



The $1/N$ investment strategy is optimal under high model ambiguity

Georg Ch. Pflug^a, Alois Pichler^a, David Wozabal^{b,*}

^a Department of Statistics and Operations Research, University of Vienna, Universitätsstraße 5/9, Vienna 1010, Austria

^b Department of Business Administration, University of Vienna, Brünner Straße 72, Vienna 1210, Austria

ARTICLE INFO

Article history:

Received 4 April 2011

Accepted 25 July 2011

Available online 3 August 2011

JEL classification:

C44

D14

D81

G11

Keywords:

Model uncertainty

Risk aware planning

Robust optimization

ABSTRACT

The $1/N$ investment strategy, i.e. the strategy to split one's wealth uniformly between the available investment possibilities, recently received plenty of attention in the literature. In this paper, we demonstrate that the uniform investment strategy is rational in situations where an agent is faced with a sufficiently high degree of model uncertainty in the form of ambiguous loss distributions. More specifically, we use a classical risk minimization framework to show that, for a broad class of risk measures, as the uncertainty concerning the probabilistic model increases, the optimal decisions tend to the uniform investment strategy.

To illustrate the theoretical results of the paper, we investigate the Markowitz portfolio selection model as well as Conditional Value-at-Risk minimization with ambiguous loss distributions. Subsequently, we set up a numerical study using real market data to demonstrate the convergence of optimal portfolio decisions to the uniform investment strategy.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

The uniform investment strategy is interesting for researchers as well as practitioners for two reasons. Firstly, comparative studies show that naive diversification is hard to outperform as an investment strategy in a portfolio management context. Secondly, behavioral studies show that it is applied by agents in many situations. This is explained in the literature either by an inherent psychological bias, leading to potentially irrational decisions, or by the presence of some fundamental uncertainty in the decision model of the agent, making uniform diversification a rational strategy to follow. The contribution of this paper falls into the latter category as we argue that uniform diversification is an optimal strategy for certain types of risk averse investors facing model uncertainty in a stochastic programming context.

The authors do not want to imply that uniform diversification is a recommendable investment strategy in general. However, based on the results of the paper, one can explain the relative success of the $1/N$ rule in a stochastic portfolio optimization context as the result of an inaccurate specification of the data generating process, i.e. a lack of accuracy in the modeling of the distributions of the random asset returns. If the true model remains sufficiently ambig-

uous, uniform diversification may outperform more sophisticated approaches.

We start our exposition by a literature review.

The uniform investment strategy can be traced back to the 4th century, when Rabbi Issac bar Aha gave the following advice: "One should always divide his wealth into three parts: a third in land, a third in merchandise, and a third ready to hand".¹

Of course, an asset allocation strategy as simple as the rule to divide the available capital evenly among some (or even all) investment opportunities falls short of the sophistication of modern portfolio theory, which in broad terms states that a portfolio should strike an *optimal* balance between the prospective return of an investment and the possible risks of investing. The optimal decision depends on the risk preferences of the investor. It can be seen as an irony that Markowitz, arguably the father of modern portfolio theory, answered the question how he manages his own funds by stating: "My intention was to minimize my future regret. So I split my contributions fifty-fifty between bonds and equities." (see Zweig, 1998) – an application of the $1/N$ rule on an aggregate level.

In a recent paper, DeMiguel et al. (2009b) use the $1/N$ strategy as a benchmark in a rolling horizon setting and compare it against several portfolio optimization strategies. The models include the classical Markowitz portfolio selection rule as well as its most prominent extensions like Bayesian-Shrinkage type estimators,

* Corresponding author. Tel.: +43 1 4277 38105; fax: +43 1 4277 38104.

E-mail addresses: georg.pflug@univie.ac.at (G.C. Pflug), alois.pichler@univie.ac.at (A. Pichler), david.wozabal@univie.ac.at (D. Wozabal).

¹ Babylonian Talmud: Tractate Baba Mezi'a, folio 42a.

aimed at dampening the effects of estimation error, and more recent approaches based on the investors beliefs about several competing asset pricing models. Furthermore, the authors include approaches that try to minimize the influence of estimation errors by restricting the asset weights or entirely focussing on the risk minimal portfolio (ignoring the expected loss dimension altogether). The results show that the benchmark $1/N$ rule outperforms most of the other more involved strategies in terms of Sharpe ratio, certainty equivalent, and turnover and is not consistently outperformed by any of the models considered in the study. The authors explain the results by stating that the errors in estimation of the parameters of the optimization models outweigh the gains of the more advanced methodology. Chan et al. (1999) and Jagannathan and Ma (2003) conduct similar studies and also conclude that it is hard to find an investment policy that consistently outperforms the uniform investment strategy. Several authors try to incorporate this finding in their proposed portfolio selection framework, see for example DeMiguel et al. (2009a) and Tu and Zhou (2011).

Apart from the success of the $1/N$ rule in empirical studies, there is evidence that uniform investment strategies are actually used in a multitude of situations where agents have to decide on a mix of different alternatives. Benartzi and Thaler (2001) conduct experiments, where subjects are asked to allocate money to different funds available in hypothetical defined contribution pension plans. The authors find that a significant share of the investors use the $1/N$ rule. This choice seems to be independent of the variety of funds offered, i.e. subjects that were offered more equity funds invested more money in equity than subjects that were confronted with an asset universe consisting of relatively fewer equity funds and more bonds. This leads the authors to the conclusion that there is a natural psychological bias towards the $1/N$ strategy, which may result in clearly irrational and even contradictory decisions. This can be interpreted as a *cognitive bias* in the sense of Tversky and Kahneman (1981) and Kahneman (2003). In Huberman and Jiang (2006), a paper motivated by the work of Benartzi and Thaler (2001), data on the choice of consumers in actual 401(k) plans is analyzed. The authors find that there is a significant share of investors (roughly two thirds) that follow the uniform investment rule. However, there is no statistical evidence of irrational behavior of the type found in the experimental studies by Benartzi and Thaler (2001).

Other studies investigating the same phenomena in different situations, under the name of *diversification heuristic*, *diversification bias*, or *variety seeking*, arrive at similar conclusions. Simonson (1990) observes variety seeking behavior in setups where multiple decisions on future consumption have to be taken as opposed to sequential decisions on immediate consumption. In Simonson and Winer (1992), an analysis of yoghurt purchases of families reveals that larger purchases (representing simultaneous decisions on future consumption) are significantly more diverse than purchases of smaller quantities by the same families. The larger purchases contain varieties which are otherwise not bought at all. The authors explain their findings by rational risk minimizing behavior of the subjects facing uncertain future preferences. On the contrary, Read and Loewenstein (1995) explain variety seeking behavior in simultaneous decisions for future consumption by cognitive deficits termed *time contraction* and *choice bracketing*. The former refers to a situation where the consumer underestimates the time between the consumption of goods and thereby overestimates the satiation effect resulting from consuming the same product, while the latter describes the phenomena that simultaneous choices are often framed as a single portfolio choice encouraging diversification.

As mentioned before, the explanations offered in the literature for the empirical prevalence of $1/N$ heuristics can be divided into papers conjecturing that there are inherent psychological patterns

which encourage the use of uniform investment decisions, even in situations where it is disadvantageous, and approaches which try to find a rationale for this behavior. The latter usually refers to some kind of fundamental uncertainty about the optimization problem involved in the decision situation, making simple uniform diversification a rational strategy to follow. The contribution of this paper is to show that this is indeed the case in portfolio optimization problems under uncertainty if the distribution of the returns is ambiguous.

