the optimal value of (3.43) vanishes iff the problem admits a minimizer $\boldsymbol{\xi}$ with $\prod_{t=1}^{T}\xi_t = 0$. More precisely, by permutation symmetry, the minimum of (3.43) vanishes iff there exists $\boldsymbol{\xi}$ with $\xi_T = 0$, $\frac{1}{T}\sum_{t=1}^{T-1}\xi_t = m_1$ and $\frac{1}{T}\sum_{t=1}^{T-1}\xi_t^2 = m_2$. By assertion (i), however, the last two inequalities are satisfiable iff

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

and thus the claim follows. ■

## 3.4 Right-Sided Chebyshev Bounds

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

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

where the ambiguity set $\mathscr{P}$ is defined in (3.8). We first present the main result of this section.

**Theorem 3.4** (Right-Sided Chebyshev Bound)**.** *Let $\gamma > 0$. For all $T \geq 3$ the right-sided Chebyshev bound* $\mathrm{R}(\gamma)$ *coincides with the optimal objective value of the semidefinite* *program*

$$
\begin{array}{ll}
\inf & \alpha + T\mu\beta + T(\mu^2+\sigma^2)\gamma_1 + T(T\mu^2+\sigma^2+(T-1)\rho\sigma^2)\gamma_2 \\
\text{s.t.} & \alpha,\beta,\gamma_1,\gamma_2 \in \mathbb{R},\ \lambda_1,\lambda_2,\lambda_3 \geq 0,\ \boldsymbol{p}\in\mathbb{R}^{2T+1},\ \boldsymbol{P}\in\mathbb{S}_+^{T+1},\ \boldsymbol{q}\in\mathbb{R}^{2T-1},\ \boldsymbol{Q}\in\mathbb{S}_+^{T} \\
& \alpha \geq 0, \quad \alpha \geq 1-\lambda_3 T\gamma^{1/T}, \quad \gamma_1 + T\gamma_2 \geq 0, \quad \gamma_1+\gamma_2 \geq 0 \\
& \gamma_2 + \frac{\gamma_1}{T} + \alpha \geq \left\| \left( \beta-\lambda_1, \gamma_2 + \frac{\gamma_1}{T} - \alpha \right) \right\|_2 \\
& \gamma_2+\gamma_1+\alpha \geq \left\| (\beta-\lambda_2, \gamma_2+\gamma_1-\alpha) \right\|_2 \\
& \gamma_2 + \frac{\gamma_1}{T} + \lambda_3 T\gamma^{1/T} + \alpha - 1 \geq \left\| \left( \beta-\lambda_3, \gamma_2 + \frac{\gamma_1}{T} - \lambda_3 T\gamma^{1/T} - \alpha + 1 \right) \right\|_2 \\
& p_0 = (T-1)\gamma_1\gamma^{\frac{2}{T-1}} + (T-1)^2\gamma_2\gamma^{\frac{2}{T-1}}, \quad p_1 + q_0 = (T-1)\beta\gamma^{\frac{1}{T-1}} \\
& p_2 + q_1 = \alpha - 1, \quad p_T + q_{T-1} = 2(T-1)\gamma_2\gamma^{\frac{1}{T-1}}, \quad p_{T+1} + q_T = \beta \\
& p_{2T} = \gamma_1 + \gamma_2, \quad p_t + q_{t-1} = 0 \quad \forall t = 3,\ldots,T-1, T+2, \ldots, 2T-1 \\
& p_t = \sum_{i+j=t} P_{i,j} \quad \forall t = 0,\ldots,2T, \quad q_t = \sum_{i+j=t} Q_{i,j} \quad \forall t = 0,\ldots,2T-2,
\end{array}
\tag{3.44}
$$

*where we use the convention that the entries of* $\boldsymbol{p}$, $\boldsymbol{P}$, $\boldsymbol{q}$ *and* $\boldsymbol{Q}$ *are numbered starting from* 0. *For* $T=2$, $\mathrm{R}(\gamma)$ *is given by a variant of* (3.44) *where the constraints* $p_2+q_1=\alpha-1$ *and* $p_T+q_{T-1}=2(T-1)\gamma_2\gamma^{\frac{1}{T-1}}$ *are combined to* $p_2+q_1=\alpha-1+2(T-1)\gamma_2\gamma^{\frac{1}{T-1}}$.

*Proof.* Using similar arguments as in the proof of Theorem 3.2, one first shows that the worst-case probability problem $\sup_{\mathbb{P}\in\mathscr{P}} \mathbb{P}(\prod_{t=1}^T \tilde{\xi}_t \geq \gamma)$ admits a strong dual which constitutes a semi-infinite optimization problem. Exploiting this problem's permutation symmetry, one can further show that its optimal value amounts to

$$
\begin{array}{rll}
\mathrm{R}(\gamma) = & \inf & \alpha + T\mu\beta + T(\mu^2+\sigma^2)\gamma_1 + T\left[T\mu^2+\sigma^2+(T-1)\rho\sigma^2\right]\gamma_2 \\
& \text{s.t.} & \alpha,\beta,\gamma_1,\gamma_2 \in \mathbb{R} \\
& & \alpha + \beta\|\boldsymbol{\xi}\|_1 + \gamma_1\|\boldsymbol{\xi}\|_2^2 + \gamma_2\|\boldsymbol{\xi}\|_1^2 \geq 0 \quad \forall \boldsymbol{\xi} \geq \mathbf{0} \\
& & \alpha + \beta\|\boldsymbol{\xi}\|_1 + \gamma_1\|\boldsymbol{\xi}\|_2^2 + \gamma_2\|\boldsymbol{\xi}\|_1^2 \geq 1 \quad \forall \boldsymbol{\xi} \geq \mathbf{0} : \prod_{t=1}^T \xi_t \geq \gamma.
\end{array}
\tag{3.45}
$$

Details are omitted for brevity of exposition. Lemma 3.2 then implies that (3.45) reduces to

$$
\begin{aligned}
\mathrm{R}(\gamma) = \inf \quad & \alpha + T\mu\beta + T(\mu^2+\sigma^2)\gamma_1 + T(T\mu^2+\sigma^2+(T-1)\rho\sigma^2)\gamma_2 \\
\text{s.t.} \quad & \alpha,\beta,\gamma_1,\gamma_2 \in \mathbb{R} \\
& \inf_{s\geq 0} \alpha + \beta s + \gamma_2 s^2 + \frac{\gamma_1}{T} s^2 \geq 0 \\
& \inf_{s\geq 0} \alpha + \beta s + \gamma_2 s^2 + \gamma_1 s^2 \geq 0 \\
& \inf_{s\geq T\gamma^{1/T}} \alpha + \beta s + \gamma_2 s^2 + \frac{\gamma_1}{T} s^2 \geq 1 \\
& \inf_{s\geq T\gamma^{1/T}} \alpha + \beta s + \gamma_2 s^2 + \gamma_1 s^2 g_T\left(\frac{\gamma}{s^T},\infty\right) \geq 1.
\end{aligned} \tag{3.46}
$$

By leveraging the $\mathscr{S}$-lemma and a well-known reformulation of hyperbolic constraints as second-order cone constraints, one can use similar arguments as in the proof of Theorem 3.2 to show that the first three constraints in (3.46) hold iff there exist $\lambda_1,\lambda_2,\lambda_3 \geq 0$ satisfying

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

By Lemma 3.6 below, the last semi-infinite constraint in (3.46) can be re-expressed as

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

which is identical to (3.19). The claim then follows by replacing this constraint with its explicit semidefinite reformulation familiar from Theorem 3.2. ∎

The proof of Theorem 3.4 relies on 2 auxiliary lemmas, which we prove next.

**Lemma 3.6.** *For $T \geq 2$, $\overline{\gamma} = \infty$ and $\underline{\gamma} \geq 0$, the optimal value $g_T(\underline{\gamma},\infty)$ of* (3.25b) *equals*

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

*Proof.* If $\underline{\gamma} > T^{-T}$, then the maximization problem (3.25b) is infeasible due to the inequality of arithmetic and geometric means, and thus we have $g_T(\underline{\gamma},\infty) = -\infty$. For $\underline{\gamma} = T^{-T}$, the unique feasible solution of (3.25b) is $\boldsymbol{\xi} = \frac{1}{T}\mathbf{1}$, which implies that $g_T(\underline{\gamma},\infty) = \frac{1}{T}$. Moreover, for $\underline{\gamma} = 0$, the last constraint in (3.25b) becomes redundant. In this case $g_T(\underline{\gamma},\infty)$ is optimized by $\boldsymbol{\xi} = \mathbf{e}_i$, and thus we find $g_T(\underline{\gamma},\infty) = 1$. Lastly, for $0 < \underline{\gamma} < T^{-T}$, the maximization problem (3.25b) is feasible, and every feasible solution has strictly positive components. In addition, the product constraint $\prod_{t=1}^{T} \xi_t \geq \underline{\gamma}$ is binding at optimality for otherwise convex combinations of the optimal solution $\boldsymbol{\xi}$ with $\mathbf{e}_i$, where $i \in \arg\max\{\boldsymbol{\xi}_j : j = 1,\ldots,T\}$, would improve the objective function of (3.25b), which is a contradiction. We thus conclude that

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

As in the proof of Lemma 3.3, for $0 < \underline{\gamma} < T^{-T}$ one can use the optimality conditions of (3.25b) to show that

$$g_T(\underline{\gamma},\infty) = \max_{k\in\{1,\ldots,\lfloor \frac{T}{2} \rfloor\}} g_{T,k}(\underline{\gamma}), \tag{3.47}$$

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

$$g_{T,k}(\underline{\gamma}) = \max_{\underline{\xi}>0,\overline{\xi}>0} \left\{ k\underline{\xi}^2 + (T-k)\overline{\xi}^2 : k\underline{\xi} + (T-k)\overline{\xi} = 1,\ \underline{\xi}^k \overline{\xi}^{T-k} = \underline{\gamma} \right\}. \tag{3.48}$$

Lemma 3.7 below asserts that the maximum in (3.47) is attained at $k = 1$. We thus obtain

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

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

**Lemma 3.7.** *For $T \geq 2$ and $0 < \underline{\gamma} < T^{-T}$, the optimal value of* (3.47) *is given by $g_{T,1}(\underline{\gamma})$.*

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

We now show that in the extreme case, the weak-sense geometric random walk $\tilde{\boldsymbol{\pi}} = \{\tilde{\pi}_T\}_{T\in\mathbb{N}}$ defined through $\tilde{\pi}_T = \prod_{t=1}^{T} \tilde{\xi}_t$ weakly exceeds the deterministic growth process $\{\mu^T\}_{T\in\mathbb{N}}$ with certainty for any time horizon $T$, assuming that $\rho \geq 0$. The result can be viewed as the right-sided analogue of Theorem 3.3.

**Proposition 3.2.** *If $\rho \geq 0$, then $\mathrm{R}(\gamma) = 1$ for all $\gamma \leq \mu^T$.*

*Proof.* The objective function of problem (3.46) can be reformulated as

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

For $\gamma \leq \mu^T$, the first term equals the left hand side of the third semi-infinite constraint in (3.46) if we set $s = T\mu$, and it must therefore be greater than or equal to 1. In the second term, the factor $(\gamma_1 + (1 + (T-1)\rho)\gamma_2)$ can be expressed as the linear combination $\rho\cdot(\gamma_1 + T\gamma_2) + (1-\rho)\cdot(\gamma_1 + \gamma_2)$. For $\rho \geq 0$, this linear combination becomes a convex combination, and the claim follows since $\gamma_1 + T\gamma_2 \geq 0$ and $\gamma_1 + \gamma_2 \geq 0$ are explicit constraints in the equivalent reformulation (3.44). ∎

We highlight that Proposition 3.2 breaks down for $\rho < 0$.

## 3.5 Covariance Bounds

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

$$\mathscr{P}' = \left\{ \mathbb{P} \in \mathscr{M}_+(\mathbb{R}^T_+) \; : \; \mathbb{P}\left(\tilde{\boldsymbol{\xi}} \geq \mathbf{0}\right) = 1,\, \mathbb{E}_{\mathbb{P}}\left(\tilde{\boldsymbol{\xi}}\right) = \boldsymbol{\mu},\, \mathbb{E}_{\mathbb{P}}\left(\tilde{\boldsymbol{\xi}}\tilde{\boldsymbol{\xi}}^\intercal\right) \preceq \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal \right\},$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are defined as in Section 3.1. For $\gamma > 0$, we are then interested in quantifying *relaxed left-sided and right-sided Chebyshev bounds* of the form

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

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

**Theorem 3.5** (Relaxed Left-Sided Chebyshev Bound)**.** *The relaxed left-sided Chebyshev bound satisfies* $\mathrm{L}'(\gamma) = \mathrm{L}(\gamma)$ *for all* $\gamma > 0$.

*Proof.* By repeating the first few steps of the proof of Theorem 3.2, one can show that $\mathrm{L}'(\gamma)$ coincides with the optimal value of (3.15) with the extra constraint $\boldsymbol{\Gamma} \geq \mathbf{0}$. In this case Lemma 3.1 remains valid and implies that we can restrict attention to permutation-symmetric solutions of the form $\boldsymbol{\Gamma} = \gamma_1\mathbb{I} + \gamma_2\mathbf{1}\mathbf{1}^\intercal$ for some $\gamma_1, \gamma_2 \in \mathbb{R}$. As $\boldsymbol{\Gamma} = \gamma_1\mathbb{I} + \gamma_2\mathbf{1}\mathbf{1}^\intercal \succeq \mathbf{0}$ iff $\gamma_1 + T\gamma_2 \geq 0$ and $\gamma_1 \geq 0$ by virtue of Proposition 2.4, we may then conclude that $\mathrm{L}'(\gamma)$ coincides with the optimal value of (3.16) with the extra constraints $\gamma_1 + T\gamma_2 \geq 0$ and $\gamma_1 \geq 0$. Note that (3.16) is equivalent to (3.13) and (3.17). As $\gamma_1 + T\gamma_2 \geq 0$ is an explicit constraint of problem (3.13), it is necessarily an implicit constraint of the problems (3.16) and (3.17). Thus, $\mathrm{L}'(\gamma)$ coincides with the optimal value of (3.17) with the extra constraint $\gamma_1 \geq 0$. To prove the identity $\mathrm{L}(\gamma) = \mathrm{L}'(\gamma)$, it is therefore sufficient to show that appending the extra constraint $\gamma_1 \geq 0$ has no impact on the optimal value of (3.17).

To this end, fix any feasible solution of problem (3.17) with $\gamma_1 < 0$. As this solution must satisfy the constraint $\alpha + s\beta + s^2\gamma_2 + s^2\gamma_1 \geq 1$ for every $s \geq 0$ and as $s = T\mu > 0$, we have

$$\alpha + T\mu\beta + T^2\mu^2\gamma_2 + T^2\mu^2\gamma_1 \geq 1. \tag{3.49}$$

Moreover, the objective function of (3.17) can be reformulated as

$$
\begin{aligned}
&\alpha + T\mu\beta + T(\mu^2+\sigma^2)\gamma_1 + T\left[T\mu^2+\sigma^2+(T-1)\rho\sigma^2\right]\gamma_2 \\
&= \left(\alpha + T\mu\beta + T^2\mu^2\gamma_2 + T^2\mu^2\gamma_1\right) + T(1-T)\left(\mu^2+\rho\sigma^2\right)\gamma_1 + T\sigma^2(1+(T-1)\rho)(\gamma_1+\gamma_2),
\end{aligned}
$$

which constitutes a sum of three terms. The first term in the sum is greater than or equal to 1 because of (3.49), and the second term is strictly positive because $T \geq 2$, $\gamma_1 < 0$ and $\mu^2 + \rho\sigma^2 > 0$. The third term is non-negative because $\rho > -1/(T-1)$ and $\gamma_1 + \gamma_2 \geq 0$ is an explicit constraint of (3.13) and thus an implicit constraint of (3.17). In summary, we have shown that the objective value of any feasible solution of (3.17) with $\gamma_1 < 0$ is strictly greater than 1. As the optimal value $\mathrm{L}(\gamma)$ of (3.17) represents a probability, however, we conclude that no feasible solution with $\gamma_1 < 0$ can optimize (3.17). Thus, the extra constraint $\gamma_1 \geq 0$ does not change the optimal value of (3.17), and the claim follows. ∎

**Theorem 3.6** (Relaxed Right-Sided Chebyshev Bound)**.** *The relaxed right-sided Chebyshev bound admits the analytical solution*

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

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

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

$$
\begin{aligned}
\mathrm{R}'(\gamma) = \inf \quad & \alpha + \mu\mathbf{1}^\intercal\boldsymbol{\beta} + \langle(1-\rho)\sigma^2\mathbb{I} + \left(\mu^2+\rho\sigma^2\right)\mathbf{1}\mathbf{1}^\intercal, \boldsymbol{\Gamma}\rangle \\
\text{s.t.} \quad & \alpha \in \mathbb{R},\ \boldsymbol{\beta} \in \mathbb{R}^T,\ \boldsymbol{\Gamma} \in \mathbb{S}_+^T \\
& \alpha + \overline{\boldsymbol{\xi}}^\intercal\boldsymbol{\beta} + \overline{\boldsymbol{\xi}}^\intercal\boldsymbol{\Gamma}\overline{\boldsymbol{\xi}} \geq 0 \quad \forall \overline{\boldsymbol{\xi}} \geq \mathbf{0} \\
& \alpha + \underline{\boldsymbol{\xi}}^\intercal\boldsymbol{\beta} + \underline{\boldsymbol{\xi}}^\intercal\boldsymbol{\Gamma}\underline{\boldsymbol{\xi}} \geq 1 \quad \forall \underline{\boldsymbol{\xi}} \geq \mathbf{0} : \textstyle\prod_{t=1}^T \underline{\xi}_t \geq \gamma
\end{aligned}
\tag{3.50}
$$

Without loss of generality, we use different symbols $\overline{\boldsymbol{\xi}}$ and $\underline{\boldsymbol{\xi}}$ to denote the uncertain parameters in the two semi-infinite constraints, respectively. Note that (3.50) can be viewed as the robust counterpart of an uncertain convex program with constraint-wise uncertainty sets (Ben-Tal et al. 2009). As the left hand sides of the robust constraints are convex in the respective uncertainties, the '*primal worst equals dual best*' duality scheme portrayed in Beck and Ben-Tal (2009, Theorem 4.1) implies that (3.50) is equivalent to

$$\begin{aligned} \mathrm{R}'(\gamma) = \sup \quad & q \\ \text{s.t.} \quad & p, q \in \mathbb{R}_+,\ \overline{\boldsymbol{\xi}}, \underline{\boldsymbol{\xi}} \in \mathbb{R}^+_T,\ \textstyle\prod_{t=1}^T \underline{\xi}_t \geq \gamma \\ & p + q = 1 \\ & p\overline{\boldsymbol{\xi}} + q\underline{\boldsymbol{\xi}} = \mu \mathbf{1} \\ & p\overline{\boldsymbol{\xi}}\,\overline{\boldsymbol{\xi}}^\top + q\underline{\boldsymbol{\xi}}\,\underline{\boldsymbol{\xi}}^\top \preceq (1-\rho)\sigma^2 \mathbb{I} + \left(\mu^2 + \rho\sigma^2\right) \mathbf{1}\mathbf{1}^\top, \end{aligned} \tag{3.51}$$

where $p$ and $q$ represent dual variables assigned to the two robust constraints in (3.50). Thus, the primal uncertain convex program (3.50) is solved under the worst possible realizations of $\overline{\boldsymbol{\xi}}$ and $\underline{\boldsymbol{\xi}}$, while the dual uncertain convex program (3.51) is solved under the best possible realizations, in which case $\overline{\boldsymbol{\xi}}$ and $\underline{\boldsymbol{\xi}}$ become decision variables. Problem (3.51) has intuitive appeal as it can be interpreted as a restriction of the original worst-case probability problem that minimizes over all two-point distributions in the ambiguity set $\mathscr{P}'$ with scenarios $\overline{\boldsymbol{\xi}}$ and $\underline{\boldsymbol{\xi}}$ and corresponding probabilities $p$ and $q$, respectively. Note that (3.51) constitutes a non-convex program because it involves multilinear terms in the decisions. Using the variable transformations $\boldsymbol{u} \leftarrow p\overline{\boldsymbol{\xi}}$ and $\boldsymbol{v} \leftarrow q\underline{\boldsymbol{\xi}}$ we can reformulate (3.51) as

$$\begin{aligned} \mathrm{R}'(\gamma) = \sup \quad & q \\ \text{s.t.} \quad & p, q \in \mathbb{R}_+,\ \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^T_+ \\ & \textstyle\prod_{t=1}^T v_t \geq q^T \gamma \\ & p + q = 1 \\ & \boldsymbol{u} + \boldsymbol{v} = \mu \mathbf{1} \\ & \tfrac{1}{p}\boldsymbol{u}\boldsymbol{u}^\top + \tfrac{1}{q}\boldsymbol{v}\boldsymbol{v}^\top \preceq (1-\rho)\sigma^2 \mathbb{I} + \left(\mu^2 + \rho\sigma^2\right) \mathbf{1}\mathbf{1}^\top. \end{aligned} \tag{3.52}$$

Note that if $p = 0$ ($q = 0$), then $\boldsymbol{u} = \mathbf{0}$ ($\boldsymbol{v} = \mathbf{0}$) for otherwise the matrix inequality is not satisfiable. In (3.52) and below we adhere to the convention that $0/0 = 0$, which reflects the idea that a scenario with zero probability mass should have zero weight in the covariance matrix. Observe that problem (3.52) is a convex program. In particular, the first constraint is convex because of the concavity of geometric means, and the last constraint is convex due to a standard Schur complement argument. Exploiting the problem's permutation symmetry and convexity, one can proceed as in Lemma 3.1 to show that (3.52) has a permutation symmetric minimizer of the form $\boldsymbol{u} = u\mathbf{1}$ and $\boldsymbol{v} = v\mathbf{1}$ for some scalar decision variables $u, v \in \mathbb{R}_+$. Restricting the search to permutation symmetric solutions, problem (3.52) can therefore be reformulated as

$$\begin{aligned} \mathrm{R}'(\gamma) = \sup \quad & q \\ \text{s.t.} \quad & p, q, u, v \in \mathbb{R}_+ \\ & v \geq q\gamma^{1/T} \\ & p + q = 1 \\ & u + v = \mu \\ & (1-\rho)\sigma^2\mathbb{I} + \left(\mu^2 + \rho\sigma^2 - \tfrac{u^2}{p} - \tfrac{v^2}{q}\right)\mathbf{1}\mathbf{1}^\intercal \succeq \mathbf{0}. \end{aligned} \tag{3.53}$$

