

A Review on Ambiguity in Stochastic Portfolio Optimization

Georg Ch. Pflug¹ · Mathias Pohl¹ 

Received: 8 February 2017 / Accepted: 16 October 2017
© The Author(s) 2017. This article is an open access publication

Abstract In mean-risk portfolio optimization, it is typically assumed that the assets follow a known distribution P_0 , which is estimated from observed data. Aiming at an investment strategy which is robust against possible misspecification of P_0 , the portfolio selection problem is solved with respect to the worst-case distribution within a Wasserstein-neighborhood of P_0 . We review tractable formulations of the portfolio selection problem under model ambiguity, as it is called in the literature. For instance, it is known that high model ambiguity leads to equally-weighted portfolio diversification. However, it often happens that the marginal distributions of the assets can be estimated with high accuracy, whereas the dependence structure between the assets remains ambiguous. This leads to the problem of portfolio selection under dependence uncertainty. We show that in this case portfolio concentration becomes optimal as the uncertainty with respect to the estimated dependence structure increases. Hence, distributionally robust portfolio optimization can have two very distinct implications: Diversification on the one hand and concentration on the other hand.

Keywords Stochastic optimization · Robustness · Portfolio optimization · Model uncertainty · Dependence uncertainty

Mathematics Subject Classification (2010) 90C15 · 91B28

1 Introduction

The history of optimization is characterized by weakening the assumptions about the available information. Back in the old days, the paradigm was that both the objective function

✉ Georg Ch. Pflug
georg.pflug@univie.ac.at

Mathias Pohl
mathias.pohl@univie.ac.at

¹ University of Vienna, Wien, Austria

and the constraints are well known, deterministic functions. This strict assumption has been weakened even in the early work of George Dantzig [9] by allowing some parameters to be unknown, but assuming that they are realizations of known distributions and thus random. Other pioneering works in the area of stochastic programming include Charnes and Cooper [7], Tinter [53] and Wets [57]. More recently, the assumption of known distributions has been replaced by the weaker assumption that only a set of distributions is known which includes the unknown *true* distribution. We call this set of distributions *ambiguity set*. Optimizing with respect to the *worst case* distribution, i.e., the distribution contained in the ambiguity set which yields the worst value of the objective function, leads to distributionally robust decisions.

In this context, the portfolio selection problem is arguably one of the most prominent problems. Introduced by Markowitz in the early 1950s, more than 25000¹ citations of his celebrated work [34] speak for themselves. We refer to Brandt [6] and Steinbach [52] for surveys listing some of the most important contributions in the vast field of portfolio selection.

Adapting the ideas mentioned above, this paper discusses portfolio selection strategies which are robust with respect to misspecification of the assets' distribution. In other words, we account for model uncertainty in the portfolio selection problem. Hence, our approach falls into the field of robust portfolio optimization. Two perceptive reviews of the developments in this area were written by Fabozzi et al. [17] and Kim et al. [28]. The present paper differs from most of the existing literature in that we rely on a purely non-parametric approach which builds on the Wasserstein distance to construct the ambiguity set. Within this setting, we contrast two fundamentally different ways to address the problem: Firstly, we consider the case of model uncertainty with respect to the model for the joint distribution of the portfolio's assets, which we will refer to as *model ambiguity*. Secondly, we assume that the marginal distribution of the assets are known and the model uncertainty lies solely on the level of the dependence structure. The latter case is known as *dependence uncertainty*. The gist of our analysis is that when increasing the level of model uncertainty, portfolio diversification becomes optimal, while in the latter case of dependence uncertainty portfolio concentration becomes optimal.

The remainder of this paper is structured as follows. Section 2 introduces ambiguity in stochastic optimization in a general context. The portfolio selection problem under model ambiguity is studied in Section 3, whereas Section 4 looks at dependence uncertainty in the context of portfolio selection. The empirical study in Section 5 compares the two presented approaches. Section 6 concludes.

2 Ambiguity Sets

The concept that probability models are given by ambiguity sets is quite popular in statistics as there is a plethora of parametric and semi-parametric models, which are based on confidence regions of parameters. The basic construction for parametric models, i.e., probability models of the form $(P_\theta, \theta \in \Theta)$, allows us to define an ambiguity set as a subset $(P_\theta, \theta \in \hat{\Theta})$, which is determined by a parametric estimate $\hat{\theta}$ of θ and some confidence region $\hat{\Theta}$ around it. Typical examples include:

¹according to Google Scholar.

- The scenario distribution stems from a normal distribution $N(\mu, \sigma^2)$, where μ and σ are only known to lie within given bounds, i.e., $\mu_1 \leq \mu \leq \mu_2$ and $\sigma_1^2 \leq \sigma^2 \leq \sigma_2^2$.
- The multivariate distributions possess first and second moments, where the mean vector μ satisfies $\mu_1 \leq \mu \leq \mu_2$ (componentwise) and the covariance matrix Σ satisfies $\Sigma_1 \preceq \Sigma \preceq \Sigma_2$ (in the sense of positive semidefinite ordering).

When it comes to semi-parametric and fully nonparametric ambiguity models, we have seen different approaches emerge in recent years. For instance, ambiguity set containing all distributions with given constraints on the moments are studied e.g. in [10, 13, 31]. This approach can be extended to ambiguity sets which are characterized, additionally to the moment bounds, by constraints on structural properties such as symmetry, unimodality or independence patterns, see Hanasusanto et al. [24]. Another field of research deals with sample average approximations, where ambiguity sets appear as confidence regions, see e.g. [5, 37, 59]. An overview of these and related topics can be found in the book by Consigli, Kuhn and Brandimarte [8].

In contrast to the cited concepts above, this paper relies on a very intuitive approach, where ambiguity sets are given by some baseline model P_0 and a ball around P_0 with respect to some (pseudo-) distance for probability measures. Examples for such distances are:

- the bounded Lipschitz distance

$$d_{BL}(P, P_0) = \sup \left\{ \left| \int h(x) dP(x) - \int h(x) dP_0(x) \right| : |h(x) - h(y)| \leq d(x, y); |h(x)| \leq 1 \right\};$$

= the Wasserstein distance (here of order 1)²

$$d_W(P, P_0) = \sup \left\{ \left| \int h(x) dP(x) - \int h(x) dP_0(x) \right| : |h(x) - h(y)| \leq d(x, y) \right\}; \tag{1}$$

- the Fortet-Mourier distance

$$d_{FM}(P, P_0) = \sup \left\{ \left| \int h(x) dP(x) - \int h(x) dP_0(x) \right| : |h(x) - h(y)| \leq L(x, y) \cdot d(x, y) \right\},$$

where $L(x, y) = \max(\|x\|^r, \|y\|^r, 1)$;

- the relative entropy (Kullback-Leibler pseudo-distance)

$$d_{KL}(P, P_0) = \int \log \left(\frac{dP}{dP_0} \right) (x) dP_0(x);$$

- the Vapnik-Cervonenkis distance

$$d_{VC}(P, P_0) = \sup \{ P(C) - P_0(C) : C \in \mathcal{K} \},$$

where \mathcal{K} is a Vapnik-Cervonenkis class, see Shorack and Wellner [51].

²Wasserstein distance of order 1 is sometimes referred to as Kantorovich distance.

More details on the ideas behind the above definitions can be found in [41]. The following facts are well known: The bounded Lipschitz distance metricizes the weak topology of probability measures and therefore the empirical measure converges to the true one a.s. in this distance. If the true measure has a finite first moment, then a.s. convergence also holds with respect to the Wasserstein distance. Existence of higher moments is required for the convergence in Fortet-Mourier distance. It was shown in [40] that the Fortet-Mourier distance is a special case of the Wasserstein distance with a different metric on the real line. In contrast, the relative entropy and the VC distance generate stronger topologies implying that the empirical measure does not converge to the true one in these topologies. For this reason, these distances are excluded from our considerations. If a first moment exists, then the a.s. convergence of the empirical measure to the true measure in Wasserstein distance is an easy consequence of the law of large numbers and the dominated convergence theorem. However, rates of convergence have only been established recently, see Fournier and Guillin [21].

While the relative entropy as a concept for closedness of models was introduced by Hansen and Sargent [25], Wasserstein neighborhoods were considered for the first time by Pflug and Wozabal [44] and recently followed by Esfahani and Kuhn [16] as well as Gao and Klowegit [22] and Ji and Lejeune [26]. In the remainder of this paper we restrict our considerations to ambiguity sets defined by Wasserstein neighborhoods. In particular, we set the distance $d(x, y) = \|x - y\|_1 := \sum_{i=1}^m |x_i - y_i|$ in (1). This choice is motivated by the fact that the expected return as well as the Average-Value-at-Risk functional, defined below, are Lipschitz continuous with respect to this distance. This property will prove to be useful when we discuss the portfolio selection problem below.

In all the approaches cited above, the model uncertainty aspect in decision making is taken into account by the following concept: While for a single model P one considers the standard stochastic program

$$\min_x \{\mathbb{E}_P[Q(x, \xi)] : x \in \mathbb{X}\},$$

where Q is a cost function depending on the decision x and the random part ξ , in case of an ambiguity set \mathcal{P} one solves the minimax stochastic program

$$\min_x \max_{P \in \mathcal{P}} \{\mathbb{E}_P[Q(x, \xi)] : x \in \mathbb{X}\}. \quad (2)$$

Such a problem formulation is called risk-neutral. As an extension, we might also look at the risk-averse formulation

$$\min_x \max_{P \in \mathcal{P}} \{\mathcal{R}_P[Q(x, \xi)] : x \in \mathbb{X}\}. \quad (3)$$

where \mathcal{R} is some risk-functional.

The existence of a solution of (3) is related to the existence of a saddle point of $f(x, P) : (x, P) \mapsto \mathcal{R}_P[Q(x, \cdot)]$. Recall that the pair (x^*, P^*) is called a saddle point for f , if for any $x \in \mathbb{X}$ and $P \in \mathcal{P}$

$$f(x^*, P) \leq f(x^*, P^*) \leq f(x, P^*).$$

If (x^*, P^*) is a saddle point, then x^* is the minimizer of $x \mapsto \max_P f(x, P)$ and $P^* \in \operatorname{argmax}_P f(x^*, P)$. The converse is not true: by searching for all saddle points, one in general does not obtain all possible solutions of the minimax problem. In order to characterize all solutions of the minimax problem as saddle points, one has to impose stronger global conditions, such as quasi-convexity and quasi-concavity: A function $(x, P) \mapsto f(x, P)$ is quasi convex-concave if $\{x : f(x, P) \leq c\}$ is convex or empty for all $P \in \mathcal{P}$ and all $c \in \mathbb{R}$ and at the same time $\{P : f(x, P) \geq c\}$ is convex for all $x \in \mathbb{X}$ and all $c \in \mathbb{R}$. For such functions, Sion's famous saddle point Theorem holds, which is given in Appendix B.

3 Portfolio Selection Under Model Ambiguity

We start this section by reviewing the classical mean-risk framework, which goes back to Markowitz [34] and his mean-variance approach. In the literature, alternatives to Markowitz' choice of the variance as a risk functional have been proposed. For example, the semi-variance [35], the mean absolute deviation [29], the Value-at-Risk [4], the Average-Value-at-Risk [47] or, more generally, convex risk measures [19] as well as coherent risk measures [3] were proposed.