We consider a rational investor who tries to minimize her risk by choosing a portfolio of assets with uncertain returns. While the investor has some prior information about the possible distributions of the asset returns, the distribution is not exactly known. Hence, additional to the uncertainty about the return, there is another layer of model uncertainty present, which we will call ambiguity (also called epistemic or Knightian uncertainty after Frank Knight). Note that, this kind of uncertainty is similar to the uncertainty used as justification of the $1/N$ rule in Simonson (1990) and DeMiguel et al. (2009b) as it involves uncertainty about the nature of the optimization problem faced by the decision maker.

The investor deals with this uncertainty by adopting a worst case approach and minimizing the worst case risk under all distributions which seem plausible given the available information. In accordance with the terminology in Ben-Tal et al. (2009), we call this set of distributions the *ambiguity set*. We construct ambiguity sets as non-parametric neighborhoods of the prior in a way which is natural from a mathematical statistics' viewpoint. Subsequently, we show that under weak conditions on the risk preferences of the investor, the optimal decisions approach portfolios which obey the $1/N$ rule as the amount of model uncertainty increases.

The idea of robustifying portfolio selection problems with respect to ambiguity about the distribution of future returns is not new and is mostly pursued in the Operations Research literature. See Maenhout (2004), Calafiore (2007), Pflug and Wozabal (2007), Garlappi et al. (2007), Quaranta and Zaffaroni (2008), Vrontos et al. (2008), Kerkhof et al. (2010), Lutgens and Schotman (2010), Tarashev (2010) and Wozabal (2010) for recent advances in this direction. The proposed approaches differ in the way the ambiguity sets are defined and in the methods applied to solve the resulting optimization problems. Most of the papers make strong assumptions on the nature of the ambiguity to be able to deal with the robustified problems. Other papers that use non-parametric methods similar to our approach are Calafiore (2007), Pflug and Wozabal (2007) and Wozabal (2010). A comprehensive summary is beyond the scope of this paper.

The paper is organized as follows: in Section 2 we set up portfolio optimization problems under ambiguity and discuss how to quantify the degree of model uncertainty by the use of probability metrics. Furthermore, we discuss how the Markowitz functional as well as the Conditional Value-at-Risk fit in this framework. Section 3 contains the main theoretical results of the paper, which permit us to identify the uniform investment strategy as optimal strategy as model uncertainty increases. In Section 4, we demonstrate the theoretical results in numerical studies based on real market data. We study the ambiguous Markowitz portfolio selection model as well as the Conditional Value-at-Risk in detail. Section 5 concludes the paper by summarizing the findings as well as outlining the implications of the results.

2. Investing under ambiguity

We consider an asset universe of N financial assets with random future losses and analyze the decision problem of an agent who wants to invest a fixed amount of money in a combination of these assets for one period of time. We model the investment decision as

relative, possibly negative, weights $w = (w_1, \dots, w_N)^T \in \mathbb{R}^N$ assigned to the assets. The investor has beliefs about the joint distribution of future losses, which we describe by a prior distribution on \mathbb{R}^N .

Let (Ω, σ, μ) be a fixed probability space which admits a random variable $X^P : (\Omega, \sigma, \mu) \rightarrow \mathbb{R}^N$ with image measure P for each Borel measure on \mathbb{R}^N (see Lemma 2 in the Appendix for a justification of this assumption). This assumption permits us to use the terms *distribution* and *probability measure* synonymously. We will denote by $\|\cdot\|_{L^p}$ the norm in $L^p(\Omega, \sigma, \mu)$ to distinguish it from the vector norm $\|\cdot\|_p$ in \mathbb{R}^N .

Assume that the risk preferences of the investor can be described by a risk functional $\mathcal{R} : L^p(\Omega, \sigma, \mu) \rightarrow \mathbb{R}$, which assigns a real value to random variables $X : (\Omega, \sigma, \mu) \rightarrow \mathbb{R}$, representing random future losses. The risk functional quantifies the riskiness of X , i.e. higher values indicate more risk and thereby less desirable random variables. There is a plethora of risk functionals discussed in the literature. However, in this paper, we mostly concentrate on the following two well known functionals:

1. The Markowitz functional

$$M_\gamma(X) = \mathbb{E}(X) + \gamma \sqrt{\text{Var}(X)}, \quad (1)$$

where $\mathbb{E}(X)$ is the expectation of X and Var is the variance, while the parameter $\gamma > 0$ represents the risk aversion of the decision maker.

2. The Conditional Value-at-Risk (also called Average Value-at-Risk)

$$\text{CVaR}_\alpha(X) = \frac{1}{1-\alpha} \int_\alpha^1 F_X^{-1}(t) dt,$$

where F_X is the cumulative distribution function of the random variable X , and F_X^{-1} denotes its inverse distribution function. Note that since we define the Conditional Value-at-Risk as a risk functional, we are concerned with the values in the upper tail of the loss distribution, i.e. α will typically be chosen close to 1.

If the investor was sure that P is an accurate description of the future distribution of losses, then she would decide on a portfolio composition $w \in \mathbb{R}^N$ by solving the following single stage stochastic programming problem

$$\begin{aligned} \inf_{w \in \mathbb{R}^N} \quad & \mathcal{R}(\langle X^P, w \rangle) \\ \text{s.t.} \quad & \langle w, \mathbb{1} \rangle = 1, \end{aligned} \quad (2)$$

where $\langle \cdot, \cdot \rangle : \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}$ is the inner product, and $\mathbb{1} \in \mathbb{R}^N$ is a vector of ones. Note that $\langle X^P, w \rangle = \sum_{n=1}^N w_n X_n^P : (\Omega, \sigma, \mu) \rightarrow \mathbb{R}$, and the risk $\mathcal{R}(\langle X^P, w \rangle)$ depends on the probability measure P on \mathbb{R}^N as well as the portfolio decision $w \in \mathbb{R}^N$. We will assume throughout the paper that problem (2) is well-posed – in particular, we require that (2) is bounded from below.

However, in most real life situations the measure P is not known to the decision maker. While statistical methods, analysis of fundamentals, and expert opinion can help to form a belief about the measure P , the true distribution remains uncertain. It is, therefore, reasonable to assume that the decision maker takes the available information into account, but also accounts for model uncertainty in her decisions. As mentioned before, we model this uncertainty by specifying a set of possible loss distributions given the prior information represented by a distribution P . This set of distributions is referred to as ambiguity set, and P is called the reference probability measure. The ambiguity set consists of measures whose distance to the reference measure does not exceed a certain threshold. To this end, we denote by $\mathcal{P}(\mathbb{R}^N)$ the space of all Borel probability measures on \mathbb{R}^N , and by

$$d(\cdot, \cdot) : \mathcal{P}(\mathbb{R}^N) \times \mathcal{P}(\mathbb{R}^N) \rightarrow \mathbb{R}^+ \cup \{0\}$$

a metric on this space (see Gibbs and Su (2002) for a short introduction to the subject of probability metrics). The ambiguity set can then be defined as

$$\mathcal{B}_\kappa(P) = \{Q \in \mathcal{P} : d(P, Q) \leq \kappa\},$$

i.e. the ball of radius κ around the reference measure P .