It can be shown that the eigenvalues of the matrix $(1-\rho)\sigma^2\mathbb{I} + (\mu^2 + \rho\sigma^2 - \frac{u^2}{p} - \frac{v^2}{q})\mathbf{1}\mathbf{1}^\intercal$ are given by $(1-\rho)\sigma^2$ and $(1-\rho)\sigma^2 + T(\mu^2 + \rho\sigma^2 - \frac{u^2}{p} - \frac{v^2}{q})$; see e.g. Proposition 2.4. Since $(1-\rho)\sigma^2 > 0$ by assumption, the matrix inequality in (3.53) is equivalent to the scalar constraint

$$(1-\rho)\sigma^2 + T\left(\mu^2 + \rho\sigma^2 - \tfrac{u^2}{p} - \tfrac{v^2}{q}\right) \geq 0. \tag{3.54}$$

Any feasible solution of (3.53) satisfies $q\gamma^{1/T} \leq v \leq \mu$, implying that the optimal value of (3.53) is bounded above by $\min\{1, \mu\gamma^{-1/T}\}$. For $0 < \gamma^{1/T} \leq \mu$, an optimal solution of (3.53) is then given by $(p, q, u, v) = (0, 1, 0, \mu)$, and the optimal value is equal to 1. For $\mu < \gamma^{1/T} < \mu + \frac{(1+(T-1)\rho)\sigma^2}{T\mu}$, on the other hand, an optimal solution is given by $(p, q, u, v) = (1 - \mu\gamma^{-1/T}, \mu\gamma^{-1/T}, 0, \mu)$ with corresponding optimal value $\mu\gamma^{-1/T}$. Indeed, any larger value of $q$ would require a larger value of $v$, which in turn would violate the non-negativity of $u$ as $u + v = \mu$. One can show that the constraint (3.54) is always *in*active at this solution. For $\gamma^{1/T} \geq \mu + \frac{(1+(T-1)\rho)\sigma^2}{T\mu}$, finally, the constraint (3.54) implies that $q$ must not exceed $\mu\gamma^{-1/T}$, which in turn implies that the constraint must be binding. Furthermore, $q$ has to be strictly positive for otherwise (3.53) would be solved by $(p, q, u, v) = (1, 0, \mu, 0)$, which contradicts our earlier finding that the constraint (3.54) is binding. Substituting $p = 1 - q$ and $u = \mu - v$, the left hand side of (3.54) becomes a quadratic function of $v$ parametric in $q$. We denote the two roots of this function by $v^+$ and $v^-$ and define $u^+ = \mu - v^+$ and $u^- = \mu - v^-$. A direct calculation yields

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

By construction, both $(u^+, v^+)$ and $(u^-, v^-)$ satisfy (3.54) as an equality. However, there is no $q \in (0,1]$ for which $(u^+, v^+)$ is feasible in (3.53). Indeed, a direct calculation reveals that the constraint $v^+ \geq q\gamma^{1/T}$ from (3.53) can hold only if

$$q(\mu-\gamma^{1/T}) \geq \sigma\sqrt{1+(T-1)\rho}\sqrt{\frac{q(1-q)}{T}}. \tag{3.55}$$

However, (3.55) is not satisfiable as its left hand side is strictly negative by assumption, whereas its right hand side is non-negative. Therefore, $(u^+, v^+)$ is infeasible in (3.53).

In contrast, the second solution $(u^-, v^-)$ is feasible in (3.53) if we select $q \in (0,1]$ with

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

and

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

Problem (3.53) aims to maximize $q$, which is tantamount to setting

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

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

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

**Corollary 3.1** (Extremal Distribution)**.** *A distribution* $\mathbb{P}^\star \in \mathscr{P}'$ *attaining the relaxed* *right-sided Chebyshev bound* $\mathrm{R}'(\gamma)$ *is given by* $\mathbb{P}^\star = p^\star\delta_{[u^\star/p^\star]\mathbf{1}} + q^\star\delta_{[v^\star/q^\star]\mathbf{1}}$, *where*

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

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

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

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

*Proof.* The proof follows directly from that of Theorem 3.6 and is thus omitted. ∎

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

**Proposition 3.3.** *If* $\mu > \sqrt{\frac{1-\rho}{T}}\sigma$, *then* $\mathrm{R}'(\gamma) = \mathrm{R}(\gamma)$ *for all* $\gamma \ge \overline{\gamma}$, *where*

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

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

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

*Proof.* We first show that $\overline{\gamma}^{1/T} > \mu + \frac{\sigma^2\theta}{T\mu}$ (Step 1), which allows us to invoke Theorem 3.6 to conclude that $\mathrm{R}'(\gamma) = \frac{\sigma^2\theta}{\sigma^2\theta + T(\mu-\gamma^{1/T})^2}$. We then employ Corollary 3.1 to construct a distribution $\mathbb{P}^\star \in \mathscr{P}'$ that satisfies $\mathbb{P}^\star\left(\prod_{t=1}^T \tilde{\xi}_t \ge \gamma\right) = \mathrm{R}'(\gamma)$ (Step 2), and we show that a suitable perturbation of $\mathbb{P}^\star$ results in a distribution $\mathbb{P} \in \mathscr{P}$ that satisfies $\mathbb{P}\left(\prod_{t=1}^T \tilde{\xi}_t \geq \gamma\right) = \mathbb{P}^\star\left(\prod_{t=1}^T \tilde{\xi}_t \geq \gamma\right)$ (Step 3). The statement then follows from the fact that $\mathrm{R}(\gamma)$ is bounded above by $\mathrm{R}'(\gamma)$.

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

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

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

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

and replacing $a$ with its definition inside the square root reveals that $x^\star = \overline{\gamma}^{1/T}$. To show that $\overline{\gamma}^{1/T} > \mu + \frac{\sigma^2\theta}{T\mu}$, we observe that

$$\begin{aligned} q\left(\mu + \frac{\sigma^2\theta}{T\mu}\right) &= \left(\mu - \sqrt{\frac{1-\rho}{T}}\sigma\right)\left(\sigma^2\theta + \frac{\sigma^4\theta^2}{T\mu^2}\right) - \sigma^2\theta\left(\mu + \frac{\sigma^2\theta}{T\mu}\right) \\ &= \sigma^2\theta(\sigma^2\theta + T\mu^2)\left(\frac{\mu - \sqrt{(1-\rho)/T}\sigma}{T\mu^2} - \frac{1}{T\mu}\right) < 0, \end{aligned}$$

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

**Step 2:** The distribution $\mathbb{P}^\star$ in Corollary 3.1 satisfies $\mathbb{P}^\star\left(\prod_{t=1}^T \tilde{\xi}_t \geq \gamma\right) = \mathrm{R}'(\gamma)$. For later reference, we remark that $\mathbb{P}^\star = p^\star\delta_{[u^\star/p^\star]\mathbf{1}} + q^\star\delta_{[v^\star/q^\star]\mathbf{1}}$ satisfies the properties

$$v^\star = q^\star\gamma^{1/T}, \quad u^\star + v^\star = \mu \quad \text{and} \quad \frac{(u^\star)^2}{p^\star} + \frac{(v^\star)^2}{q^\star} = \mu^2 + \frac{1}{T}(1 + (T-1)\rho)\sigma^2. \tag{3.56}$$

Note that the last condition holds because (3.54) is binding when $\gamma^{1/T} \geq \mu + \frac{\sigma^2\theta}{T\mu}$.

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

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

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

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

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

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}\left(\tilde{\boldsymbol{\xi}}\right) &= \frac{p^\star}{T}\sum_{i=1}^{T}\left(\left(\frac{u^\star}{p^\star} - \lambda\right)\mathbf{1} + T\lambda\mathbf{e}_i\right) + v^\star\mathbf{1} = (u^\star + v^\star)\mathbf{1} = \boldsymbol{\mu} \\
\mathbb{E}_{\mathbb{P}}\left(\tilde{\boldsymbol{\xi}}\tilde{\boldsymbol{\xi}}^\intercal\right) &= \frac{p^\star}{T}\sum_{i=1}^{T}\left(\left(\frac{u^\star}{p^\star} - \lambda\right)\mathbf{1} + T\lambda\mathbf{e}_i\right)\left(\left(\frac{u^\star}{p^\star} - \lambda\right)\mathbf{1} + T\lambda\mathbf{e}_i\right)^\intercal + \frac{(v^\star)^2}{q^\star}\mathbf{1}\mathbf{1}^\intercal \\
&= \frac{p^\star}{T}\left(\left(T\left(\frac{u^\star}{p^\star} - \lambda\right)^2 + 2\left(\frac{u^\star}{p^\star} - \lambda\right)T\lambda\right)\mathbf{1}\mathbf{1}^\intercal + T^2\lambda^2\mathbb{I}\right) + \frac{(v^\star)^2}{q^\star}\mathbf{1}\mathbf{1}^\intercal \\
&= \left(\frac{(u^\star)^2}{p^\star} + \frac{(v^\star)^2}{q^\star} - p^\star\lambda^2\right)\mathbf{1}\mathbf{1}^\intercal + p^\star T\lambda^2\mathbb{I} \\
&= (\mu^2 + \rho\sigma^2)\mathbf{1}\mathbf{1}^\intercal + (1-\rho)\sigma^2\mathbb{I} = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal,
\end{aligned}$$

where the last row is due to (3.56) and our definition of $\lambda$. It remains to be shown that $\tilde{\boldsymbol{\xi}}$ is non-negative $\mathbb{P}$-a.s. By construction of $\mathbb{P}$, this is the case iff $u^\star \geq p^\star\lambda$. We now observe that

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

where the first identity follows from (3.56), the second one is due to the definition of $q^\star$ in Corollary 3.1, and the inequality holds since there is $C > 0$ such that

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

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

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

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

## 3.6 Extensions

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

**Theorem 3.7.** *For any permutation-symmetric continuous functional $h : \mathbb{R}_+^T \to \mathbb{R}$, we have*

$$\begin{aligned} \sup_{\mathbb{P}\in\mathscr{P}} \mathbb{P}(h(\tilde{\boldsymbol{\xi}}) \leq 0) \;=\; \inf \;\; & \alpha + T\mu\beta + T(\mu^2+\sigma^2)\gamma_1 + T\left[T\mu^2+\sigma^2+(T-1)\rho\sigma^2\right]\gamma_2 \\ \text{s.t.} \;\; & \alpha, \lambda_1, \lambda_2 \in \mathbb{R}_+, \; \beta, \gamma_1, \gamma_2 \in \mathbb{R} \\ & \gamma_1+\gamma_2 \geq 0, \; \gamma_2+\gamma_1+\alpha \geq \left\| \left(\beta-\lambda_1, \gamma_2+\gamma_1-\alpha\right) \right\|_2 \\ & \tfrac{\gamma_1}{T}+\gamma_2 \geq 0, \; \gamma_2+\tfrac{\gamma_1}{T}+\alpha \geq \left\| \left(\beta-\lambda_2, \gamma_2+\tfrac{\gamma_1}{T}-\alpha\right) \right\|_2 \\ & \alpha+\beta s+\gamma_2 s^2+\gamma_1 \underline{\phi}(s) \geq 1 \quad \forall s \in \mathscr{S} \\ & \alpha+\beta s+\gamma_2 s^2+\gamma_1 \overline{\phi}(s) \geq 1 \quad \forall s \in \mathscr{S}, \end{aligned} \tag{3.57}$$

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

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

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

*Proof.* The proof is largely based on arguments familiar from Theorems 3.2 and 3.4. Details are omitted for brevity of exposition. ∎

The significance of Theorem 3.7 is that it enables us to compute $\sup_{\mathbb{P}\in\mathscr{P}} \mathbb{P}(h(\tilde{\boldsymbol{\xi}}) \leq 0)$ by solving a semidefinite program whenever $\underline{\phi}(s)$ and $\overline{\phi}(s)$ are piecewise polynomials. In this case the last two constraints in (3.57) reduce to the requirement that a univariate piecewise polynomial, whose coefficients depend affinely on the decision variables, must be non-negative uniformly on $\mathscr{S}$. Such conditions can systematically be reformulated as linear matrix inequalities (Nesterov 2000).

Table 3.1 lists examples of permutation-symmetric functionals $h(\boldsymbol{\xi})$ that lead to piecewise polynomial mappings $\underline{\phi}(s)$ and $\overline{\phi}(s)$ and thus to computable Chebyshev bounds. Theorems 3.8 and 3.9 below present two special cases in which these bounds can be evaluated analytically.

**Theorem 3.8** (Left-Sided Chebyshev Bound for Sums)**.** *For any $\gamma > 0$ we have*

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

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

*Proof.* By Theorem 3.7 the Chebyshev bound $\sup_{\mathbb{P}\in\mathscr{P}} \mathbb{P}(\sum_{t=1}^{T} \tilde{\xi}_t \leq \gamma)$ can be reformulated as the semi-infinite program (3.57), where the functions $\underline{\phi}(s)$ and $\overline{\phi}(s)$ are specified in Table 3.1. Distinguishing the cases $\gamma_1 \geq 0$ and $\gamma_1 < 0$, this semi-infinite program can be reduced to a robust optimization problem with a scalar uncertain parameter. By using the '*primal worst equals dual best*' duality scheme from robust optimization (Beck and Ben-Tal 2009), one can further show that the optimal value of this problem coincides with the univariate Chebyshev bound $\sup_{\mathbb{P}\in\mathscr{P}_1} \mathbb{P}(\tilde{\xi} \leq \gamma)$, where $\mathscr{P}_1$ contains all distributions of $\tilde{\xi}$ supported on $\mathbb{R}_+$ with mean $T\mu$ and variance $\sigma^2 T(1 + (T-1)\rho)$. The latter Chebyshev bound has an analytical formula, which can be derived based on arguments familiar from Section 3.1. ∎

**Theorem 3.9** (Right-Sided Chebyshev Bound for Sums)**.** *For any $\gamma > 0$ we have*

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

| Chebyshev bound | $h(\boldsymbol{\xi})$ | $\underline{\phi}(s)$ | $\overline{\phi}(s)$ |
|---|---|---|---|
| $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\Big(\min_{t=1,\dots,T} \tilde{\xi}_t \leq \gamma\Big)$ | $\min_{t=1,\dots,T} \xi_t - \gamma$ | $\begin{cases} \frac{s^2}{T} & \text{if } 0 \leq s \leq \gamma T \\ \gamma^2 + \frac{1}{T-1}(s-\gamma)^2 & \text{if } s > \gamma T \end{cases}$ | $s^2$ |
| $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\Big(\min_{t=1,\dots,T} \tilde{\xi}_t \geq \gamma\Big)$ | $\gamma - \min_{t=1,\dots,T} \xi_t$ | $\begin{cases} \frac{s^2}{T} & \text{if } s \geq \gamma T \\ +\infty & \text{if } 0 \leq s < \gamma T \end{cases}$ | $\begin{cases} (s-\gamma T)^2 + 2\gamma(s-\gamma T) + T\gamma^2 & \text{if } s \geq \gamma T \\ -\infty & \text{if } 0 \leq s < \gamma T \end{cases}$ |
| $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\Big(\max_{t=1,\dots,T} \tilde{\xi}_t \leq \gamma\Big)$ | $\max_{t=1,\dots,T} \xi_t - \gamma$ | $\begin{cases} \frac{s^2}{T} & \text{if } 0 \leq s \leq \gamma T \\ +\infty & \text{if } s > \gamma T \end{cases}$ | $\begin{cases} s^2 & \text{if } 0 \leq s \leq \gamma \\ \gamma^2 + (s-\gamma)^2 & \text{if } \gamma < s \leq 2\gamma \\ \vdots & \vdots \\ (T-1)\gamma^2 + (s-(T-1)\gamma)^2 & \text{if } (T-1)\gamma < s \leq T\gamma \\ -\infty & \text{if } s > T\gamma \end{cases}$ |
| $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\Big(\max_{t=1,\dots,T} \tilde{\xi}_t \geq \gamma\Big)$ | $\gamma - \max_{t=1,\dots,T} \xi_t$ | $\begin{cases} \frac{s^2}{T} & \text{if } s \geq \gamma T \\ \gamma^2 + \frac{1}{T-1}(s-\gamma)^2 & \text{if } \gamma \leq s < \gamma T \\ +\infty & \text{if } 0 \leq s < \gamma \end{cases}$ | $\begin{cases} s^2 & \text{if } s \geq \gamma \\ -\infty & \text{if } 0 \leq s < \gamma \end{cases}$ |
| $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\Big(\sum_{t=1}^{T} \tilde{\xi}_t \leq \gamma\Big)$ | $\sum_{t=1}^{T} \tilde{\xi}_t - \gamma$ | $\begin{cases} \frac{s^2}{T} & \text{if } 0 \leq s \leq \gamma \\ +\infty & \text{if } s > \gamma \end{cases}$ | $\begin{cases} s^2 & \text{if } 0 \leq s \leq \gamma \\ -\infty & \text{if } s > \gamma \end{cases}$ |
| $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\Big(\sum_{t=1}^{T} \tilde{\xi}_t \geq \gamma\Big)$ | $\gamma - \sum_{t=1}^{T} \tilde{\xi}_t$ | $\begin{cases} \frac{s^2}{T} & \text{if } s \geq \gamma \\ +\infty & \text{if } 0 \leq s < \gamma \end{cases}$ | $\begin{cases} s^2 & \text{if } s \geq \gamma \\ -\infty & \text{if } 0 \leq s < \gamma \end{cases}$ |

Table 3.1: Chebyshev bounds equivalent to $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}(h(\tilde{\boldsymbol{\xi}}) \leq 0)$ for some permutation symmetric functional $h(\boldsymbol{\xi})$. These bounds coincide with the optimal value of (3.57), instantiated with the respective piecewise polynomials $\underline{\phi}(s)$ and $\overline{\phi}(s)$.

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

*Proof.* The proof is widely parallel to that of Theorem 3.8 and is thus omitted for brevity. ∎

## 3.7 Numerical Experiments

We first compare our Chebyshev bounds $\mathrm{R}(\gamma)$ and $\mathrm{L}(\gamma)$ with alternative bounds proposed in the literature, as well as the relaxed Chebyshev bound $\mathrm{R}'(\gamma)$ from Section 3.5. We then present a case study that employs our left-sided Chebyshev bound $\mathrm{L}(\gamma)$ to select financial portfolios under imprecise knowledge of the asset return distributions. All optimization problems are solved with the SDPT3 optimization software using the YALMIP interface (Löfberg 2004; Toh et al. 1999).

### 3.7.1 Comparison of Chebyshev Bounds

Instead of employing the bounds $\mathrm{R}(\gamma)$ and $\mathrm{L}(\gamma)$ from Sections 3.3 and 3.4, which are exact but may result in computationally challenging optimization problems, one can employ existing results to derive approximate bounds on the tail probabilities of a product of non-negative, permutation-symmetric random variables. In the following, we compare our bounds with two such approximations based on earlier results of Marshall and Olkin (1960) and Vandenberghe et al. (2007). Both approximations rely on the larger ambiguity set

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

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

Marshall and Olkin (1960) derive a convex optimization problem that provides a tight upper bound on the probability that the random vector $\tilde{\boldsymbol{\xi}}$ is contained in a closed convex set $\mathscr{C}$, assuming that $\tilde{\boldsymbol{\xi}}$ can be governed by any distribution from the ambiguity set $\mathscr{P}^0$. The choice $\mathscr{C} = \left\{\boldsymbol{\xi} \in \mathbb{R}^T : \prod_{t=1}^T \xi_t \geq \gamma\right\}$ allows us to approximate the right-sided Chebyshev bound $\mathrm{R}(\gamma)$. For this special case, the bound of Marshall and Olkin has the analytical solution

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

which follows from Bertsimas and Popescu (2005, Theorem 6.1). By construction, $\mathrm{R}^{\mathrm{MO}}(\gamma) \ge \mathrm{R}(\gamma)$ since $\mathscr{P} \subset \mathscr{P}^0$. Note that $\mathrm{R}^{\mathrm{MO}}(\gamma)$ coincides with our relaxed Chebyshev bound $\mathrm{R}'(\gamma)$ for $\gamma \ge \left(\mu + \frac{\sigma^2\theta}{T\mu}\right)^T$, see Theorem 3.6. Thus, $\mathrm{R}^{\mathrm{MO}}(\gamma)$ also coincides with our right-sided Chebyshev bound $\mathrm{R}(\gamma)$ for large values of $\gamma$, see Proposition 3.3. Note that the bound of Marshall and Olkin cannot be used to approximate our left-sided Chebyshev bound $\mathrm{L}(\gamma)$ since the complement of $\mathscr{C}$ fails to be convex.

