Dynamic mean-variance portfolio analysis under model risk

Daniel Kuhn
Department of Computing, Imperial College of Science, Technology, and Medicine, 180 Queen's Gate, London SW7 2AZ, UK; email: dkuhn@doc.ic.ac.uk

Panos Parpas
Department of Computing, Imperial College of Science, Technology, and Medicine, 180 Queen's Gate, London SW7 2AZ, UK; email: pp500@doc.ic.ac.uk

Berç Rustem
Department of Computing, Imperial College of Science, Technology, and Medicine, 180 Queen's Gate, London SW7 2AZ, UK; email: br@doc.ic.ac.uk

Raquel Fonseca
Department of Computing, Imperial College of Science, Technology, and Medicine, 180 Queen's Gate, London SW7 2AZ, UK; email: rfonseca@doc.ic.ac.uk

The classical Markowitz approach to portfolio selection is compromised by two major shortcomings. First, there is considerable model risk with respect to the distribution of asset returns. Particularly, mean returns are notoriously difficult to estimate. Moreover, the Markowitz approach is static in that it does not account for the possibility of portfolio rebalancing within the investment horizon. We propose a robust dynamic portfolio optimization model to overcome both shortcomings. The model arises from an infinite-dimensional min–max framework. The objective is to minimize the worst-case portfolio variance over a family of dynamic investment strategies subject to a return target constraint. The worst-case variance is evaluated with respect to a set of conceivable return distributions. We develop a quantitative approach to approximate this intractable problem by a tractable one and report on numerical experiments.

1 INTRODUCTION

Let us consider a market consisting of N investment opportunities or assets. A vector of price relatives $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_N)$ is a vector of non-negative valued random variables. The nth component $\boldsymbol{\xi}_n$ expresses the ratio of the terminal and initial prices of asset n over a given time interval $[0, T]$. Put differently, $\boldsymbol{\xi}_n$ is the factor by which capital invested in asset n grows from time 0 up to time T. All random objects appearing in this paper are defined on a measurable space

The authors would like to thank Steve Zymler and Kai Ye for helpful discussions. Support by EPSRC grants GR/T02560/01 and EP/C513584/1 is gratefully acknowledged.

$(\Omega, \mathcal{S})$. By convention, random objects (that is, random variables, random vectors, or stochastic processes) appear in boldface, while their realizations are denoted by the same symbols in normal face.

An investor can diversify his or her capital at the beginning of the planning horizon by choosing a portfolio vector $w = (w_1, \ldots, w_N)$. The component w_n of w denotes the amount of money invested in asset n. In his Nobel Prize-winning work, Markowitz (1952) recognized that any such investor faces two conflicting objectives: portfolio "risk" should be minimized while, at the same time, "performance" should be maximized. Using a benchmark-relative approach, we measure portfolio risk and performance as the variance and expectation of the portfolio excess return $(w - \hat{w})^\top \boldsymbol{\xi}$, respectively. Here, the fixed non-negative vector $\hat{w}$ characterizes a benchmark portfolio we seek to outperform. As with any multi-criterion optimization problem, the Markowitz problem does not have a unique solution. Instead, it has a family of Pareto-optimal solutions, which are found by solving the parametric risk minimization problem:

$$\begin{aligned} \text{minimize} \quad & \mathrm{Var}_Q((w - \hat{w})^\top \boldsymbol{\xi}) \\ \text{s.t.} \quad & \mathrm{E}_Q((w - \hat{w})^\top \boldsymbol{\xi}) \geq \rho \\ & (w - \hat{w})^\top e = 0 \\ & w \geq 0 \end{aligned} \tag{$\mathcal{P}_1$}$$

for different values of the return target ρ. Note that Q stands for a probability measure on $(\Omega, \mathcal{S})$ which "objectively" describes the distribution of the asset returns, $\mathrm{E}_Q(\cdot)$ and Var_Q denote the expectation and variance operators under Q, respectively, and e represents the element of $\mathbb{R}^N$ with all components equal to 1. The set of Pareto-optimal portfolios gives rise to the efficient frontier, which is defined as the graph of the mapping $\rho \mapsto \min \mathcal{P}_1$ in the risk–return plane. It is easy to see that the probability measure Q impacts problem $\mathcal{P}_1$ only through the mean value μ and the covariance matrix Σ of the asset return vector $\boldsymbol{\xi}$.

Instead of solving $\mathcal{P}_1$, we can determine the efficient frontier also by solving the parametric return maximization problem:

$$\begin{aligned} \text{maximize} \quad & (w - \hat{w})^\top \mu \\ \text{s.t.} \quad & (w - \hat{w})^\top \Sigma (w - \hat{w}) \leq \sigma^2 \\ & (w - \hat{w})^\top e = 0 \\ & w \geq 0 \end{aligned} \tag{$\mathcal{P}_2$}$$

for different values of the risk target σ^2. Sometimes, this alternative model formulation is more convenient. The Markowitz model in a single period setting and with alternative risk measures is described in Markowitz (1959, Section XIII).

Unfortunately, the original Markowitz approach to portfolio selection suffers from two major shortcomings. Firstly, the resulting efficient frontier is not robust with respect to errors in the estimation of μ and Σ. Secondly, the approach is static and does not account for the possibility of portfolio adjustments in response to changing market conditions. While the lack of robustness leads to an overestimate, neglecting the dynamic nature of the investment process leads to an underestimate of the achievable portfolio performance. Recently, several worst-case optimization approaches have been suggested to immunize Markowitz-type models against estimation errors in the distributional input parameters, see, for example, Ceria and Stubbs (2006), El Ghaoui *et al* (2003), Goldfarb and Iyengar (2002) or Rustem *et al* (2000). A comprehensive survey of this line of research is provided in Fabozzi *et al* (2007). Notice that worst-case robust portfolio models are usually formulated for a single investment period. Multiperiod extensions of the mean-variance portfolio problem have been studied, for example, by Frauendorfer and Siede (2000), Gülpınar *et al* (2004) and Steinbach (2001) in a stochastic programming framework, by Li and Ng (2000) and Leippold *et al* (2004) in a dynamic programming setting, and by Zhou and Li (2000) and Bielecki *et al* (2005) from a stochastic calculus perspective. The robustification of dynamic mean-variance portfolio problems has received little attention so far; preliminary results are presented in Gülpınar and Rustem (2007).

The main contribution of this paper is to propose a framework for the computation of investment strategies that are dynamic as well as robust with respect to uncertain input parameters. Our algorithmic procedure provides provable error bounds and a dynamic portfolio strategy which is implementable with probability one (without the necessity to reoptimize the problem at later stages).

Section 2 further highlights the problems associated with estimation errors in the input parameters and reviews the concept of robust portfolio optimization. This modeling paradigm has been designed to counter the adverse effects of estimation errors. However, robust portfolio optimization is usually cast in a single-period framework. A generalized robust model involving several rebalancing periods is elaborated in Section 3. This model can be seen as a special instance of a multistage stochastic program under model risk, that is, a min–max problem where minimization is over a set of dynamic trading strategies and maximization is over a family of rival return distributions within a given ambiguity set. Min–max problems of this type are extremely difficult to solve. They represent functional optimization problems since investment decisions are functions of the past asset returns. Even worse, the distribution of the asset returns is partially unknown because of estimation errors, and finding the worst-case distribution is part of the problem. The approximation of functional min–max problems by finite-dimensional tractable models proceeds in two steps. In Section 4 the constrained min–max problem is reformulated as an unconstrained problem by dualizing the explicit constraints. Section 5 exploits the favorable structural properties of the reformulated model to discretize all rival return distributions simultaneously. We will show that the approximate problem in which the true return distributions are replaced by their discretizations provides an upper bound on the original problem. Moreover, we will establish a transformation that maps any optimal solution of the approximate problem to a near-optimal solution of the original problem. The mathematical tools developed in Sections 4 and 5 will then be used in Section 6 to study a particular investment situation. We will compare Markowitz efficient frontiers obtained under different assumptions about the underlying return distributions and under different rebalancing schemes.

2 ROBUST PORTFOLIO OPTIMIZATION

Several authors have observed that the efficient frontier obtained from the classical Markowitz model is extremely sensitive to the input parameters μ and Σ: see, for example, Black and Litterman (1990), Broadie (1993) or Chopra and Ziemba (1993). This implies that a small estimation error in μ or Σ can have a substantial effect on the efficient frontier. In an influential article, Michaud argues that the Markowitz approach to portfolio selection essentially amplifies the impact of the estimation errors (Michaud (1989)). While the entries of the covariance matrix Σ can be determined quite accurately, it is virtually impossible to measure μ to within workable precision. The reason for this is that return fluctuations over a given time period scale with the square root of period length, which generates a "mean blur" effect (Luenberger (1998, Section 8.5)). Hence, the main impediment for the Markowitz approach to produce reasonable results is that the expected returns are flawed with an estimation error, while the efficient frontier is very sensitive to such errors.

To shed light on Michaud's error maximization property, we assume that the covariance matrix Σ is perfectly known, while the expected return vector is affected by an estimation error. We denote by $v_2(\mu)$ the optimal value of $\mathcal{P}_2$. This representation emphasizes the dependence on the return vector μ. Next, we assume that $\hat{\boldsymbol{\mu}}$ is an unbiased estimator for μ, that is, $\hat{\boldsymbol{\mu}}$ is a random variable with $\mathrm{E}_Q(\hat{\boldsymbol{\mu}}) = \mu$. By using a simple result about the interchange of expectation and maximization operators, we then conclude that $v_2(\hat{\boldsymbol{\mu}})$ is an upward biased estimator for $v_2(\mu)$, the best portfolio return that can be achieved when μ is perfectly known:

$$\begin{aligned}\mathrm{E}_Q(v_2(\hat{\boldsymbol{\mu}})) &= \mathrm{E}_Q(\max\{(w - \hat{w})^\top \hat{\boldsymbol{\mu}} \mid \text{constraints of } \mathcal{P}_2\}) \\ &\geq \max\{(w - \hat{w})^\top \mathrm{E}_Q(\hat{\boldsymbol{\mu}}) \mid \text{constraints of } \mathcal{P}_2\} \\ &= v_2(\mu) \qquad (2.1)\end{aligned}$$

Thus, using the (unbiased) sample average estimator to approximate μ, for instance, leads to a systematic overestimation of the optimal portfolio return and to an upshift of the Markowitz efficient frontier in the risk–return plane.

The notorious mean blur effect is a special case of model risk. In more general terms, one could argue that the objective probability measure Q, that is, the "true" probabilistic model underlying the Markowitz problem, is unknown to the decision maker. In fact, it is questionable whether a true model exists at all. Different experts may have different views about the prospects of different investments, and there may be no objective method (eg, a statistical test) to decide which model is the best. This kind of non-probabilistic model uncertainty is referred to as ambiguity in the decision theory literature.