In the remainder of this paper, let the random vector $\xi = (\xi_1, \dots, \xi_m)^\top$ describe the random return of the m risky assets in the capital market under consideration. Let the decision variable $x = (x_1, \dots, x_m)^\top \in \mathbb{X} := \{x \in \mathbb{R}_+^m : \sum_{j=1}^m x_j = 1\}$ denote the proportion of the portfolio to be invested in the m risky assets. Notice that we do not allow for short-selling. Moreover, assume that $\lambda \in \mathbb{R}_+$ quantifies the investor's risk-aversion and \mathcal{R} is her preferred risk functional. The portfolio selection problem can then be formulated as the following single-stage stochastic program

$$\max_{x \in \mathbb{X}} \mathbb{E} \left(x^\top \xi^P \right) - \lambda \mathcal{R} \left(-x^\top \xi^P \right), \tag{4}$$

where P refers to the known probability distribution of the random vector ξ representing the asset returns. To put problem (4) into words, we aim to maximize the risk adjusted expected return of the portfolio, where the risk adjustment is controlled by the risk aversion coefficient λ as well as the risk functional \mathcal{R} . Together λ and \mathcal{R} are assumed to capture the investor's risk preferences. Notice that for a fixed probability space (Ω, σ, μ) the risk functional $\mathcal{R} : (\Omega, \sigma, \mu) \rightarrow \mathbb{R}$ assigns a real value to the random variable X , which represent the random future losses. To be more precise, $\mathcal{R}(X)$ quantifies the riskiness of X which means high values of \mathcal{R} indicate more risk and hence a less desirable situation.

We choose the Average-Value-at-Risk as a risk functional throughout most of our analysis:

$$AV@R_\alpha(X) = \frac{1}{1-\alpha} \int_\alpha^1 F_X^{-1}(t) dt,$$

where F_X^{-1} denotes the inverse distribution function of the random variable X . Notice that this definition implies that values of α are typically chosen close to 1 as we consider the upper tail of the loss distribution. As shown in Rockafellar and Uryasev [47], the AV@R allows to reformulate problem (4) as a linear program in case P is a discrete probability distribution.

In practice, the distribution P of the future returns is typically not known. However, statistical methods allow us to come up with an - hopefully accurate - estimate P_0 of the true, unknown probability distribution P . Therefore, it seems reasonable to assume that the decision maker takes the available information in form of P_0 into account, while incorporating the ambiguity with respect to P_0 into her decision. We model this situation by defining the ambiguity set as the set of distributions, whose Wasserstein distance to the so-called *reference distribution* P_0 does not exceed a certain threshold κ . That is, we define the ambiguity set as

$$\mathcal{B}_\kappa(P_0) = \{Q \in \mathcal{P}(\mathbb{R}^m) : d_W(P_0, Q) \leq \kappa\},$$

where $\mathcal{P}(\mathbb{R}^m)$ denotes the space of all Borel probability measures on \mathbb{R}^m . We can now formulate the distributionally robust counterpart of the stochastic program (4) as the following maximin problem

$$\max_{x \in \mathbb{X}} \min_{Q \in \mathcal{B}_\kappa(P_0)} \mathbb{E}(x^\top \xi^\mathcal{Q}) - \lambda \text{AV@R}_\alpha(-x^\top \xi^\mathcal{Q}). \tag{5}$$

As stated in Wozabal [58], problem (5) is algorithmically intractable for the following three reasons: Firstly, it contains infinitely many constraints. Secondly, the solutions of the maximin problem (5) are elements of an infinite dimensional space. Lastly, solving for a saddle point of the objective function is usually harder than simply finding a maximum or a minimum. Nevertheless, three distinct approaches to solve problem (5) overcome the mentioned difficulties: Pflug and Wozabal [44], who first considered the problem, restricted their analysis to all the discrete measures with atoms identical to those of P_0 , which is assumed to be the empirical measure \hat{P}_n constructed from n observations of the random vector ξ . They then proposed a successive convex programming solution method to solve the resulting semi-infinite program. As an alternative, Pflug et al. [42] directly characterized the worst case distributions Q^* , which satisfies $d_W(P_0, Q^*) = \kappa$. Hence, problem (5) can be solved by a linear program. Recently, Esfahani and Kuhn [16] derived a tractable reformulation of (5) as an LP by relying on state-of-the-art global optimization techniques. Loosely speaking, the authors change the location of the atoms of the empirical measure \hat{P}_n rather than their probability weights, as done in [44]. Figure 1 illustrates this difference. In the remaining part of this section, we review and compare these three approaches to solve problem (5).

Proceeding in chronological order, we start our literature review with the paper by Pflug and Wozabal [44]. The authors proposed the following successive convex programming algorithm:

1. Set $i = 0$ and $\mathcal{P}_0 = \{\hat{P}_n\}$.
2. Solve the outer problem

$$\begin{aligned} & \max_{x \in \mathbb{X}, t \in \mathbb{R}} t \\ & \text{s.t.: } \mathbb{E}(x^\top \xi^\mathcal{Q}) - \lambda \text{AV@R}_\alpha(-x^\top \xi^\mathcal{Q}) \geq t \quad \forall Q \in \mathcal{P}_n, \end{aligned}$$

and call the optimizer (x_i, t_i) .

3. Set $Y_x = (x_i)^\top \xi$ and solve the inner problem

$$\min_{Q \in \mathcal{B}_\kappa(\hat{P}_n)} \mathbb{E}(Y_x^\mathcal{Q}) - \lambda \text{AV@R}(-Y_x^\mathcal{Q}),$$

call the optimizer P_i and let $\mathcal{P}_{i+1} = \mathcal{P}_i \cup \{P_i\}$.

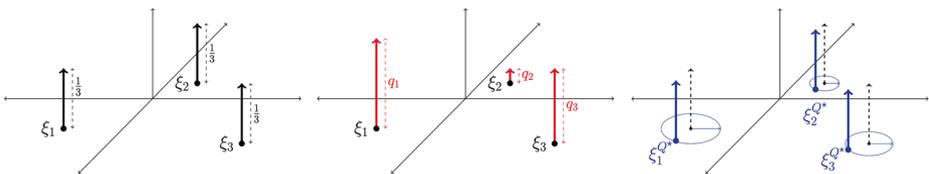


Fig. 1 The left panel illustrates an empirical measure \hat{P}_3 in \mathbb{R}^2 with support points (ξ_1, ξ_2, ξ_3) . The middle panel illustrates how [44] find the worst case probability measure, namely by changing the probability weights (q_1, q_2, q_3) of the support points (ξ_1, ξ_2, ξ_3) . The right panel illustrates the approach in [42] as well as [16]: they compute the worst case distribution by changing the support points (ξ_1, ξ_2, ξ_3) to $(\xi_1^{Q^*}, \xi_2^{Q^*}, \xi_3^{Q^*})$ while leaving their probability weights unchanged

4. If either $\mathcal{P}_{i+1} = \mathcal{P}_i$ or the optimal value of the inner problem equals t_i , then a saddle point of problem (5) is found and the algorithm stops. Otherwise, set $i = i + 1$ and go to step 2.

As \mathcal{P}_n only contains finitely many objects, the outer problem can be solved as an LP. The inner problem however has a more complicated structure. As indicated above, [44] restrict their analysis to probability measures Q which live on the same n points as the empirical measure \hat{P}_n . Hence, in the inner problem one optimizes over the probability weights $(q_i)_{i=1,\dots,n}$ of the n sample points, which have equal weight under the empirical measure \hat{P}_n . Still, the inner problem is non-convex due to bilinear terms defining the Wasserstein distance and the AV@R. We refer to section 4.2 of the follow-up paper by Wozabal [58], which proposes a method to approximately solve this program rewriting it in terms of differences of convex (D.C.) functions.

Let us now come to the paper by Pflug, Pichler and Wozabal [42], who explicitly derived the worst case distribution of $Q^* \in \mathcal{B}_\kappa(P_0)$ for fixed portfolio weight x . In order to adapt the more general results in [42], notice that the maximin problem (5) is equivalent to the following minimax problem

$$\min_{x \in \mathbb{X}} \max_{Q \in \mathcal{B}_\kappa(P_0)} (\lambda + 1) \mathcal{A}_\alpha(-x^\top \xi^Q) \tag{6}$$

where \mathcal{A}_α is a mixture of AV@R's with different α 's, i.e., it has the Choquet representation:

$$\mathcal{A}_\alpha(X) = \frac{1}{\lambda + 1} \text{AV@R}_0(X) + \frac{\lambda}{\lambda + 1} \text{AV@R}_\alpha(X).$$

We denote by Q^* the worst case distribution in the Wasserstein-ball $\mathcal{B}_\kappa(P_0)$ for a fixed x , i.e., $Q^* = \operatorname{argmax}_{Q \in \mathcal{B}_\kappa(P_0)} \{ \mathcal{A}_\alpha(-x^\top \xi^Q) \}$. Applying the result in [42] to \mathcal{A}_α , we find that the random vector of the asset returns $\xi^{Q^*} = (\xi_1^{Q^*}, \dots, \xi_m^{Q^*})^\top$ under the worst case distribution Q^* is given by

$$\xi_j^{Q^*} = \begin{cases} \xi_j^{P_0} + \operatorname{sign}(x_j) \operatorname{sign}(Z) \frac{\kappa}{k(1+\lambda)} \left(\frac{\lambda}{1-\alpha} + 1 \right) & \text{if } |x_j| = \|x\|_\infty \\ \xi_j^{P_0} & \text{otherwise} \end{cases},$$

for all $j = 1, \dots, m$, where Z is in the set $\partial \mathcal{A}_\alpha(-x^\top \xi^{P_0})$ of subgradients of \mathcal{A}_α at $-x^\top \xi^{P_0}$ and $k = |\{j \in \{1, \dots, m\} : |x_j| = \|x\|_\infty\}|$. Setting the reference distribution P_0 to the empirical distribution \hat{P}_n , problem (6) can be written as the following linear program

$$\begin{aligned} \min_{a, W, x, z} \quad & \lambda a + \frac{\lambda}{1-\alpha} \frac{1}{n} \sum_{i=1}^n z_i - \frac{1}{n} \sum_{i=1}^n x^\top \xi_i + W \\ \text{s.t.} \quad & \sum_{j=1}^m x_j = 1 \\ & z_i \geq -x^\top \xi_i + W - a \quad i = 1, \dots, n \\ & x_j \frac{\kappa}{\lambda + 1} \left(\frac{\lambda}{1-\alpha} + 1 \right) \leq W \quad j = 1, \dots, m \\ & x_j \geq 0, z_i \geq 0 \quad i = 1, \dots, n; j = 1, \dots, m. \end{aligned} \tag{7}$$

Notice that a very important implication of the above is that under high model ambiguity (i.e., $\kappa \rightarrow \infty$) portfolio diversification becomes optimal. We refer to [42] for a detailed and more general proof of this result. It should be mentioned that this result is confirmed by Esfahani and Kuhn [16], which we review next.