Vandenberghe et al. (2007) derive a semidefinite program that provides a tight upper bound on the probability that $\tilde{\boldsymbol{\xi}} \in \mathscr{C}$ for a (not necessarily convex) set $\mathscr{C} = \{\boldsymbol{\xi} \in \mathbb{R}^T : \boldsymbol{\xi}^\intercal \boldsymbol{A}_i \boldsymbol{\xi} + 2\boldsymbol{b}_i^\intercal \boldsymbol{\xi} + c_i < 0 \ \forall i = 1, \ldots, m\}$, assuming that the random vector $\tilde{\boldsymbol{\xi}}$ can be governed by any distribution from the ambiguity set $\mathscr{P}^0$. Employing a second-order Taylor approximation of $\prod_{t=1}^T \xi_t$ around $\mu\mathbf{1}$,

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

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

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

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

Figure 3.2 compares our Chebyshev bounds $\mathrm{L}(\gamma)$ and $\mathrm{R}(\gamma)$ with the approximate

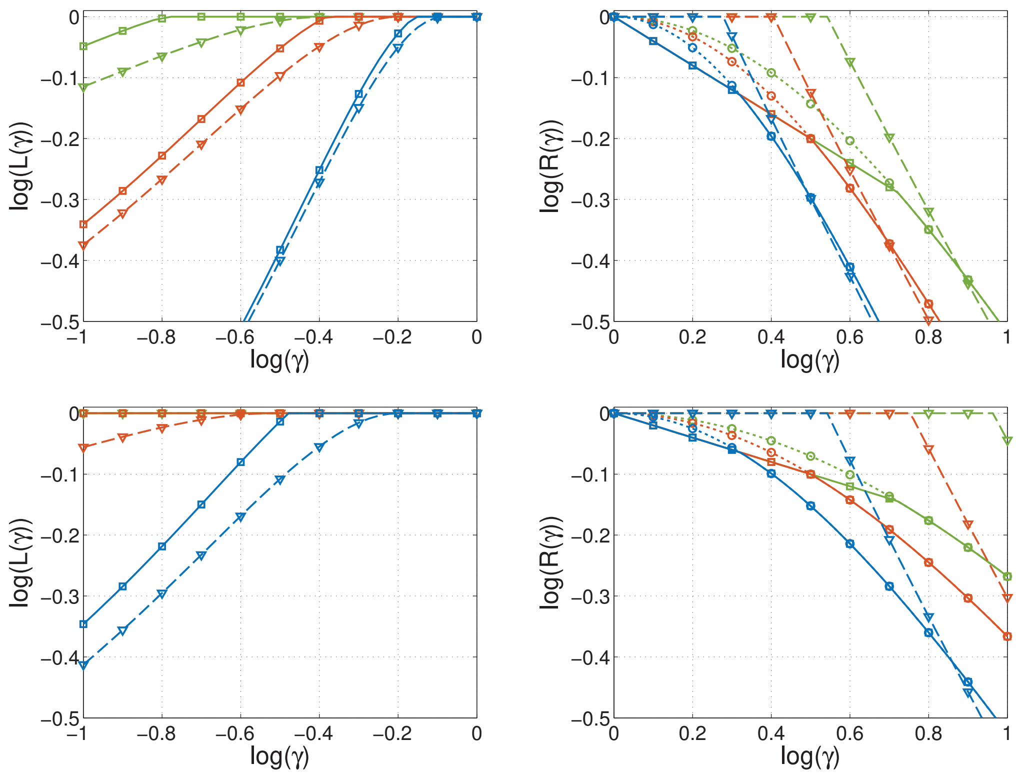



Figure 3.2: Comparison of the left-sided (left) and right-sided (right) Chebyshev bounds for the products of $T = 5$ (top) and $T = 10$ (bottom) random variables with $\mu = 1$ and $\rho = 0$. The solid lines with squares, the dashed lines with triangles and the dotted lines with circles represent our bounds, the VBC bounds and the MO bounds, respectively. From bottom to top, the blue, red and green lines correspond to $\sigma = 0.2$, 0.3 and 0.4 (left) and $\sigma = 0.4$, 0.5 and 0.6 (right), respectively.

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

The MO bound has an analytical solution and can therefore be computed in negligible time. In contrast, the VBC bounds and our bounds require the solution of semidefinite programs with two LMIs of size $\mathcal{O}(T^2)$. Table 3.2 compares the computation times

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

Table 3.2: Runtimes (secs) required to calculate the Chebyshev bounds. Each runtime is averaged over 10 instances with randomly selected $\mu$, $\sigma$ and $\gamma$, and it includes the calculation of both the left-sided and the right-sided bounds.

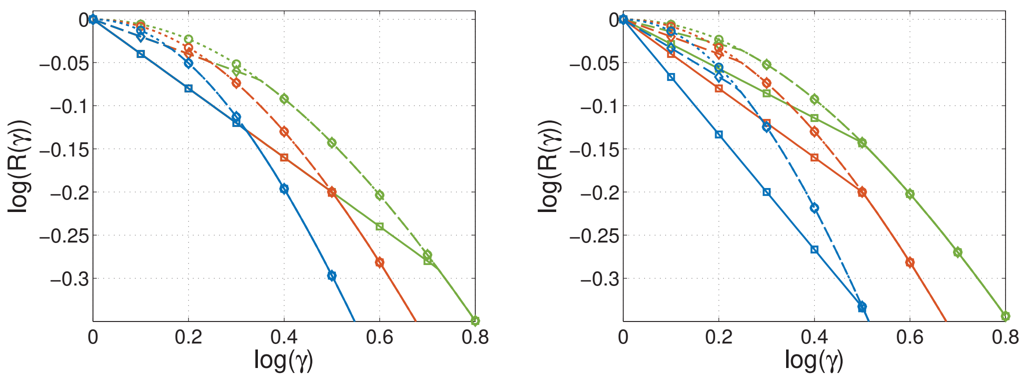



Figure 3.3: Comparison of the right-sided Chebyshev bounds $\mathrm{R}(\gamma)$ (solid lines with squares), $\mathrm{R}'(\gamma)$ (dashed lines with diamonds) and $\mathrm{R}^{\mathrm{MO}}(\gamma)$ (dotted lines with circles) with $\mu = 1$ and $\rho = 0$. From bottom to top, the blue, red and green lines correspond to $\sigma = 0.4$, 0.5 and 0.6 in the left graph (with $T = 5$ fixed) and to $T = 3$, 5 and 7 in the right graph (with $\sigma = 0.5$ fixed), respectively.

of both bounds for products of different size $T$ on a computer with a 3.40GHz i7 CPU and 16GB RAM. While both bounds can be computed within seconds, the VBC bounds require significantly less runtime than our bounds. We attribute this to the LMI reformulations of the polynomial constraints in Theorems 3.2 and 3.4, which seem to lack structure that can be exploited by SDPT3.

Figure 3.3 compares the right-sided Chebyshev bound $\mathrm{R}(\gamma)$ with the relaxed right-sided bound $\mathrm{R}'(\gamma)$ and the MO bound $\mathrm{R}^{\mathrm{MO}}(\gamma)$. The figure illustrates that $\mathrm{R}^{\mathrm{MO}}(\gamma)$ coincides with $\mathrm{R}'(\gamma)$ for $\gamma \geq \left(\mu + \frac{\sigma^2\theta}{T\mu}\right)^T$, and subsequently both bounds coincide with $\mathrm{R}(\gamma)$ for large values of $\gamma$. The gaps between the bounds increase with larger variances $\sigma^2$, and they decrease with larger numbers of random variables $T$.

### 3.7.2 Case Study: Financial Risk Management

Consider an investor who allocates a limited budget to a fixed pool of $n$ assets over a time horizon of $T$ periods. We denote by $\tilde{r}_{t,i} \geq -1$, $t = 1,\ldots,T$ and $i = 1,\ldots,n$, the relative price change of asset $i$ between periods $t$ and $t+1$. We assume that the investor pursues a fixed-mix (or constant proportions) strategy which rebalances the portfolio composition to a pre-selected set of weights $\boldsymbol{w} \in \mathscr{W} = \{\boldsymbol{z} \in \mathbb{R}^n_+ : \mathbf{e}^\intercal \boldsymbol{z} = 1\}$ at the beginning of each period. Note that despite being memoryless, fixed-mix strategies are dynamic since they recapitalize those assets whose returns were below average (*'buy low'*) and divest assets whose returns were above average (*'sell high'*). Fixed-mix strategies generalize the well-known $1/N$-portfolio (DeMiguel et al. 2009), and they have received significant attention among both academics and practitioners.

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

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

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

Calculating the value-at-risk of a portfolio's terminal wealth requires perfect knowledge of the joint asset return distribution $\mathbb{P}$, which is unavailable in practice. Following Chapter 2, we will assume that it is only known that the asset returns $(\tilde{\boldsymbol{r}}_t)_{t=1}^T$ follow a weak-sense white noise process with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$, that is, the asset returns are serially uncorrelated and have period-wise identical first and second-order moments. In that case, the wealth evolution $(\tilde{\xi}_t)_{t=1}^T = (1 + \boldsymbol{w}^\intercal \tilde{\boldsymbol{r}}_t)_{t=1}^T$ also follows a weak-sense stochastic process governed by a distribution $\mathbb{P}_{\boldsymbol{w}}$ supported on $\mathbb{R}^T_+$, under which the $\tilde{\xi}_t$ have mean $\boldsymbol{w}^\intercal \boldsymbol{\mu}$ and variance $\boldsymbol{w}^\intercal \boldsymbol{\Sigma} \boldsymbol{w}$ and are serially uncorrelated. We denote the set of all these distributions by $\mathscr{P}_{\boldsymbol{w}}$. In this setting, an ambiguity-averse investor may assess the fixed-mix strategy $\boldsymbol{w}$ in view of the *worst-case* value-at-risk of the portfolio's terminal wealth over all distributions $\mathbb{P}_{\boldsymbol{w}} \in \mathscr{P}_{\boldsymbol{w}}$:

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

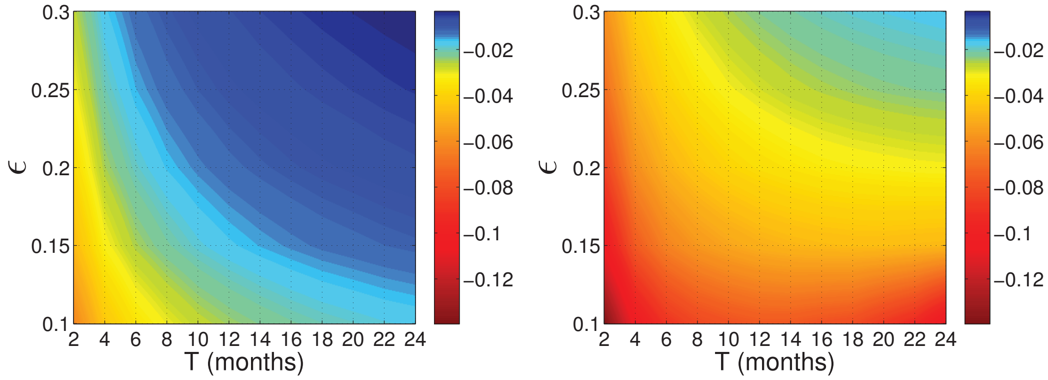



Figure 3.4: Worst-case value-at-risk of the growth rates of the minimum-variance (left) and maximum-expectation (right) portfolios for different investment horizons $T$ and risk tolerances $\epsilon$.

In Chapter 2, the worst-case value-at-risk of the portfolio's terminal wealth is replaced with a quadratic approximation. The Chebyshev bounds proposed in this chapter allow us to calculate the worst-case value-at-risk exactly without resorting to any approximation. Indeed, one verifies that

$$\mathrm{WVaR}_{\epsilon}(\boldsymbol{w}) = \sup_{\gamma \in \mathbb{R}} \left\{ \gamma : \sup_{\mathbb{P}_{\boldsymbol{w}} \in \mathscr{P}_{\boldsymbol{w}}} \mathbb{P}_{\boldsymbol{w}} \left( \prod_{t=1}^{T} \tilde{\xi}_t \leq \gamma \right) \leq \epsilon \right\} = \sup_{\gamma \in \mathbb{R}} \left\{ \gamma : L(\gamma; \boldsymbol{w}^\top \boldsymbol{\mu}, \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}) \leq \epsilon \right\},$$

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

Figure 3.4 reports the worst-case value-at-risk of two portfolios over different time horizons $T$, where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are calibrated to the 2003–2012 period of Fama and French's 10 Industry Portfolios data set.[2] The minimum-variance portfolio (left graph) corresponds to the weight vector $\boldsymbol{w} \in \mathscr{W}$ that minimizes $\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}$, whereas the maximum-expectation portfolio (right graph) invests all wealth into the asset $i$ with the highest expected return $\mu_i$. To facilitate a fair comparison among different time horizons, the graphs report the growth rates of the portfolios, that is, the logarithms of the terminal wealth, divided by the number of investment periods $T$. As expected, the minimum-variance portfolio is less risky than the maximum-expectation portfolio,

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

and the risk of both portfolios tends to decrease when the investment horizon $T$ grows. Interestingly, however, the risk of the maximum-expectation portfolio *increases* with large $T$ for low risk tolerances $\epsilon \lesssim 0.15$. This seemingly counter-intuitive effect is explained by Theorem 3.3, which states that the wealth evolution $\prod_{t=1}^{T} \tilde{\xi}_t$ is absorbed at 0 for large investment horizons $T$.

In addition to *evaluating* the worst-case value-at-risk of a pre-selected portfolio $\boldsymbol{w}$, an investor often seeks to determine a portfolio $\boldsymbol{w}^\star$ that *optimizes* the worst-case value-at-risk. The search for optimal portfolios is greatly simplified by the observation that there is always a portfolio $\boldsymbol{w}^\star$ on the mean-variance efficient frontier that maximizes $\mathrm{WVaR}_\epsilon(\boldsymbol{w})$ over (subsets of) $\mathscr{W}$. Indeed, Theorem 3.5 implies that $\mathrm{L}(\gamma; \boldsymbol{w}^\intercal \boldsymbol{\mu}, \boldsymbol{w}^\intercal \boldsymbol{\Sigma} \boldsymbol{w}) = \mathrm{L}'(\gamma; \boldsymbol{w}^\intercal \boldsymbol{\mu}, \boldsymbol{w}^\intercal \boldsymbol{\Sigma} \boldsymbol{w})$, and one readily verifies that $\mathrm{L}'(\gamma; \boldsymbol{w}^\intercal \boldsymbol{\mu}, \boldsymbol{w}^\intercal \boldsymbol{\Sigma} \boldsymbol{w})$ is non-decreasing in both $\gamma$ and $\boldsymbol{w}^\intercal \boldsymbol{\Sigma} \boldsymbol{w}$. This implies that

$$\sup_{\gamma \in \mathbb{R}} \left\{ \gamma : \mathrm{L}'(\gamma; \boldsymbol{w}^\intercal \boldsymbol{\mu}, \boldsymbol{w}^\intercal \boldsymbol{\Sigma} \boldsymbol{w}) \le \epsilon \right\} \le \sup_{\gamma \in \mathbb{R}} \left\{ \gamma : \mathrm{L}'(\gamma; \boldsymbol{w}'^\intercal \boldsymbol{\mu}, \boldsymbol{w}'^\intercal \boldsymbol{\Sigma} \boldsymbol{w}') \le \epsilon \right\}$$

for two portfolios $\boldsymbol{w}$ and $\boldsymbol{w}'$ that satisfy $\boldsymbol{w}^\intercal \boldsymbol{\mu} = \boldsymbol{w}'^\intercal \boldsymbol{\mu}$ and $\boldsymbol{w}^\intercal \boldsymbol{\Sigma} \boldsymbol{w} \ge \boldsymbol{w}'^\intercal \boldsymbol{\Sigma} \boldsymbol{w}'$. We thus conclude that among all portfolios $\boldsymbol{w} \in \mathscr{W}$ that achieve the same mean return $\boldsymbol{w}^\intercal \boldsymbol{\mu}$, the portfolio with smallest variance $\boldsymbol{w}^\intercal \boldsymbol{\Sigma} \boldsymbol{w}$ provides the best worst-case value-at-risk. We can therefore identify an optimal portfolio through a one-dimensional line search over the mean-variance efficient frontier.

# 4 Multi-Market Multi-Reservoir Management

Peak/off-peak spreads on European electricity spot markets are eroding due to the nuclear phaseout and the recent growth in photovoltaic capacity. The reduced profitability of peak/off-peak arbitrage thus forces hydropower producers to participate in the reserve markets. We propose a two-layer stochastic programming model for the optimal operation of a cascade of hydropower plants selling energy on both the spot and reserve markets. The planning problem optimizes the reservoir management over a yearly horizon with weekly granularity, and the trading subproblems optimize the market transactions over a weekly horizon with hourly granularity. We solve both the planning and trading problems in linear decision rules, and we exploit the inherent parallelizability of the trading subproblems to achieve computational tractability.

## 4.1 Introduction

Electricity from renewable sources, e.g. wind, geothermal, solar and hydropower, has seen its share growing in European electricity markets in recent years. The increase of renewable energies is resulting in numerous environmental and economic benefits. However, electricity generation from some of these sources, especially wind and solar, is intermittent because of its reliance on weather and sunlight conditions. Hence, there is a growing need to invest in power plants with storage capacities that can produce or consume electricity on a short notice. For example, pumped-storage hydropower plants are capable of buffering short-term fluctuations in demand and supply because of their storage capabilities and negligible start-up times.

In terms of wholesale electricity markets, most generation companies in Switzerland, France, Germany, Luxembourg and Austria participate in the European Energy Exchange (EEX), which is one of the largest electricity markets in central Europe.[1] Among the markets offered by EEX, European Power Exchange Spot (EPEX SPOT) is an exchange for power spot trading. It consists of different forward markets, with the main component being the day-ahead market. Generation companies operating pumped-storage hydropower plants typically trade in this market. In the remainder of the chapter, we use the terminology *spot market* to refer to this day-ahead market. Pumped-storage hydropower plants benefit from participating in the spot market by releasing the water downstream for electricity generation at peak times and by pumping the water upstream during off-peak periods for future generation ('*buy low & sell high*'). In doing so, the generation companies exploit the gaps between peak and off-peak electricity prices to make immediate profits. However, these price gaps have been shrinking since 2008 (Mayer 2014). This phenomenon occurs because of two main reasons: (i) the phaseout of nuclear power plants from European electricity markets and (ii) the rapid growth in photovoltaic capacity; see Morris and Pehnt (2015); Wirth (2016). Nuclear power plants are important sources of base load power. As a result, their withdrawal from the electricity markets increases base load electricity prices. On the other hand, the growth in photovoltaic capacity increases the amount of electricity supply during daytime, which significantly overlaps weekdays' peak hours, and thus reduces the peak electricity prices.

As the spot market is a day-ahead market, electricity supply and demand are settled on the day before delivery. In practice, however, this cannot be achieved without errors for many reasons. Examples include operational outages, withdrawal of power plants, and sudden rises in demand. While small errors are usually corrected by trading in intraday markets, bigger errors need to be handled separately. Moreover, since wind and sunlight conditions can change abruptly and are difficult to predict with high accuracy, wind and solar energy is highly volatile. Thus, as the percentage of the renewable energy supply increases, the resulting fluctuations can be large, and they cannot be absorbed completely in the spot and intraday markets.

In order for the frequency of the electricity grid to be maintained at 50Hz, the im-

[1] https://www.eex.com

balances between demand and supply have to be diminished. To achieve this, the transmission system operators procure additional ancillary services (in this case, control energy) in advance on separate markets.[2] These markets, depending on their geographical locations, have different names, for example, balancing markets, reserve markets, regulation markets, and control markets. To avoid terminological confusion, we will consistently use the term *reserve market* in the remainder of the chapter. For a succinct overview of how the reserve markets work and what role the hydropower producers play in these markets, we refer interested readers to Beck and Scherer (2015); Hirth and Ziegenhagen (2015).

With reserve markets, when electricity demand and supply differ, the transmission system operator of the control area responds by requesting reserve market participants to increase or decrease their electricity output. For generation companies, the benefit of trading in the reserve markets is two-fold. First, they receive capacity fees in advance regardless of whether the reserve capacities are activated. Second, they also earn money proportional to the increase or decrease in their production levels when the transmission system operator triggers the reserves. Since the amount of activated reserves can be positive (upward regulation) or negative (downward regulation), we distinguish between the two cases by decomposing the reserve market into reserve-up and reserve-down markets, respectively. Figure 4.1 visualizes situations when reserve-down and reserve-up capacities are activated by the transmission system operator. The existence of the reserve markets should ease the pressure on the hydropower plant operators who are struggling to recover their capital costs or their original profitability on the spot market because of the eroding peak/off-peak spreads.

The focus of this chapter is to develop, for hydropower producers, a stochastic program that maximizes their total revenues earned from simultaneously trading in both the spot and reserve markets. The resulting optimization model is computationally challenging because it involves a large number of decision stages as well as significant uncertainty in electricity prices and water inflows. For example, consider the setting where the planning horizon is one year, and where electricity is traded daily. In this case, the number of decision stages already exceeds a few hundreds. Furthermore, in a

[2] The European Network of Transmission System Operators for Electricity (ENTSOE) publishes the list of transmission system operators in Europe, available at https://www.entsoe.eu.

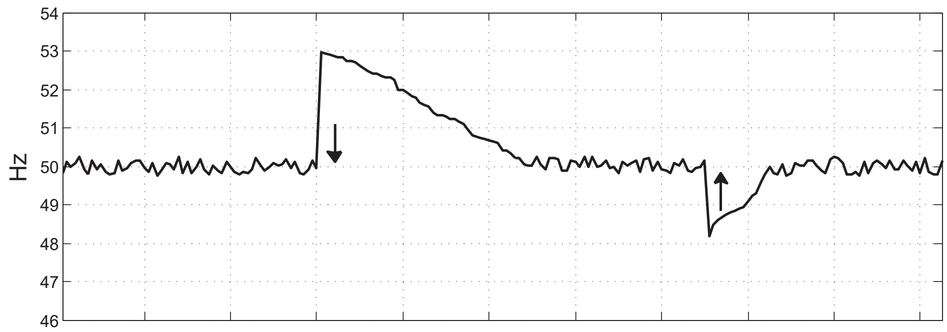



Figure 4.1: Illustration of downward regulation (↓) and upward regulation (↑).

system with multiple connected reservoirs, a coordinated water release and pumping policy is required because the water releases from an upstream reservoir contribute to the inflows of its downstream reservoir(s).

As pointed out by Shapiro and Nemirovski (2005), multi-stage stochastic programs *'generically are computationally intractable already when medium-accuracy solutions are sought.'* It would appear hopeless for us to directly solve the formulated stochastic program. Inspired by Pritchard et al. (2005), we decompose the problem temporally, which in our case is achieved by splitting the planning horizon into weekly periods. At the beginning of each week, the generation company sets a target release for each reservoir. This target is obtained by solving a reservoir management problem, which can also be formulated as a stochastic program. Then, for a predetermined target, the generation company solves another stochastic program to determine an optimal trading policy for both the spot and reserve markets over the course of one week. To gain tractability, all arising stochastic programs are solved approximately in linear decision rules, which we discuss next.