The classical Markowitz approach to portfolio selection completely disregards model risk. Put differently, no information about the reliability or the accuracy of the input parameters is passed to the optimization model. To overcome this deficiency, the paradigm of robust portfolio optimization was proposed; a recent survey of the field is provided by Fabozzi *et al* (2007). The basic idea is to introduce an ambiguity set $\mathcal{A}$ which contains several probabilistic models or rival probability measures. In order to avoid uninteresting technical complications, we assume that every two measures $P_1, P_2 \in \mathcal{A}$ are mutually equivalent, that is, P_2 has a density with respect to P_1 and vice versa. For the time being, however, we make no additional assumptions about the structure of $\mathcal{A}$. If each $P \in \mathcal{A}$ is interpreted as a model suggested by an independent expert, then a robust version of the portfolio problem can be formulated as:

$$\begin{aligned} \underset{w \in \mathbb{R}^N}{\text{minimize}} \quad & \Big\{ \sup_{P \in \mathcal{A}} \operatorname{Var}_P((w - \hat{w})^\top \boldsymbol{\xi}) \Big\} \\ \text{s.t.} \quad & \mathrm{E}_P((w - \hat{w})^\top \boldsymbol{\xi}) \geq \rho \quad \forall P \in \mathcal{A} \\ & (w - \hat{w})^\top e = 0 \\ & w \geq 0 \end{aligned} \qquad (\mathcal{P}_3)$$

This model is robust with respect to all expert opinions. The objective is to minimize the worst-case variance while satisfying the return target constraint under each rival probability measure. We are not so much interested in the optimal value of problem $\mathcal{P}_3$, but rather in its optimal solution. The latter exhibits a non-inferiority property in the following sense: unless the most pessimistic model turns out to be the true one, implementing a strategy from arg min $\mathcal{P}_3$ leads to a portfolio variance that is smaller (ie, better) than the predicted variance min $\mathcal{P}_3$.

It is easy to imagine that the solution quality strongly depends on the specification of the ambiguity set. If $\mathcal{A}$ is large enough to also cover remotely conceivable models, then the resulting robust portfolio strategy may be extremely conservative. On the other hand, if $\mathcal{A}$ collapses to a singleton, we recover the classical Markowitz model $\mathcal{P}_1$ with its well-known shortcomings.

Rustem *et al* (2000) consider situations in which the ambiguity set is finite. They assume, for instance, that $\mathcal{A}$ contains models obtained from different estimation methods (eg, by using implied versus historical information). A specialized algorithm is employed to solve the arising discrete min–max problem. El Ghaoui *et al* (2003) as well as Goldfarb and Iyengar (2002) specify the ambiguity set in terms of confidence intervals for certain moments of the return distribution. They reformulate the resulting robust optimization problem as a second-order cone program, which can be solved efficiently. A similar approach is pursued by Ceria and Stubbs (2006), who also report on extensive numerical experiments. Pflug and Wozabal (2007) determine a "most likely" reference measure $\hat{P}$ and define $\mathcal{A}$ as some ε-neighborhood of $\hat{P}$. Distances of probability measures are computed by using the Wasserstein metric, and the arising continuous min–max problem is addressed by a semi-infinite programming-type algorithm.

3 DYNAMIC PORTFOLIO OPTIMIZATION

So far we have assumed that a portfolio selected at time 0 may not be restructured or "rebalanced" at any time within the planning horizon. In reality, however, an investor will readjust the portfolio allocation whenever significant price changes are observed. In order to capture the dynamic nature of the investment process, we choose a set of ordered time points[1] $0 = t_1 < \cdots < t_H = T$ and assume that portfolio rebalancing is restricted to these discrete dates.

A dynamic extension of the single-period portfolio models studied so far involves intertemporal return vectors. By convention, $\boldsymbol{\xi}_h$ characterizes the asset returns over the rebalancing interval from t_{h-1} to t_h. Below, we will always assume that the $\boldsymbol{\xi}_h$ are serially independent.[2] We define $\mathcal{F}_h = \sigma(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_h)$ to be the σ-algebra induced by the first h return vectors and set $\mathcal{F} = \mathcal{F}_H$. Moreover, we introduce the shorthand notation $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_H)$ to describe the stochastic return process.

Let us now specify the decision variables needed to formulate a dynamic version of the portfolio problem. First, we set $\boldsymbol{w}_h^- = (\boldsymbol{w}_{1,h}^-, \ldots, \boldsymbol{w}_{N,h}^-)$, where $\boldsymbol{w}_{n,h}^-$ denotes the capital invested in asset n at time t_h before reallocation of funds. Similarly, we define $\boldsymbol{w}_h^+ = (\boldsymbol{w}_{1,h}^+, \ldots, \boldsymbol{w}_{N,h}^+)$, where $\boldsymbol{w}_{n,h}^+$ is the capital invested in asset n at time t_h after portfolio rebalancing. Notice that these decisions represent random variables since they may depend on past stock price observations. We denote by $\boldsymbol{b}_{n,h}$ and $\boldsymbol{s}_{n,h}$ the amount of money used at time t_h to buy and sell assets of type n, respectively. As for the asset holdings, we combine these variables to decision vectors $\boldsymbol{b}_h = (\boldsymbol{b}_{1,h}, \ldots, \boldsymbol{b}_{N,h})$ and $\boldsymbol{s}_h = (\boldsymbol{s}_{1,h}, \ldots, \boldsymbol{s}_{N,h})$.

In analogy to $\boldsymbol{w}_h^-$ and $\boldsymbol{w}_h^+$, the random vectors $\hat{\boldsymbol{w}}_h^-$ and $\hat{\boldsymbol{w}}_h^+$ specify the asset holdings in the benchmark portfolio at time t_h before and after rebalancing, respectively. However, $\hat{\boldsymbol{w}}_h^-$ and $\hat{\boldsymbol{w}}_h^+$ are not decision variables. Instead, they constitute prescribed functions of the past asset returns $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_h$. Real investors frequently choose an exchange-traded stock index as their benchmark portfolio. Without much loss of generality, we may assume that this stock index is given by the Nth asset within our market. In this case, the $N-1$ first components of the initial portfolio vector $\hat{\boldsymbol{w}}_1^-$ vanish, while we have:

$$\hat{\boldsymbol{w}}_h^+ = \hat{\boldsymbol{w}}_h^- \quad \text{and} \quad \hat{\boldsymbol{w}}_{h+1}^- = \hat{\boldsymbol{w}}_h^+ \, \boldsymbol{\xi}_{N,h+1} \quad \text{for} \quad 1 \le h < H \tag{3.1}$$

Moreover, it is reasonable to normalize the benchmark portfolio in such a way that its initial value corresponds to our investor's initial endowment, that is:[3]

$$(\boldsymbol{w}_1^- - \hat{\boldsymbol{w}}_1^-)^\top e = 0$$

[1] For notational convenience, we sometimes make reference to an additional time point $t_0 = 0$.
[2] Serial independence is assumed to hold with respect to all measures $P \in \mathcal{A}$.
[3] Observe that $\boldsymbol{w}_1^-$ is an input parameter to our model and not a decision variable.

Using the above terminology, we can introduce a dynamic version of the robust portfolio optimization problem $\mathcal{P}_3$:

$$\begin{array}{lll} \underset{\boldsymbol{w}^-,\boldsymbol{w}^+,\,\boldsymbol{b},\,\boldsymbol{s}}{\text{minimize}} & \left\{ \sup\limits_{P\in\mathcal{A}} \operatorname{Var}_P((\boldsymbol{w}_{H-1}^+ - \hat{\boldsymbol{w}}_{H-1}^+)^\top \boldsymbol{\xi}_H) \right\} & \\ \text{s.t.} & \mathrm{E}_P((\boldsymbol{w}_{H-1}^+ - \hat{\boldsymbol{w}}_{H-1}^+)^\top \boldsymbol{\xi}_H) \geq \varrho & \forall P \in \mathcal{A} \\ & \boldsymbol{w}_h^+ = \boldsymbol{w}_h^- + \boldsymbol{b}_h - \boldsymbol{s}_h & 1 \leq h < H \\ & (1+c_b)\, e^\top \boldsymbol{b}_h = (1-c_s)\, e^\top \boldsymbol{s}_h & " \\ & \boldsymbol{w}_{h+1}^- = \boldsymbol{\xi}_{h+1} \odot \boldsymbol{w}_h^+ & " \\ & \boldsymbol{w}_h^-,\, \boldsymbol{w}_h^+,\, \boldsymbol{b}_h,\, \boldsymbol{s}_h \geq 0, \quad \mathcal{F}_h\text{-measurable} & " \end{array} \qquad (\mathcal{P}_4)$$

As in the single-period case, the objective is to minimize the worst-case variance of terminal wealth, while satisfying a return target constraint for each rival model $P \in \mathcal{A}$. The second constraint plays the role of an asset-wise balance equation. Notice that proportional transaction costs are incurred whenever the portfolio is rebalanced. We denote by c_s and c_b the transaction costs per unit of currency for sales and purchases of the assets, respectively. The third constraint thus captures the idea that only a certain percentage of the money received from selling assets can be spent on new purchases. The evolution of money (on a per-asset basis) between the rebalancing dates is described by the fourth constraint. It is the only dynamic constraint coupling successive decision stages. Note that the symbol "$\odot$" stands for the element-wise multiplication operator. Finally, all decision variables must be non-negative[4] and non-anticipative. The latter requirement means that decisions taken at time t_h must be $\mathcal{F}_h$-measurable, that is, they may only depend on information about return vectors observed in the past. All constraints are assumed to hold almost surely with respect to some probability measure $P \in \mathcal{A}$. The choice of P is irrelevant since the measures in $\mathcal{A}$ are pairwise equivalent.

The standard approach to address multi-period portfolio problems is via dynamic programming, see, for example, Ingersoll (1987). However, there are three major obstacles that complicate the use of dynamic programming for solving the robust mean-variance portfolio problem under consideration. First, $\mathcal{P}_4$ is not separable in the sense of dynamic programming due to the variance objective. Indeed, unlike conditional expectations, conditional variances fail to satisfy a tower property that could be used to set up sensible dynamic programming recursions. The return target constraint is problematic as well since it is associated with the first stage but involves (expected) decisions of the last stage. Methods to circumvent these difficulties have been explored in Leippold *et al* (2004) and Li and Ng (2000). Next, standard dynamic programming breaks down under ambiguity of the probability

[4]The no-short-sales constraints prevent the investor from exploiting potential arbitrage opportunities. In an arbitrage-free market, however, these constraints are not necessary to ensure the boundedness of problem $\mathcal{P}_4$.

measure. A possible remedy to overcome this obstacle would be to employ robust dynamic programming as suggested in Iyengar (2005). However, robust dynamic programming is applicable only if the ambiguity set exhibits a stagewise separability or rectangularity property that is unlikely to hold in the current context. Finally, dynamic programming is very inefficient if the state space has a high dimension (eg, $\gtrsim 4$). In the presence of transaction costs, however, we have to maintain a separate state variable for each asset. Hence, dynamic programming becomes computationally unmanageable if there are more than about four assets.