In the above discussion of [44], we have seen that quite evolved techniques are needed to solve problem (5), when restricting the feasible set to discrete measures with the same number of atoms as P_0 , fixing the location of these atoms and optimizing over their probability. But what happens when one is more ambitious and instead fixes the probability and optimizes over the location of the atoms? Esfahani and Kuhn [16] show that this leads to a tractable reformulation of the problem as an LP. We only restate their result here and refer to the original paper [16] for their derivation of more general results than needed in our context. First, we provide yet another equivalent reformulation of the original problem (5):

$$\min_{x \in \mathbb{X}, \tau \in \mathbb{R}} \max_{Q \in \mathcal{B}_\kappa(P_0)} \mathbb{E} \left[\max_{k \leq 2} a_k \left(x^\top \xi^Q \right) + b_k \tau \right],$$

where $a_1 = -1, a_2 = (1 - \alpha + \lambda)/(\alpha - 1), b_1 = \lambda$ and $b_2 = \lambda\alpha/(\alpha - 1)$. Assuming that $P_0 = \hat{P}_n$ and that the support of the *true* distribution P of ξ is contained in the set $\Xi := \{\xi \in \mathbb{R}^m : C\xi \leq d\}$, where C and d are of appropriate size, Esfahani and Kuhn [16] show that the above minimax problem can be written as the following linear program

$$\begin{aligned} \min_{x, \tau, \mu, z, \gamma_{ik}} \quad & \mu\kappa + \frac{1}{n} \sum_{i=1}^n z_i & (8) \\ \text{s.t.:} \quad & \sum_{j=1}^m x_j = 1 \\ & b_k \tau + a_k \left(x^\top \xi_i \right) + \gamma_{ik}^\top (d - C\xi_i) \leq z_i & k = 1, 2; i = 1, \dots, n \\ & |C^\top \gamma_{ik} - a_k x_j| \leq \mu & k = 1, 2; i = 1, \dots, n, j = 1, \dots, m \\ & \gamma_{ik} \geq 0, x \geq 0 & k = 1, 2; i = 1, \dots, n, j = 1, \dots, m. \end{aligned}$$

Comparing the three discussed approaches, it is clear that solving the LP (7) takes the least computational effort as the worst case distribution is known a priori. The LP (8) given yields the same solution as (7), while being computationally considerably more expensive.³ Still the results in Esfahani and Kuhn [16] allow us to solve the problem surprisingly fast given that the worst case distribution is not computed a priori. Lastly, we should mention that the results in [44] cannot keep up. The SCP algorithm is computationally very intense and a large sample size n is required if the algorithm should produce results similar to those of the other two approaches. Here the follow-up paper by Wozabal [58] should be mentioned, which slightly improves the requirement of a large sample size n but does so at the cost of even more computational complexity.

We refer to Section 5, where we present an empirical study based on problem (7) or equivalently (8) and consider the optimal portfolio decomposition when the level κ of model ambiguity increases.

4 Portfolio Selection under Dependence Uncertainty

In many cases of practical interest, one might be able to determine the distribution of a portfolio’s individual assets but fail to estimate their dependence structure accurately. In such a situation only the dependence structure of the asset returns is ambiguous and not their joint

³Notice that the LP in (8) has $2 + m + 3n$ variables and $2n + 4mn$ inequality constraints compared with $2 + m + n$ variables and $m + n$ inequality constraints in the LP (7).

distribution. This situation is known as *dependence uncertainty*. The main motivation to study this topic comes from risk management, as it might lie in the nature of an application at hand, that the dependence structure of different risks cannot be estimated with high accuracy. Still the dependence structure is crucial when it comes to aggregating the individual risks. We refer to Aas and Puccetti [1] for an illustrative example. Since risk measurement is the corner stone of any portfolio selection strategy, the methods and results concerning dependence uncertainty are particularly relevant in our context. As it turns out, computing bounds for aggregated risks with given marginal distributions is a challenging task on its own. Therefore, we provide a short literature review and some examples in the following.

4.1 Background

First of all, let us introduce the notion of *copulas*. The intriguing yet simple idea underlying copulas is that they allow us to split the modeling of multivariate random variables into two parts: Firstly, the modeling of the marginal distributions and secondly, the modeling of the dependence between the univariate random variables. The latter is done via copula functions. Thus, in the presence of dependence uncertainty the first task is straight forward whereas the second task is complicated due to the lack of information. To put it in a nutshell: dependence uncertainty means marginals known and copula unknown.

Formally, let H be an m -dimensional cumulative distribution function with marginals F_1, F_2, \dots, F_m . Then there exists a function C , which is called m -copula such that

$$H(t_1, t_2, \dots, t_m) = C(F_1(t_1), F_2(t_2), \dots, F_m(t_m)), \quad \forall t_1, \dots, t_m \in \mathbb{R}.$$

This basic connection is known as Sklar’s Theorem. For a precise definition of copulas and an introduction to the topic, the reader is referred to Nelsen [38].

Let us consider a simple example to understand the essence of dependence uncertainty and its connection to the concept of copulas: What is the Average-Value-at-Risk $AV@R_\alpha(U + V)$ of the sum of two standard uniforms U and V , i.e., $U, V \sim \text{Uniform}[0, 1]$? The answer depends on the copula linking the uniforms U and V :

1. Assume U and V are counter-monotonic, i.e., perfectly negative dependent. Their copula is given by the lower Fréchet Hoeffding bound $W(u, v) = \max(u + v - 1, 0)$ for all $u, v \in [0, 1]$. In this case $AV@R_\alpha^W(U + V) = 1$ for all α .⁴
2. Assume U and V are comonotonic, i.e., perfectly positive dependent. Their copula is given by the upper Fréchet Hoeffding bound $M(u, v) = \min(u, v)$ for all $u, v \in [0, 1]$. In this case $AV@R_\alpha^M(U + V) = 1 + \alpha$.
3. Assume U and V are independent and hence coupled with the product copula $\Pi(u, v) = uv$ for all $u, v \in [0, 1]$. A tedious calculation shows that if $\alpha > 0.5$, $AV@R_\alpha^\Pi(U + V) = 2 - 2\sqrt{2(1 - \alpha)}/3$.

Given that we have no information regarding the copula linking U and V , the best we can do is compute bounds for $AV@R_\alpha(U + V)$:

$$1 = AV@R_\alpha^W(U + V) \leq AV@R_\alpha^C(U + V) \leq AV@R_\alpha^M(U + V) = 1 + \alpha,$$

for any bivariate copula C . Note that these bounds are very far apart. This issue is addressed below. Let us first discuss the upper bound. A nice property of the $AV@R$ is that the copula which yields the upper bound turns out to be the comonotonic copula M . This property

⁴Notice that only in this example we indicate the copula linking U and V in the superscript of $AV@R$.

holds true in arbitrary dimensions due to the fact that the AV@R is subadditive, see [39]. Intuitively, it is clear that perfect positive dependence implies no diversification benefit when trying to minimize the overall risk of your portfolio. Hence, this dependence should be the least favored by any investor. The AV@R reflects this, whereas other risk functionals, like the Value-at-Risk (V@R), do not. In fact, finding upper bounds for the V@R is much harder. Makarov [33] and Rüschendorf [48] solved this problem in two dimensions, showing that the following copula is the one maximizing the V@R with confidence level α :

$$C_\alpha(u, v) = \begin{cases} \max(W(u, v), \alpha) & (u, v) \in [\alpha, 1] \times [\alpha, 1] \\ M(u, v) & \text{otherwise} \end{cases}. \quad (9)$$

As can be seen, counter-monotonicity in the upper tail is needed for the construction. This is precisely the reason why the result cannot be generalized to higher dimensions and leads us directly to the discussion of the difficulties that arise when computing lower bounds in dimensions strictly larger than 2. As a matter of fact, the concept of counter-monotonicity does not naturally extend to higher dimensions. To give some intuition, perfect negative dependence between two random variables implies that whenever one random variable moves into one direction, the other one moves into the opposite direction. Arguably, considering perfect negative dependence between three random variables is not as simple since it is not clear when three directions stand in maximal contrast. More formally, this is reflected by the fact that the multivariate extension of the lower Fréchet Hoeffding bound $W(u_1, \dots, u_n) = \max(\sum_{i=1}^n u_i - n, 0)$ ceases to be a copula, see [38]. We refer to Wang and Wang [55] for a discussion of possible ways to define perfect negative dependence in arbitrary dimensions.

In the context of dependence uncertainty, the best and worst case value of the V@R as well as the worst case value of the AV@R are in general unknown. Nevertheless, these bounds have been extensively studied in recent years and partial solutions have been obtained, see Puccetti and Wang [46] for an overview. Most importantly, Puccetti and Rüschendorf [45] introduced the so-called *rearrangement algorithm* which is a fast procedure to numerically compute the bounds of interest. Under quite restrictive assumptions, analytical bounds can be computed based on the notion of *complete mixability*, which was introduced in Wang and Wang [54].

In practical applications, these lower and upper bounds cease to be useful as they are too far apart. Thus, the case where additional dependence information is available, received some attention in recent years. The literature often refers to this topic as *partial dependence uncertainty* in contrast to *complete dependence uncertainty*, which we discussed above. Bounds for aggregated risks have been computed in the case where (a) the copula is given on a subset of the unit cube, (b) bounds for the copula itself are given or (c) bounds for the variance of the aggregated risks are given. We refer to the survey by Rüschendorf [50] and references therein. Most interestingly, we want to point out the work by Lux and Papapanoleon [32], who among other intriguing results derived V@R-bounds for the case where the copula is known to be close in the sense of a statistical distance to some reference copula.

This short detour to the existing literature on dependence uncertainty should serve as a motivation to model the situation where the marginal distribution of the individual assets is assumed to be known, whereas their dependence structure (i.e., the copula linking them) is only partially known in the context of portfolio selection. The fact, that the combination of portfolio optimization with dependence uncertainty is of practical and theoretical interest, has of course been noted by others before. For instance, Kakouris and Rustem [27] introduce an investment strategy which is robust against possible misspecification of the chosen parametric copula family. On the contrary, the more involved framework by Doan et al.

[12] does not need any parametric assumptions. The authors derive AV@R bounds for the Fréchet class of discrete, multivariate and overlapping marginals with finite support. Optimizing these bounds leads to an investment strategy based on a minimum of distributional assumptions.