In this paper, we focus on the Kantorovich or Wasserstein metric, i.e. we choose $d(\cdot, \cdot)$ as

$$d_p(P, Q) = \inf \left\{ \left(\int_{\mathbb{R}^N \times \mathbb{R}^N} \|x - y\|_p^p d\pi(x, y) \right)^{\frac{1}{p}} : \text{proj}_1(\pi) = P, \text{proj}_2(\pi) = Q \right\}, \quad (3)$$

where the infimum runs over all transportation plans, viz. joint distributions π on $\mathbb{R}^N \times \mathbb{R}^N$ and $\text{proj}_1(\pi)$, $\text{proj}_2(\pi)$ are the marginal distribution of the first N and the last N components respectively. It is well known that the infimum in the above definition is always attained (see Villani, 2003).

One reason for choosing the Kantorovich distance is that it plays an important role in stability results in stochastic programming, see for example Mirkov and Pflug (2007) and Heitsch and Römisch (2009).

Furthermore, the Kantorovich metric d_p metrizes weak convergence on sets of probability measures on \mathbb{R}^N for which $x \mapsto \|x\|_p^p$ is uniformly integrable (see Villani, 2003). In particular, the empirical measure \hat{P}_m based on m observations, satisfies

$$d_p(P, \hat{P}_m) \xrightarrow{m \rightarrow \infty} 0,$$

if the p th moment of P exists. This property justifies the use of d_p to construct ambiguity sets: a stronger metric would not necessarily allow to reduce the degree of ambiguity by collecting more data, while a weaker metric would lead to a topology which permits too many convergent sequences. A particularly interesting alternative would be the Kullback–Leibler distance, which is used in Calafiore (2007) as well as in Kovacevic (2011) in a robust programming context. Since this metric is stronger than the Kantorovich distance, the results of this paper do not ensure that the optimal portfolio for a high level of ambiguity is the uniform portfolio.

Since d_p is closely related to the concept of weak convergence, there exist a range of finite sample results making it possible to interpret Kantorovich balls as confidence sets around the empirical measure. See for example Dudley (1968) for completely nonparametric bounds, or Kersting (1978) for bounds which require certain smoothness properties of the true measure. Ideas on how to use these results to construct ambiguity sets can be found in Pflug and Wozabal (2007).

Given the above definition of the ambiguity set and $\kappa > 0$, we arrive at the robustified problem, the robust counterpart of (2):

$$\begin{aligned} \inf_{w \in \mathbb{R}^N} \quad & \sup_{Q \in \mathcal{B}_\kappa(P)} \mathcal{R}(\langle X^Q, w \rangle) \\ \text{s.t.} \quad & \langle w, \mathbb{1} \rangle = 1. \end{aligned} \quad (4)$$

The parameter κ signifies the degree of ambiguity, i.e. the uncertainty about the probability model P . In problem (4), the decision maker deals with the ambiguity by adopting a worst case approach, i.e. choosing the portfolio weights in such a way that the resulting decision is *robust* with respect to the model uncertainty present in the problem.

If $\kappa = 0$, the problem reduces to the minimization of $\mathcal{R}(\langle X^P, w \rangle)$ in w , i.e. the nominal problem (2). On the other hand, if κ increases, the decision will become more conservative as the supremum in (4) is taken over a growing set of measures. It seems plausible to conjecture that as $\kappa \rightarrow \infty$, the weight of the information, represented by the measure P , diminishes and the optimal decisions tend to a more diversified portfolio, approaching the uniform investment strategy $w^\mu = (N^{-1}, \dots, N^{-1})^T$ in the limit.

Purpose of this paper is to prove the correctness of this conjecture for a large class of risk functionals, which includes the examples mentioned at the beginning of this section.

3. Uniform investment strategy as a robust risk minimizing strategy

We will focus on convex, version independent risk functionals $\mathcal{R} : L^p(\Omega, \sigma, \mu) \rightarrow \mathbb{R}$ with $p < \infty$, which admit a dual characterization of the form

$$\mathcal{R}(X) = \max\{\mathbb{E}(XZ) - R(Z) : Z \in L^q\} \quad (5)$$

where q is such that $\frac{1}{p} + \frac{1}{q} = 1$ and $R : L^q(\Omega, \sigma, \mu) \rightarrow \mathbb{R}$ is convex. Note that if \mathcal{R} is lower semi-continuous, then it admits a representation of the form (5), with $R = \mathcal{R}^*$, where \mathcal{R}^* is the convex conjugate of \mathcal{R} . However, we do not require $R = \mathcal{R}^*$ for the purpose of this paper; see Pflug and Römisch (2007) for a discussion. We call a risk measure *version independent* or *law invariant* if $\mathcal{R}(X_1) = \mathcal{R}(X_2)$ for all random variables X_1 and X_2 which have the same distribution. Note that if $R = \mathcal{R}^*$ and X is in the interior of the domain $\{X \in L^p(\Omega, \sigma, \mu) : \mathcal{R}(X) < \infty\}$, then

$$\operatorname{argmax}_Z \{\mathbb{E}(XZ) - R(Z)\} = \partial \mathcal{R}(X)$$

where $\partial \mathcal{R}(X)$ is the set of subgradients of \mathcal{R} at X . For ease of notation, we will, therefore, denote the set of maximizers of (5) at X by $\partial \mathcal{R}(X)$, even though $\partial \mathcal{R}(X)$ does not have to be the set of subgradients for the case $R \neq \mathcal{R}^*$.

We start by proving the following Lemma, which investigates how much the riskiness of a decision w can change with changes in the distributions of the losses.

Lemma 1. *Let $1 \leq p < \infty$ and $\mathcal{R} : L^p(\Omega, \sigma, \mu) \rightarrow \mathbb{R}$ be a convex, version independent risk measure with dual representation (5). Let further $q > 1$ be such that $\frac{1}{p} + \frac{1}{q} = 1$ and $w \in \mathbb{R}^N$, then*

$$\left| \mathcal{R}(\langle X^{P_1}, w \rangle) - \mathcal{R}(\langle X^{P_2}, w \rangle) \right| \leq \sup_{Z: R(Z) < \infty} \|Z\|_{L^q} \|w\|_q d_p(P_1, P_2) \quad (6)$$

for arbitrary measures P_1 and P_2 on \mathbb{R}^N .

Proof. Let π be the optimal transport plan between P_1 and P_2 and choose $Y : (\Omega, \sigma, \mu) \rightarrow \mathbb{R}^N \times \mathbb{R}^N$ such that the image measure of Y on $\mathbb{R}^N \times \mathbb{R}^N$ is π . Call the projections of Y on the first and the second component X^{P_1} and X^{P_2} respectively. Note that, as suggested by the notation, the image measure of X^{P_i} is P_i , $i = 1, 2$.