If the ambiguity set $\mathcal{A}$ contains only one probability measure, then $\mathcal{P}_4$ reduces to a standard dynamic mean-variance portfolio selection problem, which has been extensively studied in several modeling frameworks. Frauendorfer and Siede (2000), Steinbach (2001) and Gülpınar *et al* (2004) address this problem by using stochastic programming techniques. Their approach is very flexible in that it allows the modeler to incorporate portfolio constraints and market frictions with relative ease. However, it is computationally expensive if the number of rebalancing periods becomes large. Li and Ng (2000) as well as Leippold *et al* (2004) study mean-variance problems in a dynamic programming framework. They reduce the multi-period mean-variance problem to a stochastic linear quadratic regulator problem, whose analytical solution is well understood. Zhou and Li (2000) use a similar approach to address mean-variance problems in continuous time. A further generalization based on martingale techniques is due to Bielecki *et al* (2005). Applicability of these analytical approaches is limited to situations in which portfolio constraints, transaction costs, and other complicating factors are of minor importance.

Cases in which the ambiguity set $\mathcal{A}$ contains more than one model have hardly been addressed in the literature. A first approach to tackle such problems is presented in Gülpınar and Rustem (2007). In that work, $\mathcal{A}$ is assumed to be a finite set of finitely supported models. The present work goes one step further by allowing $\mathcal{A}$ to contain probability measures with an infinite support. As in stochastic programming, these models need to be approximated by finitely-supported probability measures, which are representable as scenario trees. This amounts to approximating the original min–max problem $\mathcal{P}_4$ with infinite-dimensional functional decisions by a simpler min–max problem with finite-dimensional vectorial decisions. Stochastic programming research has explored a vast number of algorithms for scenario tree construction or scenario generation, see, for example, the survey of Dupacova *et al* (2000). Different scenario generation methods apply to problems with different structural properties; they provide different error estimates and asymptotic guarantees. The scenario generation method to be used here has the following properties, which make it attractive for portfolio optimization problems of the type $\mathcal{P}_4$:

- Implementability: The solution of the discretized approximate problem allows us to construct a trading strategy that is implementable in each conceivable return scenario and not only in the (relatively few) scenarios of the underlying scenario trees. We can thus use Monte Carlo sampling to obtain an unbiased a posteriori estimate of this trading strategy's performance.

- Conservatism: The approximation overestimates the worst-case portfolio risk under the optimal trading strategy, that is, the solution of the discretized approximate problem provides an upper bound on the solution of the original problem $\mathcal{P}_4$. This form of conservatism is desirable since exaggerated optimism can be disastrous in financial decision making.
- Asymptotic consistency: If the number of scenarios in the scenario trees is increased, the solution of the discretized approximate problem converges to the solution of the original problem.

4 LAGRANGIAN REFORMULATION

Our ultimate goal is to devise a portfolio strategy which is near-optimal and, a fortiori, feasible in $\mathcal{P}_4$. The objective value of this strategy will provide a tight upper bound on the minimum of problem $\mathcal{P}_4$. In order to facilitate the derivation of a quantitative approximation scheme in Section 5, we now slightly increase the level of abstraction and reformulate $\mathcal{P}_4$ as a general stochastic optimization problem under model risk. To this end, we introduce the decision process $\boldsymbol{x} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_H)$ which is defined through:[5]

$$\boldsymbol{x}_h = \operatorname{vec}(\boldsymbol{w}_h^-, \boldsymbol{w}_h^+, \hat{\boldsymbol{w}}_h^-, \hat{\boldsymbol{w}}_h^+, \boldsymbol{b}_h, \boldsymbol{s}_h) \tag{4.1}$$

The operator "vec" returns the concatenation of its arguments. Thus, $\boldsymbol{x}_h$ has dimension $n_h = 6N$. As the constraints in $\mathcal{P}_4$ preclude negative $\boldsymbol{x}_h$, the state space of the process $\boldsymbol{x}$ can be identified with the non-negative orthant of $\mathbb{R}^{6NH}$. In standard terminology, such decision processes are also referred to as strategies, policies, or decision rules. We define the set of admissible decision processes as:

$$X(\mathbb{F}) = \bigcap_{P \in \mathcal{A}} \{\boldsymbol{x} \in \times_{h=1}^{H} \mathcal{L}^\infty(\Omega, \mathcal{F}_h, P; \mathbb{R}^{n_h}) \mid \boldsymbol{x} \geq 0 \ P\text{-a.s.}\} \tag{4.2}$$

Note that the strategies in $X(\mathbb{F})$ are essentially bounded with respect to all models $P \in \mathcal{A}$ and adapted to the filtration $\mathbb{F} = \{\mathcal{F}_h\}_{h=1}^H$, that is, they are non-anticipative with respect to the underlying return process. Next, we let Ξ denote the state space of the return process, while X denotes the state space of the (class of admissible) decision processes. Using a suitable cost function $c\colon X \times X \to \mathbb{R}$ as well as a sequence of suitable constraint functions $f_h\colon X \times \Xi \to \mathbb{R}^{m_h}$, the dynamic portfolio problem $\mathcal{P}_4$ is representable as an abstract multistage stochastic program under model risk:

$$\begin{aligned} &\underset{\boldsymbol{x} \in X(\mathbb{F})}{\text{minimize}} \quad \left\{ \sup_{P \in \mathcal{A}} \mathrm{E}_P(c(\boldsymbol{x}, \mathrm{E}_P(\boldsymbol{x}))) \right\} \\ &\qquad \text{s.t.} \quad \mathrm{E}_P(f_h(\boldsymbol{x}, \boldsymbol{\xi}) \mid \mathcal{F}_h) \leq 0 \quad \text{a.s. } \forall P \in \mathcal{A}, \ h = 1, \dots, H \end{aligned} \tag{$\mathcal{P}$}$$

[5]For technical reasons, it is useful to consider $\hat{\boldsymbol{w}}_h^-$ and $\hat{\boldsymbol{w}}_h^+$ as (degenerate) decision variables. Note, however, that their values are known a priori for each scenario, since they represent unambiguous functions of $\boldsymbol{\xi}$.

Here, the rules for updating the benchmark portfolio vectors (see, for example, (3.1)) are incorporated as additional constraints. It should be highlighted that the generalized problem $\mathcal{P}$ accommodates expected value constraints. The corresponding constraint functions are required to be nonpositive in expectation (instead of almost everywhere), while expectation is conditional on the stagewise information sets. As for problem $\mathcal{P}_4$, it is easy to see that the cost and constraint functions can be chosen in such a way that the following regularity conditions hold:

(C1) c is convex and continuous in its first argument;
(C2) f_h is representable as $f_h(x, \xi) = \tilde{f}_h((1, x) \otimes (1, \xi))$, where $\tilde{f}_h$ is convex, continuous, and constant in $x_i \otimes \xi_j$ for all $1 \leq j \leq i \leq H$, $1 \leq h \leq H$.

The operator "$\otimes$" stands for the usual dyadic product of vectors. Note that condition (C2) follows from the fact that our dynamic portfolio problem has fixed recourse (Birge and Louveaux (1997)). Besides the above, we impose an additional regularity condition on the return process:

(C3) the process $\boldsymbol{\xi}$ is serially independent with respect to all models $P \in \mathcal{A}$, and its state space Ξ is a compact polyhedron.

The compactness requirement in (C3) can always be enforced by truncating certain extreme scenarios of the return process that have a negligible effect on the solution of $\mathcal{P}$. Serial independence of returns, on the other hand, is a widely used standard assumption in finance literature. Apart from that, we impose no further restrictions on the return distribution. The conditions (C1)–(C3) are assumed to hold throughout the rest of this paper.

REMARK 4.1 If the stagewise decision vectors are of the form (4.1), we can set:

$$c(\boldsymbol{x}, \mathrm{E}_P(\boldsymbol{x})) = [e^\top(\boldsymbol{w}_H^- - \hat{\boldsymbol{w}}_H^-)]^2 - [e^\top \mathrm{E}_P(\boldsymbol{w}_H^-) - e^\top \mathrm{E}_P(\hat{\boldsymbol{w}}_H^-)]^2$$

to obtain the desired variance objective:

$$\mathrm{E}_P(c(\boldsymbol{x}, \mathrm{E}_P(\boldsymbol{x}))) = \mathrm{Var}_P(e^\top(\boldsymbol{w}_H^- - \hat{\boldsymbol{w}}_H^-)) = \mathrm{Var}_P((\boldsymbol{w}_{H-1}^+ - \hat{\boldsymbol{w}}_{H-1}^+)^\top \boldsymbol{\xi}_H)$$

Note that, in accordance with condition (C1), the cost function c is indeed convex and continuous in its first argument.

In complete analogy to the set $X(\mathbb{F})$ of primal decision processes , we introduce a set of dual decision processes:

$$Y(\mathbb{F}) = \bigcap_{P \in \mathcal{A}} \{\boldsymbol{y} \in \times_{h=1}^{H} \mathcal{L}^1(\Omega, \mathcal{F}_h, P; \mathbb{R}^{m_h}) \mid \boldsymbol{y} \geq 0 \ P\text{-a.s.}\} \tag{4.3}$$

By construction, a strategy $\boldsymbol{y} \in Y(\mathbb{F})$ constitutes a non-negative $\mathbb{F}$-adapted stochastic process that is P-integrable with respect to all models $P \in \mathcal{A}$. Furthermore, we denote by Y the state space of the (class of admissible) dual decision processes and define the Lagrangian density L: $X \times X \times Y \times \Xi \to \mathbb{R}$ associated with the problem data through:

$$L(x, \bar{x}, y, \xi) = c(x, \bar{x}) + \sum_{h=1}^{H} y_h^\top f_h(x, \xi)$$

Using the above terminology, we can prove that the stochastic program $\mathcal{P}$ under model risk has an equivalent reformulation in terms of the Lagrangian density.