In dynamic optimization problems, future decisions are representable as measurable functions of the observable data. One major challenge for solving such problems is that optimizing over functions is generally much harder than optimizing over finite vectors. The linear decision rule approximation simplifies the problem by focusing on the subclass of affine functions only. By focusing on decision rules in affine forms, we obtain a conservative approximation of the true optimization problem. The main advantage of solving dynamic optimization problems in linear decision rules is tractability as highlighted by Shapiro and Nemirovski (2005); Ben-Tal et al. (2004). Often, such an approximation is scalable to industrial-size problems. However, there are two main issues that may limit the relevance of the linear decision rule approximation in practice. First, it is possible that the approximation may result in a non-negligible degree of suboptimality. Second, some feasible stochastic programs, even those with complete recourse, may fail to admit feasible linear decision rules (Chen et al. 2008). To address the first shortcoming, Kuhn et al. (2011) outline a primal-dual approach to numerically quantify the loss of optimality incurred by the linear decision rule approximation, and Bertsimas et al. (2010b) further show that linear decision rules are optimal for some instances of one-dimensional dynamic problems. In contrast, the second shortcoming ceases to be relevant to the reservoir management problem if the generation company sets the target storage levels at the end of the planning horizon to be the same as the initial storage levels. In this case, there always exists a feasible water discharge policy in affine form, one of which is to immediately spill the inflows. In any case, more flexible decision rules can be used if high-accuracy solutions are sought; see Georghiou et al. (2015); Chen et al. (2008). The linear decision rule approximation (in a simpler form) has been previously applied to the reservoir management problem in different settings. We refer the readers to Yeh (1985) and the references therein.

The main contributions of the chapter may be summarized as follows.

(i) We propose a stochastic program for maximizing the net revenue of a hydropower producer who simultaneously trades in both the spot and reserve markets. The proposed model accounts for uncertainty in electricity prices, reserve capacity fees, inflows, etc. Accounting for price uncertainty usually leads to intractability because of the high frequency at which market prices fluctuate. Therefore, we develop a complexity reduction scheme (ii).

(ii) To achieve tractability, we propose a planner-trader decomposition, which leads to a two-layer stochastic program. The decomposition is achieved by separating fast and slow dynamics. In particular, the inflow uncertainty is accounted for in the planning problem, which is a reservoir management problem with weekly granularity. On the other hand, the market uncertainty is absorbed in the the trading subproblems, which optimize intra-week transactions in both the spot and reserve markets.

(iii) We prove that if the reservoirs in the considered cascade have seasonal storages in the sense that it takes at least a few weeks to fully replenish or deplete them, then the planner-trader decomposition generically provides a conservative approximation of the stochastic program proposed in (i).

(iv) We consider a cascade of three connected reservoirs operating in the control area of the Austrian Power Grid AG (APG).[3] The data for this case study is provided by the energy consultancy Decision Trees GmbH.[4] We solve our stochastic programs conservatively using a linear decision rule approximation to determine a near-optimal operation of the generation company under two circumstances: (a) the generation company participates in the spot market only and (b) the generation company participates in both the spot and reserve markets. Our experimental results suggest that participating in the reserve markets increases total revenues by 48.3% on average. Furthermore, it also reduces variation in storage levels.

The remainder of the chapter is structured as follows. In Section 4.2, we introduce the notation and the operational constraints for the hydropower producers, while Section 4.3 presents the revenue-maximizing stochastic program. Our decomposition scheme and numerical solution procedure are described in Section 4.4 and Section 4.5, respectively. In Section 4.6, we discuss a common heuristic used in hydropower scheduling and extend it to cater for additional investments in the reserve markets. Finally, we quantify the benefits of trading in the reserve markets in Section 4.7.

**Notation.** Basic matrix operations used in this chapter follow from MATLAB symbols. In particular, for two matrices with the same number of rows (columns), we use a comma (semicolon) to concatenate them horizontally (vertically). For a column vector $\boldsymbol{x}$, we let $\mathrm{diag}(\boldsymbol{x})$ denote a square diagonal matrix with $\boldsymbol{x}$ on its main diagonal.

[3] https://www.apg.at/

[4] http://dtrees.com/

## 4.2 Hydropower Scheduling Model

We consider a generation company that operates a cascade of reservoirs, indexed by $i \in \mathscr{I}$, and trades hydroelectricity in both the spot and reserve markets. For notational convenience, we represent the topology of the interconnected reservoirs by a directed acyclic graph with a set of nodes $\mathscr{I}$ and a set of arcs $\mathscr{A} \subset \mathscr{I} \times \mathscr{I}$ where reservoirs are represented as nodes. A tuple $(i, j)$ is an arc in $\mathscr{A}$ if $i$ $(j)$ is an upstream (downstream) reservoir of $j$ $(i)$. Without loss of generality, we assume that $\mathscr{I}$ contains a unique sink node $\otimes$ which represents a dummy reservoir below the cascade. In this way, all reservoirs except $\otimes$ have at least one outdegree. The topology of the cascade can be encoded conveniently with an incidence matrix $\mathbf{M}$, where for each $(i, a) \in \mathscr{I} \times \mathscr{A}$ we have that

$$M_{i,a} = \begin{cases} 1 & \text{if } i \text{ is the tail of arc } a, \\ -1 & \text{if } i \text{ is the head of arc } a, \\ 0 & \text{otherwise.} \end{cases}$$

Last but not least, we denote the cardinality of $\mathscr{A}$ and $\mathscr{I}$ by $A$ and $I$, respectively.

We split the entire planning horizon into trading hours indexed by $t \in \mathscr{T} := \{1, \ldots, T\}$. We remark that the planning horizon should span at least a year in order for the generation company to fully capture the seasonality of electricity prices and water inflows. To begin our discussion on the constraints of hydropower scheduling problems, we consider the case where the company participates only in the spot market. At the beginning of each trading hour $t$, the generation company commits, for each arc $a = (i, j) \in \mathscr{A}$, $g_{t,a}$ and $p_{t,a}$ which represent the amount of water released from $i$ to $j$ (in $\mathrm{m}^3$) and the amount of water pumped up from $j$ to $i$ (in $\mathrm{m}^3$), respectively. We denote the storage level of reservoir $i \in \mathscr{I}$ at the end of trading hour $t$ by $v_{t,i}$. For the sake of transparent exposition, we assume that the delays in water flows between reservoirs are negligible. Aggregating these decisions across the arcs as

$$\boldsymbol{g}_t := [g_{t,a}]_{a \in \mathscr{A}} \in \mathbb{R}^A, \quad \boldsymbol{p}_t := [p_{t,a}]_{a \in \mathscr{A}} \in \mathbb{R}^A, \quad \boldsymbol{v}_t := [v_{t,i}]_{i \in \mathscr{I}} \in \mathbb{R}^I,$$

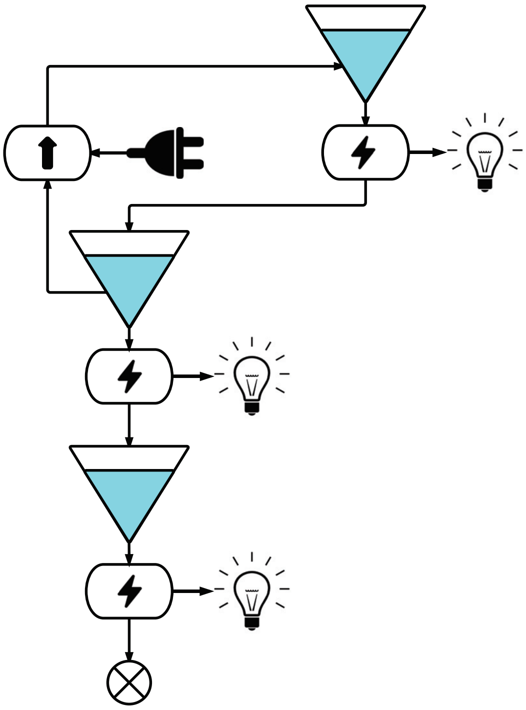


Figure 4.2: An example cascade of three reservoirs. From top to bottom, we denote them by $\mathscr{U}$, $\mathscr{M}$ and $\mathscr{L}$, respectively. Hence in this case, $\mathscr{I} = \{\mathscr{U}, \mathscr{M}, \mathscr{L}, \otimes\}$ and $\mathscr{A} = \{(\mathscr{U}, \mathscr{M}), (\mathscr{M}, \mathscr{L}), (\mathscr{L}, \otimes)\}$. A turbine-pump pair is attached to arc $(\mathscr{U}, \mathscr{M})$, whereas the remaining two arcs have only turbines.

we can then represent the dynamics of the reservoirs' storage levels as

$$\boldsymbol{v}_t = \boldsymbol{v}_{t-1} + \boldsymbol{\psi}_t - \mathbf{M}\big(\boldsymbol{s}_t + \boldsymbol{g}_t - \boldsymbol{p}_t\big) \quad \forall\, t \in \mathscr{T},$$

where $\boldsymbol{s}_t = [s_{t,a}]_{a\in\mathscr{A}} \in \mathbb{R}^A$ represents the vector of spilling decisions, while $\psi_{t,i}$ and $v_{0,i}$ denote the hourly natural inflow and the initial storage level of reservoir $i$, respectively. Both $g_{t,a}$ and $s_{t,a}$ represent water release quantities. However, the difference between them lies in the purpose of the release. On the one hand, $g_{t,a}$ represents the amount of water discharged to a turbine for hydroelectricity generation and thus contributes to an hourly revenue the company earns on the spot market. On the other hand, $s_{t,a}$ represents the amount of water released to adjust the reservoir storage level when necessary, for example, when the upstream reservoir of arc $a$ is overfilled, and hence it does not contribute to any profit. We may refer to $g_{t,a}$ and $s_{t,a}$ as the *productive* and *non-productive* releases, respectively. For the equation of the reservoir dynamics, we use the matrix **M** to ensure that the water released from an upstream reservoir flows into the corresponding downstream reservoir(s). The decisions $\boldsymbol{g}_t, \boldsymbol{p}_t, \boldsymbol{s}_t, \boldsymbol{v}_t$ have to be taken subject to the following physical constraints.

1. Non-negativity:

$$\boldsymbol{g}_t \geq \mathbf{0}, \quad \boldsymbol{p}_t \geq \mathbf{0}, \quad \boldsymbol{s}_t \geq \mathbf{0} \quad \forall t \in \mathcal{T}.$$

2. Maximum generating and pumping levels:

$$\boldsymbol{g}_t \leq \overline{\boldsymbol{g}}, \quad \boldsymbol{p}_t \leq \overline{\boldsymbol{p}} \quad \forall t \in \mathcal{T},$$

   where $\overline{g}_a$ (m$^3$/h) and $\overline{p}_a$ (m$^3$/h) are the maximum release and pumping rates of a turbine-pump pair attached to arc $a$. If the pump (turbine) is absent, we set $\overline{p}_a = 0$ ($\overline{g}_a = 0$).

3. Bounds on storage levels:

$$\underline{\boldsymbol{v}}_t \leq \boldsymbol{v}_t \leq \overline{\boldsymbol{v}}_t \quad \forall t \in \mathcal{T},$$

   where $\underline{v}_{t,i}$ (m$^3$) and $\overline{v}_{t,i}$ (m$^3$) represent the lower and upper bounds on the storage levels of reservoir $i$, respectively. Typically, the upper bounds coincide with the reservoirs' full capacities, whereas the lower bounds can vary throughout the year. For example, they can be higher in the summer months than in the rest of the year for environmental or touristic purposes.

The generation company aims to find a policy to distribute water among reservoirs over the entire planning horizon such that it maximizes the total revenue from the spot market which amounts to

$$\sum_{t \in \mathcal{T}} \pi_t^{\mathrm{s}} \boldsymbol{c}^\top \left( \boldsymbol{g}_t - \mathbf{D} \boldsymbol{p}_t \right),$$

where $\pi_t^{\mathrm{s}}$ is the electricity spot price (€/MWh) at time $t$, whereas $\boldsymbol{c}$ and $\mathbf{D}$ represent the collections of conversion rates and cycling deficiencies of all hydrological arcs

$a \in \mathscr{A}$ in the form

$$\boldsymbol{c} := [c_a]_{a\in\mathscr{A}} \quad \text{and} \quad \mathbf{D} := \operatorname{diag}([d_a]_{a\in\mathscr{A}}).$$

For each hydrological arc $a$, the conversion rate $c_a > 0$ (MWh/m$^3$) translates the water amount (m$^3$) to energy output (MWh). In reality, this conversion rate $c_a$ depends on the net hydraulic head which is defined as the difference of the water levels between endpoint reservoirs; see e.g. Tester et al. (2012, § 12.3). For tractability reasons, however, we follow the literature (see e.g. Baillo et al. (2004); Löhndorf et al. (2013)) in assuming that $c_a$ is a constant. This typically offers a good approximation for a cascade of weekly and annual reservoirs, especially when the net heads are high. Besides, the cycling deficiency $d_a > 1$ determines the rate of energy loss when producing electricity from pumped water.

In addition to trading in the spot market, the generation company may simultaneously participate in the reserve market (which is decomposed into reserve-up and reserve-down markets). The transmission system operator regulates these reserve markets to maintain the network frequency in its control area. The mechanisms underlying the spot and reserve markets are different in the following senses.

1. The spot market is a day-ahead market, i.e., it is cleared on the day before energy delivery, whereas the reserve markets are cleared week-ahead.[5]
2. Trading decisions for the spot market have to be implemented regardless of the state of the control area, whereas reserve capacities may or may not be activated by the transmission system operator, whose action depends on total real-time demand and supply within the control area.

From the transmission system operator's perspective, the main purpose of reserve markets is to smooth out the differences between real-time electricity demand and supply. These differences can be caused by several reasons, for example, a production failure, a sudden rise in demand, or, more importantly, an inaccurate forecast of the production levels of other renewable energy sources. The reserve-up capacities

[5] We consider the electricity markets in the control area of APG, Austria. The tendering period can be different in other control areas. However, this does not fundamentally change our model.

may be activated when the demand is higher than the supply and more electricity is needed. On the other hand, the reserve-down capacities may be activated when there is an oversupply of electricity in the transmission system operator's control area. Upon the request to activate the reserve-up capacity, the generation company has two choices to honor the reserve commitments. It can either increase the electricity production or decrease the electricity consumption. In the case that a significant amount of additional electricity is required, it may also be possible that the generation company has to take both actions. Similar arguments in the opposite direction apply to the reserve-down capacities.

From the generation company's perspective, the benefit of participating in the reserve markets is two-fold. First, the transmission system operator offers compensations for the reserve capacities in the form of an advance payment (i.e., reserve capacity fees). Second, the company is also paid for the hydroelectricity production/consumption when the reserve capacities are activated.

Participation in the reserve markets complicates the producer's planning problem because it involves more decisions. Such decisions include $u_{t,a} \geq 0$ and $d_{t,a} \geq 0$ (for all $t \in \mathscr{T}$ and $a \in \mathscr{A}$) representing the reserve-up and reserve-down capacities, respectively. Similarly to previously introduced decision variables, we aggregate these decisions over $a \in \mathscr{A}$ as

$$\boldsymbol{u}_t := [u_{t,a}]_{a\in\mathscr{A}} \quad \text{and} \quad \boldsymbol{d}_t := [d_{t,a}]_{a\in\mathscr{A}}.$$

Moreover, additional constraints have to be incorporated in order to ensure that the company is able to honor the commitment in both the spot and reserve markets. These new constraints are:

$$\begin{aligned}
\boldsymbol{0} \leq \boldsymbol{u}_t \leq \overline{\boldsymbol{g}} - \boldsymbol{g}_t + \boldsymbol{p}_t, \quad \forall t \in \mathscr{T}, \\
\boldsymbol{0} \leq \boldsymbol{d}_t \leq \overline{\boldsymbol{p}} - \boldsymbol{p}_t + \boldsymbol{g}_t, \quad \forall t \in \mathscr{T}.
\end{aligned}$$

As discussed above, the upper bound on the reserve-up capacity $u_{t,a}$ (reserve-down capacity $d_{t,a}$) is given by the sum of the production buffer $\overline{g}_a - g_{t,a}$ (consumption buffer $\overline{p}_a - p_{t,a}$) and the consumption level $p_{t,a}$ (production level $g_{t,a}$). Furthermore to take into account the effects of reserve capacities, the dynamics of the reservoirs' storage levels for all $t \in \mathcal{T}$ and $i \in \mathcal{I}$ is amended to

$$\boldsymbol{v}_t = \boldsymbol{v}_{t-1} + \boldsymbol{\psi}_t - \mathbf{M}\left(\boldsymbol{s}_t + \boldsymbol{g}_t - \boldsymbol{p}_t + \rho_t^{\mathrm{u}} \boldsymbol{u}_t - \rho_t^{\mathrm{d}} \boldsymbol{d}_t\right),$$

where $\rho_t^u = 1$ implies an activation for the reserve-up energy (= 0 otherwise) and $\rho_t^d = 1$ implies an activation for the reserve-down energy (= 0 otherwise). A generation company that participates in both the spot and reserve markets earns the total revenue of

$$\sum_{t \in \mathcal{T}} \boldsymbol{c}^\top \left((\phi_t^{\mathrm{u}} + \rho_t^{\mathrm{u}} \pi_t^{\mathrm{u}}) \boldsymbol{u}_t + (\phi_t^{\mathrm{d}} + \rho_t^{\mathrm{d}} \pi_t^{\mathrm{d}}) \boldsymbol{d}_t + \pi_t^{\mathrm{s}} (\boldsymbol{g}_t - \mathbf{D} \boldsymbol{p}_t)\right),$$

where $(\phi_t^{\mathrm{u}}, \pi_t^{\mathrm{u}})$ and $(\phi_t^{\mathrm{d}}, \pi_t^{\mathrm{d}})$ are the capacity fees (€/MWh) and the electricity prices (€/MWh) in the reserve-up and reserve-down markets, respectively.

Lastly, we assume that the generation company is a price taker, that is, its operation is not large enough to have a non-negligible influence on the electricity prices, be it spot prices ($\pi_t^{\mathrm{s}}$), capacity fees ($\phi_t^{\mathrm{u}}, \phi_t^{\mathrm{d}}$) or reserve prices ($\pi_t^{\mathrm{u}}, \pi_t^{\mathrm{d}}$).

## 4.3 Revenue Maximization

Based on the discussion in Section 4.2, the generation company may formulate an optimization problem to identify a revenue maximizing policy subject to the constraints previously described. In practice, though, it is impossible to have a perfect forecast of the future inflows ($\psi_{t,i}$), the electricity prices in the different markets ($\pi_t^{\mathrm{s}}, \pi_t^{\mathrm{u}}, \pi_t^{\mathrm{d}}$), the reserve capacity fees ($\phi_t^{\mathrm{u}}, \phi_t^{\mathrm{d}}$) and the activation sequences of the reserve capacities ($\rho_t^{\mathrm{u}}, \rho_t^{\mathrm{d}}$). The optimization problem thus needs to account for uncertainty in the market and inflow information.

We highlight that it is of great importance to explicitly account for uncertainty in this problem. This is the case, for example, if hydroelectricity generation is profitable in only a few hours with high spot prices within a week. Assuming a constant spot price throughout the week may lead to a highly suboptimal solution, that is, not to generate at all. Henceforth, we assume that the generation company is risk-neutral. Therefore, it aims to identify a reservoir management policy that is feasible almost surely under the joint probability distribution $\mathbb{P}$ of the relevant random variables and maximizes the expected total revenue. In order to respect non-anticipativity, future decisions can exploit knowledge of the past, but not that of the future. Thus, it is necessary for us to introduce the sequence of revealed information which evolves as time passes.

For the sake of concise notation, we denote by $\boldsymbol{\xi}_t$ the information that is revealed to the generation company at the beginning of trading hour $t$. Furthermore, we denote the first hour of the day and the first hour of the week containing $t$ by $d(t)$ and $w(t)$, respectively, and we let $t(w)$ denote the first trading hour of the operational week $w$. Last but not least, we denote by $\boldsymbol{\xi}$ (without subscript) the information available at the end of the planning horizon, that is, $\boldsymbol{\xi} = (\boldsymbol{\xi}_1;\ldots;\boldsymbol{\xi}_{T+1})$.

In our hydropower scheduling problems, natural inflows $\boldsymbol{\psi}_t$ as well as electricity prices $\pi^{\mathrm{u}}_t, \pi^{\mathrm{d}}_t$ and activations $\rho^{\mathrm{u}}_t, \rho^{\mathrm{d}}_t$ in the reserve markets are assumed to materialize hourly. On the other hand, daily spot prices are revealed at the beginning of the day, i.e., when $t = d(t)$, whereas weekly reserve capacity fees are revealed at the beginning of the week, i.e., when $t = w(t)$. We summarize the information that is revealed to the generation company at the beginning of hour $t \in \mathcal{T}$ as follows.

$$\boldsymbol{\xi}_t = \begin{cases} \left(\rho^{\mathrm{u}}_{t-1}, \rho^{\mathrm{d}}_{t-1}, \pi^{\mathrm{u}}_{t-1}, \pi^{\mathrm{d}}_{t-1}, \boldsymbol{\psi}^{\mathsf{T}}_{t-1}\right)^{\mathsf{T}} & \text{if } t \neq d(t), \\ \left(\rho^{\mathrm{u}}_{t-1}, \rho^{\mathrm{d}}_{t-1}, \pi^{\mathrm{u}}_{t-1}, \pi^{\mathrm{d}}_{t-1}, \boldsymbol{\psi}^{\mathsf{T}}_{t-1}, \pi^{\mathrm{s}}_{t}, \ldots, \pi^{\mathrm{s}}_{t+23}\right)^{\mathsf{T}} & \text{if } t = d(t) \text{ and } t \neq w(t), \\ \left(\rho^{\mathrm{u}}_{t-1}, \rho^{\mathrm{d}}_{t-1}, \pi^{\mathrm{u}}_{t-1}, \pi^{\mathrm{d}}_{t-1}, \boldsymbol{\psi}^{\mathsf{T}}_{t-1}, \pi^{\mathrm{s}}_{t}, \ldots, \pi^{\mathrm{s}}_{t+23}, \right. & \\ \left. \phi^{\mathrm{u}}_{t}, \ldots, \phi^{\mathrm{u}}_{t+167}, \phi^{\mathrm{d}}_{t}, \ldots, \phi^{\mathrm{d}}_{t+167}\right)^{\mathsf{T}} & \text{if } t = w(t) \end{cases}$$

Recall that the spot market is a day-ahead market. Therefore, when $t = d(t)$, the generation company decides on the trading volumes in the spot market for the next 24 hours, i.e., $\boldsymbol{g}_\tau$ and $\boldsymbol{p}_\tau$ for $\tau = t, \ldots, t+23$. On the other hand, the reserve markets are cleared week-ahead. Thus, when $t = w(t)$, the generation company has to decide on the reserve capacities for the next whole week, i.e., $\boldsymbol{u}_\tau$ and $\boldsymbol{d}_\tau$ for $\tau = t, \ldots, t+167$.