Now choose a Z as a maximizer of (5) at the point $\langle X^{P_1}, w \rangle$, then

$$\begin{aligned} \mathcal{R}(\langle X^{P_1}, w \rangle) - \mathcal{R}(\langle X^{P_2}, w \rangle) &\leq \mathbb{E}(\langle X^{P_1}, w \rangle Z) - R(Z) - \mathbb{E}(\langle X^{P_2}, w \rangle Z) + R(Z) \\ &\leq \|Z\|_{L^q} \left(\int_{\Omega} |\langle X^{P_1} - X^{P_2}, w \rangle|^p d\mu \right)^{\frac{1}{p}} \\ &\leq \|Z\|_{L^q} \|w\|_q \left(\int_{\Omega} \sum_{n=1}^N |X_n^{P_1} - X_n^{P_2}|^p d\mu \right)^{\frac{1}{p}} \\ &= \|Z\|_{L^q} \|w\|_q \left(\int_{\mathbb{R}^n \times \mathbb{R}^n} \sum_{n=1}^N |x_n^1 - x_n^2|^p d\pi(x^1, x^2) \right)^{\frac{1}{p}} \\ &= \|Z\|_{L^q} \|w\|_q d_p(P_1, P_2) \end{aligned}$$

where the second and third step follow from Hölders inequality, while the last two follow from the definition of the variables X^{P_1} , X^{P_2} , and π . The result finally follows by repeating the argument for $\mathcal{R}(\langle X^{P_2}, w \rangle) - \mathcal{R}(\langle X^{P_1}, w \rangle)$ and taking the supremum over all the Z . \square

Obviously, the statement only makes sense if the upper bounds are finite. In this case, the Lipschitz continuity of a class of risk

measures with respect to the Kantorovich metric is established. Since

$$w^\mu = \operatorname{argmin}_{w \in \mathbb{R}^N : \langle w, 1 \rangle = 1} \|w\|_q, \quad \text{for all } q \geq 1, \quad (7)$$

inspecting the right hand side in (6), we see that the bound is the smallest for $w = w^\mu$. Hence, showing that the bound is always achieved would establish that the difference in risk for different measures is always the smallest for the uniform investment strategy. To show that this is indeed the case, we fix P and a radius $\kappa > 0$ and construct a measure Q for which $d_p(P, Q) = \kappa$ and (6) holds with equality. We formalize this in the next Proposition.

Proposition 1. *Let $\mathcal{R} : L^p(\Omega, \sigma, \mu) \rightarrow \mathbb{R}$ be a convex, version independent risk measure as in Lemma 1 and let $1 < p < \infty$ and q be defined by $\frac{1}{p} + \frac{1}{q} = 1$. Let further P be a probability measure on \mathbb{R}^N and assume that*

$$\|Z\|_{L^q} = C \quad \text{for all } Z \in \bigcup_{X \in L^p} \partial \mathcal{R}(X) \text{ with } R(Z) < \infty. \quad (8)$$

Then it holds that for every $\kappa > 0$ and every $w \in \mathbb{R}^N$, there is a measure Q on \mathbb{R}^N such that $d_p(P, Q) = \kappa$ and

$$\left| \mathcal{R}(\langle X^Q, w \rangle) - \mathcal{R}(\langle X^P, w \rangle) \right| = C\kappa \|w\|_q,$$

i.e. the bound of Lemma 1 holds with equality.

Proof. Fix a $Z \in \partial \mathcal{R}(\langle X^P, w \rangle)$ with $R(Z) < \infty$ and define a random variable $X^Q = (X_1^Q, \dots, X_N^Q)$ by setting $X_n^Q = X_n^P + c_1(n)|w_n|^{\frac{q}{p}}$ with

$$c_1(n) = \frac{\operatorname{sign}(w_n)\operatorname{sign}(Z)c_2}{\|w\|_q^q} |Z|^{\frac{q}{p}}$$

for all $n: 1 \leq n \leq N$ and $c_2 > 0$. If we set $c_1 = |c_1(n)|$, it is easily verified that

$$c_1^p |w_n|^q = |X_n^Q - X_n^P|^p, \quad \forall n: 1 \leq n \leq N \quad (9)$$

holds. Furthermore, we have

$$\begin{aligned} \left| \sum_{i=1}^N w_n (X_n^Q - X_n^P) \right|^p &= \left| \sum_{n=1}^N w_n c_1(n) |w_n|^{\frac{q}{p}} \right|^p = c_1^p \left| \sum_{n=1}^N |w_n|^q \right|^p \\ &= c_1^p \|w\|_q^{pq} = c_2^p |Z|^q. \end{aligned} \quad (10)$$

Note that the choice of the parameter $c_2 > 0$ determines the distance $d_p(P, Q)$ of the image measure Q of X^Q to P , i.e. bigger values yield a bigger distance, and for every $\kappa > 0$, there is a $c_2 > 0$ such that $d_p(P, Q) = \kappa$ for the respective image measure Q . Assume that c_2 is chosen like that, then

$$\mathcal{R}(\langle X^Q, w \rangle) - \mathcal{R}(\langle X^P, w \rangle) \geq \mathbb{E}(\langle X^Q, w \rangle Z) - R(Z) - \mathbb{E}(\langle X^P, w \rangle Z) + R(Z) \quad (11)$$

$$\begin{aligned} &= \mathbb{E}(\langle X^Q - X^P, w \rangle Z) \\ &= \|Z\|_{L^q} \left(\int_{\Omega} |\langle X^Q - X^P, w \rangle|^p d\mu \right)^{\frac{1}{p}} \end{aligned} \quad (12)$$

$$\begin{aligned} &= \|Z\|_{L^q} \|w\|_q \left(\int_{\Omega} \sum_{n=1}^N |X_n^Q - X_n^P|^p d\mu \right)^{\frac{1}{p}} \\ &\geq \|Z\|_{L^q} \|w\|_q \kappa \geq 0, \end{aligned} \quad (13)$$

where inequality (11) follows from the choice of Z . Equality in (12) follows from (10) and

$$\operatorname{sign}(Z) = \operatorname{sign}(\langle X^Q - X^P, w \rangle),$$

which in turn is a consequence of the choice of c_1 . Finally, (13) follows from (9) and last inequality by the definition of the Kantorovich

distance. The assumptions on the subgradients, together with Lemma 1, yield the desired result with Q the image measure of X^Q . \square

Note that it follows from (13) and Lemma 1, that

$$d_p(P, Q) = \left(\int_{\Omega} \sum_{n=1}^N |X_n^Q - X_n^P|^p d\mu \right)^{\frac{1}{p}} \quad (14)$$

for the worst case measure Q defined in the proof of Proposition 1.

Although slightly different, the case $p = 1$ can be handled in a similar fashion.

Proposition 2. Let $\mathcal{R} : L^1(\Omega, \sigma, \mu) \rightarrow \mathbb{R}$ be a convex, version independent risk measure like in Lemma 1. Assume that

$$\|\mathcal{R}\|_{L^\infty} = C \quad \text{and} \quad |Z| = C \quad \text{or} \quad |Z| = 0 \quad (15)$$

almost everywhere for all possible subgradients of \mathcal{R} . Then it holds that for every probability measure P on \mathbb{R}^N and $\kappa > 0$, there is a measure Q on \mathbb{R}^N such that $d_1(P, Q) = \kappa$ and

$$|\mathcal{R}(\langle X^P, w \rangle) - \mathcal{R}(\langle X^Q, w \rangle)| = C \|w\|_\infty \kappa,$$

i.e. the bound of Lemma 1 holds with equality.

Proof. The proof proceeds along the same lines as the proof of Proposition 1, with the only difference that the definition of $X^Q = (X_1^Q, \dots, X_N^Q)$ changes to $X_n^Q = X_n^P + c_1(n)$ for $n:1 \leq n \leq N$ with

$$c_1(n) = \begin{cases} \text{sign}(w_n) \text{sign}(Z) c_2, & |w_n| = \|w\|_\infty \\ 0, & \text{otherwise,} \end{cases} \quad (16)$$

where we define $\text{sign}(0) = 0$. \square

The conditions (8) and (15) on the subgradients in Propositions 1 and 2 might seem restrictive at the first glance. However, the conditions in Propositions 1 and 2 are valid for most of the common risk measures. Two important examples are given below.