LEMMA 4.2 *By dualizing the explicit constraints, problem $\mathcal{P}$ can be rewritten as the following unconstrained min–max problem:*

$$\underset{\boldsymbol{x} \in X(\mathbb{F})}{\text{minimize}} \sup_{\boldsymbol{y} \in Y(\mathbb{F})} \sup_{P \in \mathcal{A}} \mathrm{E}_P(L(\boldsymbol{x}, \mathrm{E}_P(\boldsymbol{x}), \boldsymbol{y}, \boldsymbol{\xi}))$$

PROOF We generalize an argument due to Wright (1994, Section 4), which applies when $\mathcal{A}$ is a singleton while the cost and constraint functions are linear. The claim follows if we can show that the two functionals:

$$g(\boldsymbol{x}) = \begin{cases} \sup_{P \in \mathcal{A}} \mathrm{E}_P(c(\boldsymbol{x}, \mathrm{E}_P(\boldsymbol{x})) & \text{if } \mathrm{E}_P(f_h(\boldsymbol{x}, \boldsymbol{\xi}) \mid \mathcal{F}_h) \leq 0 \text{ a.s.} \\ & \forall P \in \mathcal{A}, \ h = 1, \dots, H \\ +\infty & \text{else} \end{cases}$$

and:

$$\tilde{g}(\boldsymbol{x}) = \sup_{\boldsymbol{y} \in Y(\mathbb{F})} \sup_{P \in \mathcal{A}} \mathrm{E}_P(L(\boldsymbol{x}, \mathrm{E}_P(\boldsymbol{x}), \boldsymbol{y}, \boldsymbol{\xi}))$$

are equal on their common domain $X(\mathbb{F})$. Thus, we fix an arbitrary $\boldsymbol{x} \in X(\mathbb{F})$. Employing standard manipulations, we find:

$$\begin{aligned}
\tilde{g}(\boldsymbol{x}) &= \sup_{\boldsymbol{y} \in Y(\mathbb{F})} \sup_{P \in \mathcal{A}} \mathrm{E}_P\left(c(\boldsymbol{x}, \mathrm{E}_P(\boldsymbol{x})) + \sum_{h=1}^{H} \boldsymbol{y}_h^\top f_h(\boldsymbol{x}, \boldsymbol{\xi})\right) \\
&= \sup_{P \in \mathcal{A}} \sup_{\boldsymbol{y} \in Y(\mathbb{F})} \mathrm{E}_P\left(c(\boldsymbol{x}, \mathrm{E}_P(\boldsymbol{x})) + \sum_{h=1}^{H} \boldsymbol{y}_h^\top \mathrm{E}_P(f_h(\boldsymbol{x}, \boldsymbol{\xi}) \mid \mathcal{F}_h)\right) \\
&\leq \sup_{P \in \mathcal{A}} \mathrm{E}_P\left(c(\boldsymbol{x}, \mathrm{E}_P(\boldsymbol{x})) + \sum_{h=1}^{H} \sup_{y \geq 0} y^\top \mathrm{E}_P(f_h(\boldsymbol{x}, \boldsymbol{\xi}) \mid \mathcal{F}_h)\right) \\
&= g(\boldsymbol{x})
\end{aligned}$$

The inequality in the third line reflects the relaxation of the measurability constraints which require $\boldsymbol{y}_h$ to be $\mathcal{F}_h$-measurable for each $h = 1, \dots, H$. To prove the converse inequality, we consider a sequence of dual decision strategies $\{\tilde{\boldsymbol{y}}^{(i)}\}_{i=1}^{\infty}$ defined through:

$$\tilde{\boldsymbol{y}}_h^{(i)} = \begin{cases} 0 & \text{if } \mathrm{E}_P(f_h(\boldsymbol{x}, \boldsymbol{\xi}) \mid \mathcal{F}_h) \leq 0 \text{ a.s.} \quad \forall P \in \mathcal{A} \\ i & \text{else} \end{cases}$$

By construction, we have $\tilde{\boldsymbol{y}}^{(i)} \in Y(\mathbb{F})$ for all $i \in \mathbb{N}$, which implies that:

$$\tilde{g}(\boldsymbol{x}) \geq \sup_{i \in \mathbb{N}} \sup_{P \in \mathcal{A}} \mathrm{E}_P \left(c(\boldsymbol{x}, \mathrm{E}_P(\boldsymbol{x})) + \sum_{h=1}^{H} (\tilde{\boldsymbol{y}}_h^{(i)})^\top f_h(\boldsymbol{x}, \boldsymbol{\xi}) \right) = g(\boldsymbol{x})$$

This observation completes the proof. □

5 APPROXIMATION

The stochastic optimization problem $\mathcal{P}$ under model risk represents a large-scale min–max problem. Minimization is over a set of measurable functions $\boldsymbol{x} \in X(\mathbb{F})$, while maximization is over a set of rival models $P \in \mathcal{A}$. In the remainder, we will focus on situations in which the ambiguity set $\mathcal{A}$ is finite. However, even in this favorable case, problem $\mathcal{P}$ remains computationally intractable unless the stochastic return process is discrete (ie, finitely supported), implying that $\mathcal{P}$ reduces to a finite-dimensional min–max problem. If the return process fails to be discrete, we must approximate it by a suitable discrete process in order to obtain a computationally tractable problem.

The aim of this section is to introduce an approximate return process $\boldsymbol{\xi}^u$ which is supported on a finite subset of Ξ and which relates to the original process $\boldsymbol{\xi}$ in a quantitative way. In fact, we strive for finding a $\boldsymbol{\xi}^u$ such that the following holds: the optimal value of the min–max problem $\mathcal{P}^u$, which results from the original problem $\mathcal{P}$ by substituting $\boldsymbol{\xi}^u$ for $\boldsymbol{\xi}$, represents a tight upper bound on the optimal value of $\mathcal{P}$. Our construction of $\boldsymbol{\xi}^u$ is inspired by Birge and Louveaux (1997, Section 11.1) and Kuhn (2008, Section 4). In the sequel, we use the shorthand notation:

$$\boldsymbol{\xi}^h = (\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_h) \quad \text{and} \quad \boldsymbol{\xi}^{u,h} = (\boldsymbol{\xi}_1^u, \dots, \boldsymbol{\xi}_h^u), \quad h = 1, \dots, H$$

to denote the outcome histories of the processes $\boldsymbol{\xi}$ and $\boldsymbol{\xi}^u$, respectively. Moreover, we denote by Ξ_h the projection of Ξ on the space spanned by the realizations of $\boldsymbol{\xi}_h$, while Ξ^h stands for the projection of Ξ on the space spanned by the realizations of the return history $\boldsymbol{\xi}^h$.

In the following discussion we explicitly specify the sample space Ω. It is convenient to define $\Omega = \Xi \times \Xi$ and let $\mathcal{S}$ be the Borel field on $\Xi \times \Xi$. By definition, the stochastic processes $\boldsymbol{\xi}$ and $\boldsymbol{\xi}^u$ are functions from the sample space Ω to the state space Ξ. In our setting, we let $\boldsymbol{\xi}$ and $\boldsymbol{\xi}^u$ be coordinate projections:

$$\boldsymbol{\xi}\colon \begin{cases} \Xi \times \Xi \to \Xi \\ (\xi, \xi^u) \mapsto \xi \end{cases} \qquad \boldsymbol{\xi}^u\colon \begin{cases} \Xi \times \Xi \to \Xi \\ (\xi, \xi^u) \mapsto \xi^u \end{cases}$$

For every measure $P \in \mathcal{A}$, we denote by $P_{\boldsymbol{\xi}}$ the marginal distribution of $\boldsymbol{\xi}$. Since the sample space is identified with $\Xi \times \Xi$, the joint distribution $P_{\boldsymbol{\xi},\boldsymbol{\xi}^u}$ of $\boldsymbol{\xi}$ and $\boldsymbol{\xi}^u$ coincides with P. In the following, we assume that only $P_{\boldsymbol{\xi}}$ is given a priori, while the conditional distribution of $\boldsymbol{\xi}^u$ given $\boldsymbol{\xi}$, ie, $P_{\boldsymbol{\xi}^u \mid \boldsymbol{\xi}}$, is selected at our discretion. The joint distribution $P_{\boldsymbol{\xi},\boldsymbol{\xi}^u}$ is then obtained by using the product measure theorem (Ash (1972, Theorem 2.6.2)) to merge $P_{\boldsymbol{\xi}}$ and $P_{\boldsymbol{\xi}^u \mid \boldsymbol{\xi}}$. Although the measures $P \in \mathcal{A}$

are partly unknown before we specify the conditional distribution of $\boldsymbol{\xi}^u$, there is no problem assuming that they were known already at the outset.

We let the conditional distribution $P_{\boldsymbol{\xi}^u \mid \boldsymbol{\xi}}$ be model-independent, that is, we require it to be equal for all models $P \in A$. The construction of $P_{\boldsymbol{\xi}^u \mid \boldsymbol{\xi}}$ relies on a decomposition into conditional probability distributions P_h^u for $1 \leq h \leq H$. By definition, P_h^u stands for the distribution of $\boldsymbol{\xi}_h^u$ conditional on $\boldsymbol{\xi} = \xi$ and $\boldsymbol{\xi}^{u,\,h-1} = \xi^{u,\,h-1}$. By using the product measure theorem (Ash (1972, Theorem 2.6.2)), the building blocks P_h^u can later be reassembled to yield $P_{\boldsymbol{\xi}^u \mid \boldsymbol{\xi}}$.

Thus, instead of specifying $P_{\boldsymbol{\xi}^u|\boldsymbol{\xi}}$, it is sufficient to prescribe the conditional probability distributions P_h^u for $1 \leq h \leq H$, which will be equal for all $P \in \mathcal{A}$. In order to construct P_h^u, we select a family of measurable multifunctions (Rockafellar and Wets (1998, Section 14)):

$$\Xi_{h,l}\colon \Xi^{h-1} \rightrightarrows \Xi_h\,, \quad l = 1, \ldots, L$$

such that the following conditions hold for all return histories $\xi^{u,h-1} \in \Xi^{h-1}$:

(i) simpliciality: the set $\Xi_{h,\,l}(\xi^{u,\,h-1})$ is either empty or represents a bounded but not necessarily closed simplex for all l;
(ii) disjointness: $\Xi_{h,\,k}(\xi^{u,\,h-1}) \cap \Xi_{h,\,l}(\xi^{u,\,h-1}) = \emptyset$ for all $k \neq l$;
(iii) exhaustiveness: $\cup_{l=1}^{L} \Xi_{h,\,l}(\xi^{u,\,h-1}) = \Xi_h$.

Suitable multifunctions with the postulated properties exist since Ξ_h is a coordinate projection of Ξ, thus representing a compact polyhedron; see (C3). In the sequel, we let $\{\xi_{h,\,l,\,n}(\xi^{u,\,h-1})\}_{n=0}^{N}$ be a set of $N+1$ vectors in Ξ_h which correspond to the vertices of the simplex $\Xi_{h,\,l}(\xi^{u,\,h-1})$ if this set is nonempty, or to some arbitrary constant vectors, otherwise. Each $\xi_{h,\,l,\,n}$ can be interpreted as a function from Ξ^{h-1} to Ξ_h. Since $\Xi_{h,\,l}$ is a measurable multifunction, the vector-valued functions $\xi_{h,\,l,\,n}$ can be chosen measurably. Next, for any return vector $\xi_h \in \Xi_h$ we introduce a set of $N+1$ real numbers $\{\lambda_{h,\,l,\,n}(\xi_h \mid \xi^{u,\,h-1})\}_{n=0}^{N}$ which satisfy:

$$\sum_{n=0}^{N} \lambda_{h,\,l,\,n}(\xi_h \mid \xi^{u,\,h-1}) = 1 \quad \text{and} \quad \sum_{n=0}^{N} \lambda_{h,\,l,\,n}(\xi_h \mid \xi^{u,\,h-1})\, \xi_{h,\,l,\,n}(\xi^{u,\,h-1}) = \xi_h$$

Each $\lambda_{h,\,l,\,n}$ can be interpreted as a real-valued function on $\Xi_h \times \Xi^{h-1}$. Since the vector-valued functions $\xi_{h,\,l,\,n}$ are measurable, the $\lambda_{h,\,l,\,n}$ can be chosen measurably as well. We are now prepared to specify the conditional distribution P_h^u. For $\xi \in \Xi$ and $\xi^{u,\,h-1} \in \Xi^{h-1}$ we set:

$$P_h^u(\cdot \mid \xi, \xi^{u,\,h-1}) = \sum_{l=1}^{L} 1_{\Xi_{h,\,l}(\xi^{u,\,h-1})}(\xi_h) \sum_{n=0}^{N} \lambda_{h,\,l,\,n}(\xi_h \mid \xi^{u,\,h-1})\, \delta_{\xi_{h,\,l,\,n}(\xi^{u,\,h-1})}(\cdot) \tag{5.1}$$

where $1_{\hat{\Xi}}$ denotes the characteristic function of a set $\hat{\Xi}$, while $\delta_{\hat{\xi}}$ denotes the Dirac measure concentrated at a point $\hat{\xi}$. The distribution $P_{\boldsymbol{\xi},\boldsymbol{\xi}^u}$ obtained by combining the marginal distribution of $\boldsymbol{\xi}$ and the conditional distributions P_h^u in the appropriate way has several intriguing properties. Some of these will be explored in Lemmas 5.1 through 5.3.

