Ambiguity in portfolio selection

Georg Pflug* David Wozabal

Abstract

In this paper, we consider the problem of finding optimal portfolios in cases when the underlying probability model is not perfectly known. For the sake of robustness, a maximin approach is applied which uses a "confidence set" for the probability distribution. The approach shows the tradeoff between return, risk and robustness in view of the model ambiguity. As a consequence, a monetary value of information in the model can be determined.

1 Introduction: The ambiguity problem

The decision about optimal composition of a portfolio is a complex process, not just a single optimization task. It comprises of the selection of a statistical model, the collection of data, the estimation of the model in a parametric, semi-parametric or nonparametric way, the choice of an appropriate optimization criterion and finally the numerical solution of an optimization problem. It is well understood that the precision of the final decision depends on the quality of the described complete chain of subtasks.

In his 1921 book Knight [1921], the American economist Frank Knight made a famous distinction between "risk" and "uncertainty". In Knight's view, "risk" refers to situations where the decision-maker can assign mathematical probabilities to the randomness, which he is faced with. In contrast, Knight's "uncertainty" refers to situations when this randomness cannot be expressed in terms of specific mathematical probabilities.

Since the days of Knight, the terms have changed. As introduced by Ellsberg [1961], we refer today to the *ambiguity problem* if the probability model is unknown and to the *uncertainty problem*, if the model is known, but the realizations of the random variables are unknown. While large classes of portfolio optimization problems under uncertainty have been successfully solved (see e.g. the surveys by Yu et al. [2003] or Ziemba and Vickson [1973] and references therein), there is a common observation that these solutions lack of stability with respect to the chosen parameters (see e.g. Klein and Bawa [1976], Chopra and Ziemba [1993]).

*Department of Statistics and Decision Support Systems, University of Vienna, Universitätsstraße 5, 1090 Wien-Vienna, Austria, Contact e-mail: georg.pflug@univie.ac.at

The aim of this paper is to discuss an approach that explicitly takes into account the ambiguity in choosing the probability model and therefore is robust in following sense: The selected portfolio is slightly suboptimal for the given basic probability model, but performs also well under neighboring models. In contrast, non-robust portfolio decisions show a dramatic drop in performance, when deviating from the underlying model.

The organization of this paper is follows. In section 2, the decision model is formulated. The successive convex program as a solution technique is presented in section 3. In sections 4 and 5, the choice of ambiguity sets is discussed. Section 6 illustrates the approach by an Example.

2 Problem formulation

As introduced in the seminal work of Markowitz (see Markowitz [1959]), the basic portfolio selection problem in this paper is to minimize the risk under a constraint for the expected return or - equivalently - maximize the expected return under a risk constraint. We follow the latter approach here.

Let $Y_x = \sum_{m=1}^{M} x_i \xi_i$ be the random return of a portfolio consisting of M assets with individual returns ξ_i and portfolio weights x_i. The chosen or estimated probability model P determines the distribution of the return vector $(\xi_1, \ldots, \xi_M)$ on $\mathbb{R}^M$.

Once a probability model is specified, the expected return is well defined. There are however several ways of quantifying the risk. In this paper, risk is measured as negative acceptability. Acceptability in turn is measured by acceptability functionals $\mathcal{A}$ defined on the random asset returns.

Let $(\Omega, \mathcal{F})$ be a measure space, $\mathcal{X}$ a linear space of $\mathcal{F}$ measurable functions, $X : \Omega \to \mathbb{R}$, then a coherent acceptability functional $\mathcal{A}$ from $\mathcal{X}$ to $\bar{\mathbb{R}}$ is required to fulfill the following axioms (see Pflug [2006] for a discussion of coherent acceptability functionals)

[A1] Concavity: $\forall X, Y \in \mathcal{X}, \lambda \in [0, 1]$

$$\mathcal{A}(\lambda X + (1 - \lambda)Y) \geq \lambda \mathcal{A}(X) + (1 - \lambda)\mathcal{A}(Y)$$

[A2] Monotonicity: If $X \leq Y$ then $\mathcal{A}(X) \leq \mathcal{A}(Y)$

[A3] Translation equivariance: $\mathcal{A}(X + a) = \mathcal{A}(X) + a$

[A4] Positive Homogeneity: If $\lambda > 0$ then $\mathcal{A}(\lambda X) = \lambda \mathcal{A}(X)$

We interpret the negative acceptability as risk, i.e. a risk functional is of the form $\mathcal{R} = -\mathcal{A}$. The resulting risk functionals have properties coinciding with the convex risk functionals commonly used in the literature (see for example Delbaen [2002] or Föllmer and Schied [2002]). The goal in this paper is to find decisions with high acceptability and thereby low risk. Note that since acceptability functionals and risk functionals differ only by their sign, they have the same level sets, which are called acceptance sets. For a given return Y, the acceptability value $\mathcal{A}(Y)$ indicates the maximal shift of the distribution in the negative direction, which still keeps the return acceptable. Note that since the acceptability functions are concave, the level sets are convex sets. We will further assume, that the acceptability functionals $\mathcal{A}$ are continuous, which assures that the respective level sets are closed.