The company may then solve

$$
\begin{aligned}
\max\ & \mathbb{E}\left(\sum_{t\in\mathcal{T}} \boldsymbol{c}^\top \left(\left(\phi_t^{\mathrm{u}}+\rho_t^{\mathrm{u}}\pi_t^{\mathrm{u}}\right)\boldsymbol{u}_t + \left(\phi_t^{\mathrm{d}}+\rho_t^{\mathrm{d}}\pi_t^{\mathrm{d}}\right)\boldsymbol{d}_t + \pi_t^{\mathrm{s}}\left(\boldsymbol{g}_t - \mathbf{D}\boldsymbol{p}_t\right)\right)\right) \\
\text{s.t.}\ & \boldsymbol{g}_t,\boldsymbol{p}_t \sim \mathcal{F}_{d(t)},\ \boldsymbol{u}_t,\boldsymbol{d}_t \sim \mathcal{F}_{w(t)},\ \boldsymbol{s}_t,\boldsymbol{v}_t \sim \mathcal{F}_{t+1} \quad \forall t\in\mathcal{T} \\
& \left.\begin{aligned}
& \mathbf{0}\le \boldsymbol{g}_t \le \overline{\boldsymbol{g}}, \quad \mathbf{0}\le \boldsymbol{p}_t \le \overline{\boldsymbol{p}}, \quad \mathbf{0}\le \boldsymbol{s}_t \\
& \boldsymbol{u}_t \le \overline{\boldsymbol{g}} - \boldsymbol{g}_t + \boldsymbol{p}_t \\
& \boldsymbol{d}_t \le \overline{\boldsymbol{p}} - \boldsymbol{p}_t + \boldsymbol{g}_t \\
& \boldsymbol{v}_t = \boldsymbol{v}_{t-1} + \boldsymbol{\psi}_t - \mathbf{M}\left(\boldsymbol{s}_t + \boldsymbol{g}_t - \boldsymbol{p}_t + \rho_t^{\mathrm{u}}\boldsymbol{u}_t - \rho_t^{\mathrm{d}}\boldsymbol{d}_t\right) \\
& \underline{\boldsymbol{v}}_t \le \boldsymbol{v}_t \le \overline{\boldsymbol{v}}_t
\end{aligned}\right\} \ \forall t\in\mathcal{T},\ \mathbb{P}\text{-a.s.,}
\end{aligned}
\tag{4.1}
$$

which represents a multistage stochastic program. Here, the objective is to maximize the expected revenue accumulated over the planning horizon from the reserve-up, reserve-down, and spot markets. The constraints were previously discussed in Section 4.2. It is worth noting again that all decisions must be non-anticipative, and we capture the non-anticipativity requirements by using a $\sigma$-algebra

$$\mathcal{F}_t = \sigma(\boldsymbol{\xi}_s | 1 \le s \le t)$$

generated by the observation history up to time $t\in\mathcal{T}$. Since the reserve market is a week-ahead market, $\boldsymbol{u}_t$ and $\boldsymbol{d}_t$ are adapted to the $\sigma$-algebra $\mathcal{F}_{w(t)}$. On the other hand, the spot market is a day-ahead market; hence, $\boldsymbol{g}_t$ and $\boldsymbol{p}_t$ are adapted to $\mathcal{F}_{d(t)}$. Lastly, $\boldsymbol{s}_t$ and $\boldsymbol{v}_t$ are adapted to $\mathcal{F}_{t+1}$ as the storage levels are monitored and adjusted hourly.

Multistage stochastic programs are difficult to solve numerically. In fact, determining an exact solution of a two-stage stochastic program is already computationally intractable when the random variables follow independent uniform distributions; see Dyer and Stougie (2006); Hanasusanto et al. (2015a). To achieve tractability, we propose in Section 4.4 a framework to decompose the stochastic program (4.1) into a two-layer stochastic programming problem. After the decomposition, a number of smaller stochastic programs, which consist of much fewer decision stages and random variables, emerge. Moreover, many of them can be solved in parallel. Nonetheless, they still constitute multistage stochastic programs. To simplify the problem even more, we solve these stochastic programs approximately in linear decision rules, that is, we express future decisions as affine functions of prior observations. In many cases, including ours, a linear decision rule approximation ensures the tractability of dynamic optimization problems under uncertainty.

## 4.4 Planner-Trader Decomposition

The stochastic program (4.1) consists of a large number of decision stages as the planning horizon usually comprises at least one year. As a first step to simplify the problem, we decompose (4.1) into a collection of smaller and more tractable stochastic programs. Our decomposition is motivated by the following observation concerning the problem's dynamics. Electricity prices are volatile and can differ greatly from one hour to another. On the other hand, the dynamics of the reservoirs' storage levels change on a much coarser scale. For a cascade of seasonal reservoirs, it takes at least weeks to fully deplete or to fully replenish each reservoir, and hourly changes in the storage levels are often marginal. This contrast between electricity prices and hydrological dynamics can be used to simplify (4.1) as follows.

Since the reservoirs' storage levels change slowly, we split the entire planing horizon $\mathcal{T}$ into weeks $\mathcal{W} := \{1, \ldots, W\}$, and we monitor the storage levels only at the end of each week. Denote by $\mathcal{T}(w)$ the set of trading hours within week $w$. Hence, the entire planning horizon can be expressed as $\mathcal{T} = \bigcup_{w \in \mathcal{W}} \mathcal{T}(w)$. By the tower property of conditional expectations, we may then rewrite the objective function of (4.1) as

$$\sum_{w \in \mathcal{W}} \sum_{a \in \mathcal{A}} \mathbb{E}\left( \sum_{t \in \mathcal{T}(w)} \left( \hat{\pi}^{\mathrm{u}}_{t,a} u_{t,a} + \hat{\pi}^{\mathrm{d}}_{t,a} d_{t,a} + \hat{\pi}^{\mathrm{s}}_{t,a} (g_{t,a} - d_a p_{t,a}) \right) \right)$$

$$= \sum_{w \in \mathcal{W}} \sum_{a \in \mathcal{A}} \mathbb{E}\left( \mathbb{E}\left( \sum_{t \in \mathcal{T}(w)} \left( \hat{\pi}^{\mathrm{u}}_{t,a} u_{t,a} + \hat{\pi}^{\mathrm{d}}_{t,a} d_{t,a} + \hat{\pi}^{\mathrm{s}}_{t,a} (g_{t,a} - d_a p_{t,a}) \right) \middle| \boldsymbol{\xi}_{t(w)} \right) \right),$$

where $\hat{\pi}^{\mathrm{u}}_{t,a}$, $\hat{\pi}^{\mathrm{d}}_{t,a}$ and $\hat{\pi}^{\mathrm{s}}_{t,a}$ are introduced for notational simplicity as net revenues earned from the reserve-up, reserve-down, and spot markets, respectively, for comitting 1 $\mathrm{m}^3$/h of water release, i.e.,

$$\hat{\pi}^{\mathrm{u}}_{t,a} := c_a(\phi^{\mathrm{u}}_t + \rho^{\mathrm{u}}_t \pi^{\mathrm{u}}_t), \quad \hat{\pi}^{\mathrm{d}}_{t,a} := c_a(\phi^{\mathrm{d}}_t + \rho^{\mathrm{d}}_t \pi^{\mathrm{d}}_t), \quad \hat{\pi}^{\mathrm{s}}_{t,a} := c_a \pi^{\mathrm{s}}_t.$$

The outer expectation is taken with respect to $\boldsymbol{\xi}_{t(w)}$, and the inner expectation is taken with respect to $\hat{\pi}^{\mathrm{u}}_t$, $\hat{\pi}^{\mathrm{d}}_t$ and $\hat{\pi}^{\mathrm{s}}_t$ (for $t \in \mathcal{T}(w)$) conditionally on $\boldsymbol{\xi}_{t(w)}$. We are now ready to discuss the planner-trader decomposition which is used to simplify the full model (4.1).

At the beginning of each week $w \in \mathcal{W}$, the generation company solves a reservoir management problem, that is, it sets a weekly target of productive release, say $q_{w,a}$, for every arc $a \in \mathcal{A}$ of the hydrological system. We highlight that $q_{w,a}$ is not necessarily positive as it can take a negative value when the net volume of the pumped water exceeds the net volume of the released water. Subsequently, after deciding on $q_{w,a}$, the generation company has to optimize the intra-week market transactions, that is, it has to trade in both the spot and reserve markets on an hourly basis while respecting the weekly targets of water consumption $\boldsymbol{q}_w := [q_{w,a}]_{a\in\mathcal{A}}$. At the end of the week, the weekly aggregate inflows, from both natural sources and upstream reservoirs, then materialize.

This explains the main idea behind the planner-trader decomposition where the planner optimizes the reservoir management with weekly granularity, whereas the trader optimizes the intra-week market transactions with hourly granularity. Referring to the graph representation of the cascade's topology, the planner optimizes the allocation of water over the set of nodes $\mathcal{I}$ to maintain the reservoirs' storage levels, and the trader optimizes hourly transactions locally for each individual arc $a \in \mathcal{A}$.

Before we formulate the planning and the trading problems mathematically, we impose the following statistical assumptions about the relevant random variables.

- (A1) The process $[\boldsymbol{\psi}_t]_{t\in\mathcal{T}}$ is statistically independent of the other exogenous stochastic processes.
- (A2) The stochastic processes $[\pi^{\mathrm{u}}_t]_{t\in\mathcal{T}}, [\pi^{\mathrm{d}}_t]_{t\in\mathcal{T}}, [\rho^{\mathrm{u}}_t]_{t\in\mathcal{T}}, [\rho^{\mathrm{d}}_t]_{t\in\mathcal{T}}$ are serially independent.
- (A3) The stochastic process $[\pi^{\mathrm{s}}_t]_{t\in\mathcal{T}}$ is Markovian.

We remark that the planner-trader decomposition is still implementable even if Assumptions (A1)–(A3) fail to hold. However, these assumptions strengthen the approximation quality and improve the tractability of the decomposition. We are now ready to state our decomposition scheme.

### 4.4.1 Trading Subproblems

The information crucial to the decision making process of the trader is that from within the week. This observation is due to Assumptions (A2) and (A3). For trading hour $t \in \mathscr{T}$, we define a new $\sigma$-algebra $\mathscr{F}'_t \subset \mathscr{F}_t$ as

$$\mathscr{F}'_t = \sigma(\boldsymbol{\xi}_s | w(t) \leq s \leq t). \tag{4.2}$$

To maximize the expected weekly revenue with the hydroelectricity production and consumption from hydrological arc $a$, the trader solves

$$\begin{aligned}
\max \quad & \mathbb{E}\left(\sum_{t\in\mathscr{T}(w)} \left(\hat{\pi}^{\mathrm{u}}_{t,a} u_{t,a} + \hat{\pi}^{\mathrm{d}}_{t,a} d_{t,a} + \hat{\pi}^{\mathrm{s}}_{t,a}\left(g_{t,a} - d_a p_{t,a}\right)\right) \middle| \boldsymbol{\xi}_{t(w)}\right) \\
\text{s.t.} \quad & g_{t,a}, p_{t,a} \sim \mathscr{F}'_{d(t)},\ u_{t,a}, d_{t,a} \in \mathbb{R}_+ \quad \forall t \in \mathscr{T}(w) \\
& \left.\begin{aligned} & 0 \leq g_{t,a} \leq \overline{g}_a, \quad 0 \leq p_{t,a} \leq \overline{p}_a \\ & u_{t,a} \leq \overline{g}_a - g_{t,a} + p_{t,a}, \quad d_{t,a} \leq \overline{p}_a - p_{t,a} + g_{t,a} \end{aligned}\right\} \forall t \in \mathscr{T}(w),\ \mathbb{P}\text{-a.s.,} \\
& \sum_{t\in\mathscr{T}(w)} \left(g_{t,a} - p_{t,a} + \rho^{\mathrm{u}}_t u_{t,a} - \rho^{\mathrm{d}}_t d_{t,a}\right) \leq q_{w,a}
\end{aligned} \tag{4.3}$$

where the weekly target of water release $q_{w,a}$ is determined by the planner and is therefore exogenous to the trader's problem. We then denote the optimal objective value of this trading problem by $\Pi_{w,a}\left(q_{w,a}, \boldsymbol{\xi}_{t(w)}\right)$. A detailed description of the stochastic modeling of the market uncertainty $[\pi^{\mathrm{u}}_t]_{t\in\mathscr{T}}$, $[\pi^{\mathrm{d}}_t]_{t\in\mathscr{T}}$, $[\rho^{\mathrm{u}}_t]_{t\in\mathscr{T}}$, $[\rho^{\mathrm{u}}_t]_{t\in\mathscr{T}}$ and $[\pi^{\mathrm{s}}_t]_{t\in\mathscr{T}}$ is relegated to Section 4.7, where we report on the experimental results of our model. Note that the last constraint in (4.3) is imposed as an inequality (as opposed to an equality) constraint. Indeed, we will see in the proof of Theorem 4.1 that imposing the water target as an equality would disallow exploitation of the information materializing within a week.

### 4.4.2 Planning Problem

The planner receives information about the expected weekly revenues from the trader through the functionals $\Pi_{w,a}$ and manages the water resources in the cascade in order to maximize the expected earnings of the generation company over the entire planning horizon $\mathscr{T}$. Particularly, the planner solves

$$
\begin{aligned}
&\max\ \mathbb{E}\left(\textstyle\sum_{w\in\mathscr{W}}\sum_{a\in\mathscr{A}}\Pi_{w,a}\left(q_{w,a},\boldsymbol{\xi}_{t(w)}\right)\right)\\
&\text{s.t.}\quad \boldsymbol{q}_w\sim\mathscr{F}_{t(w)},\ \boldsymbol{s}_w\sim\mathscr{F}_{t(w+1)}\quad \forall w\in\mathscr{W}\\
&\qquad\left.\begin{aligned}&\underline{\boldsymbol{q}}_w\le\boldsymbol{q}_w\le\overline{\boldsymbol{q}}_w,\quad \boldsymbol{s}_w\ge\mathbf{0}\\ &\underline{\boldsymbol{v}}_w^+\le\boldsymbol{v}_0+\textstyle\sum_{\omega=1}^{w}\left(\boldsymbol{\psi}_\omega^+-\mathbf{M}\left(\boldsymbol{s}_\omega+\boldsymbol{q}_\omega\right)\right)\le\overline{\boldsymbol{v}}_w^+\end{aligned}\right\}\ \forall w\in\mathscr{W},\ \mathbb{P}\text{-a.s.}
\end{aligned}
\tag{4.4}
$$

Here, the planner optimizes weekly target release quantities $q_{w,a}$ for every $(w,a)\in\mathscr{W}\times\mathscr{A}$. The target $q_{w,a}$ attains its upper bound when the turbine operates at a maximum level throughout the week, whereas it attains its lower bound when the pump operates maximally. Hence, to make (4.4) complete, we set $\underline{\boldsymbol{q}}_w$ and $\overline{\boldsymbol{q}}_w$ to $-168\overline{\boldsymbol{p}}$ and $+168\overline{\boldsymbol{g}}$, respectively. Note that, for notational convenience, in the formulation above we introduce $\boldsymbol{\psi}_w^+$ as an aggregate weekly inflow, i.e., $\boldsymbol{\psi}_w^+=\sum_{t\in\mathscr{T}(w)}\boldsymbol{\psi}_t$. Similarly, we introduce $\underline{\boldsymbol{v}}_w^+$ and $\overline{\boldsymbol{v}}_w^+$ as the lower and upper bounds on the storage levels at the end of week $w$, i.e., $\underline{\boldsymbol{v}}_w^+=\underline{\boldsymbol{v}}_t$ and $\overline{\boldsymbol{v}}_w^+=\overline{\boldsymbol{v}}_t$ for $t$ such that $t+1=t(w+1)$. This corresponds to the earlier observation that the storage levels may not need updating every hour. By a slight abuse of notation, $\boldsymbol{s}_w$ in (4.4) represents weekly spilling decisions, not to be confused with hourly spilling decisions $\boldsymbol{s}_t$ in (4.1).

We end this section by discussing the accuracy of the planner-trader decomposition. Theorem 4.1 below asserts that the planner-trader decomposition provides a conservative approximation to a variant of (4.1) when the intra-week bounds on the storage levels are omitted.

**Theorem 4.1** (Suboptimality of planner-trader decomposition)**.** *If* $\underline{v}_{t,a}=-\infty$ *and* $\overline{v}_{t,a}=+\infty$ *for every* $t\in\mathscr{T}:t+1\neq w(t+1)$*, then the planner-trader decomposition* (4.4) *provides a conservative approximation for* (4.1).

*Proof.* Under the assumption that the reservoir bounds are imposed only at the end of each week, there is no need for us to keep track of the intra-week storage levels. Therefore, we can simplify (4.1) by aggregating the inflows and the spilling decisions as well as the productive release quantities over the trading hours in week $w$ as

$$\boldsymbol{\psi}_w^+ = \sum_{t\in\mathcal{T}(w)} \boldsymbol{\psi}_t, \quad \boldsymbol{s}_w^+ = \sum_{t\in\mathcal{T}(w)} \boldsymbol{s}_t, \quad \text{and} \quad \boldsymbol{q}_w = \sum_{t\in\mathcal{T}(w)} \boldsymbol{g}_t - \boldsymbol{p}_t + \rho_t^{\mathrm{u}}\boldsymbol{u}_t - \rho_t^{\mathrm{d}}\boldsymbol{d}_t.$$

Recall that since $\boldsymbol{s}_t \sim \mathcal{F}_{t+1}$, we have that $\boldsymbol{s}_w^+ \sim \mathcal{F}_{t(w+1)}$. Moreover, since both $\boldsymbol{g}_t$ and $\boldsymbol{p}_t$ are adapted to $\mathcal{F}_{d(t)}$ and $\rho_t^{\mathrm{u}}, \rho_t^{\mathrm{d}}$ are contained in $\boldsymbol{\xi}_{t+1}$, it follows that $\boldsymbol{q}_w \sim \mathcal{F}_{t(w+1)}$. If $\boldsymbol{v}_w^+$ denotes the vector of storage levels at the end of week $w$, we find

$$\boldsymbol{v}_w^+ = \boldsymbol{v}_{w-1}^+ + \boldsymbol{\psi}_w^+ - \mathbf{M}\left(\boldsymbol{s}_w^+ + \boldsymbol{q}_w\right),$$

and therefore $\boldsymbol{v}_w^+ \sim \mathcal{F}_{t(w+1)}$. We can then re-express (4.1) as

$$\begin{aligned}
\max \quad & \textstyle\sum_{w\in\mathcal{W}} \sum_{a\in\mathcal{A}} \mathbb{E}\Big(\mathbb{E}\Big(\sum_{t\in\mathcal{T}(w)} \big(\hat{\pi}_{t,a}^{\mathrm{u}} u_{t,a} + \hat{\pi}_{t,a}^{\mathrm{d}} d_{t,a} + \hat{\pi}_{t,a}^{\mathrm{s}}(g_{t,a} - d_a p_{t,a})\big) \Big| \boldsymbol{\xi}_{w(t)}\Big)\Big) \\
\text{s.t.} \quad & \boldsymbol{g}_t, \boldsymbol{p}_t \sim \mathcal{F}_{d(t)}, \ \boldsymbol{u}_t, \boldsymbol{d}_t \sim \mathcal{F}_{w(t)} \quad \forall t \in \mathcal{T} \\
& \left.\begin{aligned}
& \mathbf{0} \le \boldsymbol{g}_t \le \overline{\boldsymbol{g}}, \quad \mathbf{0} \le \boldsymbol{p}_t \le \overline{\boldsymbol{p}} \\
& \boldsymbol{u}_t \le \overline{\boldsymbol{g}} - \boldsymbol{g}_t + \boldsymbol{p}_t \\
& \boldsymbol{d}_t \le \overline{\boldsymbol{p}} - \boldsymbol{p}_t + \boldsymbol{g}_t
\end{aligned}\right\} \ \forall t \in \mathcal{T}, \ \mathbb{P}\text{-a.s.} \\
& \boldsymbol{q}_w, \boldsymbol{s}_w^+, \boldsymbol{v}_w^+ \sim \mathcal{F}_{t(w+1)} \quad \forall w \in \mathcal{W} \\
& \left.\begin{aligned}
& \mathbf{0} \le \boldsymbol{s}_w^+, \quad \underline{\boldsymbol{v}}_w^+ \le \boldsymbol{v}_w^+ \le \overline{\boldsymbol{v}}_w^+ \\
& \boldsymbol{q}_w = \textstyle\sum_{t\in\mathcal{T}(w)} \boldsymbol{g}_t - \boldsymbol{p}_t + \rho_t^{\mathrm{u}}\boldsymbol{u}_t - \rho_t^{\mathrm{d}}\boldsymbol{d}_t \\
& \boldsymbol{v}_w^+ = \boldsymbol{v}_{w-1}^+ + \boldsymbol{\psi}_w^+ - \mathbf{M}\left(\boldsymbol{s}_w^+ + \boldsymbol{q}_w\right)
\end{aligned}\right\} \ \forall w \in \mathcal{W}, \ \mathbb{P}\text{-a.s.,}
\end{aligned}$$

where the measurability properties of the aggregate decisions follow directly from their construction. Next, we argue that the equality constraint which assigns a value to the weekly net productive release $\boldsymbol{q}_w$ can be equivalently rewritten as a greater than or equal to constraint. Indeed, if there is a positive slack between the left hand side and the right hand side of this constraint, we can add this slack to the corresponding spilling decision $\boldsymbol{s}_w$ as there are no upper bounds on spilling. In this way, the weekly total releases (both productive and non-productive) $\boldsymbol{q}_w + \boldsymbol{s}_w^+$ remain unchanged.