We, however, take a different and novel approach to the portfolio selection problem under dependence uncertainty, which is very much in the spirit of Section 3:

$$\max_{x \in \mathbb{X}} \min_{C \in \mathcal{C}_\epsilon(C_0)} \mathbb{E}(x^\top \xi) - \lambda \mathcal{R}(-x^\top \xi^C), \tag{10}$$

where \mathcal{C} denotes the set of all m -dimensional copulas and the ambiguity sets $\mathcal{C}_\epsilon(C_0) := \{C \in \mathcal{C} : d_W(C_0, C) \leq \epsilon\}$ refers to the set of m -copulas which are close to some reference copula C_0 with respect to the Wasserstein distance. Since the copula has no impact on the expected value of $x^\top \xi$, we omit the letter C in the superscript of the first ξ in (10). We use the Wasserstein distance as a distance measure between copulas in order to guarantee that problem (10) is defined similarly to the portfolio selection problem under model ambiguity introduced in Eq. 4. In fact, the only difference is that in (10) we assume that the marginal distributions of the asset returns are known whereas in (4) this assumption is not made. Still the question may arise why the Wasserstein distance is a meaningful distance measure for copulas. The reason lies in the following fact: Suppose that the pairs of random variables (X_1, X_2) resp. (Y_1, Y_2) have the same marginals F_1 and F_2 but differ in their copulas C_X and C_Y . If h is a Lipschitz function with two arguments and Lipschitz constant L and both quantile functions F_1^{-1} and F_2^{-1} are Lipschitz with constant K , then

$$\mathbb{E} [|h(X_1, X_2) - h(Y_2, Y_2)|] \leq LKd_W(C_X, C_Y).$$

More generally, take a distribution P on \mathbb{R}^m with non-Lipschitz quantile functions, which are assumed not to grow faster than $-x^\gamma$ near zero and $(1-x)^\gamma$ near 1, then one may define β -Lipschitz functions h , which satisfy

$$|h(x_1, \dots, x_m) - h(y_1, \dots, y_m)| \leq L \sum_{i=1}^m \left| \text{sgn}(x_i) \cdot |x_i|^\beta - \text{sgn}(y_i) \cdot |y_i|^\beta \right|,$$

for $\beta < 1$. If $\beta < 1/\gamma$, then $P \mapsto \int h \, dP$ is, for fixed marginals, Lipschitz in the Wasserstein distance of the copula.

Concerning the choice of the risk functional \mathcal{R} , we want to point out that the AV@R is *aggregation-robust*, see Embrechts et al. [14], which implies that the AV@R is less sensitive to model uncertainty at the level of the dependence structure than, for instance, the Value-at-Risk. Hence, the AV@R will be of special interest in our subsequent analysis.

4.2 Complete Dependence Uncertainty

Before discussing how to approach problem (10) computationally, we first want to consider the case of *complete dependence uncertainty*: Assume that the degree of dependence uncertainty, given by the radius ϵ of the ambiguity set in (10), converges to infinity. Then the ambiguity set $\mathcal{C}_\epsilon(C_0)$ extends to the set of all copulas \mathcal{C} . Thus, we obtain the problem

$$\max_{x \in \mathbb{X}} \min_{C \in \mathcal{C}} \mathbb{E}(x^\top \xi) - \lambda \mathcal{R}(-x^\top \xi^C). \tag{11}$$

In the following we show that the above problem can be simplified significantly when we restrict our considerations to (a) subadditive, (b) comonotone additive and (c) positively homogeneous risk functionals \mathcal{R} . A risk functional \mathcal{R} is said to be

- (a) subadditive, if for all random variables X and Y , $\mathcal{R}(X + Y) \leq \mathcal{R}(X) + \mathcal{R}(Y)$.

- (b) comonotone additive, if for any comonotonic⁵ random variables X and Y it holds that $\mathcal{R}(X + Y) = \mathcal{R}(X) + \mathcal{R}(Y)$.
- (c) positively homogeneous, if for all random variables X and all constants $c \geq 0$ it holds that $\mathcal{R}(cX) = c\mathcal{R}(X)$.

We refer to [43] for a detailed discussion of these and other properties of risk functionals. In the following we use the notation ξ_j to denote the j -th component of the m -dimensional (random) vector ξ .

Proposition 1 *If the risk functional \mathcal{R} is subadditive, comonotone additive and positively homogeneous, then*

$$\max_{x \in \mathbb{X}} \min_{C \in \mathcal{C}} \mathbb{E}(x^\top \xi) - \lambda \mathcal{R}(-x^\top \xi^C) = \mathbb{E}(\xi_{j^*}) - \lambda \mathcal{R}(-\xi_{j^*}),$$

for any $j^* \in \operatorname{argmax}_j \mathbb{E}(\xi_j) - \lambda \mathcal{R}(-\xi_j)$.

Proof First notice that

$$\mathcal{R}(-x^\top \xi^C) \stackrel{(a)}{\leq} \sum_{j=1}^m \mathcal{R}(-x_j \xi_j) \stackrel{(b)}{=} \mathcal{R}(-x^\top \xi^M),$$

for all copulas C . Hence, we have that $\max_{C \in \mathcal{C}} \mathcal{R}(-x^\top \xi^C) = \mathcal{R}(-x^\top \xi^M)$ and we obtain

$$\begin{aligned} & \max_{x \in \mathbb{X}} \min_{C \in \mathcal{C}} \mathbb{E}(x^\top \xi) - \lambda \mathcal{R}(-x^\top \xi^C) \\ &= \max_{x \in \mathbb{X}} \mathbb{E}(x^\top \xi) - \lambda \mathcal{R}(-x^\top \xi^M) \\ & \stackrel{(b)}{=} \max_{x \in \mathbb{X}} \sum_{j=1}^m x_j \mathbb{E}(\xi_j) - \lambda \sum_{j=1}^m \mathcal{R}(-x_j \xi_j) \\ & \stackrel{(c)}{=} \max_{x \in \mathbb{X}} \sum_{j=1}^m x_j (\mathbb{E}(\xi_j) - \lambda \mathcal{R}(-\xi_j)) \\ &= \mathbb{E}(\xi_{j^*}) - \lambda \mathcal{R}(-\xi_{j^*}), \end{aligned}$$

for any $j^* \in \operatorname{argmax}_{j \in \{1, \dots, m\}} \mathbb{E}(\xi_j) - \lambda \mathcal{R}(-\xi_j)$. In the last equality we use the fact that \mathbb{X} is the m -dimensional unit simplex. □

Note that due to the high precision of modern data processors the argmax set will in practice only contain one element. Thus, we find that complete dependence uncertainty implies that concentration of the portfolio into one single asset is optimal. At first, this insight might be somewhat counterintuitive since putting all the capital into a single asset does not seem to be a robust decision. In fact, the decision when facing high model uncertainty on the level of the joint distribution could not differ more from the decision which is optimal under complete dependence uncertainty: the first implies portfolio diversification, the latter portfolio concentration. However, looking at it from the technical side, on the one hand

⁵Two random variables X and Y are said to be comonotonic if there exists a real-valued random variable Z and non-decreasing functions f_1 and f_2 such that $X = f_1(Z)$ and $Y = f_2(Z)$ or equivalently if the copula linking them is given by the comonotonic copula M .

subadditivity rewards portfolio diversification but on the other hand comonotone additivity does not attribute any diversification benefits in case of comonotone losses. Hence, without any knowledge of the dependence between the asset returns, the worst case is that all asset prices rise and fall simultaneously. Basing our decision on this worst case, portfolio diversification cannot be beneficial. Rather, investing into a single asset, namely one of those which performs best, is optimal.

Alternatively, we can approach the result from an information theoretical point of view: a decision is better the more informed it is. The equal weighted portfolio investment strategy requires no information at all. Thus, equally-weighted portfolio diversification is optimal in case of high model ambiguity and suboptimal under complete dependence uncertainty. Portfolio concentration, however, requires full information on the marginal distributions to determine which asset is the best. Therefore, it is not surprising that portfolio concentration is optimal when the model uncertainty lies solely on the level of the dependence structure.

We conclude that subadditivity, comonotone additivity and positively homogeneity are indeed reasonable properties for risk functionals in the portfolio selection problem under dependence uncertainty. Let us therefore take a closer look at risk functionals with these three properties. Spectral risk measures, introduced by Acerbi [2], and distortion risk measures, which developed from research on premium principles by Wang [56], are well-known families of risk functionals with these properties as they are coherent (in the sense of Artzner et al. [3]) and comonotone additive and law invariant, see [20, 49]. Arguably, the most relevant representative of these families is the AV@R, as it can be seen as the basic building block for spectral risk measures, see Kusuoka [30].

Nevertheless, well known risk functionals violate at least one of the mentioned properties: the variance (which has neither of these properties), the standard deviation (which is not comonotone additive), the Value-at-Risk (which is not subadditive) and expectiles (which are in general not comonotone additive), see Emmer et al. [15]. The questions arises what happens if one of the latter risk functionals is chosen in the portfolio selection problem (11) under complete dependence uncertainty. The following example addresses this question.

Example 1 Let some portfolio be composed of two assets. We represent the two assets by random variables U and W , where $W = 2V^2$ and U and V are both Uniform[0,1]. Thus, the second asset promises a higher return at a higher risk. For $0 \leq x \leq 1$, the portfolio with weights $1 - x$ resp. x has return

$$Y_x^C = (1 - x)U + x2V^2,$$

where C denotes the copula linking U and V . Assuming the copula C is unknown leads to the following portfolio selection problem:

$$\max_{x \in [0,1]} \min_{C \in \mathcal{C}} \mathbb{E}(Y_x^C) - \lambda \mathcal{R}(Y_x^C). \tag{12}$$

We analyze four different risk measures \mathcal{R} : i) the variance (Var), ii) the standard deviation (Std), iii) the Average-Value-at-Risk (AV@R) and iv) the Value-at-Risk (V@R). In case i) and ii), $\mathcal{R}(Y_x^C)$ is maximal if $\text{Cov}(U, V^2)$ is maximal. According to Theorem 5.25 in [36], this means that U and V^2 are comonotone, i.e., the worst case copula is given by the comonotonic copula M and we can set $V = U$. The same statement holds in case iii) where $\max_{C \in \mathcal{C}} \text{AV@R}(Y_x^C) = \text{AV@R}(Y_x^M)$. In case iv), it was mentioned above that $\max_{C \in \mathcal{C}} \text{V@R}(Y_x^C) = \text{V@R}(Y_x^{C_\alpha})$ where the worst case copula C_α is given in Eq. 9. Note

that C_α depends on the confidence level α but not on the portfolio weight x . Problem (12) can now be solved explicitly for the four different choices of \mathcal{R} :

- i) $\max_{x \in [0,1]} \mathbb{E}(Y_x) - \lambda \text{Var}(Y_x^M) = \max_{x \in [0,1]} \frac{1}{2} + \frac{x}{6} - \lambda \left(\frac{1}{12} + \frac{x}{6} + \frac{19x^2}{180} \right)$.
- ii) $\max_{x \in [0,1]} \mathbb{E}(Y_x) - \lambda \text{Std}(Y_x^M) = \max_{x \in [0,1]} \frac{1}{2} + \frac{x}{6} - \lambda \sqrt{\frac{1}{12} + \frac{x}{6} + \frac{19x^2}{180}}$.
- iii) $\max_{x \in [0,1]} \mathbb{E}(Y_x) - \lambda \text{AV@R}_\alpha(Y_x^M) = \max_{x \in [0,1]} \frac{1}{2} + \frac{x}{6} - \lambda \left(\frac{4\alpha^2 x + \alpha x + x + 3\alpha + 3}{6} \right)$.
- iv) $\max_{x \in [0,1]} \mathbb{E}(Y_x) - \lambda \text{V@R}_\alpha(Y_x^{C_\alpha}) = \max_{x \in [0,1]} \frac{1}{2} + \frac{x}{6} - \lambda \min(2x\alpha^2 + (1-x), \alpha(1-x) + 2x)$ for $\alpha \geq 0.5$.⁶

The optimal portfolio decomposition for the four different risk functionals is illustrated in Fig. 2 as a function of λ : For low risk aversion λ the portfolio is concentrated in second asset which promises a higher return, whereas in case of high risk aversion, the portfolio is concentrated in the less risky first asset. As indicated by Proposition 1, when using the AV@R as a risk measure, the optimal portfolio is concentrated for all λ (see the second panel from the bottom in Fig. 2). The panel at the bottom of Fig. 2 shows that the same also holds true when using the V@R. In case i) and ii) we observe that the portfolio is composed for certain values of risk aversion. For instance, in case ii) $\lambda = \sqrt{695/49} \approx 0.54$ implies that portfolio diversification into half/half is optimal.