Examples for such acceptability functionals $\mathcal{A}$ are for instance

- the average-value-at risk (conditional value-at-risk, expected shortfall)

$$\mathcal{A}(Y) = \mathbb{AV@R}_\alpha(Y) = \frac{1}{\alpha}\int_0^\alpha G_Y^{-1}(p)\,dp, \tag{1}$$

  where $G_Y^{-1}(p) = \inf\{v : P\{Y \le v\} \ge p\}$ is the quantile function of Y. An alternative representation is

$$\mathbb{AV@R}_\alpha(Y) = \max\{a - \frac{1}{\alpha}\mathbb{E}([Y-a]^-) : a \in \mathbb{R}\}.$$

  (see Rockafellar and Uryasev [2000]). A dual representation of $\mathbb{AV@R}_\alpha(Y)$ is

$$\mathbb{AV@R}_\alpha(Y) = \min\{\mathbb{E}(Y\,Z) : \mathbb{E}Z = 1, 0 \le Z \le 1/\alpha\}.$$

- also utility functionals of the following types

  1. $\mathcal{A}(Y) = \mathbb{E}(Y) - \lambda\mathbb{V}\mathrm{ar}(Y)$
  2. $\mathbb{E}(Y) - \lambda\mathbb{S}\mathrm{td}(Y)$
  3. $\mathcal{A}(Y) = \mathbb{E}(Y) - \mathbb{V}\mathrm{ar}([Y - \mathbb{E}Y]^-)$
  4. $\mathbb{E}(Y) - \lambda\mathbb{S}\mathrm{td}([Y - \mathbb{E}Y]^-)$

  with $0 < \lambda \le 1$ respectively, fall in the category of acceptability functionals (see Tokat et al. [2003] for a more detailed discussion of functionals of this form).

To translate acceptability back into risk, one may use $-\mathcal{A}(Y)$, or, as we will do in the numerical examples, $1 - \mathcal{A}(Y)$ as risk functional.

Having identified the criteria for the optimization and the risk bound, we can now write the problem as an optimization problem under uncertainty using the basic probability model $\hat{P}$ which can be regarded as the "best guess" for the real model. In most cases $\hat{P}$ will be determined by empirical data, but also other ways of obtaining $\hat{P}$ like the incorporation of expert opinion may be considered (see for example Clemen and Winkler [1999]).

As mentioned before the portfolio selection model we consider is a simple Markowitz type of model without shortselling. The asset returns are modeled to be uncertain. Assuming there is no ambiguity about the statistical model of the returns the model reads

$$\left\| \begin{array}{l} \text{Maximize (in } x\text{)}: \ \mathbb{E}_{\hat{P}}(Y_x) \\ \text{subject to} \\ \mathcal{A}_{\hat{P}}(Y_x) \geq q \\ x^\top \mathbb{1} = 1 \\ x \geq 0 \end{array} \right. \tag{2}$$

Note that the restriction to positive portfolio weights is chosen just for simplicity of exposition. It would pose neither theoretical nor computational problems to allow short selling in this setting as long as the feasible set of asset compositions remains bounded. In fact it might be rewarding to study the effects of ambiguity in the more risky setting where shortselling is permitted.

Let $\hat{x}^*$ be the optimal portfolio composition found by solving (2). The robustness of this solution is often checked by *stress testing*. A stress test consists in finding an alternative probability model P and calculating $\mathbb{E}_P(Y_{\hat{x}^*})$ as well as $\mathcal{A}_P(Y_{\hat{x}^*})$ to judge the change in the return dimension as well as in the risk dimension under model variation. While stress testing helps in assessing the robustness of a given portfolio selection, it does not help to find a robust portfolio.

For the latter goal, one has to replace the basic model (2) by its ambiguity extension. To this end, let $\mathcal{P}$ be an ambiguity set, i.e. the set of probability models, to which the modeler is indifferent. The portfolio selection model under $\mathcal{P}$-*ambiguity* is of maximin type and reads

$$\left\| \begin{array}{l} \text{Maximize (in } x\text{)}: \ \min\{\mathbb{E}_P(Y_x) : P \in \mathcal{P}\} \\ \text{subject to} \\ \mathcal{A}_P(Y_x) \geq q \ \text{ for all } P \in \mathcal{P} \\ x^\top \mathbb{1} = 1 \\ x \geq 0 \end{array} \right. \tag{3}$$

Let us comment on the structure of the problem (3). It is a combination of a robust and a stochastic problem (see Zackova [1966] for the earliest occurrence of problems of the form (3)). Recall the definition of a robust optimization problem (see Ben-Tal and Nemirovski [2002]):