Example 1 (Conditional Value-at-Risk). The dual representation of CVaR is given by

$$\text{CVaR}_\alpha(X) = \sup \left\{ \mathbb{E}(XZ) : \mathbb{E}(Z) = 1, 0 \leq Z \leq \frac{1}{1-\alpha} \right\}$$

for $0 < \alpha \leq 1$ (see Pflug and Römisch, 2007). We apply Proposition 2, since the CVaR is defined on $L^1(\Omega, \sigma, \mu)$. If we choose a set $A \subseteq \Omega$ such that $\mu(A) = 1 - \alpha$ and $X(\omega) \geq F_X^{-1}(\alpha)$ for all $\omega \in A$, then it is easy to see that

$$Z(\omega) = \begin{cases} \frac{1}{1-\alpha}, & \omega \in A \\ 0, & \text{otherwise} \end{cases} \in \partial \text{CVaR}_\alpha(X).$$

Hence, condition (15) of Proposition 2 is fulfilled.

Example 2 (Markowitz Functional). The natural domain of the Markowitz functional is $L^2(\Omega, \sigma, \mu)$. To derive its dual formulation, note that

$$\begin{aligned} \sqrt{\text{Var}(X)} &= \|X - \mathbb{E}(X)\|_{L^2} = \sup \{ \mathbb{E}((X - \mathbb{E}(X))Z) : \|Z\|_{L^2} = 1 \} \\ &= \sup \{ \mathbb{E}(X(Z - \mathbb{E}(Z))) : \|Z\|_{L^2} \leq 1 \} \\ &= \sup \{ \mathbb{E}(XZ) : \mathbb{E}(Z) = 0, \|Z\|_{L^2} = 1 \}. \end{aligned}$$

Therefore, we obtain

$$\begin{aligned} M_\gamma(X) &= \mathbb{E}(X) + \gamma \sup \{ \mathbb{E}(XZ) : \mathbb{E}(Z) = 0, \|Z\|_{L^2} = 1 \} \\ &= \sup \{ \mathbb{E}(X(\gamma Z + 1)) : \mathbb{E}(Z) = 0, \|Z\|_{L^2} = 1 \} \\ &= \sup \left\{ \mathbb{E}(XZ) : \mathbb{E}(Z) = 1, \|Z\|_{L^2} = \sqrt{1 + \gamma^2} \right\}. \end{aligned}$$

Hence, it is immediate that assumption (8) in Proposition 1 is fulfilled.

Propositions 1 and 2 show that, for given portfolio weights w ,

$$\sup_{Q \in \mathcal{B}_\kappa(P)} \mathcal{R}(\langle X^Q, w \rangle) = \mathcal{R}(\langle X^P, w \rangle) + C \|w\|_q \kappa. \quad (17)$$

The solution Q of (17) can be found as the image measure of X^Q . By (7), given the budget constraint $\langle w, \mathbb{1} \rangle = 1$, the smallest change occurs for the uniform portfolio w^μ . To find w^* which solves (4) for a given $\kappa > 0$, we have to consider the tradeoff between choosing a portfolio which fares well under the original measure P and the robustness of the choice with respect to the ambiguity. However, it can be immediately seen that for every admissible w there is a level κ , such that

$$\mathcal{R}(\langle X^P, w^\mu \rangle) + C \|w^\mu\|_q \kappa < \mathcal{R}(\langle X^P, w \rangle) + C \|w\|_q \kappa.$$

Hence, as $\kappa \rightarrow \infty$, the optimal portfolio converges to w^μ . We formalize this finding in the next Proposition.

Proposition 3. Let $1 \leq p < \infty$ and \mathcal{R} be a convex risk measure as in Proposition 1 or Proposition 2, then, as $\kappa \rightarrow \infty$, in problem (4), the optimal portfolios converge to the uniform portfolio w^μ . More specifically:

1. If $p = 1$ then w^μ is the optimal solution to problem (4) for $\kappa > \kappa^*$ with

$$\kappa^* = (N-1) \mathbb{E} \left(\|X^P\|_1 \mathbb{1}_{\{Z \neq 0\}} \right).$$

2. If $p = 2$, then the optimal portfolio w^* solving (4) satisfies $\|w^* - w^\mu\|_2 \leq D$, if

$$\kappa \geq \left(\left(\frac{1}{ND^2} + 1 \right)^{\frac{1}{2}} + \frac{1}{\sqrt{ND}} \right) \mathbb{E} \left(\|X^P\|_2^2 \mathbb{1}_{\{Z \neq 0\}} \right)^{\frac{1}{2}}.$$

3. If $p \notin \{1, 2\}$, then for every $\epsilon > 0$, there is a κ_ϵ such that for $\kappa > \kappa_\epsilon$ the optimal solution w^* for (4) fulfills $\|w^* - w^\mu\|_q < \epsilon$.

Proof. We start by stating the following inequality

$$\mathcal{R}(\langle X^P, w^1 \rangle) - \mathcal{R}(\langle X^P, w^2 \rangle) \leq C \|w^1 - w^2\|_q \mathbb{E} \left(\|X^P\|_p^p \mathbb{1}_{\{Z \neq 0\}} \right)^{\frac{1}{p}} \quad (18)$$

for all $Z \in \partial \mathcal{R}(\langle X^P, w^1 \rangle)$. Eq. (18) can be proven using a similar argument as employed in Lemma 1. By Eq. (17), the uniform portfolio is optimal for problem (4) among a given set of portfolios B , iff

$$\mathcal{R}(\langle X^P, w^\mu \rangle) + C \|w^\mu\|_q \kappa \leq \mathcal{R}(\langle X^P, w \rangle) + C \|w\|_q \kappa, \quad \forall w \in B$$

which, using (18), is implied by

$$\kappa \geq \frac{\|w - w^\mu\|_q}{\|w\|_q - \|w^\mu\|_q} \mathbb{E} \left(\|X^P\|_p^p \mathbb{1}_{\{Z \neq 0\}} \right)^{\frac{1}{p}}, \quad \forall w \in B. \quad (19)$$

For the case $p = 1$, let $n^* = \arg\max_{1 \leq n \leq N} |w_n - 1/N|$. If $w_{n^*} > 1/N$, then $\|w - w^\mu\|_\infty = \|w\|_\infty - \|w^\mu\|_\infty$. If, on the other hand, $w_{n^*} < 1/N$, then $w_{n^*} = \min_n w_n$ and we conclude that

$$\max_n w_n \geq \frac{1}{N} + \frac{1/N - w_{n^*}}{N-1}.$$

It follows that

$$(N-1)(\|w\|_\infty - 1/N) \geq 1/N - w_{n^*} = \|w - w^\mu\|_\infty$$

establishing the first part of the Proposition.