This example shows us that independently of the chosen risk functionals, optimal portfolios under complete dependence uncertainty are typically concentrated. Only for specific values of the risk aversion coefficient λ it might happen that the portfolio is composed. For subadditive, comonotone additive and positively homogeneous risk functionals, composed portfolios can even be ruled out a priori, which constitutes another advantage of this class of risk measures. In fact, we can go even further and conjecture that for any risk functional the optimal portfolio under complete dependence uncertainty is never composed of more than two assets.

Summing up, portfolio concentration is a robust decision when facing complete dependence uncertainty and the choice of subadditive, comonotone additive and positively homogeneous risk functionals is particularly convenient in this context.

4.3 Partial Dependence Uncertainty

Let us now discuss problem (10) for $\epsilon > 0$. In particular, we assume that a reference copula C_0 is given and we consider an ambiguity set of copulas which are close to this reference copula. In the spirit of Section 3, we avoid all kinds of parametric assumptions and choose the empirical copula \hat{C}_n to be the reference copula C_0 . The empirical m -copula \hat{C}_n generated by n points is defined as

$$\hat{C}_n(u_1, \dots, u_m) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(U_{i,1} \leq u_1, \dots, U_{i,m} \leq u_m), \tag{13}$$

with $u_1, \dots, u_m \in [0, 1]$ and $U_{i,j} = n\hat{F}_j(\xi_{i,j})/(n + 1)$ for all $i = 1, \dots, n$ and $j = 1, \dots, m$, where \hat{F}_j denotes the empirical marginal distribution function of asset j , the matrix $(\xi_{i,j})_{i=1, \dots, n; j=1, \dots, m}$ contains n observations of the m asset returns and the scaling

⁶For the computation of $\text{V@R}_\alpha(Y_x^{C_\alpha})$ Proposition 1 in Denuit et al. [11] is very useful.

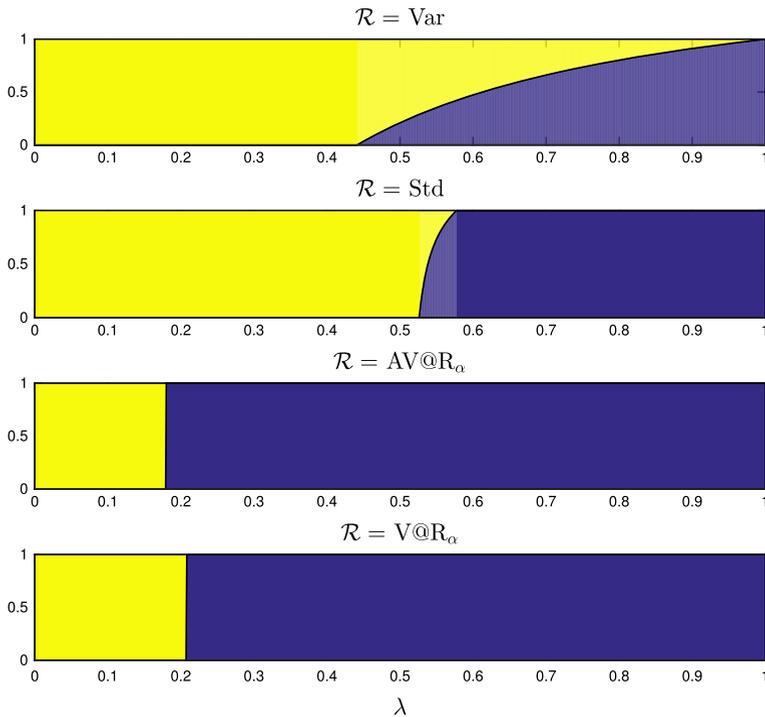


Fig. 2 The optimal portfolio decomposition of the portfolio selection problem under complete dependence uncertainty, which is given in Eq. 12, is shown for four different risk functionals \mathcal{R} as a function of the risk aversion $\lambda \in [0, 1]$ and for fixed $\alpha = 0.95$. For a fixed λ , the bright area corresponds to the fraction x which should be invested in the second asset whereas the dark area gives the optimal weight $1 - x$ of the first asset

factor $n/(n + 1)$ is only introduced to avoid potential problems at the boundary of $[0, 1]^m$, see Genest et al. [23]. Various conditions under which \hat{C}_n is a consistent estimator of the true underlying copula are given in Fermanian et al. [18]. In the formulation of problem (10) the ambiguity set $\mathcal{C}_\epsilon(C_0)$ contains infinitely many objects for all $\epsilon > 0$. This reminds us of the more general, yet similar problem (5) and the ambiguity set $\mathcal{B}_\kappa(P_0)$. As shown in [16, 42] and discussed in Section 3, when setting $P_0 = \hat{P}_n$, the worst case distribution also lives on n points. Thus, we conjecture that when setting $C_0 = \hat{C}_n$, the worst case copula also lives on n points. We therefore restrict our considerations to the smaller set $\mathcal{C}_{\epsilon,n}(\hat{C}_n) := \{C \in \mathcal{C}_n : d_W(C, \hat{C}_n) \leq \epsilon\}$ of all empirical m -copulas resulting from n observations which are close to \hat{C}_n . This restriction as well as choosing the AV@R as a risk functional, leads to the following problem:

$$\max_{x \in \mathbb{X}} \min_{C \in \mathcal{C}_{\epsilon,n}(\hat{C}_n)} \mathbb{E} \left(x^\top \xi \right) - \lambda \text{AV@R}_\alpha \left(-x^\top \xi^C \right). \tag{14}$$

In order to solve this problem, we have to understand the objects in the set $\mathcal{C}_{\epsilon,n}(\hat{C}_n)$. From definition (13) we know that any empirical copula $C \in \mathcal{C}_n$ is determined by its support points $U_{i,j} = n\hat{F}_j(\xi_{i,j})/(n + 1)$, where $i = 1, \dots, n$ and $j = 1, \dots, m$. The construction in (13) guarantees that the margins of the distribution induced by $(U_{i,j})$ are uniformly distributed on $[0, 1]$. Figure 3 shows an example of points $(U_{i,j})_{i=1,\dots,n; j=1,2}$ generating

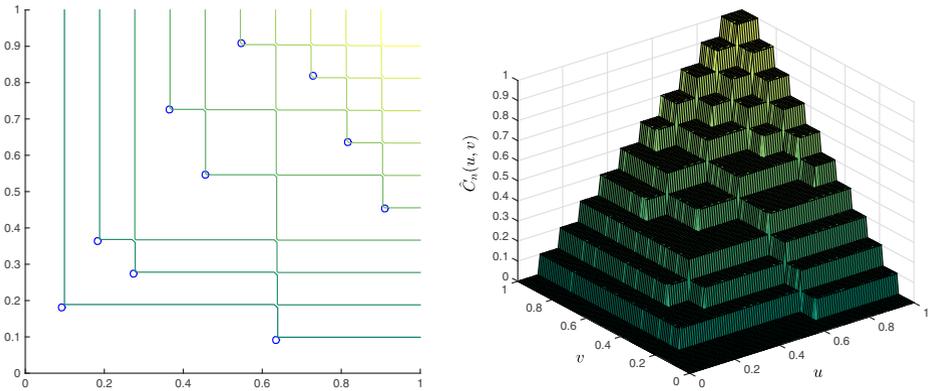


Fig. 3 The left panel shows a scatter plot of points $(U_{i,j})_{i=1,\dots,10,j=1,2}$ generating the empirical copula \hat{C}_n with $n = 10$ plotted in the right panel

an empirical bivariate copula. This graph illustrates the so-called *chess-tower property*, which is generally satisfied by all points $(U_{i,j})_{i=1,\dots,n;j=1,\dots,m}$ generating an empirical copula $C \in \mathcal{C}_n$: For fixed $j \in \{1, \dots, m\}$ and fixed $k \in \{1, \dots, n\}$ there exists a unique $i \in \{1, \dots, n\}$ such that $U_{i,j} = k/(n + 1)$, since $U_{i,j} = n\hat{F}_j(\xi_{i,j})/(n + 1) = R_{i,j}/(n + 1)$ where $(R_{i,j})$ denotes the ranks corresponding to the observations $(\xi_{i,j})$. This implies that there are $(m - 1)!n!$ elements in the set \mathcal{C}_n .

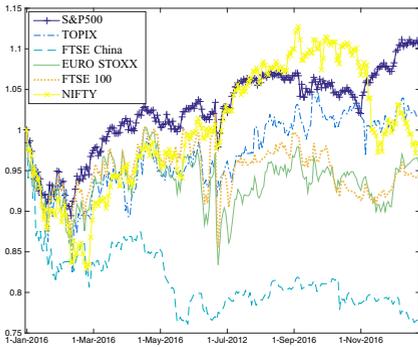
Hence, solving problem (14) for fixed portfolio weights x amounts to a combinatorial problem. Solving it by enumeration requires to compute the Wasserstein distance $d_W(C, \hat{C}_n)$ for all $(m - 1)!n!$ empirical copulas C in \mathcal{C}_n . Even for fixed portfolio weights x , this task is computationally infeasible for reasonable values of m and n . However, we can propose a heuristic, which allows us to rapidly compute a lower bound $C_L \in \mathcal{C}_{\epsilon,n}(\hat{C}_n)$ and an upper bound $C_U \in \mathcal{C}_n \setminus \mathcal{C}_{\epsilon,n}(\hat{C}_n)$ such that

$$AV@R_\alpha \left(-x^\top \xi^{C_L} \right) \leq \max_{C \in \mathcal{C}_{\epsilon,n}(\hat{C}_n)} AV@R_\alpha \left(-x^\top \xi^C \right) \leq AV@R_\alpha \left(-x^\top \xi^{C_U} \right), \quad (15)$$

for all $x \in \mathbb{X}$. An illustration of the “true” worst case copula and the corresponding bounds can be seen in Fig. 8 in the appendix. Let us sketch the main idea behind the proposed heuristic: First, note that the discrete counterpart M_n of the comonotonic copula M is defined by the points $U_{i,j} = i/(n + 1)$ for all i and j , i.e., M_n lives on the main diagonal of the unit cube. Therefore, our heuristic iteratively transports points to the main diagonal of the unit cube starting with the smallest points in the lower tail of the copula until the bound ϵ of the maximal feasible Wasserstein distance is exceeded. The last feasible discrete copula is set to be C_L and the first infeasible copula defines C_U . Appendix A describes this heuristic for the case $m = 2$. The procedure extends naturally to higher dimensions m . We omit a further discussion here.