From now on we denote by $\mathbb{F}^u$ the filtration generated by $\boldsymbol{\xi}^u$, that is, $\mathbb{F}^u = \{\mathcal{F}_h^u\}_{h=1}^H$ where $\mathcal{F}_h^u = \sigma(\boldsymbol{\xi}^{u,h})$, and we use the convention $\mathcal{F}^u = \mathcal{F}_H^u$. Moreover, we introduce spaces of approximate primal and dual strategies $X(\mathbb{F}^u)$ and $Y(\mathbb{F}^u)$, respectively. These are defined in the obvious way by replacing the original filtration $\mathbb{F}$ by the approximate filtration $\mathbb{F}^u$ in (4.2) and (4.3).

LEMMA 5.1 *The following relations hold for suitable versions of the conditional expectations and for all $P \in \mathcal{A}$, respectively:*

$$E_P(\boldsymbol{x} \mid \mathcal{F}) \in X(\mathbb{F}) \quad \textit{for all } \boldsymbol{x} \in X(\mathbb{F}^u) \tag{5.2a}$$

$$E_P(\boldsymbol{y} \mid \mathcal{F}^u) \in Y(\mathbb{F}^u) \quad \textit{for all } \boldsymbol{y} \in Y(\mathbb{F}) \tag{5.2b}$$

$$E_P(\boldsymbol{\xi}^u \mid \mathcal{F}) = \boldsymbol{\xi} \tag{5.2c}$$

PROOF Let us fix a probability measure $P \in \mathcal{A}$. The inclusion (5.2a) is related to the fact that the random vectors $\{\boldsymbol{\xi}_i^u\}_{i\leq h}$ and $\{\boldsymbol{\xi}_i\}_{i>h}$ are conditionally independent given $\{\boldsymbol{\xi}_i\}_{i\leq h}$ for all stage indices h. Conditional independence, in turn, follows from the serial independence of $\boldsymbol{\xi}$ and the recursive construction of $\boldsymbol{\xi}^u$. A rigorous proof of the preceding qualitative arguments is provided in Kuhn (2008, Section 4). The inclusion (5.2b), on the other hand, is related to the fact that $\{\boldsymbol{\xi}_i\}_{i\leq h}$ and $\{\boldsymbol{\xi}_i^u\}_{i>h}$ are conditionally independent given $\{\boldsymbol{\xi}_i^u\}_{i\leq h}$ for all h, see Kuhn (2008, Section 4).

Next, the relation (5.2c) follows immediately from the definition of P_h^u:

$$\begin{aligned}
&\mathrm{E}_P(\boldsymbol{\xi}_h^u \mid \boldsymbol{\xi}, \boldsymbol{\xi}^{u,h-1}) \\
&\quad = \int_{\Xi_h} \xi_h^u \, P_h^u(\mathrm{d}\xi_h^u \mid \boldsymbol{\xi}, \boldsymbol{\xi}^{u,h-1}) \\
&\quad = \sum_{l=1}^{L} 1_{\Xi_{h,l}(\xi^{u,h-1})}(\boldsymbol{\xi}_h) \sum_{n=0}^{N} \lambda_{h,l,n}(\boldsymbol{\xi}_h \mid \xi^{u,h-1})\, \xi_{h,l,n}(\xi^{u,h-1}) \\
&\quad = \boldsymbol{\xi}_h \quad P\text{-a.s.}
\end{aligned}$$

The law of iterated conditional expectations then allows us to conclude that:

$$\mathrm{E}_P(\boldsymbol{\xi}_h^u \mid \boldsymbol{\xi}) = \mathrm{E}_P(\mathrm{E}_P(\boldsymbol{\xi}_h^u \mid \boldsymbol{\xi}, \boldsymbol{\xi}^{u,h-1}) \mid \boldsymbol{\xi}) = \boldsymbol{\xi}_h \quad P\text{-a.s.} \quad \forall 1 \leq h \leq H$$

Thus, assertion (5.2c) is established. □

The relations (5.2) are crucial for our main result on the approximation of the min–max problem $\mathcal{P}$; see Theorem 5.4 below. It should be emphasized that there is considerable flexibility in the construction of the joint distribution of $\boldsymbol{\xi}$ and $\boldsymbol{\xi}^u$ since there are many different ways to specify the multifunctions $\Xi_{h,l}$. In particular, if the diameters of the simplices $\Xi_{h,l}(\xi^{u,h-1})$ become uniformly small over all admissible triples $(h, l, \xi^{u,h-1})$, then the approximate process $\boldsymbol{\xi}^u$ converges to the original return process $\boldsymbol{\xi}$ with respect to the $\mathcal{L}^\infty$-norm.

Replacing the original return process by $\boldsymbol{\xi}^u$ and the original filtration by $\mathbb{F}^u$ in the min–max problem $\mathcal{P}$ yields an approximate min–max problem, which will be denoted by $\mathcal{P}^u$. Observe that changing the filtration affects not only the feasible set $X(\mathbb{F})$ but also the information sets in the expected value constraints. Notice further that Lemma 4.2 remains valid for problem $\mathcal{P}^u$ with the approximate data process and filtrations. In the remainder of this section we will prove that the optimal value of $\mathcal{P}^u$ constitutes an upper bound on the optimal value of the original problem $\mathcal{P}$. We first establish two technical lemmas, both of which were proved in Kuhn *et al* (2008). The proofs are repeated here in order to keep this paper self-contained.

LEMMA 5.2 *The following relation holds for suitable versions of the conditional expectations and for all* $1 \leq i < j \leq H$, $P \in \mathcal{A}$*:*

$$\mathrm{E}_P(\boldsymbol{x}_i \otimes \boldsymbol{\xi}_j^u \mid \mathcal{F}) = \mathrm{E}_P(\boldsymbol{x}_i \mid \mathcal{F}) \otimes \boldsymbol{\xi}_j \quad \textit{for all } \boldsymbol{x} \in X(\mathbb{F}^u)$$

PROOF Select a probability measure $P \in \mathcal{A}$. Next, fix two stage indices i and j such that $1 \leq i < j \leq H$, and let $\boldsymbol{x}$ be an element of $X(\mathbb{F}^u)$. By applying elementary manipulations, we find that P-almost surely:

$$\begin{aligned}
\mathrm{E}_P(\boldsymbol{x}_i \otimes \boldsymbol{\xi}_j^u \mid \boldsymbol{\xi}) &= \mathrm{E}_P(\mathrm{E}_P(\boldsymbol{x}_i \otimes \boldsymbol{\xi}_j^u \mid \boldsymbol{\xi}, \boldsymbol{\xi}^{u,\, j-1}) \mid \boldsymbol{\xi}) \\
&= \mathrm{E}_P(\boldsymbol{x}_i \otimes \mathrm{E}_P(\boldsymbol{\xi}_j^u \mid \boldsymbol{\xi}, \boldsymbol{\xi}^{u,\, j-1}) \mid \boldsymbol{\xi}) \\
&= \mathrm{E}_P(\boldsymbol{x}_i \otimes \boldsymbol{\xi}_j \mid \boldsymbol{\xi}) \\
&= \mathrm{E}_P(\boldsymbol{x}_i \mid \boldsymbol{\xi}) \otimes \boldsymbol{\xi}_j
\end{aligned}$$

where the third equality follows from the proof of Lemma 5.1. □

LEMMA 5.3 *The following relation holds for suitable versions of the conditional expectations and for all* $1 \leq h \leq H$, $P \in \mathcal{A}$*:*

$$\mathrm{E}_P(f_h(\boldsymbol{x}, \boldsymbol{\xi}^u) \mid \mathcal{F}) \geq f_h(\mathrm{E}_P(\boldsymbol{x} \mid \mathcal{F}), \boldsymbol{\xi}) \quad \textit{for all } \boldsymbol{x} \in X(\mathbb{F}^u)$$

PROOF Select a probability measure $P \in \mathcal{A}$ and choose a strategy $\boldsymbol{x} \in X(\mathbb{F}^u)$. Then, condition (C2) and the conditional Jensen inequality imply that:

$$\begin{aligned}
\mathrm{E}_P(f_h(\boldsymbol{x}, \boldsymbol{\xi}^u) \mid \mathcal{F}) &= \mathrm{E}_P(\tilde{f}_h((1, \boldsymbol{x}) \otimes (1, \boldsymbol{\xi}^u)) \mid \mathcal{F}) \\
&\geq \tilde{f}_h(\mathrm{E}_P((1, \boldsymbol{x}) \otimes (1, \boldsymbol{\xi}^u) \mid \mathcal{F})) \\
&= \tilde{f}_h(\mathrm{E}_P((1, \boldsymbol{x}) \mid \mathcal{F}) \otimes (1, \boldsymbol{\xi})) \\
&= f_h(\mathrm{E}_P(\boldsymbol{x} \mid \mathcal{F}), \boldsymbol{\xi})
\end{aligned}$$

almost surely with respect to P. The equality in the third line follows from Lemma 5.2 and independence of $\tilde{f}_h$ in all of its arguments $x_i \otimes \xi_j$ with $i \geq j$. □

The technical Lemmas 5.1 and 5.3 constitute important ingredients for the proof of the following main result.