For $p = q = 2$, let f_2, \dots, f_N orthogonal to each other and to w^μ with $\|f_i\|_2 = 1$ for $i = 2, \dots, N$. Hence, any w with $\langle w, \mathbb{1} \rangle = 1$ can be written as $w = w^\mu + \sum_{i=2}^N c_i f_i$ with $c_2, \dots, c_N \in \mathbb{R}$ and

$$\begin{aligned}
\frac{\|w - w^\mu\|_2}{\|w\|_2 - \|w^\mu\|_2} &= \frac{\|w - w^\mu\|_2}{\left(1/N + \sum_{i=2}^N c_i^2\right)^{\frac{1}{2}} - 1/\sqrt{N}} \\
&= \frac{\|w - w^\mu\|_2}{\left(1/N + \|w - w^\mu\|_2^2\right)^{\frac{1}{2}} - 1/\sqrt{N}} \\
&= \left(\frac{1}{N\|w - w^\mu\|_2^2} + 1\right)^{\frac{1}{2}} + \frac{1}{\sqrt{N}\|w - w^\mu\|_2}.
\end{aligned}$$

Clearly, as $\|w - w^\mu\|_2 \rightarrow \infty$, the above expression tends to 1, while it approaches ∞ for $\|w - w^\mu\|_2 \rightarrow 0$. Hence, it follows that

$$\frac{\|w - w^\mu\|_2}{\|w\|_2 - \|w^\mu\|_2} \leq \left(\frac{1}{ND^2} + 1\right)^{\frac{1}{2}} + \frac{1}{\sqrt{ND}}, \quad \forall w : \|w - w^\mu\|_2 \geq D.$$

This, together with (19), establishes the second statement.

For $p \notin \{1, 2\}$, let $(x_n)_{n \in \mathbb{N}}$ be a sequence with $x_n \nearrow \infty$ and define the convex sets

$$A_n = \{w \in \mathbb{R}^N : \langle w, \mathbb{1} \rangle = 1, \mathcal{R}(\langle X^p, w^\mu \rangle) + C\|w^\mu\|_q x_n \geq \mathcal{R}(\langle X^p, w \rangle) + C\|w\|_q x_n\}$$

in \mathbb{R}^N . $A_{n+1} \subseteq A_n$ for all $n \in \mathbb{N}$ and

$$\bigcap_{n=1}^{\infty} A_n = \{w^\mu\}.$$

Since (2) is well-posed, the mapping $w \mapsto \mathcal{R}(\langle X^p, w \rangle) + C\|w\|_q x_n$ is inf-compact, i.e. the sets A_n are compact. For $\epsilon > 0$, define the compact sets $B_n^\epsilon = A_n \setminus \{w \in \mathbb{R}^N : \|w - w^\mu\| < \epsilon\}$, and note that by the above

$$\bigcap_{n=1}^{\infty} B_n^\epsilon = \emptyset$$

and by compactness, there is a $M_\epsilon \in \mathbb{N}$ such that $\bigcap_{n=1}^{M_\epsilon} B_n^\epsilon = \emptyset$. Hence, we have shown that for every $\epsilon > 0$, there is a $M_\epsilon \in \mathbb{N}$ such that the optimal solution w^* for (4) fulfills $\|w^* - w^\mu\|_q < \epsilon$ for $\kappa > x_{M_\epsilon}$. Setting $\kappa_\epsilon = x_{M_\epsilon}$ concludes the proof. \square

4. Numerical study

In this section, we will demonstrate the results of the previous section using real market data. In particular, we solve problem (4) for the Markowitz functional, and the Conditional Value-at-Risk and investigate the optimal portfolios as the degree of ambiguity increases. As a byproduct, we derive robust counterparts of the two risk functionals, which lead to ambiguous optimization problems of the same computational complexity as the nominal problems with the original measures. We demonstrate that for $p = 1$, the threshold κ , for the uniform portfolio to be optimal, is actually smaller than the bound in Proposition 3. Similarly, we show that $\|w^* - w^\mu\|_2$ is actually smaller than the bound derived in Proposition 3 for the case $p = 2$.

The asset universe for the numerical study consists of the following seven indices: the Dow Jones Industrial index (DJI), the Dow Jones CBOT Treasury Index (CBTI), SPDR Gold Shares (GLD), the Dow Jones Composite All REIT (RCIT), the Euro Stoxx 50 (STOXX50), the Nikkei 225 index (N225), and the Shanghai Stock Exchange Composite Index (SSEC). The assets are all quoted in US dollars, i.e. the assets that are originally quoted in another currency are multiplied with the respective exchange rates.

We use historical weekly return data for the period 01.01.2007 until 31.10.2010 to obtain scenarios for the joint asset returns. In all we use 151 data points, each of which we assign the same probability, i.e. the measure P equals the empirical measure constructed from these 151 historical asset returns. While in Section 4.1 the

scenarios are used directly, facilitating a scenario based approach to robustified Conditional Value-at-Risk optimization, in Section 4.2, they are used to estimate the expected return as well as the covariance matrix needed for the robustified Markowitz approach.

4.1. Conditional Value-at-Risk

We start our investigation by defining the Ambiguous Conditional Value-at-Risk as

$$A - \text{CVaR}_\alpha(\langle w, X^p \rangle, \kappa) = \max_{Q \in \mathcal{B}_\kappa(P)} \text{CVaR}_\alpha(\langle w, X^Q \rangle)$$

and consider the problem

$$\begin{aligned} \min_w \quad & A - \text{CVaR}_\alpha(\langle w, X^p \rangle, \kappa) \\ \text{s.t.} \quad & \langle w, \mathbb{1} \rangle = 1. \end{aligned} \quad (20)$$

To ensure that the worst case distribution Q is exactly at distance κ from P , we use (14) and (16) and choose

$$c_2 = \frac{1}{(1-\alpha)k} \kappa$$

where $k = |\{n : |w_n| = \|w\|_\infty\}|$. Therefore, by (16), we have that $X^Q = (X_1^Q, \dots, X_N^Q)^\top$ with

$$X_n^Q = \begin{cases} X_n^p + \text{sign}(w_n) \text{sign}(Z) \frac{\kappa}{(1-\alpha)k}, & |w_n| = \|w\|_\infty \\ X_n^p, & \text{otherwise} \end{cases}$$

for $n: 1 \leq n \leq N$ and $Z \in \partial \text{CVaR}_\alpha(\langle X^p, w \rangle)$. Furthermore,

$$\langle w, X^Q \rangle = \langle w, X^p \rangle + \mathbb{1}_{\{Z \neq 0\}} \|w\|_\infty \frac{\kappa}{1-\alpha}.$$

In a finite scenario setting with loss scenarios x_1, \dots, x_S and probabilities p_1, \dots, p_S under the measure P , problem (20) can be cast as the following linear programming problem

$$\begin{aligned} \inf_{w \in \mathbb{R}^N, M \in \mathbb{R}} \quad & a + \frac{1}{1-\alpha} \sum_{s=1}^S z_s p_s \\ \text{s.t.} \quad & z_s \geq \langle w, x_s \rangle + M - a, \quad \forall s \in \{1, \dots, S\} \\ & \langle w, \mathbb{1} \rangle = 1 \\ & w_n \frac{\kappa}{1-\alpha} \leq M, \quad \forall n \in \{1, \dots, N\} \\ & z_s \geq 0, \quad \forall s \in \{1, \dots, S\}. \end{aligned}$$

The portfolio compositions for different levels of κ are depicted in Fig. 1a. Every vertical cut in the picture represents the portfolio composition for a given level of κ . For small values of κ , some weights are negative which results in overall investment larger than one. It can be seen that as κ increases, the portfolios rapidly approach the uniform portfolio. This observation is supported by Fig. 1b, which depicts the normalized Herfindahl–Hirschman index values for the portfolios. Recall that this index is defined as $\left[\sum_{n=1}^N w_n^2 - 1/N\right]/[1 - 1/N]$. It takes the value 0 for the uniform strategy w^μ and the value 1 for the investment in just one asset.

In the above example, the lowest level of κ , for which the optimal decision is w^μ , is 0.026, and the analytical bound κ^* from Proposition 3 is equal to 0.0734.