5 Empirical Study: Concentration Versus Diversification

In the example studied in this section, the asset universe consists of the following six indices: S&P 500, TOPIX, FTSE China B35, EURO STOXX 50, FTSE 100 and NIFTY 500. The



	$E(\cdot)$ (10^{-4})	AV@R _{0.95} (\cdot) (10^{-2})
S&P500	3.67	1.94
TOPIX	0.48	3.40
FTSE China	-9.06	3.41
EURO STOXX	-0.86	3.19
FTSE 100	-1.57	3.44
NIFTY 500	0.30	2.75

Fig. 4 The graph on the left shows six equity indices in 2016, which have been normalized to 1 on January 1, 2016. The table on the right shows the expected value as well as the 95%-AV@R of the corresponding daily losses

markets corresponding to these indices are USA, Japan, China, Eurozone, UK and India. We look at daily returns from the year 2016 (i.e., January 1, 2016 to December 31, 2016) computed from the prices in USD⁷ of the indices which we normalized to one at the initial time. Figure 4 shows these normalized prices and summarizes the expected value and the AV@R of the corresponding returns. The data $(\xi_{i,j})_{i=1,\dots,n;j=1,\dots,m}$, consisting of $n = 260$ observations of daily returns of $m = 6$ assets, is retrieved from the Thomson Reuters datastream licensed at the University of Vienna.

The aim of this empirical study is to graphically compare the two distinct approaches in Sections 3 and 4: Firstly, we solve the portfolio selection problem under model ambiguity as given in (5) relying on the methods of Pflug et al. [42] and Esfahani and Kuhn [16]. Thus, we solve the LP (7) and (8) given ξ , α and λ for different values of κ . It should be mentioned once more that the two LPs yield indeed the same solution. Studying the optimal portfolio decomposition given by the portfolio weights x as the level of model ambiguity κ increases, we expect to find that portfolio diversification becomes optimal. Indeed, this is confirmed by the right plot in Fig. 5, in which we set $\alpha = 0.95$, $\lambda = 10$ and ξ as discussed in the previous paragraph and plot the optimal portfolio decomposition x for different values of κ . For small values of κ , which means that the investor trusts in the empirical distribution function computed from the observed returns, the right plot in Fig. 5 indicates that investing more than half of the capital into the S&P 500 is optimal. This fraction decreases to 1/6 as the level κ of model ambiguity increases and stands in contrast to the left plot of Fig. 5, which illustrates the optimal portfolio decomposition under dependence uncertainty. Here, we assume that the empirical marginal distribution functions indeed provide an accurate model for the marginal returns, whereas the dependence structure is ambiguous and only known to be within an ϵ -neighborhood of the empirical copula computed from the observations. In other words, we consider the maximin problem (14). If ϵ is close to zero, we are in the same situation as discussed before when $\kappa \approx 0$. For sufficiently large ϵ however, the worst case dependence structure is comonotonicity and portfolio concentration becomes optimal, as proven in Section 4.2. In our example we find that, as ϵ increases, more

⁷Note that the FX-risk is neglected in this study.

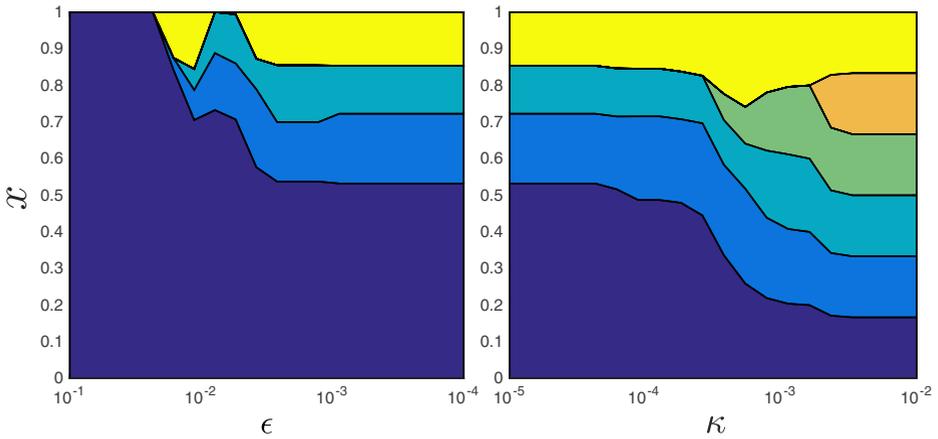


Fig. 5 In the graph on the right, the optimal portfolio decomposition is shown as a function of the level κ of model ambiguity. The graph on the left shows an approximation to the optimal portfolio decomposition as the degree ϵ of dependence uncertainty increases from right to left

and more capital should be invested into the S&P 500, which is in our example the best performing asset. It should be mentioned that we have used the heuristic algorithm presented in Section 4.3 and Appendix A to generate the left graph in Fig. 5. In fact, this graph displays the portfolio weights x which are optimal with respect to the feasible lower bound C_L in Eq. 15. The lower bound C_L corresponds to the upper bound shown in the left graph of Fig. 6, which plots the value of the objective function in Eq. 14 as a function of ϵ . The absolute difference between the upper and the lower bound of the objective function's value is 0.0051 ± 0.0037 when averaged over 20 different values of $\epsilon \in (10^{-4}, 10^{-1})$. Hence, we conclude that the results of the proposed heuristic algorithm need to be treated carefully when a fixed degree ϵ of dependence uncertainty is considered. Nevertheless, the heuristic allows us to determine the structure of the optimal portfolio decomposition as the degree of dependence uncertainty increases.

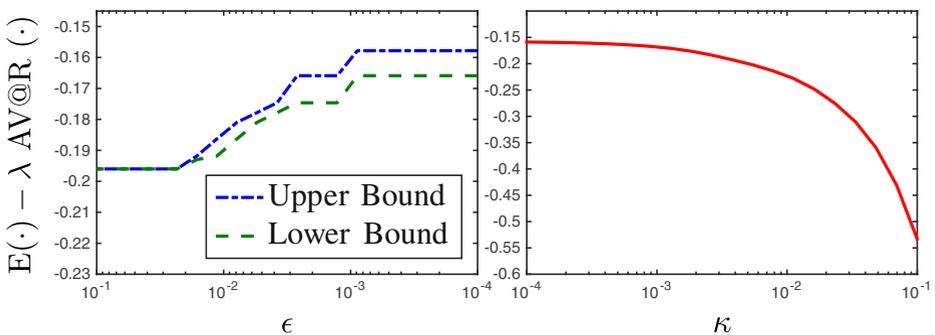


Fig. 6 The value of the objective function corresponding to the portfolio decomposition shown in Fig. 5 are plotted as a function of the level κ of model ambiguity as well as of the degree ϵ of dependence uncertainty

6 Conclusion

Figure 5 sums up the discussion of model ambiguity and dependence uncertainty in this paper. In the center of the Figure, we can see the optimal portfolio decomposition when we assume that the returns follow a known distribution P_0 . Aiming at an investment strategy which is robust with respect to this distributional assumption, we can go into two different directions. On the one hand, we can assume that the true distribution lies in a κ -neighborhood of the distribution P_0 and optimize the portfolio with respect to the worst case distribution within the κ -neighborhood. As shown in the right plot of Fig. 5, the higher the level of model ambiguity κ gets, the more diversified the portfolio becomes. On the other hand, we might only consider ambiguity with respect to the dependence structure of the distribution P_0 while fixing the marginal distributions. This leads to portfolio optimization under dependence uncertainty and is shown in the left plot of Fig. 5: The higher the degree of dependence uncertainty gets, the more concentrated the optimal portfolio becomes. Although the set-up of these two approaches is very similar as all parametric assumptions are avoided and the definition of the ambiguity set relies in both cases on the Wasserstein distance, the implication could not be more dissimilar.

Acknowledgments Open access funding provided by University of Vienna. The authors acknowledge support by the Vienna Science and Technologie Fund (WWTF) through project MA14-008.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

A Bivariate Example

Consider a portfolio consisting of $m = 2$ assets. Assume that it is known that the return of asset j is normally distributed with mean μ_j and variance σ_j^2 for $j = 1, 2$. Hence, instead of the empirical distribution function \hat{F}_n , we now use the normal cdf in the definition of \hat{C}_n . Moreover, we are given n observations of the two asset returns which allow us to compute the empirical estimate \hat{C}_n . In accordance with Section 4, we want to allow for the true dependence structure to be different from the one modeled by \hat{C}_n and make a robust investment decision in this respect. That is we want to solve problem (14). The following heuristic allows us to rapidly compute a lower bound C_L and an upper bound C_U as given in (15), which translates into bounds for problem (14):

In the following example, we set the parameter values $\mu_1 = 1\%$, $\mu_2 = 0.2\%$, $\sigma_1^2 = 2\%$, $\sigma_2^2 = 0.5\%$, $\alpha = 0.8$ and $\lambda = 10$ as well as U_0 resp. \hat{C}_n as shown in Fig. 3. Notice that the exact solution can be computed by enumeration, which is feasible only due to the small values of m and n . Figure 7 compares the optimal portfolio weights as well as the optimal objective value of the maximin problem (14) to the bounds computed with algorithm 1. As it can be seen the investment decision implied by the upper and lower bound might differ significantly from the optimal one for fixed ϵ . The reason for this observation is shown in Fig. 8, where we fix $\epsilon = 0.003$ and plot the worst case copula $C^* = \operatorname{argmax}_{C \in \mathcal{C}_{\epsilon,n}(\hat{C}_n)} AV @ R_{\alpha}(-x^T \xi^C)$ in comparison to the bounds C^L and C^U . Figure 8 shows that for this particular value of ϵ the worst case copula does not concentrate on the

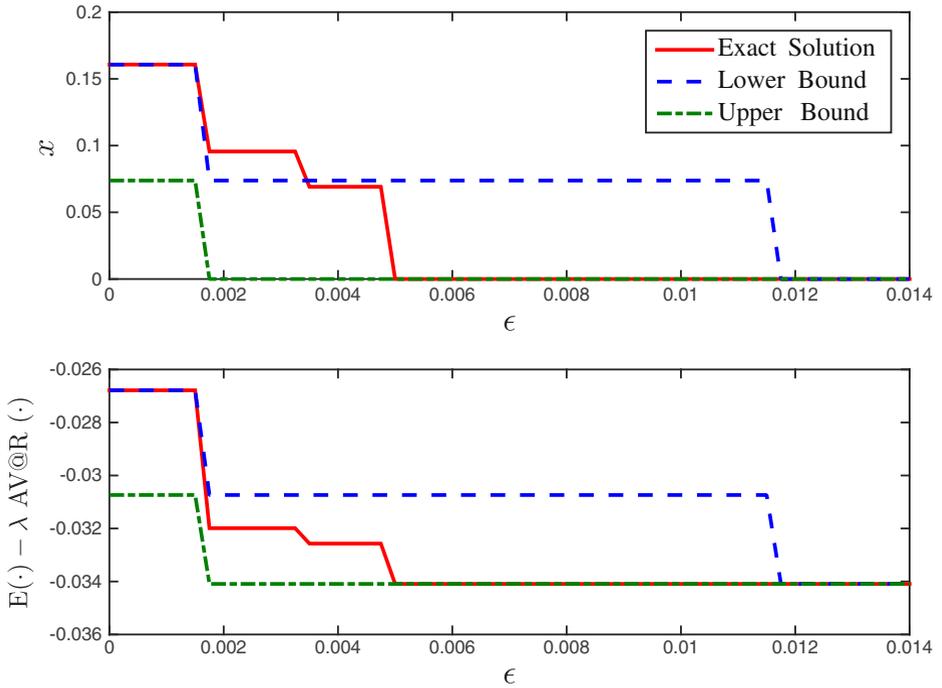


Fig. 7 These plots compare the performance of the proposed heuristic method to the exact solution. The upper panel shows the optimal weight of the first asset in the portfolio as a function of ϵ . The lower panel shows the corresponding value of the objective function (14) as a function of ϵ