THEOREM 5.4 *Assume problem $\mathcal{P}^u$ to be solvable with finite minimum. If $\boldsymbol{x}^u$ solves $\mathcal{P}^u$, then $\hat{\boldsymbol{x}} = \mathrm{E}_P(\boldsymbol{x}^u \mid \mathcal{F})$ is independent of $P \in \mathcal{A}$, feasible in $\mathcal{P}$ and satisfies:*

$$\inf \mathcal{P} \leq \sup_{P \in \mathcal{A}} \mathrm{E}_P(c(\hat{\boldsymbol{x}}, \mathrm{E}_P(\hat{\boldsymbol{x}}))) \leq \inf \mathcal{P}^u$$

PROOF Our argumentation relies on ideas from the proofs of Theorem 1 in Kuhn (2008) and Theorem 5.1 in Kuhn (2009). We use the fact that $\boldsymbol{x}^u$ is an element of $X(\mathbb{F}^u)$, implying via (5.2a) that the conditional expectation $\mathrm{E}_P(\boldsymbol{x}^u \mid \mathcal{F})$ represents the same element of $X(\mathbb{F})$ for all $P \in \mathcal{A}$:

$$\begin{aligned}
&\inf \mathcal{P} \\
&\quad \leq \sup_{\boldsymbol{y} \in Y(\mathbb{F})} \sup_{P \in \mathcal{A}} \mathrm{E}_P\left(c(\mathrm{E}_P(\boldsymbol{x}^u \mid \mathcal{F}), \mathrm{E}_P(\boldsymbol{x}^u)) + \sum_{h=1}^{H} \boldsymbol{y}_h^{\top} f_h(\mathrm{E}_P(\boldsymbol{x}^u \mid \mathcal{F}), \boldsymbol{\xi})\right) \\
&\quad \leq \sup_{\boldsymbol{y} \in Y(\mathbb{F})} \sup_{P \in \mathcal{A}} \mathrm{E}_P\left(\mathrm{E}_P(c(\boldsymbol{x}^u, \mathrm{E}_P(\boldsymbol{x}^u)) \mid \mathcal{F}) + \sum_{h=1}^{H} \boldsymbol{y}_h^{\top} \mathrm{E}_P(f_h(\boldsymbol{x}^u, \boldsymbol{\xi}^u) \mid \mathcal{F})\right) \\
&\quad = \sup_{\boldsymbol{y} \in Y(\mathbb{F})} \sup_{P \in \mathcal{A}} \mathrm{E}_P\left(c(\boldsymbol{x}^u, \mathrm{E}_P(\boldsymbol{x}^u)) + \sum_{h=1}^{H} \boldsymbol{y}_h^{\top} f_h(\boldsymbol{x}^u, \boldsymbol{\xi}^u)\right)
\end{aligned} \tag{5.3}$$

The second inequality in (5.3) uses Lemma 5.3 and the conditional Jensen inequality, while the equality relies on the law of iterated conditional expectations. Another application of the conditional Jensen inequality yields:

$$\begin{aligned}
\inf \mathcal{P} &\leq \sup_{\boldsymbol{y} \in Y(\mathbb{F})} \sup_{P \in \mathcal{A}} \mathrm{E}_P\left(c(\boldsymbol{x}^u, \mathrm{E}_P(\boldsymbol{x}^u)) + \sum_{h=1}^{H} \mathrm{E}_P(\boldsymbol{y}_h \mid \mathcal{F}^u)^{\top} f_h(\boldsymbol{x}^u, \boldsymbol{\xi}^u)\right) \\
&\leq \sup_{\boldsymbol{y} \in Y(\mathbb{F}^u)} \sup_{P \in \mathcal{A}} \mathrm{E}_P\left(c(\boldsymbol{x}^u, \mathrm{E}_P(\boldsymbol{x}^u)) + \sum_{h=1}^{H} \boldsymbol{y}_h^{\top} f_h(\boldsymbol{x}^u, \boldsymbol{\xi}^u)\right) \\
&= \inf_{\boldsymbol{x} \in X(\mathbb{F}^u)} \sup_{\boldsymbol{y} \in Y(\mathbb{F}^u)} \sup_{P \in \mathcal{A}} \mathrm{E}_P\left(c(\boldsymbol{x}, \mathrm{E}_P(\boldsymbol{x})) + \sum_{h=1}^{H} \boldsymbol{y}_h^{\top} f_h(\boldsymbol{x}, \boldsymbol{\xi}^u)\right)
\end{aligned}$$

Here, the second inequality holds by the relation (5.2b), entailing a relaxation of the dual feasible set. The last line of the above expression corresponds to $\inf \mathcal{P}^u$, which is finite by assumption. This implies that the supremum over $Y(\mathbb{F})$ in the first line of (5.3) is also finite, and $\boldsymbol{y} = \boldsymbol{0}$ is an optimal solution. Thus, $\hat{\boldsymbol{x}} = \mathrm{E}_P(\boldsymbol{x}^u \mid \mathcal{F})$ is feasible in $\mathcal{P}$, and the corresponding worst-case objective value $\sup_{P \in \mathcal{A}} \mathrm{E}_P(c(\hat{\boldsymbol{x}}, \mathrm{E}_P(\hat{\boldsymbol{x}})))$ satisfies the postulated inequalities. □

REMARK 5.5 It is possible to relax the serial independence assumption (C3) at the cost of strengthening condition (C2). In fact, Theorem 5.4 remains valid if $\boldsymbol{\xi}$ constitutes an autoregressive moving average process[6] and if the constraint

[6]By this we mean that $\boldsymbol{\xi}$ is representable as a nonanticipative affine transformation of a serially independent noise process which satisfies (C3), see Kuhn (2008, Section 4.4).

functions are restricted to be convex and continuous. Since the wealth dynamics in the portfolio problem $\mathcal{P}_4$ involves products of decision variables and random parameters, however, we cannot assume convexity of the constraint functions. As a consequence, (C3) cannot be relaxed to allow for autoregressive moving average processes without impairing the validity of Theorem 5.4. Indeed, note that Lemma 5.2 fails if $\boldsymbol{\xi}$ is of autoregressive moving average type. We conjecture that this is no serious limitation of our approach. Although asset returns are known to exhibit weak serial dependence, given the inherent ambiguity of the return distribution, it is unlikely that such dependencies could be exploited in an active trading strategy. Thus, it is justifiable to assume serial independence.

Note that the marginal distribution of the approximate return process $\boldsymbol{\xi}^u$ is discrete. Since the distribution of $\boldsymbol{\xi}^u$ conditional on $\boldsymbol{\xi}$ is the same under all rival models $P \in \mathcal{A}$, the support of $\boldsymbol{\xi}^u$ represents a finite set independent of P. Only the probabilities associated with specific discretization points may be model-dependent. Consequently, the approximate problem $\mathcal{P}^u$ constitutes a finite-dimensional min–max problem, which is principally amenable to numerical solution. By Theorem 5.4, the optimal value of the approximate problem provides an a priori upper bound on the minimal (worst-case) expected cost. However, an optimal strategy of $\mathcal{P}^u$ only prescribes the decisions corresponding to return scenarios in the support of $\boldsymbol{\xi}^u$. Such a strategy may fail to be implementable if a generic return scenario from the support of $\boldsymbol{\xi}$ materializes. Put differently, it is a priori unclear how an optimal strategy for the approximate problem can be transformed to a near-optimal strategy for the original problem. Theorem 5.4 provides a particularly satisfactory answer to this question by proposing a policy $\hat{\boldsymbol{x}}$ which is implementable in every possible scenario of the original return process and whose worst-case expected cost is bracketed by $\inf \mathcal{P}$ and $\inf \mathcal{P}^u$. Note that $\mathrm{E}_P(c(\hat{\boldsymbol{x}}, \mathrm{E}_P(\hat{\boldsymbol{x}})))$ represents an a posteriori estimate of the expected cost associated with an optimal strategy of problem $\mathcal{P}$ under the model $P \in \mathcal{A}$. This cost can conveniently be calculated by Monte Carlo simulation. Since $\boldsymbol{x}^u$ is finitely supported, evaluation of $\hat{\boldsymbol{x}}$ for an arbitrary realization of $\boldsymbol{\xi}$ reduces to the evaluation of a finite sum and poses no computational challenges.

6 COMPUTATIONAL RESULTS

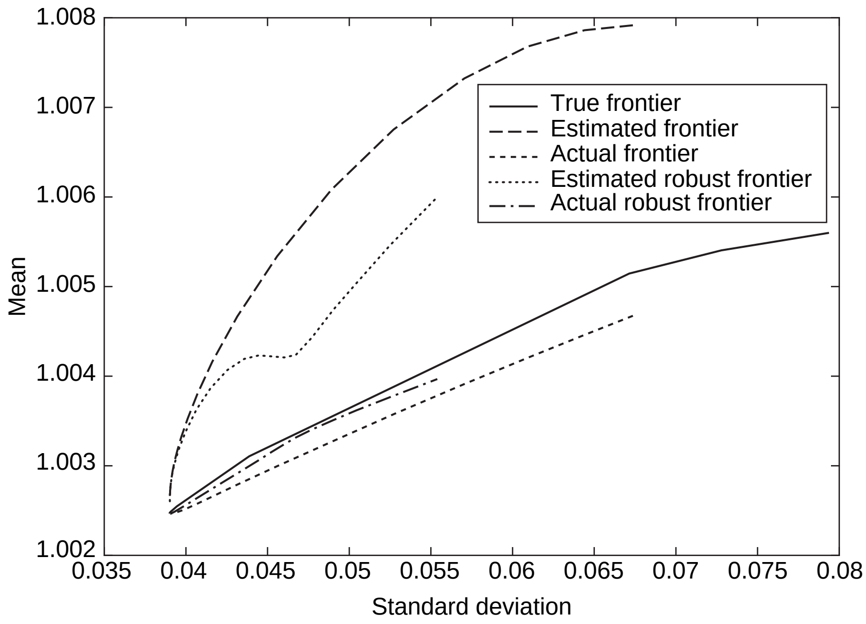
The mathematical tools developed in Section 5 will now be used to address portfolio problems of the type $\mathcal{P}_1$ (standard Markowitz), $\mathcal{P}_3$ (robust Markowitz), and $\mathcal{P}_4$ (multistage robust Markowitz). Our computational experiments are based on an asset universe consisting of $N = 5$ major stock indices: DAX, CAC 40, NASDAQ, SMI, and S&P 500. For estimating the asset return distributions, we use historical monthly time series in US dollars from October 1998 to September 2008. The means, standard deviations, and correlations of the total monthly returns are obtained via sample-average and sample-(co)variance estimators, see Table 1 (see page 108).

Our first numerical experiments are based on simulated data. In these tests, the asset prices are assumed to follow a multivariate geometric Brownian motion. The total returns over non-overlapping monthly time intervals are thus mutually

TABLE 1 Distributional parameters of monthly asset returns in US dollars (estimation period: October 1998 – September 2008).

Index	Mean	SDev	Correlations				
DAX	1.0048	0.0663	1.0000				
CAC 40	1.0036	0.0536	0.9251	1.0000			
NASDAQ	1.0056	0.0793	0.7346	0.7448	1.0000		
SMI	1.0023	0.0426	0.7584	0.8093	0.5012	1.0000	
S&P 500	1.0026	0.0406	0.8034	0.8329	0.8052	0.7602	1.0000

FIGURE 1 Mean-variance frontiers under different probabilistic models.