4.2. Markowitz functional

Analyzing the derivation of the dual representation for the Markowitz functional, we deduce that for a given X , the subgradient Z at $M_\gamma(X)$ is given by

$$Z = \gamma \frac{X - \mathbb{E}(X)}{\|X - \mathbb{E}(X)\|_2} + 1.$$

To construct a worst case measure with $d_2(P, Q) = \kappa$, we use (14) and note that

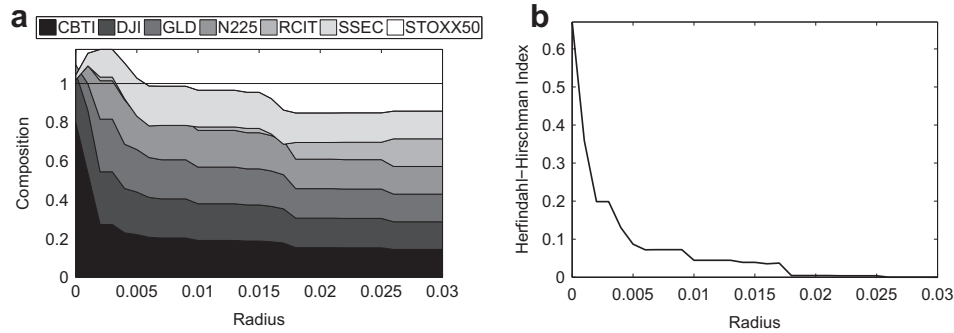


Fig. 1. In (a) the optimal portfolios in dependence of κ are depicted and (b) shows the corresponding values for the Herfindahl–Hirschman index.

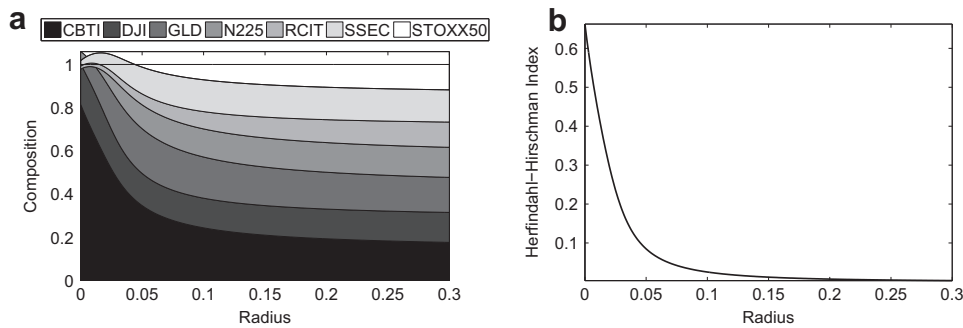


Fig. 2. In (a) the optimal portfolios in dependence of κ are depicted and (b) shows the corresponding values for the Herfindahl–Hirschman index.

$$\begin{aligned}
 d_2(P, Q) &= \left(\int_{\Omega} \sum_{n=1}^N |X_n^Q - X_n^P|^2 d\mu \right)^{\frac{1}{2}} \\
 &= \left(\int_{\Omega} \sum_{i=1}^N \left| \frac{\text{sign}(Z) \text{sign}(w_n) c_2}{\|w\|_2^2} |w_n| Z \right|^2 d\mu \right)^{\frac{1}{2}} \\
 &= \left(\int_{\Omega} \frac{c_2^2}{\|w\|_2^4} |Z|^2 \sum_{i=1}^N |w_n|^2 d\mu \right)^{\frac{1}{2}} = \frac{c_2}{\|w\|_2} \left(\int_{\Omega} |Z|^2 d\mu \right)^{\frac{1}{2}} \\
 &= \frac{c_2}{\|w\|_2} \|Z\|_{L^2}
 \end{aligned}$$

and therefore, $c_2 = \frac{\kappa \|w\|_2}{\sqrt{1+\gamma^2}}$ for a given $\kappa > 0$.

We proceed by deriving a representation of the Ambiguous-Markowitz-Functional

$$A - M_{\gamma}(\langle X^P, w \rangle, \kappa) = \max_{Q \in B_{\kappa}(P)} M_{\gamma}(\langle w, X^Q \rangle).$$

By (17), the worst case equivalent of the Markowitz risk measure is

$$A - M_{\gamma}(\langle X^P, w \rangle, \kappa) = M_{\gamma}(\langle X^P, w \rangle) + \kappa \|w\|_2 \sqrt{1 + \gamma^2}.$$

Solving the problem

$$\begin{aligned}
 \min_w \quad & A - M_{\gamma}(\langle w, X^P \rangle, \kappa) \\
 \text{s.t.} \quad & \langle w, \mathbb{1} \rangle = 1
 \end{aligned} \quad (21)$$

numerically, we obtain the portfolio weights depicted in Fig. 2a and b shows the corresponding values of the Herfindahl–Hirschman index. The results confirm the theoretical findings of Proposition 3. The optimal portfolios converge to w^H , but there seems to be no finite value of κ such that the optimal portfolios are actually equal to w^H . Nevertheless, as is evident from Fig. 2, convergence is rather fast, and even for small values of κ , the optimal portfolios are very

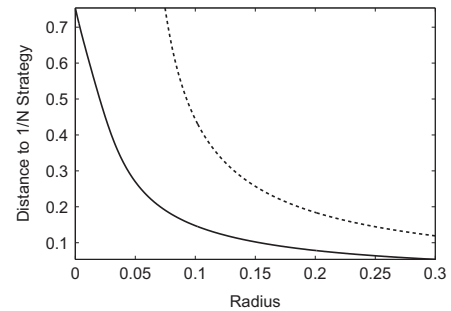


Fig. 3. Actual distance of the optimal portfolios w^* to w^H measured in the 2-norm (solid line) versus theoretical bound (dotted line).

close to w^H . In this sense, the convergence is faster than for the CVaR case. Fig. 3 depicts the actual distance of the optimal portfolios to w^H as well as the theoretical bound. In line with the results on CVaR and the discussion above, the plot shows that the actual distance turns out to be much smaller than the theoretical bound.

5. Conclusion

We showed that the uniform investment strategy or 1/N rule is a rational strategy to follow in stochastic portfolio decision problems where the distribution of asset returns is ambiguous, and the decision maker adopts a worst case approach taking into account all measures in an ambiguity set. The ambiguity set consists of all measures in a neighborhood of a reference measure, which represents the prior information of the decision maker. We use the Kantorovich metric to construct the ambiguity sets around the reference measure in a non-parametric way, i.e. we do not impose any restrictions on the measures. The choice of

the Kantorovich metric is natural since it allows the construction of ambiguity sets using statistical tools and furthermore is closely related to existing stability theory for stochastic programming problems.

In the second part of the paper, we numerically demonstrate the convergence to the uniform portfolio in portfolio optimization problems with the Markowitz functional and Conditional Value-at-Risk as the objective function. The results show that the optimal portfolio converges to the uniform portfolio even faster than suggested by the theoretical bounds established in Section 3. Furthermore, we show how the structure of the portfolio actually approaches the uniform portfolio, i.e. how even small levels of ambiguity cause diversification in the optimal portfolios. This point is illustrated by the fact that the normalized Herfindahl–Hirschman index of the portfolios is monotonically decreasing with the degree of ambiguity in the model.