By enforcing $\boldsymbol{q}_w$ to be adapted only to $\mathscr{F}_{t(w)}$ (instead of $\mathscr{F}_{t(w+1)}$), we obtain a conservative approximation of (4.1), and the emerging problem is equivalent to the proposed planner-trader decomposition because we can assign hourly constraints (those in the first bracket) to the corresponding trading subproblems and assign weekly constraints (those in the second bracket) to the planning problem. ∎

## 4.5 Multiscale Approximation

The focus of this section is to outline numerical methods for solving the planning and the trading problems efficiently. Notice that, even after the decomposition, the emerging optimization problems still constitute multistage stochastic programs, which are computationally hard to solve. Henceforth, we solve them approximately in linear decision rules, that is, we restrict the future decisions to affine functions of the observable data.

We assume that we are given a set of independent samples $\{\boldsymbol{\xi}^s : s \in \mathscr{S}\}$ of the random vector $\boldsymbol{\xi}$, where the index set $\mathscr{S}$ is defined as $\{1, \ldots, S\}$. This allows us to solve the planning problem (4.4) by replacing its objective function with the sample average

$$\frac{1}{S} \sum_{s\in\mathscr{S}} \sum_{w\in\mathscr{W}} \sum_{a\in\mathscr{A}} \Pi_{w,a}\left(q_{w,a}, \boldsymbol{\xi}^s_{t(w)}\right)$$

and by replacing all almost sure constrains by robust constraints over a polyhedral uncertainty set $\Xi$ of the form

$$\Xi = \left\{ \boldsymbol{\xi} : \begin{array}{ll} \phi^{\mathrm{u}}_t,\, \phi^{\mathrm{d}}_t,\, \pi^{\mathrm{s}}_t,\, \pi^{\mathrm{u}}_t,\, \pi^{\mathrm{d}}_t,\, \rho^{\mathrm{u}}_t,\, \rho^{\mathrm{d}}_t \geq 0,\ \boldsymbol{\psi}_t \geq \mathbf{0} & \forall t \in \mathscr{T} \\ \rho^{\mathrm{u}}_t + \rho^{\mathrm{d}}_t \leq 1,\ \underline{\phi}^{\mathrm{u}}_t \leq \phi^{\mathrm{u}}_t \leq \overline{\phi}^{\mathrm{u}}_t,\ \underline{\phi}^{\mathrm{d}}_t \leq \phi^{\mathrm{d}}_t \leq \overline{\phi}^{\mathrm{d}}_t & \forall t \in \mathscr{T} \\ \underline{\pi}^{\mathrm{s}}_t \leq \pi^{\mathrm{s}}_t \leq \overline{\pi}^{\mathrm{s}}_t,\ \underline{\pi}^{\mathrm{u}}_t \leq \pi^{\mathrm{u}}_t \leq \overline{\pi}^{\mathrm{u}}_t,\ \underline{\pi}^{\mathrm{d}}_t \leq \pi^{\mathrm{d}}_t \leq \overline{\pi}^{\mathrm{d}}_t & \forall t \in \mathscr{T} \\ \underline{\boldsymbol{\psi}}^+_w \leq \sum_{t\in\mathscr{T}(w)} \boldsymbol{\psi}_t \leq \overline{\boldsymbol{\psi}}^+_w & \forall w \in \mathscr{W} \end{array} \right\}. \tag{4.5}$$

We emphasize that any feasible solution of the resulting optimization problem is also feasible for the realizations $\boldsymbol{\xi} \in \Xi \backslash \{\boldsymbol{\xi}^s : s \in \mathscr{S}\}$. Observe that much of the description of $\Xi$ is explained by component-wise box constraints, with two exceptions. First, $\rho_t^{\mathrm{u}} + \rho_t^{\mathrm{d}} \leq 1$ is imposed because the reserve-up and the reserve-down capacities cannot be activated in the same trading hour. The other exception is $\underline{\boldsymbol{\psi}}_w^+ \leq \sum_{t\in\mathcal{T}(w)} \boldsymbol{\psi}_t \leq \overline{\boldsymbol{\psi}}_w^+$, which corresponds to the bounds on the aggregate weekly inflows instead of individual hourly inflows. The motivation for using these aggregate bounds is that, as we will describe below, in the planning problem (4.4) we aggregate hourly inflows to weekly inflows for tractability reasons.

We then solve the resulting planning problem in linear decision rules by approximating the decisions $\boldsymbol{q}_w \sim \mathcal{F}_{t(w)}$ by affine functions of $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{t(w)}$. The decisions $\boldsymbol{s}_w$ can be approximated analogously using the slightly richer information base $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{t(w+1)}$. To solve the planning problem efficiently, we further restrict our attention to simplified linear decision rules of the form

$$\boldsymbol{q}_w = \mathbf{Q}_w \left(1; \boldsymbol{\psi}_1^+; \ldots; \boldsymbol{\psi}_{w-1}^+\right) \quad \text{and} \quad \boldsymbol{s}_w = \mathbf{S}_w \left(1; \boldsymbol{\psi}_1^+; \ldots; \boldsymbol{\psi}_w^+\right),$$

for some matrices $\mathbf{Q}_w$ and $\mathbf{S}_w$ with appropriate dimensions, where in the information bases we aggregate hourly inflows within the same week to weekly inflows and we exclude the market information from the planning problem. Hence, the planning problem can be approximated by

$$\begin{aligned}
\max \quad & \frac{1}{S} \sum_{s\in\mathcal{S}} \sum_{w\in\mathcal{W}} \sum_{a\in\mathcal{A}} \Pi_{w,a} \left( \mathbf{e}_a^\top \mathbf{Q}_w \left(1; \boldsymbol{\psi}_1^+; \ldots; \boldsymbol{\psi}_{w-1}^+\right), \boldsymbol{\xi}_{t(w)}^s \right) \\
\text{s.t.} \quad & \mathbf{Q}_w \in \mathbb{R}^{A\times[(w-1)I+1]}, \mathbf{S}_w \in \mathbb{R}^{A\times[wI+1]} \\
& \left.\begin{aligned}
& \underline{\boldsymbol{q}}_w \leq \mathbf{Q}_w \left(1; \boldsymbol{\psi}_1^+; \ldots; \boldsymbol{\psi}_{w-1}^+\right) \leq \overline{\boldsymbol{q}}_w \\
& \mathbf{S}_w \left(1; \boldsymbol{\psi}_1^+; \ldots; \boldsymbol{\psi}_w^+\right) \geq \mathbf{0} \\
& \underline{\boldsymbol{v}}_w^+ \leq \boldsymbol{v}_0 + \sum_{\omega=1}^{w} \left( \boldsymbol{\psi}_\omega^+ - \mathbf{M} \left( \mathbf{S}_\omega \left(1; \boldsymbol{\psi}_1^+; \ldots; \boldsymbol{\psi}_\omega^+\right) + \right.\right. \\
& \qquad\qquad \left.\left. \mathbf{Q}_\omega \left(1; \boldsymbol{\psi}_1^+; \ldots; \boldsymbol{\psi}_{\omega-1}^+\right)\right)\right) \leq \overline{\boldsymbol{v}}_w^+
\end{aligned}\right\} \quad \begin{aligned} & \forall w \in \mathcal{W} \\ & \forall \boldsymbol{\xi} \in \Xi, \end{aligned}
\end{aligned} \tag{4.6}$$

where $\mathbf{e}_a$ represents the $a$-th standard basis vector. The above optimization problem involves robust constraints over a polyhedral uncertainty set. Therefore, they admit a linear reformulation by using strong duality of linear programs; see, e.g., Ben-Tal et al. (2009); Kuhn et al. (2011). However, the objective function is not yet tractable because the trader's objectives $\Pi_{w,a}$ remain unknown. Nonetheless, it follows from perturbation and sensitivity analysis (see, e.g., Boyd and Vandenberghe (2004)) that $\Pi_{w,a}$ is concave in its first argument $q_{w,a}$.[6] We can exploit this concavity to convert (4.6) into a tractable linear program, if we further approximate the function $\Pi_{w,a}$ by a concave piecewise linear function

$$\Pi_{w,a}\left(q_{w,a}, \boldsymbol{\xi}^s_{t(w)}\right) \approx \min_{l \in \mathscr{L}} \alpha^{w,s,l} + \beta^{w,s,l} q_{w,a},$$

where $l \in \mathscr{L}$ is the index of the affine function characterized by a slope $\beta^{w,s,l}$ and a vertical intercept $\alpha^{w,s,l}$. Then, we can rewrite the objective function of (4.6) using an epigraphical reformulation. To find this piecewise approximation, we solve each trading subproblem $\Pi_{w,a}\big(q_{w,a}, \boldsymbol{\xi}^s_{t(w)}\big)$ for 7 different values of

$$q_{w,a} \in \left\{-168\overline{p}_a, -112\overline{p}_a, -56\overline{p}_a, 0, 56\overline{g}_a, 112\overline{g}_a, 168\overline{g}_a\right\}$$

a priori (in this case $|\mathscr{L}| = 6$ because each affine piece is identified by two consecutive discretized values of $q_{w,a}$). Note that, since the trading subproblems still constitute multistage stochastic programs, we also solve them in linear decision rules. This can be done in a similar manner as before, and thus the details are omitted for brevity. The number of trading subproblems to be formulated is equal to $S \times W \times A$, and each subproblem has to be resolved $|\mathscr{L}| + 1$ times. Moreover, all of them can be solved in parallel, which is an important benefit of the planner-trader decomposition.

We highlight that, in the trading subproblems, the weekly decisions $u_{t,a}$ and $d_{t,a}$ are modeled as here-and-now decisions, whereas the daily decisions $g_{t,a}$ and $p_{t,a}$ are adapted to $\mathscr{F}'_{d(t)}$, which is defined in (4.2). This is possible because of Assumptions (A2) and (A3). Last but not least, the inflow information $\boldsymbol{\psi}_t$ can be dropped from the information bases of the decisions $g_{t,a}$ and $p_{t,a}$ because of Assumption (A1).

## 4.6 Bang-Bang Strategy

The purpose of this section is to derive a heuristic for controlling the storage levels while trading in the electricity markets. This is an important issue to address because

[6]This observation remains valid when we solve the trading problems in linear decision rules.

wait-and-see decisions, unlike here-and-now decisions, are not readily available as they constitute functions and not numbers. Even after the planner-trader decomposition, the generation and pumping decisions in the spot market are still wait-and-see decisions in the trading subproblems (4.3). This motivates us to develop a heuristic to retrieve the values of these decisions at little computational costs. For the development of this heuristic, we require that each reservoir has at most one downstream reservoir, that is, for reservoirs $i, j_1, j_2 \in \mathscr{I}$ we impose that

(A4) $(i, j_1), (i, j_2) \in \mathscr{A} \implies j_1 = j_2.$

If the generation company participates only in the spot market, then a bang-bang strategy is nearly optimal and often employed; see e.g. Kovacevic et al. (2013, § 4.3) and Näsäkkälä and Keppo (2008). Under the bang-bang strategy, for each trading hour $t \in \mathscr{T}$ and for each hydrological arc $a \in \mathscr{A}$, one of the following three actions is taken, namely, the generation company either operates the turbine (if exists) at full power, operates the pump (if exists) at full power, or switches off both the turbine and the pump. The selection criterion is based on a comparison between spot prices and water values, where a water value (in €/m$^3$) measures the opportunity cost of releasing a cubic meter of water. We denote these water values by $\vartheta_{t,a}$.

To make spot prices and water values comparable, we divide the water value $\vartheta_{t,a}$ by the corresponding conversion rate $c_a$ in order to measure the opportunity cost of producing 1 MWh of hydroelectricity. In this case, Table 4.1 below presents the selection criterion and the corresponding actions underlying the bang-bang strategy. Simply put, when the spot price is high in comparison to the water value, then it is profitable for the generation company to generate and sell hydroelectricity. On the other hand, when the spot price is low, then the generation company should purchase electricity from the spot market and use it to pump water upstream.

| $\vartheta_{t,a}/c_a \le \pi^s_t$ | $\vartheta_{t,a}/c_a \in (\pi^s_t, d_a\pi^s_t]$ | $\vartheta_{t,a}/c_a > d_a\pi^s_t$ |
|---|---|---|
| $(\overline{g}_a, 0)$ | $(0, 0)$ | $(0, \overline{p}_a)$ |

Table 4.1: Description of the bang-bang strategy explained in terms of $(g_{t,a}, p_{t,a})$ for engaging in the spot market when the water value is given by $\vartheta_{t,a}$.

Incidentally, the bang-bang strategy can also be derived from a short-term variant of our planning model, which maximizes a net revenue in trading hour $t$ locally for each hydrological arc $a$:

$$\begin{aligned}(g_{t,a}, p_{t,a}) \in \text{argmax} \quad & g_{t,a}(c_a\pi_t^{\text{s}} - \vartheta_{t,a}) + p_{t,a}(\vartheta_{t,a} - d_a c_a \pi_t^{\text{s}}) \\ \text{s.t.} \quad & g_{t,a}, p_{t,a} \in \mathbb{R}_+ \\ & g_{t,a} \le \overline{g}_a,\ p_{t,a} \le \overline{p}_a.\end{aligned} \tag{4.7}$$

This linear program incorporates altogether the immediate revenues from trading in the spot market as well as the incurred opportunity costs. Furthermore, it can be shown to admit an analytical solution that coincides with the expressions presented in Table 4.1. In the remainder of this section, we outline how to obtain the water values, which are instrumental to implementing the bang-bang strategy, from our stochastic programs and how to extend the bang-bang strategy to the multiple-market case.

### 4.6.1 Determination of Water Values

Water values may be extracted from the optimization problem (4.1). To achieve this, we first consider the shadow prices (i.e., dual variables) of the storage level constraint

$$\boldsymbol{v}_t = \boldsymbol{v}_{t-1} + \boldsymbol{\psi}_t - \mathbf{M}\left(\boldsymbol{s}_t + \boldsymbol{g}_t - \boldsymbol{p}_t + \rho_t^{\text{u}}\boldsymbol{u}_t - \rho_t^{\text{d}}\boldsymbol{d}_t\right).$$

We denote the dual variables of this $I$-dimensional constraint by $\boldsymbol{\lambda}_t \in \mathbb{R}^I$. For trading hour $t$, $\lambda_{t,i}$ determines the monetary value of one additional cubic meter of water stored in reservoir $i$. Note that, for the generation company opting out of investment in the reserve markets, $\boldsymbol{u}_t$ and $\boldsymbol{d}_t$ can be restricted to $\mathbf{0}$. In this case, the following result still applies.

To determine the water value $\vartheta_{t,a}$, we denote the upstream reservoir of arc $a$ by $i \in \mathscr{I}$. Moreover, by Assumption (A4), reservoir $i$ has exactly one downstream reservoir (which can be either a real reservoir or the sink node $\otimes$), and we denote it by $i^-$. If $i$ has no downstream reservoir, i.e., $a = (i, \otimes)$, then $\vartheta_{t,a}$ coincides with $\lambda_{t,i}$. Assume now that $i^- \neq \otimes$. Then, the water value $\vartheta_{t,a}$ is thus given by the difference of the the shadow prices $\lambda_{t,i} - \lambda_{t,i^-}$ because a water outflow from $i$ constitutes to a water inflow to $i^-$.

This method of determining water values is still applicable when we approximate (4.1) by the planner-trader decomposition. In the planner-trader decomposition, we update the reservoirs' storage levels on a weekly basis in the planning problem (4.4), which implies that the shadow prices are also updated weekly. In other words, the water values $\vartheta_{t,a}$ are fixed throughout the week because of the planner-trader decomposition. Figure 4.3 illustrates the intra-week operation of a hydrological arc $a$ during a given week when the planner-trader decomposition and the bang-bang strategy are conjunctively used.

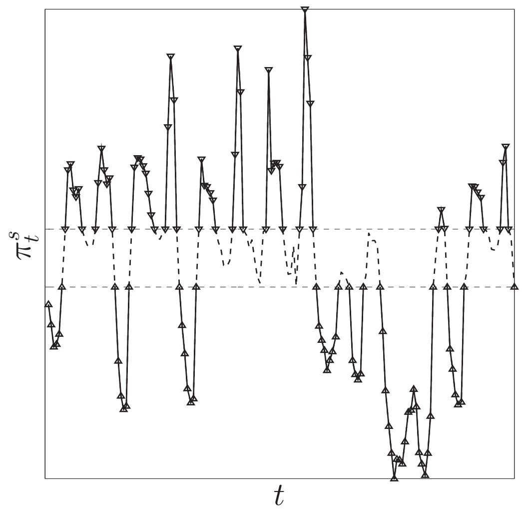



Figure 4.3: Weekly trading decisions in the spot market determined by the bang-bang strategy. The two horizontal dashed lines, from bottom to top, represent the values of $\vartheta_{t,a}/(c_a d_a)$ and $\vartheta_{t,a}/c_a$, respectively, where $\vartheta_{t,a}$ is kept constant throughout the week. When the spot price is above the top dashed line, the generation company releases water downstream (↓) to sell hydroelectricity, whereas when the spot price is under the bottom dashed line, the generation company purchases electricity to pump water upstream (↑).

### 4.6.2 Extension of the Bang-Bang Strategy

We are now ready to generalize the bang-bang strategy explained in Table 4.1 for the intra-week operation of the generation company wishing to participate in both the spot and reserve markets. To achieve this, we modify the linear program (4.7) by introducing additional constraints as follows.

$$\begin{aligned}(g_{t,a}, p_{t,a}) \in \operatorname{argmax} \quad & g_{t,a}(c_a\pi_t^{\mathrm{s}} - \vartheta_{t,a}) + p_{t,a}(\vartheta_{t,a} - d_a c_a \pi_t^{\mathrm{s}}) \\ \text{s.t.} \quad & g_{t,a}, p_{t,a} \in \mathbb{R}_+ \\ & g_{t,a} \le \overline{g}_a,\ p_{t,a} \le \overline{p}_a \\ & g_{t,a} \le \overline{g}_a - u_{t,a} + p_{t,a} \\ & p_{t,a} \le \overline{p}_a - d_{t,a} + g_{t,a}\end{aligned} \tag{4.8}$$

The new constraints model the obligations of the generation company arising from the reserve markets. We highlight that the reserve market decisions ($u_{t,a}$ and $d_{t,a}$) are chosen at the beginning of the week. They represent here-and-now decisions in the view of the trading subproblems. Therefore, they can be readily extracted upon solving the stochastic program (4.3). The trading decisions in the spot market, on the other hand, can instead be determined using the linear progrem (4.8). Henceforth, when considering this linear program, we assume that $u_{t,a}$ and $d_{t,a}$ are already determined.

Similarly to (4.7), we will now demonstrate that (4.8) also admits an analytical solution. Keeping in mind the results in Table 4.1, we begin our derivation by distinguishing three regimes for the water values: (i) $\vartheta_{t,a}/c_a \le \pi_t^s$, (ii) $\pi_t^s < \vartheta_{t,a}/c_a \le d_a\pi_t^s$ and (iii) $\vartheta_{t,a}/c_a > d_a\pi_t^s$. For the case (i), in view of the linear program (4.8), $g_{t,a}$ and $p_{t,a}$ should be set to their respective maximum and minimum, respectively. For the case (ii), both $g_{t,a}$ and $p_{t,a}$ should be set to their respective minima. Finally, for the case (iii), $g_{t,a}$ and $p_{t,a}$ should be set to their respective minimum and maximum. Note, however, that solving (4.8) is not as straightforward as solving (4.7) because of the coupling between the decision variables $g_{t,a}$ and $p_{t,a}$ in the last two constraints of (4.8).

To solve (4.8) analytically, we first visualize its feasible region in Figure 4.4. We highlight that the shape of the feasible set is determined by the reserve market decisions, which can be categorized into three disjoint cases: (i) $u_{t,a} \le \overline{g}_a$, $d_{t,a} \le \overline{p}_a$, (ii) $d_{t,a} > \overline{p}_a$ and (iii) $u_{t,a} > \overline{g}_a$. In particular, there is no overlap between the cases (ii) and (iii) because the last two constraints of (4.8) imply that $u_{t,a} + d_{t,a} \le \overline{g}_a + \overline{p}_a$.

Furthermore, one can show that it is always suboptimal to have the turbine and

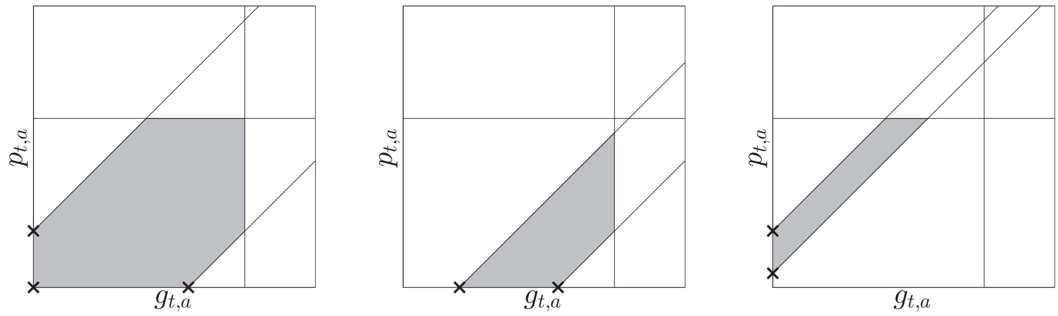



Figure 4.4: The feasible set of the linear program (4.8) is visualized by the shaded areas, from left to right, for (i) $u_{t,a} \leq \overline{g}_a$, $d_{t,a} \leq \overline{p}_a$, (ii) $d_{t,a} > \overline{p}_a$ and (iii) $u_{t,a} > \overline{g}_a$. The marked dots (×) represent the candidate optimal solutions for each case. For (i), the candidates are $(0,0)$, $(\overline{g}_a - u_{t,a}, 0)$ and $(0, \overline{p}_a - d_{t,a})$. For (ii), the candidates are $(\overline{g}_a - u_{t,a}, 0)$ and $(d_{t,a} - \overline{p}_a, 0)$. Finally, the candidates for (iii) are $(0, \overline{p}_a - d_{t,a})$ and $(0, u_{t,a} - \overline{g}_a)$.

the pump operating at the same time. Indeed, if $(g_{t,a}, p_{t,a}) = (g+\delta, p+\delta)$ for some $g, p \geq 0$ and $\delta > 0$ is feasible in (4.8), then a new solution $(g_{t,a}, p_{t,a}) = (g, p)$ is also feasible and yields a better objective value because $d_a > 1$. Hence we conclude that, at optimality, $g_{t,a}$ or $p_{t,a}$ vanishes. This observation allows us to readily solve the linear program (4.8), which in turn admits an analytical solution as displayed in Table 4.2.

| | $\vartheta_{t,a}/c_a \leq \pi_t^{\mathrm{s}}$ | $\vartheta_{t,a}/c_a \in (\pi_t^{\mathrm{s}}, d_a\pi_t^{\mathrm{s}}]$ | $\vartheta_{t,a}/c_a > d_a\pi_t^{\mathrm{s}}$ |
|---|---|---|---|
| $u_{t,a} \leq \overline{g}_a$<br>$d_{t,a} \leq \overline{p}_a$ | $(\overline{g}_a - u_{t,a}, 0)$ | $(0,0)$ | $(0, \overline{p}_a - d_{t,a})$ |
| $d_{t,a} > \overline{p}_a$ | $(\overline{g}_a - u_{t,a}, 0)$ | $(d_{t,a} - \overline{p}_a, 0)$ | $(d_{t,a} - \overline{p}_a, 0)$ |
| $u_{t,a} > \overline{g}_a$ | $(0, u_{t,a} - \overline{g}_a)$ | $(0, u_{t,a} - \overline{g}_a)$ | $(0, \overline{p}_a - d_{t,a})$ |

Table 4.2: Description of the extended bang-bang strategy explained in terms of $(g_{t,a}, p_{t,a})$ for engaging in the spot market when the water value is given by $\vartheta_{t,a}$.