The estimated and actual frontiers (both robust and non-robust) are averaged over 1,000 sample sets.

independent and identically lognormally distributed. We assume that the "true" means, standard deviations, and correlations of these monthly returns are given by the values in Table 1. This parameter choice uniquely specifies the objective model Q, that is, the "true" joint probability distribution of the asset returns. An investor who has precise knowledge of Q can solve the single-stage Markowitz problem[7] $\mathcal{P}_1$ for different return targets to obtain the true efficient frontier, see Figure 1.

[7]Throughout this section, we set the asset holdings in the benchmark portfolio to zero. This implies that the portfolio's performance and risk are measured in absolute terms.

As discussed in Section 2, however, investors are not aware of the true model. The standard deviations and correlations can be estimated with reasonable accuracy, and therefore it is acceptable to suppose that these parameters are known. Estimation of the means, however, is complicated by the mean blur effect (Luenberger (1998)). One can simulate the situation faced by a real investor who estimates the means based on ten years of monthly data: use the true model Q to sample 120 monthly returns for all five assets and determine "estimated" means by calculating the sample averages. The resulting vector $\hat{\boldsymbol{\mu}}$ of estimated mean returns is then combined with the true standard deviations and correlations from Table 1 to construct an "estimated" lognormal return distribution $\hat{Q}$. Solving model $\mathcal{P}_1$ under $\hat{Q}$ for different return targets yields an estimated efficient frontier. Note that the estimate $\hat{\boldsymbol{\mu}}$, the model $\hat{Q}$, and the corresponding estimated frontier are random objects as they depend on the random samples. Figure 1 displays the expected estimated frontier obtained from averaging 1,000 estimated frontiers (each of which is based on a different sample of 120 returns).

By Michaud's error maximization property, the expected estimated frontier must lie above the true frontier, see Equation (2.1). However, no portfolio on the true frontier can possibly be dominated by any other portfolio under the true model Q. Hence, evaluating the investment strategy corresponding to any portfolio on a given estimated frontier generates a random terminal wealth whose mean and variance (under Q) describe a point below the true frontier. The collection of all mean-variance points obtained in this manner gives rise to a random actual frontier (see, for example, Broadie (1993)). By construction, the actual frontier necessarily lies beneath the true frontier. Figure 1 shows the expected actual frontier which is obtained by averaging 1,000 actual frontiers corresponding to different sample sets.

Only the estimated frontier is observable in reality. The true and actual frontiers are not observable since their construction is based on the unknown model Q. Loosely speaking, the estimated frontier describes what the investor believes to happen, whereas the actual frontier describes what really happens. As becomes apparent from Figure 1, these two perspectives may be in severe conflict.

As pointed out by Ceria and Stubbs (2006), robust portfolio optimization may help to reduce the gap between the estimated and actual frontiers. A sophisticated investor may therefore want to solve the robust Markowitz problem $\mathcal{P}_3$ using a suitable ambiguity set $\mathcal{A}$. In this paper, we study a two-parametric family of ambiguity sets that account for the estimation error associated with $\hat{\boldsymbol{\mu}}$. In order to construct $\mathcal{A}$, we first introduce a set of conceivable mean return vectors which are within reasonable proximity to $\hat{\boldsymbol{\mu}}$:

$$\mathcal{M}_{B,\,D} = \left\{ \mu \in \mathbb{R}^n \;\middle|\; \mu_i = \hat{\mu}_i + \sigma_i D\, \lambda_i,\ \lambda_i \in \{0, \pm 1\},\ \left| \sum_{i=1}^{n} \lambda_i \right| \le B \right\}$$

Here, σ_i represents the standard deviation of the return of asset i. An ambiguity set corresponding to $\mathcal{M}_{B,\,D}$ can now be defined as follows. For every $\mu \in \mathcal{M}_{B,\,D}$, we construct a lognormal return distribution $P(\mu)$ which has mean value μ and whose standard deviations and correlations coincide with the true values in Table 1.

Then, we define $\mathcal{A}_{B,\,D} = \{P(\mu) \mid \mu \in \mathcal{M}_{B,\,D}\}$. Note that an investor using model $\mathcal{P}_3$ believes that the true return distribution belongs to $\mathcal{A}_{B,\,D}$ or, equivalently, that the true mean return vector is contained in $\mathcal{M}_{B,\,D}$.[8] Observe that every $\mu \in \mathcal{M}_{B,\,D}$ is uniquely determined by a set of integers $\{\lambda_i\}_{i=1}^n$, which are interpreted as follows:

$$\lambda_i = \begin{cases} +1 & \Rightarrow \quad \hat{\boldsymbol{\mu}}_i \ \text{underestimates the mean return of asset } i \\ 0 & \Rightarrow \quad \hat{\boldsymbol{\mu}}_i \ \text{correctly estimates the mean return of asset } i \\ -1 & \Rightarrow \quad \hat{\boldsymbol{\mu}}_i \ \text{overestimates the mean return of asset } i \end{cases}$$

The estimated mean return vector $\hat{\boldsymbol{\mu}}$ represents the center of $\mathcal{M}_{B,\,D}$, while the scale parameter D determines its size (or diameter), and B has the interpretation of an uncertainty budget. A budget of zero, for instance, expresses the investor's belief that there are roughly equally as many mean returns above their estimates as there are below. Uncertainty budgets of this type have been discussed in Ceria and Stubbs (2006). By increasing either of the two parameters B or D, the investor assigns more uncertainty to the estimate $\hat{\boldsymbol{\mu}}$, thereby increasing the robustness of model $\mathcal{P}_3$.

Solving the parametric problem $\mathcal{P}_3$ under some ambiguity set $\mathcal{A}_{B,\,D}$ yields a family of robust portfolio strategies – one for each return target. Each arising investment strategy results in a random terminal wealth, whose distribution can be evaluated under any given probabilistic model. The means and variances with respect to the estimated and true models $\hat{Q}$ and Q give rise to an estimated robust and an actual robust frontier, respectively. Robust frontiers of this type were first proposed by Ceria and Stubbs (2006). Recall that the estimate $\hat{\boldsymbol{\mu}}$ is random, and therefore $\mathcal{M}_{B,\,D}$, $\mathcal{A}_{B,\,D}$, as well as the estimated and actual robust frontiers represent random objects, as well. Figure 1 displays the expected estimated and actual robust frontiers, which are obtained by averaging 1,000 estimated and actual robust frontiers corresponding to different sample sets, respectively. The parameters specifying the ambiguity set are set to $B = 0$ and $D = 0.8$.

The estimated robust frontier describes what the investor believes to happen, whereas the actual robust frontier describes what really happens. The estimated and actual robust frontiers are much better aligned than their non-robust counterparts. Moreover, they are also closer to the true Markowitz frontier. Thus, using a robust approach may be beneficial in two respects: the investor is more realistic about the achievable outcomes than a naive Markowitz investor, and the robust portfolio typically performs better in reality (that is, under Q) than a non-robust Markowitz portfolio.

From the above preliminary tests we conjecture that robust Markowitz portfolios outperform classical Markowitz portfolios in the majority of cases. In order to verify this hypothesis and to analyze the impact of different parameter choices, we run a sequence of simulated backtests inspired by Ceria and Stubbs (2006). Our numerical experiments are designed as follows. We simulate a time series of

[8]The solution of model $\mathcal{P}_3$ does not change if $\mathcal{A}_{B,\,D}$ is replaced by its convex hull. Hence, we may make the stronger statement that the true mean return vector is believed to lie within the convex hull of $\mathcal{M}_{B,\,D}$.

TABLE 2 Simulated backtest of one-stage robust Markowitz models.

Budget (B)	Diameter (D)	Robust monthly return	Robust Win (%)	Budget (B)	Diameter (D)	Robust monthly return	Robust Win (%)
0	0.5	1.0063	71	3	0.5	1.0062	68
0	1	1.0065	75	3	1	1.0064	64
0	1.5	1.0066	77	3	1.5	1.0066	74
1	0.5	1.0063	69	4	0.5	1.0050	48
1	1	1.0064	69	4	1	1.0048	35
1	1.5	1.0064	72	4	1.5	1.0049	36
2	0.5	1.0063	61	5	0.5	1.0045	39
2	1	1.0065	73	5	1	1.0046	38
2	1.5	1.0065	73	5	1.5	1.0046	39

monthly asset returns from a multivariate lognormal distribution with the parameters in Table 1. For each month, we estimate a vector of mean returns by calculating the average return over the previous 12 months. This imprecise estimate as well as the precise standard deviations and correlations from Table 1 constitute the inputs for the classical and robust Markowitz models $\mathcal{P}_1$ and $\mathcal{P}_3$, which are used to determine the estimated and estimated robust frontiers, respectively. Next, we implement the portfolio decisions corresponding to a risk level of 20% on both frontiers and evaluate the realized portfolio returns over the subsequent month. Each backtest covers a period of 120 months, over which we calculate the geometric mean returns for both the Markowitz and robust Markowitz portfolios. In total, we conduct 100 runs of the described backtest based on different simulated time series. Table 2 reports the monthly return of the robust Markowitz portfolio averaged over the 100 backtests for different choices of the uncertainty budget (B) and the size parameter (D). The averaged monthly return of the ordinary Markowitz portfolio amounts to 1.0054 (which is independent of B and D). The column "Robust Win (%)" indicates the percentage of simulations in which the robust strategy outperformed the normal Markowitz strategy.

The results in Table 2 suggest that the geometry of the ambiguity set has a substantial impact on the performance of the robust strategy. It seems beneficial to use a small uncertainty budget $B \approx 0$, whereas a budget of the order of the number of assets leads to an overly conservative ambiguity set and results in poor portfolio performance. We further find that the portfolio performance is insensitive to the size parameter D as long as it exceeds a value of about 0.5 (note that for $D \downarrow 0$ we recover the standard Markowitz model).

Multistage portfolio models that cover more than one rebalancing interval allow the investor to anticipate long-term consequences of current decisions. Therefore, dynamic strategies are expected to be less risky than myopic strategies when subjected to equivalent return targets. To investigate the additional benefits of using a multistage model, we compare portfolio strategies obtained from the static and dynamic robust Markowitz models $\mathcal{P}_3$ and $\mathcal{P}_4$, respectively. In our numerical experiments we reuse the 100 time series of monthly asset returns generated previously. However, each time series is now subdivided into quarters (ie, blocks of three months), and the vector of mean returns $\hat{\boldsymbol{\mu}}$ is reestimated only at the beginning of a new quarter (instead of every month). This provisional reduction of the estimation frequency will facilitate the comparison of the static and dynamic models. We also assume that there are no transaction costs.

At the beginning of each quarter, we determine a monthly estimated robust frontier by solving several instances of problem $\mathcal{P}_3$ for a fixed ambiguity set $\mathcal{A}_{B,\,D}$, and we denote by $\hat{w}$ the optimal static portfolio at risk level 20%. We also construct a quarterly estimated robust frontier by solving several instances of problem $\mathcal{P}_4^u$ with $H = 4$ rebalancing times and an ambiguity set $\mathcal{A}_{B,\,D}^H$, which is defined in the obvious way as a multistage extension of $\mathcal{A}_{B,\,D}$. Note that we are forced to resort to the approximate problem $\mathcal{P}_4^u$ because the original model $\mathcal{P}_4$ is not amenable to numerical solution.