The results obtained in this paper contribute to the contemporary discussion in two ways:

1. We showed that a rational agent chooses increasingly diversified portfolios, when model uncertainty increases. This may serve as an explanation for the empirically observed use of simple diversification heuristics in portfolio selection settings. The paper therefore provides a justification of this behavioral pattern founded in the theory of rational choice.
2. The optimality of the uniform portfolio rule in the face of model uncertainty explains the good performance of this strategy in comparative studies, such as DeMiguel et al. (2009b). If naive diversification outperforms more sophisticated models, this can be seen as a clear indication that the modeling of the data generating process is not accurate enough to serve as an input for the particular model class. This, in turn, implies that the decision maker either has to improve on the statistical modeling, or if this is not possible, choose a different criterion of optimality which is more robust with respect to estimation error. The bounds derived in Proposition 3 may serve as an indication of the sensitivity of different risk measures to model uncertainty.

Further research on the topic could reveal a more systematic characterization of the different risk measures with respect to model uncertainty.

Appendix A. Random variables with given image measures

Lemma 2. Let $([0, 1], \sigma_{[0,1]}, \lambda)$ be the standard probability space with the Lebesgue measure on the Borel sets $\sigma_{[0,1]}$. Let further M be a complete, separable, uncountable metric space and P a Borel probability measure on M . Then there is a measurable function $X^P: ([0, 1], \sigma_{[0,1]}, \lambda) \rightarrow (M, \sigma_M, P)$ such that

$$P(A) = \lambda((X^P)^{-1}(A)), \quad \forall A \in \sigma_M.$$

Proof. Let $K \in \mathbb{N} \cup \{0, \infty\}$ and x_1, \dots, x_K be atoms of P with probabilities p_1, \dots, p_K (if P has no atoms then set $K = 0$) and $\bar{p} = \sum_{k=1}^K p_k$. Define $A_1 = [0, p_1]$ and

$$A_k = \left[\sum_{j=1}^{k-1} p_j, \sum_{j=1}^k p_j \right)$$

for $k = 2, \dots, K$, and note that $\lambda(A_k) = p_k$. Define the measure P' by

$$P'(A) = P(A) - \sum_{x_k: x_k \in A} P(x_k)$$

and note that under the conditions of the Lemma there is a measure preserving map $T: ([\bar{p}, 1], \sigma_{[\bar{p},1]}, \lambda) \rightarrow (M, \sigma_M, P')$ by Theorem 15.5.16 in Royden (1988). Defining

$$X^P(x) = \begin{cases} x_k, & x \in A_k, \quad k = 1, \dots, K \\ T(x), & x \in M \setminus \bigcup_{k=1}^K A_k \end{cases}$$

concludes the proof. \square

References

- Ben-Tal, A., El Ghaoui, L., Nemirovski, A., 2009. Robust Optimization. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ.
- Benartzi, S., Thaler, R., 2001. Naive diversification strategies in defined contribution saving plans. *American Economic Review* 91, 79–98.
- Calafiore, G., 2007. Ambiguous risk measures and optimal robust portfolios. *SIAM Journal on Optimization* 18, 853–877.
- Chan, L., Karceski, J., Lakonishok, J., 1999. On portfolio optimization: forecasting covariances and choosing the risk model. *Review of Financial Studies* 12, 937–974.
- DeMiguel, V., Garlappi, L., Nogales, F., Uppal, R., 2009a. A generalized approach to portfolio optimization: improving performance by constraining portfolio norms. *Management Science* 55, 798–812.
- DeMiguel, V., Garlappi, L., Uppal, R., 2009b. Optimal versus naive diversification: how inefficient is the 1/N portfolio strategy? *Review of Financial Studies* 22, 1915–1953.
- Dudley, R.M., 1968. The speed of mean Glivenko–Cantelli convergence. *Annals of Mathematical Statistics* 40, 40–50.
- Garlappi, L., Uppal, R., Wang, T., 2007. Portfolio selection with parameter and model uncertainty: a multi-prior approach. *Review of Financial Studies* 20, 41–81.
- Gibbs, A., Su, F., 2002. On choosing and bounding probability metrics. *International Statistical Review* 70, 419–435.
- Heitsch, H., Römisch, W., 2009. Scenario tree modeling for multistage stochastic programs. *Mathematical Programming* 118, 371–406.
- Huberman, G., Jiang, W., 2006. Offering versus choice in 401(k) plans: equity exposure and number of funds. *Journal of Finance* 61, 763–801.
- Jagannathan, R., Ma, T., 2003. Risk reduction in large portfolios: why imposing the wrong constraints helps. *Journal of Finance* 58, 1651–1684.
- Kahneman, D., 2003. Maps of bounded rationality: psychology for behavioral economics. *American Economic Review* 93, 1449–1475.
- Kerkhof, J., Melenberg, B., Schumacher, H., 2010. Model risk and capital reserves. *Journal of Banking & Finance* 34, 267–279.
- Kersting, G.D., 1978. Die Geschwindigkeit der Glivenko–Cantelli–Konvergenz gemessen in der Prohorov–Metrik. *Mathematische Zeitschrift* 163, 65–102.
- Kovacevic, R., 2011. Maximum-loss, Minimum-win and the Esscher Pricing Principle. Working Paper, University of Vienna.
- Lutgens, F., Schotman, P., 2010. Robust portfolio optimisation with multiple experts. *Review of Finance* 14, 343–383.
- Maenhout, P., 2004. Robust portfolio rules and asset pricing. *Review of Financial Studies* 17, 951–983.
- Mirkov, R., Pflug, G.C., 2007. Tree approximations of dynamic stochastic programs. *SIAM Journal on Optimization* 18, 1082–1105.
- Pflug, G., Wozabal, D., 2007. Ambiguity in portfolio selection. *Quantitative Finance* 7, 435–442.
- Pflug, G.C., Römisch, W., 2007. Modeling, Measuring and Managing Risk. World Scientific, Singapore.
- Quaranta, A.G., Zaffaroni, A., 2008. Robust optimization of conditional value at risk and portfolio selection. *Journal of Banking & Finance* 32, 2046–2056.
- Read, D., Loewenstein, G., 1995. Diversification bias: explaining the discrepancy in variety seeking between combined and separated choices. *Journal of Experimental Psychology: Applied* 1, 34–49.
- Royden, H., 1988. Real Analysis. Macmillan Publishing Company, New York.
- Simonson, I., 1990. The effect of purchase quantity and timing on variety-seeking behavior. *Journal of Marketing Research* 27, 150–162.
- Simonson, I., Winer, R., 1992. The influence of purchase quantity and display format on consumer preference for variety. *Journal of Consumer Research: An Interdisciplinary Quarterly* 19, 133–138.
- Tarashev, N., 2010. Measuring portfolio credit risk correctly: why parameter uncertainty matters. *Journal of Banking & Finance* 34, 2065–2076.
- Tu, J., Zhou, G., 2011. Markowitz meets talmud: a combination of sophisticated and naive diversification strategies. *Journal of Financial Economics* 99, 204–215.
- Tversky, A., Kahneman, D., 1981. The framing of decisions and the psychology of choice. *Science* 221, 453–458.
- Villani, C., 2003. Topics in Optimal Transportation. Graduate Studies in Mathematics, vol. 58. American Mathematical Society, Providence, RI.
- Vrontos, S., Vrontos, I., Giamouridis, D., 2008. Hedge fund pricing and model uncertainty. *Journal of Banking & Finance* 32, 741–753.
- Wozabal, D., 2010. A framework for optimization under ambiguity. *Annals of Operations Research*. doi:10.1007/s10479-010-0812-0.
- Zweig, J., 1998. Five investing lessons from America's top pension fund. *Money*, 115–118.