We remark that Table 4.2 generalizes Table 4.1. Indeed, one can verify that if the generation company decides to trade only in the spot market, i.e., if $u_{t,a} = d_{t,a} = 0$, then only the first row of Table 4.2 is of interest, and it coincides with the contents of Table 4.1.

| | $\mathscr{U}$ | $\mathscr{M}$ | $\mathscr{L}$ |
|---|---|---|---|
| $\overline{v}^{+}_{w,i}$ (m$^3$) | $1.85 \times 10^7$ | $6.90 \times 10^6$ | $1.6 \times 10^6$ |
| $\underline{v}^{+}_{w,i}$ (m$^3$) | $\begin{cases} 1.25 \times 10^7 & \text{if } w \leq 18, \\ 0 & \text{otherwise} \end{cases}$ | 0 | 0 |

Table 4.3: Node parameters of the cascade.

| | $(\mathscr{U}, \mathscr{M})$ | $(\mathscr{M}, \mathscr{L})$ | $(\mathscr{L}, \otimes)$ |
|---|---|---|---|
| $c_a$ (MWh/m$^3$) | $6.69 \times 10^{-4}$ | $1.03 \times 10^{-3}$ | $6.22 \times 10^{-4}$ |
| $d_a$ | 1.40 | – | – |
| $\overline{g}_a$ (m$^3$/h) | $4.63 \times 10^4$ | $4.45 \times 10^4$ | $5.18 \times 10^4$ |
| $\overline{p}_a$ (m$^3$/h) | $3.99 \times 10^4$ | – | – |

Table 4.4: Arc parameters of the cascade.

## 4.7 Case Study: Austrian Electricity Market

The purpose of this section is to evaluate how profitable it is to invest in the reserve markets. In all of our experiments, we consider a generation company operating a cascade of three reservoirs, and we assume that the planning horizon comprises one year divided into weeks $w \in \mathscr{W} = \{1, \ldots, 52\}$. Moreover, we assume that the topology of the considered cascade coincides with the one shown in Figure 4.2. We denote the upper, middle and lower reservoirs by $\mathscr{U}$, $\mathscr{M}$ and $\mathscr{L}$, respectively. In our terminology, reservoirs are represented as nodes of a graph. Their relevant parameters (i.e., bounds on storage levels) are given in Table 4.3. Reservoirs $\mathscr{M}$ and $\mathscr{L}$ are allowed to be full or empty at any point in time during the planning horizon. The storage level of reservoir $\mathscr{U}$, however, cannot be lower than 67.57% of its full capacity for the first 18 weeks (starting from 1$^{\text{st}}$ of July). Moreover, this cascade possesses three hydrological arcs: $(\mathscr{U}, \mathscr{M})$, $(\mathscr{M}, \mathscr{L})$ and $(\mathscr{L}, \otimes)$, where we use $\otimes$ to denote the sink node. The parameters of each arc are given in Table 4.4, where the first arc has an attached turbine-pump pair and the remaining ones have only turbines. Finally, we assume that initially all of the reservoirs are 80% full. The target storage levels at the end of the planning horizon are set to the initial levels. This ensures a smooth transition between operational years.

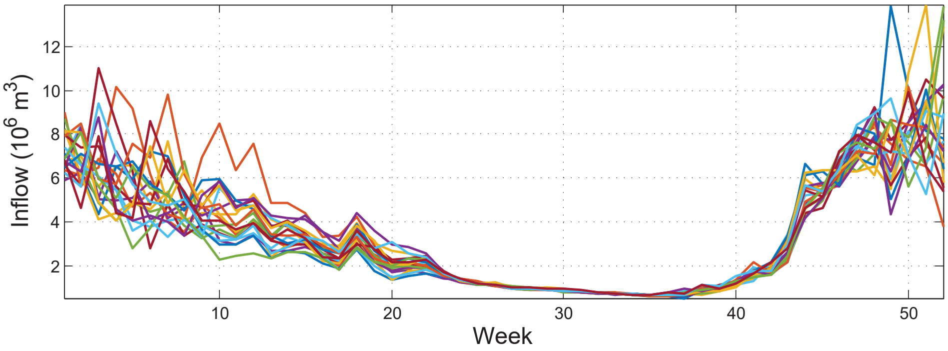



Figure 4.5: 20 realizations of the weekly inflows in the catchment area of the cascade across the year.

### 4.7.1 Experimental Setup

The planner-trader decomposition proposed in Section 4.4 requires as input $S$ realizations of the relevant stochastic processes. Here, we explain how these realizations are obtained. The information in this section has been kindly provided to us by our industrial partner, Decision Trees GmbH.

We consider 20 realizations of the total weekly inflows in the catchment area of the cascade as illustrated in Figure 4.5. These are part of the input to the planning problem (4.6). The total inflow to the different reservoirs is distributed proportionally to their surface areas. In our experiment, reservoirs $\mathscr{U}$, $\mathscr{M}$ and $\mathscr{L}$ are assumed to receive 11%, 57% and 0%, respectively, of the total inflows. We note that this constant proportion assumption speeds up the solution time of the planning problem as we can reduce the information bases for the decisions in (4.6). In particular, we can replace $\boldsymbol{\psi}_w^+ \in \mathbb{R}^I$ with a single scalar.

We now consider the stochastic processes explaining the electricity spot prices. We assume that the spot price forward curve is available to the generation company at the beginning of the planning horizon. In our experiment, we use hourly spot prices from 2014 as our price forward curve.[7] We denote this price forward curve by $\pi^{\mathrm{f}}_{0,t}$, $t \in \mathscr{T}$. We use this price forward curve to construct realizations of the spot prices as well as

[7]Historical prices are available for download at https://www.eex.com/en/market-data.

the updated price forward curve at time $\tau > 0$, denoted by $\pi^{\mathrm{f}}_{\tau,t}$, through

$$\pi^{\mathrm{s}}_t = \pi^{\mathrm{f}}_{0,t} \cdot \frac{\exp(x_t)}{\mathbb{E}\big(\exp(x_t)\big)}, \quad \forall t \geq 0 \quad \text{and} \quad \pi^{\mathrm{f}}_{\tau,t} = \pi^{\mathrm{f}}_{0,t} \cdot \frac{\mathbb{E}\big(\exp(x_t)|\mathscr{F}_\tau\big)}{\mathbb{E}\big(\exp(x_t)\big)}, \quad \forall t \geq \tau,$$

where $x_t$ and $y_t$ are exogenous stochastic processes. Their evolution is described by the stochastic differential equations

$$\mathrm{d}x_t = \alpha(y_t - x_t)\mathrm{d}t + \sigma_x \mathrm{d}w^x_t \quad \text{and} \quad \mathrm{d}y_t = \sigma_y \mathrm{d}w^y_t, \tag{4.9}$$

driven by the independent standard Wiener processes $w^x_t$ and $w^y_t$.[8] This two-factor model for electricity spot and forward prices is inspired by Pilipovic (1998); Haarbrücker and Kuhn (2009), respectively. The exact values for $\alpha, \sigma_x$ and $\sigma_y$ used in our experiments are again shared with us by Decision Trees GmbH. To further elaborate, we sort the price forward curve $[\pi^{\mathrm{f}}_{0,t}]_{t\in\mathscr{T}}$ in ascending order and divide it into five equidistant bands. Accordingly, we split the trading hours $\mathscr{T}$ into five disjoint sets. The corresponding $\alpha, \sigma_x$ and $\sigma_y$ for each price band are given in Table 4.5.

| | Spot price band | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| $\alpha$ (year$^{-1}$) | 20.767 | 391.283 | 508.168 | 333.480 | 282.830 |
| $\sigma_x$ (year$^{-1/2}$) | 4.648 | 12.260 | 9.781 | 9.758 | 9.184 |
| $\sigma_y$ (year$^{-1/2}$) | 0.123 | | | | |

Table 4.5: Parameters for the stochastic differential equations (4.9).

Note that the spot price simulation heavily relies on the calculation of $\mathbb{E}\big(\exp(x_t)\big)$. To demonstrate how these expectations are obtained from (4.9), we observe that both $x_t$ and $y_t$ are normally distributed; see Haarbrücker and Kuhn (2009). Hence, $\exp(x_t)$ follows a longnormal distribution, and its expectation can be expressed in terms of the mean and variance of $x_t$. A detailed derivation of the first two moments of $x_t$ is relegated to Appendix A.2.

Figure 4.6 below demonstrates how to use the price forward curve to simulate spot prices and how to update it in the future. The updated price forward curve can then

[8] In this setting, $[(x_t, y_t)]_{t\in\mathscr{T}}$ is Markovian but $[\pi^{\mathrm{s}}_t]_{t\in\mathscr{T}}$ is not, implying that Assumption (A3) is violated. We conjecture that this phenomenon does not greatly degrade the quality of the decomposition.

be used to simulate new updates of the spot prices as time passes. The spot price realizations in the left figure are simulated using the original price forward curve $[\pi^{\mathrm{f}}_{0,t}]_{t\in\mathcal{T}}$, and they constitute inputs for the planning problem (4.6). On the other hand, the realizations in the right figure are obtained from the updated price forward curve. These realizations represent spot prices as seen by the trading subproblems, where the trader updates the price forward curve using on-line information from the most recent week.

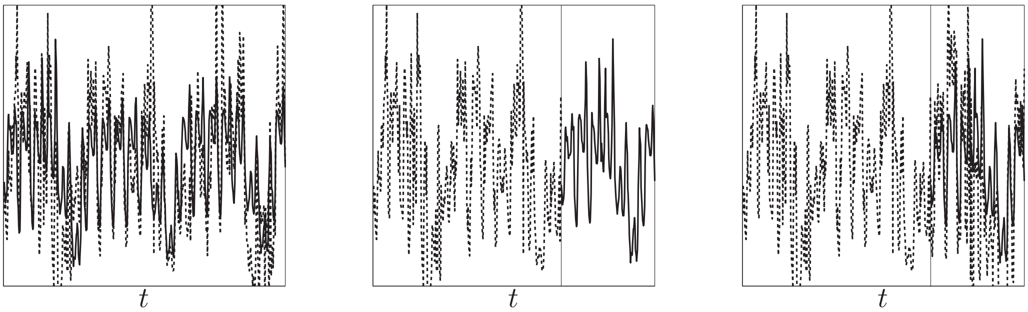



Figure 4.6: Spot price simulation using the price forward curve. In the left figure, the solid and dashed lines represent $\pi^{\mathrm{f}}_{0,t}$ and $\pi^{\mathrm{s}}_t$, respectively. In the middle figure, we update the price forward curve and obtain $\pi^{\mathrm{f}}_{\tau,t}$, which is represented by the solid line. In the right figure, we simulate new spot prices $\pi^{\mathrm{s}}_t$ for $t \geq \tau$ using the updated price forward curve.

Finally, for the reserve markets, we assume that the activations of the reserve capacities ($\rho^{\mathrm{u}}_t$ and $\rho^{\mathrm{d}}_t$) follow independent Bernoulli distributions with success probabilities of 1%. Assume further that the electricity prices in the reserve markets ($\pi^{\mathrm{u}}_t$ and $\pi^{\mathrm{d}}_t$) follow uniform distributions as described in Table 4.6. Within a week, there are 60 peaking hours and 108 off-peak hours. Peaking hours cover the periods with high electricity demands from 8:00 to 20:00 from Monday to Friday.[9] Under these assumptions, we can generate any given number of realizations of the random variables relevant to the reserve markets. To be consistent with the number of the inflow scenarios available to us, we generate 20 realizations of $\pi^{\mathrm{u}}_t, \pi^{\mathrm{d}}_t, \rho^{\mathrm{u}}_t, \rho^{\mathrm{d}}_t$, and we set all of the reserve capacity fees $\phi^{\mathrm{u}}_t, \phi^{\mathrm{d}}_t$ to zero in all of our experiments. In practice, the reserve capacity fees can be strictly positive. In that case, setting $\phi^{\mathrm{u}}_t = \phi^{\mathrm{d}}_t = 0$ for all $t \in \mathcal{T}$ is a conservative choice which underestimates the profit opportunities of the generation company.

[9] http://www.apg.at

| | Peaking hour (€/MWh) | | Off-peak hour (€/MWh) | |
|---|---|---|---|---|
| | lower bound | upper bound | lower bound | upper bound |
| $\pi_t^{\mathrm{u}}$ | 590.4 | 885.6 | 641.9 | 962.9 |
| $\pi_t^{\mathrm{d}}$ | 1518.4 | 2277.6 | 1586.6 | 2380.0 |

Table 4.6: Distributional parameters of the electricity reserve prices.

Under this setting, we then have a set $\mathscr{S}$ of 20 realizations of $\boldsymbol{\xi}$. In order to solve the planning and the trading problems, we also have to specify precisely the uncertainty set $\Xi$ defined in (4.5). The lower bounds and the upper bounds in Table 4.6 correspond to the values of $\underline{\pi}_t^{\mathrm{u}}, \overline{\pi}_t^{\mathrm{u}}, \underline{\pi}_t^{\mathrm{d}}, \overline{\pi}_t^{\mathrm{d}}$. On the other hand, the lower bounds and the upper bounds on electricity spot prices and weekly inflows are determined from the minimum and the maximum of the corresponding variables over all realizations $\boldsymbol{\xi}^1, \ldots, \boldsymbol{\xi}^S$.

Specifically to the regulations of the Austrian electricity markets, the tendering period of reserve capacities is one week. Within a week, reserve-up capacities for all peaking hours must be equal. Similarly, reserve-up capacities for all off-peak hours must also share the same value. These rules also apply to reserve-down capacities. We remark that these requirements are readily embedded in the trading subproblems (4.3).

### 4.7.2 Numerical Results

We now assess the benefit of trading in the reserve markets. To this end, we solve the planning problem in a receding horizon fashion. That is, at the beginning of each week $w$, we resolve the planning problem in order to obtain the target release quantities $\boldsymbol{q}_w^\star$. We then solve the trading subproblems to obtain the reserve capacity decisions $\boldsymbol{u}_t^\star$ and $\boldsymbol{d}_t^\star$ for $t \in \mathscr{T}(w)$. In fact, since we already solve each trading subproblem for different values of the target release $q_{w,a}$ when we approximate $\Pi_{w,a}$ by $\hat{\Pi}_{w,a}$, we may estimate $u_{t,a}^\star$ and $d_{t,a}^\star$ by linear interpolation over $q_{w,a}$. Lastly, we use the extended bang-bang strategy developed in Section 4.6 to decide on the trading decisions $\boldsymbol{g}_t^\star$ and $\boldsymbol{p}_t^\star$ in the spot market. In doing so, we only extract first-stage decisions from the planning and trading problems, and we use the heuristics to obtain high-quality wait-and-see decisions.

We then repeat the experiment with the additional restriction that all reserve capacities are zero, implying that the generation company trades in the spot market only. We then compare the results of both experiments to quantify the profitability of the reserve markets by several means: shadow prices, storage levels, and total revenues.

All linear programs are solved with CPLEX 12.6.2. For all trading subproblems that can be solved in parallel, we thank the High Performance Computing (HPC) services at École Polytechnique Fédérale de Lausanne and Imperial College London for allowing us to use their platforms.

**Shadow prices:** When we solve the planning problem at the beginning of week $w$, we can extract the shadow prices $\boldsymbol{\lambda}_{t(w)}$, which in turn give us information about water values to be used in the extended bang-bang strategy. Figure 4.7 below compares the shadow prices of the three reservoirs. By participating in the reserve markets, the shadow prices of all reservoirs increase, which is already an indicator of the additional profitability offered by the reserve markets.

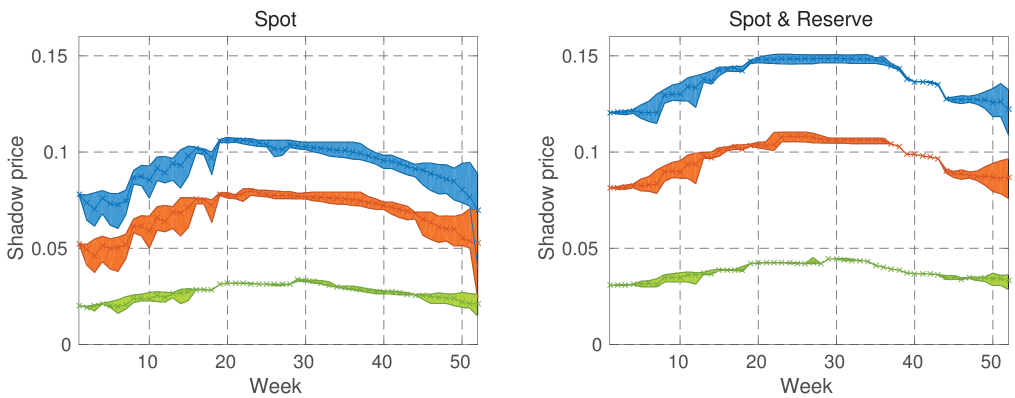



Figure 4.7: Weekly shadow prices (€/m$^3$) of the three reservoirs, from top to bottom, $\mathscr{U}$, $\mathscr{M}$ and $\mathscr{L}$. The left figure shows the shadow prices when the generation company trades in the spot market only, whereas the right figure represents the case where the generation company trades simultaneously in both the spot and reserve markets. The shaded areas cover the range between the 10%- and 90%-percentiles of the shadow prices, and the marked lines (×) in the middle of the shaded areas represent average shadow prices.

**Storage levels:** In Figure 4.8, we show how the storage levels of the three reservoirs evolve over the entire planning horizon. Recall that, for each reservoir, we set the initial and target storage levels to 80% of its full capacity. In the planning model (4.4), the storage levels are updated every week, while in reality they are updated on a much finer scale. To capture the storage levels in high resolution, Figure 4.8 shows hourly storage levels, where within a week we use the transactions (determined by the trader) on the electricity markets to determine the hourly storage levels, and we assume that natural inflows are uniform within a week, i.e., $\boldsymbol{\psi}_t = \frac{1}{168}\boldsymbol{\psi}_w^+$ for every $t \in \mathscr{T}(w)$. We can see that participation in the reserve markets reduces the variation in the storage levels.

On the other hand, it can also be seen that it is possible for the storage levels to drop below zero. This is the price that we have to pay for tractability, as in the planner-trader decomposition, we only impose bounds on the storage levels at the end of each week. However, since the considered cascade consists of only seasonal reservoirs, these events are very rare. Though, it should be pointed out that the current model is not suitable for daily or hourly reservoirs. To account for smaller reservoirs, we may have to change our decomposition scheme as we have to keep track of hourly storage levels. Alternatively, this phenomenon could be avoided by increasing the lower bounds of the storage levels to combat the effects of over-discharging.

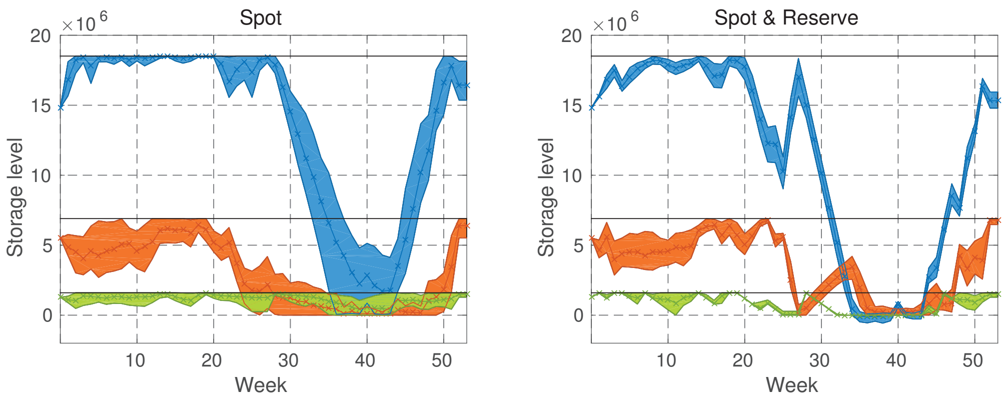



Figure 4.8: Storage levels (m$^3$) of the three reservoirs, from top to bottom, $\mathscr{U}$, $\mathscr{M}$ and $\mathscr{L}$. The left figure shows the storage levels when the generation company trades in the spot market only, whereas the right figure represents the case where the generation company trades in both the spot and reserve markets. The shaded areas cover the range between the 10%- and 90%-percentiles of the storage levels, and the marked lines (×) in the middle of the shaded areas represent the average storage levels.