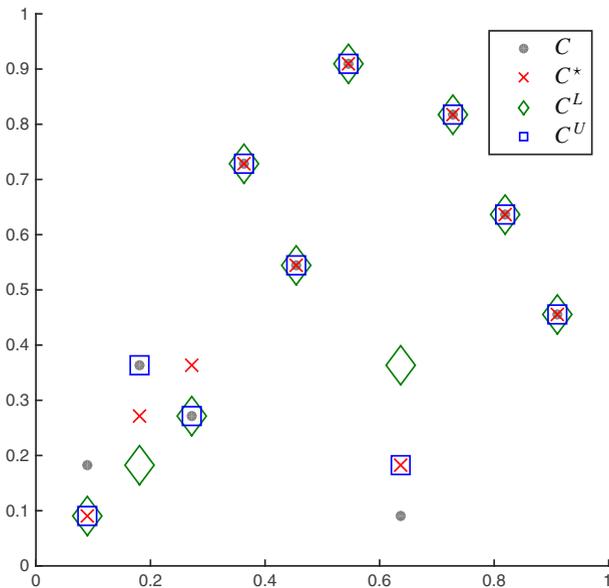


Fig. 8 The support points of the initial copula \hat{C}_n from Fig. 3 are compared with those of the modified copulas C_L and C_U , which are computed by Algorithm 1, and to those of the “true” worst case copula C^* for fixed $\epsilon = 0.003$

Algorithm 1: Heuristic for a bivariate empirical copula

INPUT:

U_0 ... $n \times 2$ matrix of support points $U_{i,j} = F_j(\xi_{i,j})$ for $i = 1, \dots, n, j = 1, 2$ of \hat{C}_n .
 ϵ ... bound for the Wasserstein distance $d_W(\cdot, \cdot)$.
 α ... confidence level of the AV@R.

OUTPUT:

U_L ... $n \times 2$ matrix of support points of the lower bound C_L .
 U_U ... $n \times 2$ matrix of support points of the upper bound C_U .

```

 $U_{NEW} = U_0$ ;
for  $l \leftarrow 1$  to  $\lceil (1 - \alpha)m \rceil$  do
    Find  $i_1$  and  $i_2$  s.t.  $U_{NEW}(i_1, 1) = U_{NEW}(i_2, 2) = l/(n + 1)$ ;
     $U_{NEW}(i_1, 1) = U_{NEW}(i_2, 1)$ ;
     $U_{NEW}(i_2, 1) = l/(n + 1)$ ;
    if  $d_W(U_{NEW}, U_0) > \epsilon$  then
         $U_L = U_{OLD}$ ;
         $U_U = U_{NEW}$ ;
        Break
    end
     $U_{OLD} = U_{NEW}$ ;
end
    
```

main diagonal in the lower tail, but C^L and C^U do so by construction. Hence, C_L and C_U differ significantly from C^* and so do the investment decision implied by these copulas. Nevertheless, it should be clear that the bounds become better for larger n .

B Minimaxy and Saddle Points

Let $f(x, y)$ be a function defined on $\mathbb{X} \times \mathbb{Y}$, where \mathbb{X} and \mathbb{Y} are compact topological spaces. We call the pair (x^+, y^+) a saddle point of f , if

$$f(x^+, y) \leq f(x^+, y^+) \leq f(x, y^+),$$

for $x \in \mathbb{X}, y \in \mathbb{Y}$. Assume that $x \mapsto f(x, y)$ is lower semicontinuous for every y and $y \mapsto f(x, y)$ is upper semicontinuous for every x . Let x^* be a minimizer of $x \mapsto \max_y f(x, y)$ and let y^* be a maximizer of $y \mapsto f(x^*, y)$. We say that (x^*, y^*) is a solution of the minimax problem. Notice that (x^*, y^*) does not need to be saddle point. Indeed, let $\mathbb{X} = \{x^*, x\}$ and $\mathbb{Y} = \{y^*, y\}$ and $f(x^*, y^*) = 2, f(x^*, y) = 1, f(x, y^*) = 1, f(x, y) = 3$. Then (x^*, y^*) is a solution of the minimax problem, but not a saddle point. On the other hand, if (x^+, y^+) is a saddle point, then it is a solution. Suppose the contrary, then (x^+, y^+) is a saddle point, but not a solution and (x^*, y^*) is a solution. Let $a = f(x^*, y^*), b = f(x^+, y^*), c = f(x^*, y^+), d = f(x^+, y^+)$. By the saddle point property of (x^+, y^+) ,

$$b \leq d \leq c,$$

and by the optimality of (x^*, y^*)

$$c \leq a < d.$$

This contradiction proves the assertion. We say that a Minimax Theorem holds, if

$$\min_{x \in \mathbb{X}} \max_{y \in \mathbb{Y}} f(x, y) = \max_{y \in \mathbb{Y}} \min_{x \in \mathbb{X}} f(x, y).$$

Notice that the existence of a saddle point (x^+, y^+) implies the validity of a Minimax Theorem: Notice first that

$$\min_{x \in \mathbb{X}} \max_{y \in \mathbb{Y}} f(x, y) \geq \max_{y \in \mathbb{Y}} \min_{x \in \mathbb{X}} f(x, y),$$

naturally holds in any case. Thus it is sufficient to prove the reverse inequality. By the saddle point property of (x^+, y^+)

$$\min_{x \in \mathbb{X}} \max_{y \in \mathbb{Y}} f(x, y) \leq \max_{y \in \mathbb{Y}} f(x^+, y) = f(x^+, y^+) = \min_{x \in \mathbb{X}} f(x, y^+) \leq \max_{y \in \mathbb{Y}} \min_{x \in \mathbb{X}} f(x, y).$$

However, a one-to-one relation between saddle points and solutions of the minimax problem can only be established under additional assumptions. We cite here Sion's famous Minimax Theorem.

Theorem 1 *Suppose that f is quasi convex-concave on a topological space $\mathbb{X} \times \mathbb{Y}$, such that $x \mapsto f(x, y)$ is lower semicontinuous for every $y \in \mathbb{Y}$ and $y \mapsto f(x, y)$ is upper semicontinuous for every $x \in \mathbb{X}$. If \mathbb{X} is compact, then*

$$\min_{x \in \mathbb{X}} \sup_{y \in \mathbb{Y}} f(x, y) = \sup_{y \in \mathbb{Y}} \min_{x \in \mathbb{X}} f(x, y).$$

If \mathbb{Y} is compact, then

$$\inf_{x \in \mathbb{X}} \max_{y \in \mathbb{Y}} f(x, y) = \max_{y \in \mathbb{Y}} \inf_{x \in \mathbb{X}} f(x, y).$$

If both are compact, then

$$\min_{x \in \mathbb{X}} \max_{y \in \mathbb{Y}} f(x, y) = \max_{y \in \mathbb{Y}} \min_{x \in \mathbb{X}} f(x, y).$$

Notice that under the assumptions of this Theorem, every solution must be a saddle point. If x^* is a solution of the minimax problem and if $y^* \in \operatorname{argmax}_y f(x^*, y)$, then

$$f(x^*, y) \leq f(x^*, y^*),$$

for all $y \in \mathbb{Y}$. Moreover, since $f(x^*, y^*) = \min_{x \in \mathbb{X}} \max_{y \in \mathbb{Y}} f(x, y) = \max_{y \in \mathbb{Y}} \min_{x \in \mathbb{X}} f(x, y)$,

$$f(x^*, y^*) = \min_{x \in \mathbb{X}} f(x, y^*) \leq f(x, y^*),$$

for all $x \in \mathbb{X}$, that is (x^*, y^*) is a saddle point. In our application, \mathbb{Y} is a subset of Borel probability measures on \mathbb{R}^m endowed with the weak topology. If \mathbb{Y} is a Wasserstein ball, then it is weakly compact (see below). \mathbb{X} is the simplex of nonnegative portfolio weights and also compact. Thus a saddlepoint theorem holds.

Proposition 2 *Let P_0 be a Borel probability measure on \mathbb{R}^m . Then closed Wasserstein balls with center P_0 are weakly compact.*

Proof Let $\mathcal{P} = \{P : d_W(P, P_0) \leq \eta\}$ be a Wasserstein ball with center P_0 . Recall Prohorov's Theorem: If a set of probability measures on a Polish space is weakly closed and tight, it is weakly compact. Closed Wasserstein balls are weakly closed and we have to prove the (uniform) tightness of such a ball, that is: for every ϵ there is a compact set K

such that $P(K) \geq 1 - 2\epsilon$ for all $P \in \mathcal{P}$. Since P_0 is Borel, there is a compact set K' such that $P_0(K') \geq 1 - \epsilon$. Let

$$h(x) = \begin{cases} 1 & x \in K' \\ \max\left(1 - \frac{\epsilon}{\eta} \cdot d(x, K'), 0\right) & x \notin K' \end{cases}.$$

Notice that h is Lipschitz with Lipschitz constant ϵ/η and therefore

$$\left| \int h(x) dP(x) - \int h(x) dP_0(x) \right| \leq \frac{\epsilon}{\eta} d(P, P_0) \leq \epsilon.$$

Let $K = \{x : d(x, K') \leq \epsilon/\eta\}$. Then K is compact and

$$P(K) \geq \int h(x) dP(x) \geq \int h(x) dP_0(x) - \left| \int h(x) dP(x) - \int h(x) dP_0(x) \right| \geq 1 - 2\epsilon,$$

for all $P \in \mathcal{P}$. □

Remark 1 With the same argument one sees that closed Wasserstein balls of copulas are weakly compact, since the set of all probability measures on $[0, 1]^n$ with uniform marginals is weakly closed and the intersection of a weakly closed and a weakly compact set is weakly compact.