In order to establish a meaningful backtest, we must decide which portfolio on the quarterly frontier should be benchmarked against the portfolio $\hat{w}$ on the monthly frontier. Since the estimate $\hat{\boldsymbol{\mu}}$ is not recalibrated intermittently, $\hat{w}$ remains optimal for $\mathcal{P}_3$ in all three months of the current quarter. The corresponding quarterly portfolio return can therefore be expressed as $\prod_{h=2}^4 \boldsymbol{\xi}_h^\top \hat{w}$, while the associated quarterly worst-case risk amounts to:

$$\begin{aligned}\sigma_H^2 &= \max_{P \in \mathcal{A}_{B,\,D}^H} \mathrm{Var}_P\left(\prod_{h=2}^4 \boldsymbol{\xi}_h^\top \hat{w}\right) \\ &= \max_{\mu \in \mathcal{M}_{B,\,D}} (\hat{w}^\top \Sigma \hat{w})^3 + 3(\hat{w}^\top \Sigma \hat{w})^2 (\mu^\top \hat{w})^2 + 3\hat{w}^\top \Sigma \hat{w} (\mu^\top \hat{w})^4\end{aligned}$$

The above formula uses the fact that the monthly asset returns are independent and identically distributed under each model $P \in \mathcal{A}_{B,\,D}^H$. Recall also that the covariance matrix $\Sigma = \mathrm{Cov}(\boldsymbol{\xi}_h)$ is determined by the data in Table 1, thus being equal for all models. For our backtests we use the dynamic portfolio strategy at risk level σ_H on the quarterly frontier. This strategy has the highest worst-case expected return over all dynamic strategies subject to the same risk. Hence, it outperforms the static fixed-mix strategy given by $\hat{w}$.

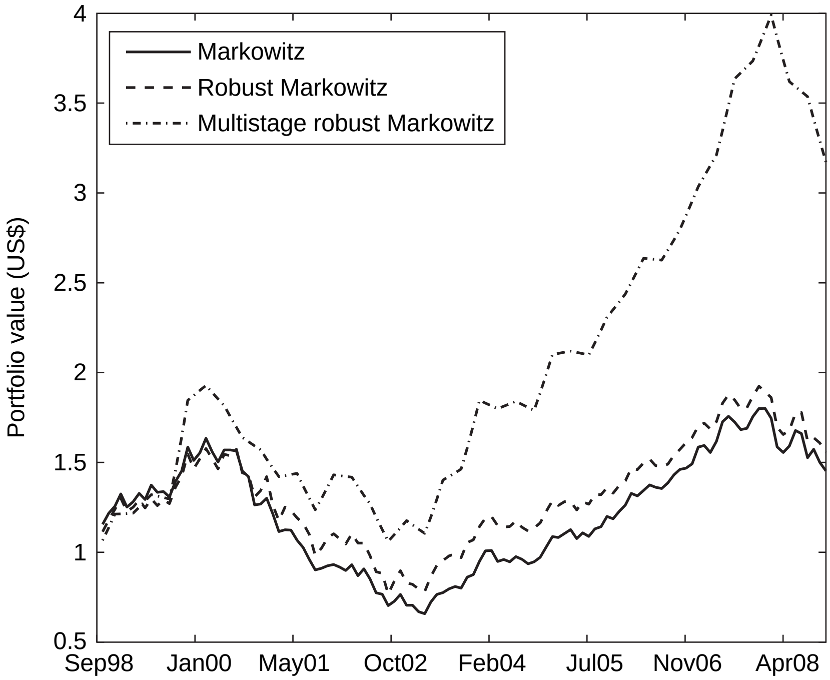
Next, we implement the described static and dynamic strategies,[9] which are constructed to result in the same worst-case portfolio variance, and evaluate the realized portfolio returns over the current quarter. Our backtest covers a period of 120 months (40 quarters), over which we calculate the geometric mean of the realized returns for both the static and dynamic robust Markowitz portfolios. This backtest is repeated for each of the 100 simulated time series. Table 3 (see page 113)

[9]Let $\boldsymbol{x}^u$ denote the dynamic strategy (interpreted as in (4.1)), which solves an instance of problem $\mathcal{P}_4^u$. Therefore, $\hat{\boldsymbol{x}} = \mathrm{E}_P(\boldsymbol{x}^u \mid \mathcal{F})$ is independent of $P \in \mathcal{A}_{B,\,D}^H$ and is near-optimal in the corresponding instance of $\mathcal{P}_4$, see Theorem 5.4. Note that $\hat{\boldsymbol{x}}$ is implementable in reality as it provides investment recommendations for all return scenarios and all rebalancing dates within the current quarter. We use $\hat{\boldsymbol{x}}$ in our backtests.

TABLE 3 Simulated backtest of multistage robust Markowitz models.

Budget (B)	Diameter (D)	Stages (H−1)	Robust monthly return (%)	Robust Win (%)
0	1	3	1.0075	82
0	1	1	1.0065	72

FIGURE 2 Historical backtest.



reports the monthly returns of the dynamic and static robust Markowitz portfolios averaged over the 100 backtests and for an ambiguity set with $B = 0$ and $D = 1$. We exclusively work with this specific parameter choice which proved to yield good results in our previous tests. In the current setting with a reduced estimation frequency, the averaged monthly return of the normal Markowitz portfolio amounts to 1.0053. The column labeled "Robust Win (%)" indicates the percentage of simulations in which the robust strategy outperformed the normal Markowitz strategy.

The results in Table 3 suggest that the dynamic robust model is superior to the static robust model, while both the static and dynamic robust models outperform the classical Markowitz model. However, this conjecture exclusively relies on backtests with simulated return data. In order to strengthen its plausibility, we repeat the previous backtest with historical returns between October 1998 and September

2008, see Figure 2. Although the static robust model performs only marginally better than the normal Markowitz model in this particular test, the results in Figure 2 are promising and seem to support the above conjecture.

Concluding remarks

In this paper, we elaborate a robust and dynamic approach to portfolio optimization based on scenario trees. We consider rival probability measures describing the future asset returns and propose a min–max model that predicts realistic portfolio performance and avoids disappointing results. We develop a computational framework for solving this problem approximately and propose methods to control the approximation errors. Our initial numerical experiments show that there are tangible benefits of integrating robust and dynamic approaches in portfolio optimization.

REFERENCES

Ash, R. (1972). Real analysis and probability. *Probability and Mathematical Statistics*. Academic Press, Berlin.

Bielecki, T. R., Jin, H., Pliska, S. R., and Zhou, X. Y. (2005). Continuous-time mean-variance portfolio selection with bankruptcy prohibition. *Mathematical Finance* **15**(2), 213–244.

Birge, J., and Louveaux, F. (1997). *Introduction to Stochastic Programming*. Springer, New York.

Black, F., and Litterman, R. (1990). Asset allocation: combining investors' views with market equilibrium. Technical report, Fixed Income Research, Goldman, Sachs, New York, September.

Broadie, M. (1993). Computing efficient frontiers using estimated parameters. *Annals of Operations Research* **45**(1–4), 21–58.

Ceria, S., and Stubbs, R. (2006). Incorporating estimation errors into portfolio selection: robust portfolio construction. *Journal of Asset Management* **7**(2), 109–127.

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

Dupacova, J., Consigli, G., and Wallace, S. (2000). Scenarios for multistage stochastic programs. *Annals of Operations Research* **100**(1–4), 25–53.

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

Fabozzi, F., Kolm, P., Pachamanova, D., and Focardi, S. (2007). *Robust Portfolio Optimization and Management*. Wiley, Hoboken.

Frauendorfer, K., and Siede, H. (2000). Portfolio selection using multi-stage stochastic programming. *Central European Journal of Operations Research* **7**, 277–290.

Goldfarb, D., and Iyengar, G. (2002). Robust portfolio selection problems. *Mathematics of Operations Research* **28**(1), 1–38.

Gülpınar, N., and Rustem, B. (2007). Worst-case robust decisions for multi-period mean-variance portfolio optimization. *European Journal of Operational Research* **183**(3), 981–1000.

Gülpınar, N., Rustem, B., and Settergren, R. (2004). Optimization and simulation approaches to scenario tree generation. *Journal of Economic Dynamics and Control* **28**(7), 1291–1315.

Ingersoll, J. E. (1987). *Theory of Financial Decision Making*. Rowman and Littlefield, Lanham, MD.

Iyengar, G. N. (2005). Robust dynamic programming. *Mathematics of Operations Research* **30**(2), 257–280.

Kuhn, D. (2008). Aggregation and discretization in multistage stochastic programming. *Mathematical Programming A* **113**(1), 61–94.

Kuhn, D. (2009). Convergent bounds for stochastic programs with expected value constraints. To appear in *Journal of Optimization Theory and Applications*.

Kuhn, D., Parpas, P., and Rustem, B. (2008). Threshold accepting approach to improve bound-based approximations for portfolio optimization. *Computational Methods in Financial Engineering*, Kontoghiorghes, E., Rustem, B., and Winker, P. (eds). Springer Verlag, Berlin, pp. 3–26.

Leippold, M., Trojani, F., and Vanini, P. (2004). A geometric approach to multiperiod mean variance optimization of assets and liabilities. *Journal of Economic Dynamics and Control* **28**(6), 1079–1113.

Li, D., and Ng, W.-L. (2000). Optimal dynamic portfolio selection: multiperiod mean-variance formulation. *Mathematical Finance* **10**(3), 387–406.

Luenberger, D. (1998). *Investment Science*. Oxford University Press, New York.

Markowitz, H. (1959). Portfolio selection: Efficient diversification of investments. *Cowles Foundation for Research in Economics at Yale University, Monograph 16*. Wiley, New York.

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

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

Pflug, G., and Wozabal, D. (2007). Ambiguity in portfolio selection. *Quantitative Finance* **7**(4), 435–442.

Rockafellar, R., and Wets, R.-B. (1998). *Variational Analysis*. *A Series of Comprehensive Studies in Mathematics*, *Vol. 317*. Springer, New York.

Rustem, B., Becker, R. G., and Marty, W. (2000). Robust min–max portfolio strategies for rival forecast and risk scenarios. *Journal of Economic Dynamics and Control* **24**(11–12), 1591–1621.

Steinbach, M. C. (2001). Markowitz revisited: mean-variance models in financial portfolio analysis. *SIAM Review* **43**(1), 31–85.

Wright, S. (1994). Primal-dual aggregation and disaggregation for stochastic linear programs. *Mathematics of Operations Research* **19**(4), 893–908.

Zhou, X., and Li, D. (2000). Continuous time mean-variance portfolio selection: A stochastic LQ framework. *Applied Mathematics and Optimization* **42**, 19–33.