**Total revenues:** Finally, Figure 4.9 compares the cumulative revenues over the entire planning horizon for two cases: (i) the generation company trades in the spot market only and (ii) the generation company trades in both the spot and reserve markets. By participating in both markets simultaneously, the generation company can increase average revenues by 48.3%. Hence, our analysis in this chapter suggests that the reserve markets offer substantial additional profit opportunities for hydropower generation companies.

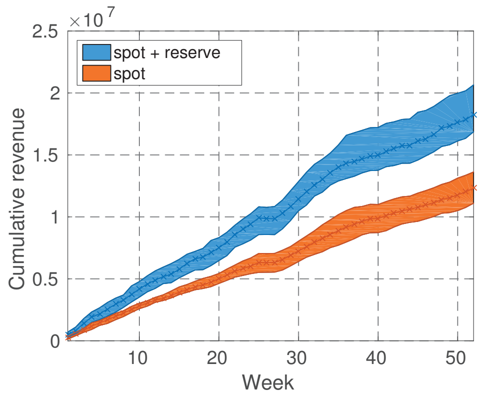



Figure 4.9: Cumulative revenues (€) from all three reservoirs over the entire planning horizon. The shaded areas cover the range between the 10%- and 90%-percentiles of the cumulative revenues, and the marked lines in the middle (×) of the shaded areas represent the averages.

# 5 Conclusions

Information in optimization problems is often inexact, and disregarding this phenomenon may lead to ill-informed decisions which perform disappointingly. In the past, researchers have identified two principles to deal with uncertain information. On the one hand, uncertainty in optimization problems is representable as random variables governed by a probability distribution. On the other hand, it can be expressed as uncertain parameters living on an uncertainty set. In the former approach, a *risk-neutral* decision maker may try to minimize an expected loss function; in this case, the arising optimization problem is a stochastic program. In the latter approach, a *risk-averse* decision maker may seek a robust solution that is optimal in the view of worst case. While both approaches are pervasively implemented in practice, they are not without shortcomings. Stochastic programs are computationally demanding to solve, and exact solution methods generally require perfect knowledge of the distribution, which is difficult to acquire in practice. On the other hand, even though typically tractable, robust programs are ignorant of any distributional information except for the support and therefore often considered to be overly conservative. To mitigate the drawbacks of each paradigm, one might consider combining both of them. Indeed, a distributionally robust optimization approach has been intensively explored in the past few years. The main results in this thesis are leveraged from distributionally robust optimization techniques and linear decision rule approximations in stochastic programs.

In this thesis, we considered applications of optimization under uncertainty in different disciplines. Particularly, we utilized techniques from distributionally robust optimization in financial and mathematical applications, in Chapter 2 and Chapter 3 respectively. In Chapter 4, on the other hand, we formulated a multi-market multi-reservoir management problem as a stochastic program and solved it approximately in linear decision rules. We highlight that the solutions obtained in Chapter 4 is distributionally robust in the sense that the expected revenue earned by the generation company remains unchanged even under perturbed probability distributions as long as their first and second moments are preserved. This is an artifact of the approximation, owing to the fact that the expectation of a quadratic function can be expressed in terms of the first two moments of the distribution.

Another important feature of this thesis lies in complexity reduction and decomposition schemes. In Chapter 2 and Chapter 3, the complexity reduction was achieved by exploiting the permutation symmetry of the first two moments of weak-sense stationary processes. In Chapter 4, the reduction is more direct as we decomposed the reservoir management problem spatially and temporally, which is possible by distinguishing between dynamics with different paces. We hope that these decomposition ideas can be reused in other problems which may appear discouraging to solve at first glance.

## 5.1 Future Research Avenues

During the course of the doctoral programme, we identified several research questions to address in the future. Some of them would complement the results in the thesis, whereas the others would answer open questions in optimization under uncertainty.

In Chapter 2 and Chapter 3, we studied variants of distributionally robust chance constrained programs with Chebyshev-type ambiguity sets. An interesting open question here is to investigate whether it is possible to extend the tractable reformulations of individual chance constraints to joint chance constraints. The extension is by no means obvious, even in the following simple setting

$$\inf_{\mathbb{P}\in\mathscr{P}} \mathbb{P}(\mathbf{A}\boldsymbol{x} \leq \tilde{\boldsymbol{\xi}}) \geq 1-\epsilon, \tag{5.1}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$ is a technology matrix with $m > 1$, $\boldsymbol{x} \in \mathbb{R}^n$ is a decision vector, and $\mathscr{P}$

is a Chebyshev-type ambiguity set. Zymler et al. (2013a) highlighted the numerical difficulties encountered upon solving these programs exactly, whereas recently Hanasusanto et al. (2015b) proposed a tractable reformulation of these programs, albeit in a different ambiguity setting. Despite these indications, it still remains unknown whether a tractable reformulation of (5.1) exists.

In Chapter 3, we derived sharp probability bounds for products of random variables. Unlike the classical Chebyshev inequalities, our bounds are not analytical and have to be obtained by solving a semidefinite program whose size scales with the number of random variables ($T$) in the product. Hence, the computational burden for calculating these bounds increases with $T$. Although, this is not as cumbersome as it sounds because we showed, for large $T$, that the bounds reduce to a trivial number or to an analytical bound for the majority of the quantiles, it would still be more insightful to have bounds that are more accessible. Perhaps one can show that the Chebyshev bounds for products admit closed-form expressions that we are not aware of. In that case, it would be interesting to relate the solutions of the semidefinite programs with the corresponding analytical bounds.

In line with Bertsimas and Popescu (2005), this chapter provides an optimization perspective for probability inequalities. There are abundant applications in machine learning, probability, and statistics which can benefit from advances in optimization under uncertainty. Recent examples include Goldenshluger et al. (2015) and Fertis (2009).

Moreover, the distributionally robust optimization with Chebyshev-type ambiguity sets can be classified as a parametric approach which assumes that the probability distribution is fully characterized by a finite set of fixed parameters. With increasing processing power and storage capabilities, however, one may be inclined towards a non-parametric (or a data-driven) approach, in which the distributional knowledge can be adapted to accumulated information; see Mohajerin Esfahani and Kuhn (2015).

Last but not least, the results in Chapter 4 are not yet settled and more experiments have to be performed before we can provide the generation companies with informed managerial insights. To get a comprehensive understanding of the problem, we should (i) quantify the loss of optimality incurred by the linear decision rule approximation,

(ii) improve the model in such a way that daily reservoirs can be included, (iii) compare the obtained results with the results from other approaches proposed in the literature, and (iv) reduce the solution times of both the planning and the trading problems.

# A Appendices

## A.1 Distributionally Robust Chance Constraints

**Theorem A.1.** *Let $\mathscr{P}$ 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 $\mathscr{P}(\Xi)$ be the subset of $\mathscr{P}$ that contains only distributions supported on the ellipsoid $\Xi = \{\boldsymbol{\xi} \in \mathbb{R}^n : (\boldsymbol{\xi} - \boldsymbol{\nu})^\intercal \boldsymbol{\Lambda}^{-1} (\boldsymbol{\xi} - \boldsymbol{\nu}) \leq \delta\}$, 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 $q^0 \in \mathbb{R}$, the following statements hold.*

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

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

*(ii) The distributionally robust chance constraint* $\inf_{\mathbb{P} \in \mathscr{P}(\Xi)} \mathbb{P}\left(\tilde{\boldsymbol{\xi}}^\intercal \mathbf{Q} \tilde{\boldsymbol{\xi}} + \tilde{\boldsymbol{\xi}}^\intercal \boldsymbol{q} + q^0 \leq 0\right) \geq 1 - \epsilon$ *with moment and support information is implied (conservatively approxi-*

*mated) by*

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

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

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

*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 (A.2) reduces to (A.1) if we set $\alpha = \lambda = 0$. ∎

## A.2 Mean and Variance of Mean-Reverting Processes

**Lemma A.1.** *Consider a two-dimensional stochastic process $\{(x_t, y_t)\}_{t\geq 0}$ defined by* (4.9) *with a given initial condition $(x_0, y_0)$. The following statements hold.*

*(i) The mean of $x_t$ is $(x_0 - y_0)\exp(-\alpha t) + y_0$.*

*(ii) The variance of $x_t$ is $(1,0,0)\int_0^t \exp(\alpha\tau\mathbf{N})\left(\sigma_x^2, 0, \sigma_y^2\right)^\top \mathrm{d}\tau$, where $\mathbf{N}$ is a constant matrix defined as*

$$
\mathbf{N} = \begin{bmatrix} -2 & 2 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix}.
$$

*Proof.* The expectation of $x_t$ has a starting point $\mathbb{E}(x_0) = x_0$ and follows a deterministic process

$$\mathrm{d}\mathbb{E}(x_t) = \mathbb{E}(\mathrm{d}x_t) = \mathbb{E}(\alpha(y_t - x_t)\mathrm{d}_t) = \alpha\left(y_0 - \mathbb{E}(x_t)\right)\mathrm{d}t,$$

where the last equality holds because $\mathbb{E}(y_t) = y_0$, which is implied by the second equation in (4.9). From elementary calculus, the solution of the above ordinary differential equation is given by

$$\mathbb{E}(x_t) = (x_0 - y_0)\exp(-\alpha t) + y_0,$$

and thus assertion (i) holds. As for assertion (ii), we consider the derivative of the second-order moment of $(x_t, y_t)$, which includes the variances of $x_t$ and $y_t$ as well as their covariance. To facilitate the calculation of these derivatives, we define $\mathrm{cov}(\cdot,\cdot)$ as the covariance between two input random variables. First, for the derivative of the variance of $x_t$, we consider

$$\begin{aligned}\mathrm{cov}(x_{t+\mathrm{d}t}, x_{t+\mathrm{d}t}) &= \mathrm{cov}\left(x_t + \alpha(y_t - x_t)\mathrm{d}t + \sigma_x \mathrm{d}w_t^x, x_t + \alpha(y_t - x_t)\mathrm{d}t + \sigma_x \mathrm{d}w_t^x\right)\\ &= \mathrm{cov}(x_t, x_t) + \alpha^2 \mathrm{cov}\left(y_t - x_t, y_t - x_t\right)(\mathrm{d}t)^2 + \sigma_x^2 \mathrm{cov}(\mathrm{d}w_t^x, \mathrm{d}w_t^x) +\\ &\quad 2\alpha\,\mathrm{cov}(x_t, y_t - x_t)\mathrm{d}t.\end{aligned}$$

Hence, we find $\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{cov}(x_t, x_t) = \sigma_x^2 + 2\alpha\,\mathrm{cov}(x_t, y_t - x_t)$. Similarly, we find

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{cov}\left(y_t, y_t\right) = \sigma_y^2 \quad \text{and} \quad \frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{cov}\left(x_t, y_t\right) = \alpha\,\mathrm{cov}\left(y_t - x_t, y_t\right).$$

Introducing a vector of second-order moments $\boldsymbol{z}_t = [\mathrm{cov}(x_t, x_t), \mathrm{cov}\left(x_t, y_t\right), \mathrm{cov}\left(y_t, y_t\right)]^\top$, the above three differential equations are representable as

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\boldsymbol{z}_t = \left(\sigma_x^2, 0, \sigma_y^2\right)^\top + \alpha\mathbf{N}\boldsymbol{z}_t.$$

Moreover, this linear system has an initial condition $\boldsymbol{z}_0 = \mathbf{0}$ because $(x_0, y_0)$ is deterministically given. It can then be shown to admit a solution of the form

$$\boldsymbol{z}_t = \int_0^t \exp(\alpha\tau\mathbf{N})\left(\sigma_x^2, 0, \sigma_y^2\right)^\top \mathrm{d}\tau,$$

where in this case 'exp' represents the matrix exponential. With this, assertion (ii) holds and the proof thus completes. ∎

# Bibliography

Aitchison, J., J. Brown. 1957. *The Lognormal Distribution with Special Reference to Its Use in Economics.* Cambridge University Press.

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.

Baillo, A., M. Ventosa, M. Rivier, A. Ramos. 2004. Optimal offering strategies for generation companies operating in electricity spot markets. *IEEE Transactions on Power Systems* **19**(2) 745–753.

Beck, A., A. Ben-Tal. 2009. Duality in robust optimization: Primal worst equals dual best. *Operations Research Letters* **37**(1) 1–6.

Beck, M., M. Scherer. 2015. Basic principles of ancillary service products. Swissgrid Ltd.

Ben-Tal, A., L. El Ghaoui, A. Nemirovski. 2009. *Robust Optimization.* Princeton University Press.

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

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

Bertsimas, D., D. Iancu, P. Parrilo. 2010b. Optimality of affine policies in multistage robust optimization. *Mathematics of Operations Research* **35**(2) 363–394.

Bertsimas, D., I. Popescu. 2005. Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal on Optimization* **15**(3) 780–804.

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.

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

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. J. Neyman, ed., *Fourth Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, 65–78.

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

Chen, X., M. Sim, P. Sun, J. Zhang. 2008. A linear decision-based approximation approach to stochastic programming. *Operations Research* **56**(2) 344–357.

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, 1–80.

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

Cover, T., J. Thomas. 1991. *Elements of Information Theory*. 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. 2015. Robustness to dependency in portfolio optimization using overlapping marginals. *Operations Research* **63**(6) 1468–1488.

Dyer, M., L. Stougie. 2006. Computational complexity of stochastic programming problems. *Mathematical Programming* **106**(3) 423–432.

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.

Fertis, A. 2009. A robust optimization approach to statistical estimation problems. Ph.D. thesis, Massachusetts Institute of Technology.

Frisch, U., D. Sornette. 1997. Extreme deviations and applications. *Journal of Physics I France* **7** 1155–1171.

Galambos, J., I. Simonelli. 2004. *Products of Random Variables: Applications to Problems of Physics and to Arithmetical Functions*. Taylor & Francis.

Georghiou, A., W. Wiesemann, D. Kuhn. 2015. Generalized decision rule approximations for stochastic programming via liftings. *Mathematical Programming* **152**(1) 301–338.

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

Godwin, H. 1955. On generalizations of Tchebycheff's inequality. *Journal of the American Statistical Association* **50** 923–945.

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

Goldenshluger, A., A. Juditsky, A. Nemirovski. 2015. Hypothesis testing by convex optimization. *Electronic Journal of Statistics* **9**(2) 1645–1712.

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, 81–118.

Haarbrücker, G., D. Kuhn. 2009. Valuation of electricity swing options by multistage stochastic programming. *Automatica* **45**(4) 889 – 899.

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.

Hanasusanto, G., D. Kuhn, W. Wiesemann. 2015a. A comment on "computational complexity of stochastic programming problems". *Mathematical Programming* (Forthcoming).

Hanasusanto, G., V. Roitch, D. Kuhn, W. Wiesemann. 2015b. Ambiguous joint chance constraints under mean and dispersion information. *Available on Optimization-Online* .

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

Hirth, L., I. Ziegenhagen. 2015. Balancing power and variable renewables: Three links. *Renewable and Sustainable Energy Reviews* **50** 1035 – 1051.

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

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

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.

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

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

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

Kovacevic, R., G. Pflug, M. Vespucci. 2013. *Handbook of Risk Management in Energy Production and Trading*. Springer.

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

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

Kuhn, D., W. Wiesemann, A. Georghiou. 2011. Primal and dual linear decision rules in stochastic and robust optimization. *Mathematical Programming* **130**(1) 177–209.

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

Löhndorf, N., D. Wozabal, S. Minner. 2013. Optimizing trading decisions for hydro storage systems using approximate dual dynamic programming. *Operations Research* **61**(4) 810–823.

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

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. L. MacLean, E. Thorp, W. Ziemba, eds., *The Kelly Capital Growth Investment Criterion: Theory and Practice*. World Scientific, 563–574.

MacLean, L., W. 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.

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

Mayer, J. 2014. Electricity production and spot-prices in Germany 2014. Fraunhofer Institute for Solar Energy Systems.

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.

Mohajerin Esfahani, P., D. Kuhn. 2015. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Available on arXiv* .

Morris, C., M. Pehnt. 2015. Energy transition: The German energiewende. Heinrich Böll Foundation.

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.

Nesterov, Y. 2000. Squared functional systems and optimization problems. H. Frenk, K. Roos, T. Terlaky, S. Zhang, eds., *High Performance Optimization*. Kluwer Academic Publishers, 405–440.

Näsäkkälä, E., J. Keppo. 2008. Hydropower with financial information. *Applied Mathematical Finance* **15**(5-6) 503–529.

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

Pilipovic, D. 1998. *Energy Risk: Valuing and Managing Energy Derivatives*. McGraw-Hill.

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

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

Powell, W. 2014. Clearing the jungle of stochastic optimization. A. Newman, J. Leung, J. Smith, H. Greenberg, eds., *Bridging Data and Decisions*. Institute for Operations Research and the Management Sciences, 109–137.

Pritchard, G., A. Philpott, P. Neame. 2005. Hydroelectric reservoir optimization in a pool market. *Mathematical Programming* **103**(3) 445–461.

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.

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

Shapiro, A., A. Nemirovski. 2005. On complexity of stochastic programming problems. V. Jeyakumar, A. Rubinov, eds., *Continuous Optimization: Current Trends and Modern Applications*. Springer, 111–146.

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

Shohat, J., J. Tamarkin. 1943. *The Problem of Moments*. American Mathematical Society.

Tester, J., E. Drake, M. Driscoll, M. Golay, W. Peters. 2012. *Sustainable Energy: Choosing Among Options*. MIT Press.

Thorp, E. 1975. Portfolio choice and the Kelly criterion. W. Ziemba, R. Vickson, eds., *Stochastic Optimization Models in Finance*. Academic Press, 599–619.

Toh, K., M. Todd, R. Tütüncü. 1999. SDPT3–a Matlab software package for semidefinite programming. *Optimization Methods and Software* **11**(1–4) 545–581.

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. *Operations Research* **62**(6) 1358–1376.

Wirth, H. 2016. Recent facts about photovoltaics in Germany. Fraunhofer Institute for Solar Energy Systems.

Ye, Y. 1997. *Interior Point Algorithms: Theory and Analysis*. Wiley.

Yeh, W. 1985. Reservoir management and operations models: A state-of-the-art review. *Water Resources Research* **21**(12) 1797–1818.

Yu, Y., Y. Li, D. Schuurmans, C. Szepesvári. 2009. A general projection property for distribution families. Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, A. Culotta, eds., *Advances in Neural Information Processing Systems 22*. Curran Associates, Inc., 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* **137**(1) 167–198.

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

# Curriculum Vitae

Napat Rujeerapaiboon ✉ napat.ruj@gmail.com

## Education

**Risk Analytics and Optimization Chair, École Polytechnique Fédérale de Lausanne**
*Doctor of Philosophy in Risk Analytics and Optimization* (2012 – 2016, Switzerland)
Thesis: "Dimensionality Reduction in Dynamic Optimization under Uncertainty"

**Department of Computing, Imperial College London**
*Master of Science in Computing* (2011 – 2012, United Kingdom)
Thesis: "Option Pricing with Linear Programming"

**Department of Computer Engineering, Chulalongkorn University**
*Bachelor of Computer Engineering* (2006 – 2010, Thailand)
Thesis: "Regular Clustering by Flexible Grammar Inference"

## Teaching Experience

**Risk Analytics and Optimization Chair, École Polytechnique Fédérale de Lausanne**
*Doctoral Assistant* (2013 – 2016, Switzerland)
List of courses: Optimization Methods and Models (PhD), Optimal Decision Making (Master), Applied Probability and Stochastic Processes (Master), Optimal Decision Analysis (Master), and Information Technology and E-Business Strategy (Master).

**Department of Computing, Imperial College London**
*Graduate Teaching Assistant* (2012, United Kingdom)

List of courses: Operations Research (Master), Computational Finance (Master), and Software Engineering (Bachelor).

## Industrial Experience

**Accenture Solutions Company Ltd.**
*Analyst* (2010, Thailand)
Developed management information system for Central Food Retail Company Ltd.

**National Electronics and Computer Technology Center**
*Research Assistant* (2008 – 2009, Thailand)
Developed a PCA-based web server anomaly detector.

## Publications

**A Multi-Scale Decision Rule Approach for Multi-Market Multi-Reservoir Management** (with D. Kuhn and W. Wiesemann). Working Paper.

**Multi-Period Portfolio Optimization: Translation of Autocorrelation Risk to Excess Variance** (with B. Choi and R. Jiang). Submitted for Publication (2016).

**Chebyshev Inequalities for Products of Random Variables** (with D. Kuhn and W. Wiesemann). Submitted for Publication (2016).

**Robust Growth-Optimal Portfolios** (with D. Kuhn and W. Wiesemann). Management Science **62**(7) 2090-2109 (2016).

## Invited Presentations

**INFORMS Optimization Society Conference** (IOS), New Jersey, United States (2016).

**International Conference on Operations Research** (OR), Vienna, Austria (2015).

**International Symposium on Mathematical Programming** (ISMP), Pittsburgh, United States (2015).

**British-French-German Conference on Optimization** (BFG), London, United Kingdom (2015).

**Society for Industrial and Applied Mathematics Conference on Optimization** (SIAM OP), San Diego, United States (2014).

**International Conference on Applied Mathematical Programming and Modeling** (APMOD), Coventry, United Kingdom (2014).

## Selected Honors

**Best Teaching Assistant Award**
École Polytechnique Fédérale de Lausanne (2015 and 2016)

**Winton Capital Applied Computing MSc Project Prize**
Imperial College London and Winton Capital Management Ltd. (2013)

**Corporate Partnership Programme Awards for Academic Excellence**
Imperial College London (2013)

**Asia Game Award (AGA) Honorable Mention and Most Creative Awards**
Software Industry Promotion Agency and Ministry of Information and Communication Technology of Thailand (2010)

**ACM International Collegiate Programming Contest (ACM-ICPC) Honorable Mention Award**
Association for Computing Machinery (2008)

**International Mathematical Olympiad (IMO) Bronze Medal Award**
International Mathematical Olympiad (2005)

**Asian Pacific Mathematics Olympiad (APMO) Silver Medal Award**
Asian Pacific Mathematics Olympiad (2005)

**National Mathematics Competition First Place Winner Awards**
Mathematical Association of Thailand under the Patronage of His Majesty the King (2002 and 2005)