References

1. Aas, K., Puccetti, G.: Bounds for total economic capital: the dnb case study. *Extremes* **17**(4), 693–715 (2014). <https://doi.org/10.1007/s10687-014-0202-0>
2. Acerbi, C.: Spectral measures of risk: a coherent representation of subjective risk aversion. *J. Bank. Financ.* **26**(7), 1505–1518 (2002). [https://doi.org/10.1016/S0378-4266\(02\)00281-9](https://doi.org/10.1016/S0378-4266(02)00281-9)
3. Artzner, P., Delbaen, F., Eber, J.M., Heath, D.: Coherent measures of risk. *Math. Financ.* **9**(3), 203–228 (1999). <https://doi.org/10.1111/1467-9965.00068>
4. Basak, S., Shapiro, A.: Value-at-risk-based risk management: optimal policies and asset prices. *Rev. Financ. Stud.* **14**(2), 371–405 (2001). <https://doi.org/10.1093/rfs/14.2.371>
5. Bertsimas, D., Gupta, V., Kallus, N.: Robust SAA. arXiv:1408.4445, p. 617 (2014)
6. Brandt, M.: Portfolio choice problems. *Handbook Financ. Economet.* **1**, 269–336 (2009). <https://doi.org/10.1016/B978-0-444-50897-3.50008-0>
7. Charnes, A., Cooper, W.W.: Chance-constrained programming. *Manag. Sci.* **6**(1), 73–79 (1959). <https://doi.org/10.1287/mnsc.6.1.73>
8. Consigli, G., Kuhn, D., Brandimarte, P.: *Optimal Financial Decision Making under Uncertainty*, pp. 255–290. Springer International Publishing, Cham (2017)
9. Dantzig, G.B.: Linear programming under uncertainty. *Manag. Sci.* **1**(3-4), 197–206 (1955). <https://doi.org/10.2307/2627159>
10. Delage, E., Ye, Y.: Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.* **58**(3), 595–612 (2010). <https://doi.org/10.1287/opre.1090.0741>
11. Denuit, M., Genest, C., Marceau: Stochastic bounds on sums of dependent risks. *Insurance: Math. Econ.* **25**(1), 85–104 (1999). [https://doi.org/10.1016/S0167-6687\(99\)00027-X](https://doi.org/10.1016/S0167-6687(99)00027-X)
12. Doan, X.V., Li, X., Natarajan, K.: Robustness to dependency in portfolio optimization using overlapping marginals. *Oper. Res.* **63**(6), 1468–1488 (2015). <https://doi.org/10.1287/opre.2015.1424>
13. Edirisinghe, N.C.P.: *Stochastic Programming Approximations Using Limited Moment Information, with Application to Asset Allocation*, pp. 97–138. Springer, New York (2011). <https://doi.org/10.1007/978-1-4419-1642-6-6>
14. Embrechts, P., Wang, B., Wang, R.: Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance Stochast.* **19**(4), 763–790 (2015). <https://doi.org/10.1007/s00780-015-0273-z>
15. Emmer, S., Kratz, M., Tasche, D.: What is the best risk measure in practice? a comparison of standard measures. *J. Risk* **18**(2), 31–60 (2015). <https://doi.org/10.2139/ssrn.2370378>
16. Esfahani, P.M., Kuhn, D. arXiv:1505.05116 (2015)

17. Fabozzi, F.J., Huang, D., Zhou, G.: Robust portfolios: contributions from operations research and finance. *Ann. Oper. Res.* **176**(1), 191–220 (2010). <https://doi.org/10.1007/s10479-009-0515-6>
18. Fermanian, J.D., Radulovic, D., Wegkamp, M.: Weak convergence of empirical copula processes. *Bernoulli* **10**(5), 847–860 (2004). <https://doi.org/10.3150/bj/1099579158>
19. Föllmer, H., Schied, A.: Convex measures of risk and trading constraints. *Finance Stochast.* **6**(4), 429–447 (2002). <https://doi.org/10.1007/s007800200072>
20. Föllmer, H., Schied, A.: *Stochastic finance: an introduction in discrete time* Walter de Gruyter (2011). <https://doi.org/10.1515/9783110463453>
21. Fournier, N., Guillin, A.: On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Relat. Fields* **162**, 707–738 (2015). <https://doi.org/10.1007/s00440-014-0583-7>
22. Gao, R., Kleywegt, A.J.: Distributionally robust stochastic optimization with Wasserstein distance. arXiv:1604.02199 (2016)
23. Genest, C., Rémillard, B., Beaudoin, D.: Goodness-of-fit tests for copulas: a review and a power study. *Insurance: Math. Econ.* **44**(2), 199–213 (2009). <https://doi.org/10.1016/j.insmatheco.2007.10.005>
24. Hanasusanto, G.A., Roitch, V., Kuhn, D., Wiesemann, W.: A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Math. Program.* **151**(1), 35–62 (2015). <https://doi.org/10.1007/s10107-015-0896-z>
25. Hansen, L.P., Sargent, T.J.: Robust control and model uncertainty. *Am. Econ. Rev.* **91**(2), 60–66 (2001). <https://doi.org/10.2307/2677734>
26. Ji, R., Lejeune, M.: Data-driven optimization of ambiguous reward-risk ratio measures. Available at SSRN 2707122 (2015)
27. Kakouris, I., Rustem, B.: Robust portfolio optimization with copulas. *Eur. J. Oper. Res.* **235**(1), 28–37 (2014). <https://doi.org/10.1016/j.ejor.2013.12.022>
28. Kim, J.H., Kim, W.C., Fabozzi, F.J.: Recent developments in robust portfolios with a worst-case approach. *J. Optim. Theory Appl.* **161**(1), 103–121 (2014). <https://doi.org/10.1007/s10957-013-0329-1>
29. Konno, H., Yamazaki, H.: Mean-absolute deviation portfolio optimization model and its applications to tokyo stock market. *Manag. Sci.* **37**(5), 519–531 (1991). <https://doi.org/10.1287/mnsc.37.5.519>
30. Kusuoka, S.: On law invariant coherent risk measures *Advances in Mathematical Economics*, pp. 83–95. Springer (2001). <https://doi.org/10.1007/978-4-431-67891-5>
31. Liu, Y., Meskarian, R., Xu, H.: Distributionally robust reward-risk ratio optimization with moment constraints. *SIAM J. Optim.* **27**(2), 957–985 (2017). <https://doi.org/10.1137/16M106114X>
32. Lux, T., Papapantoleon, A.: Model-free bounds on value-at-risk using partial dependence information. arXiv:1610.09734 (2016)
33. Makarov, G.: Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed. *Theory Probab. Appl.* **26**(4), 803–806 (1982). <https://doi.org/10.1137/1126086>
34. Markowitz, H.: Portfolio selection. *J. Financ.* **7**(1), 77–91 (1952). <https://doi.org/10.1111/j.1540-6261.1952.tb01525.x>
35. Markowitz, H.: *Portfolio Selection: Efficient Diversification of Investments*. Yale University Press, New Haven (1959). <https://doi.org/10.2307/2975974>
36. McNeil, A.J., Frey, R., Embrechts, P.: *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, Princeton (2015)
37. Meskarian, R., Xu, H., Fliege, J.: Numerical methods for stochastic programs with second order dominance constraints with applications to portfolio optimization. *Eur. J. Oper. Res.* **216**(2), 376–385 (2012). <https://doi.org/10.1016/j.ejor.2011.07.044>
38. Nelsen, R.B.: *An Introduction to Copulas*. Springer Science & Business Media, Berlin (2007). <https://doi.org/10.1007/978-1-4757-3076-0>
39. Pflug, G.C.: Some remarks on the value-at-risk and the conditional value-at-risk. In: *Probabilistic Constrained Optimization*, pp. 272–281. Springer, US (2000). <https://doi.org/10.1007/978-1-4757-3150-7>
40. Pflug, G.C., Pichler, A.: Approximations for Probability Distributions and Stochastic Optimization Problems. In: Consigli, G., Dempster, M., Bertocchi, M. (eds.) *Springer Handbook on Stochastic Optimization Methods in Finance and Energy*, International Series in OR and Management Science, vol. 163, pp. 343–387. Springer (2011). <https://doi.org/10.1007/978-1-4419-9586-5>
41. Pflug, G.C., Pichler, A.: *Multistage Stochastic Optimization*. Springer, Berlin (2014)
42. Pflug, G.C., Pichler, A., Wozabal, D.: The 1/n investment strategy is optimal under high model ambiguity. *J. Bank. Financ.* **36**(2), 410–417 (2012). <https://doi.org/10.1016/j.jbankfin.2011.07.018>
43. Pflug, G.C., Römisch, W.: *Modeling, measuring and managing risk*, vol. 190. World Scientific, Singapore (2007). <https://doi.org/10.1142/6478>
44. Pflug, G.C., Wozabal, D.: Ambiguity in portfolio selection. *Quant. Finan.* **7**(4), 435–442 (2007). <https://doi.org/10.1080/14697680701455410>

45. Puccetti, G., Rüschendorf, L.: Computation of sharp bounds on the distribution of a function of dependent risks. *J. Comput. Appl. Math.* **236**(7), 1833–1840 (2012). <https://doi.org/10.1016/j.cam.2011.10.015>
46. Puccetti, G., Wang, R.: Extremal dependence concepts. *Stat. Sci.* **30**(4), 485–517 (2015). <https://doi.org/10.1214/15-STS525>
47. Rockafellar, R.T., Uryasev, S.: Optimization of conditional value-at-risk. *J. Risk* **2**, 21–42 (2000). <https://doi.org/10.21314/JOR.2000.038>
48. Rüschendorf, L.: Random variables with maximum sums. *Adv. Appl. Probab.* **14**(3), 623–632 (1982). <https://doi.org/10.2307/1426677>
49. Rüschendorf, L.: *Mathematical Risk Analysis*. Springer, Berlin (2013). <https://doi.org/10.1007/978-3-642-33590-7>
50. Rüschendorf, L.: Risk Bounds and Partial Dependence Information. Preprint, University of Freiburg (2016)
51. Shorack, G.R., Wellner, J.A.: *Empirical processes with applications to statistics*, vol. 59 Siam (2009). <https://doi.org/10.1137/1.9780898719017>
52. Steinbach, M.C.: Markowitz revisited: mean-variance models in financial portfolio analysis. *SIAM Rev.* **43**(1), 31–85 (2001). <https://doi.org/10.1137/S0036144500376650>
53. Tintner, G.: The theory of choice under subjective risk and uncertainty. *Econometrica: J. Econ. Soc.* **9**(3/4), 298–304 (1941). <https://doi.org/10.2307/1907198>
54. Wang, B., Wang, R.: The complete mixability and convex minimization problems with monotone marginal densities. *J. Multivar. Anal.* **102**(10), 1344–1360 (2011). <https://doi.org/10.1016/j.jmva.2011.05.002>
55. Wang, B., Wang, R.: Extreme negative dependence and risk aggregation. *J. Multivar. Anal.* **136**, 12–25 (2015). <https://doi.org/10.1016/j.jmva.2015.01.006>
56. Wang, S.: Premium calculation by transforming the layer premium density. *ASTIN Bulletin: J. IAA* **26**(1), 71–92 (1996). <https://doi.org/10.2143/AST.26.1.563234>
57. Wets, R.: Programming under uncertainty: the complete problem. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **4**(4), 316–339 (1966). <https://doi.org/10.1007/bf00539117>
58. Wozabal, D.: A framework for optimization under ambiguity. *Ann. Oper. Res.* **193**(1), 21–47 (2012). <https://doi.org/10.1007/s10479-010-0812-0>
59. Xu, H., Zhang, D.: Monte carlo methods for mean-risk optimization and portfolio selection. *Comput. Manag. Sci.* **9**(1), 3–29 (2012)