If a deterministic optimization problem

$$\left\| \begin{array}{l} \text{Maximize (in } x\text{)}: \ f(x, \zeta) \\ \text{subject to} \\ f_i(x, \zeta) \leq 0, i = 1, \ldots, k \end{array} \right. \tag{4}$$

contains some parameters ζ and the decision maker only knows some range Z of these parameters, he/she may use the robust version of (4), namely

$$\left\| \begin{array}{l} \text{Maximize } \min\{f(x,\zeta) : \zeta \in Z\} \\ \text{subject to } f_i(x,\zeta) \le 0, i = 1,\dots,k \quad \text{for all } \zeta \in Z \end{array} \right. \tag{5}$$

While in stochastic optimization one has at least a distributional information about the unknown parameters, the only information one has in robust optimization is a given set of parameters. Thus one may say that stochastic programs look at the *average* situation, while robust programs look at the *worst-case* situation.

The proposed portfolio selection under ambiguity contains both aspects: While assuming that the realizations of the return vectors come from some probability distribution, we allow on the other hand to vary this distribution within a certain set $\mathcal{P}$ without further structuring it. Would we impose a probability distribution on this set of probabilities (called a *prior distribution*), we would still solve an uncertainty problem, but of Bayesian type. In our approach we do not specify a prior and our problem (3) has the structure of a robust-stochastic problem.

3 Solution Techniques

Introducing the set of constraints for the vector of portfolio weights x

$$\mathbb{X} = \{x : x^\top \mathbb{1} = 1, x \ge 0, \mathcal{A}_P(Y_x) \ge q \text{ for all } P \in \mathcal{P}\},$$

the ambiguity problem (3) reads

$$\max_{x\in\mathbb{X}} \min_{P\in\mathcal{P}} \mathbb{E}_P[Y_x]. \tag{6}$$

By continuity and concavity of $\mathcal{A}$, $\mathbb{X}$ is a compact convex set. Moreover, $(P,x) \mapsto \mathbb{E}_P[Y_x]$ is bilinear in P and x and hence convex-concave. Therefore $x^* \in \mathbb{X}$ is a solution of (6) if and only if there is a $P^* \in \mathcal{P}$ such that (P^*, x^*) is a saddle point, i.e.

$$\mathbb{E}_{P^*}[Y_x] \le \mathbb{E}_{P^*}[Y_{x^*}] \le \mathbb{E}_P[Y_{x^*}] \tag{7}$$

for all $(P,x) \in \mathcal{P} \times \mathbb{X}$ (see Rockafellar [1997]).

Since there are infinitely many constraints present, the problem (6) is a semi-infinite program. There are several solution methods for such problems. Direct methods use gradients to find the saddle point (Rockafellar [1976], Nemirovskii and Yudin [1978]). In Shapiro and Ahmed [2004] it is proposed to dualize the inner minimization problem in order to get a pure optimization problem: In particular they suppose that the set $\mathcal{P}$ is of the form

$$\mathcal{P} = \{P : \int \phi_i(u)\, dP(u) \le b_i; i = 1,\dots,k; P_1 \prec P \prec P_2\}.$$

where P_i are some measures, the ϕ_i are P integrable and $b_i \in \mathbb{R}$. Then using the dual representation

$$\min\{\mathbb{E}_P[Y_x] : P \in \mathcal{P}\} = \max\{\int [\sum_i \lambda_i \phi_i(u)]^+ \, dP_1(u) + \int [\sum_i \lambda_i \phi_i(u)]^- \, dP_2(u)\}.$$

the maximin problem is transformed into a pure maximization problem.

In this paper, we propose a successive convex programming (SCP) solution method, which uses a finitely generated inner approximation of the ambiguity set $\mathcal{P}$. To be more precise, we approximate the infinitely many constraints $\mathcal{A}_P(Y_x) \geq q$ for all $P \in \mathcal{P}$ by finitely many ones. One starts with no risk constraints. In every new step, two new probabilities enter the set of constraints. These new probabilities are chosen as current worst case probabilities and this makes the algorithm work.

In particular, the successive SCP algorithm proceeds as follows:

1. Set $n = 0$ and $\mathcal{P}_0 = \{\hat{P}\}$ with $\hat{P} \in \mathcal{P}$.

2. Solve the outer problem

$$\left\| \begin{array}{l} \text{Maximize (in } x, t) : \; t \\ \text{subject to} \\ t \leq \mathbb{E}_P(Y_x) \text{ for all } P \in \mathcal{P}_n \\ \mathcal{A}_P(Y_x) \geq q \; \text{ for all } P \in \mathcal{P}_n \\ x^\top \mathbb{1} = 1 \\ x \geq 0 \end{array} \right. \qquad (8)$$

   and call the solution (x_n, t_n).

3. Solve first inner problem

$$\left\| \begin{array}{l} \text{Minimize (in } P) : \; \mathbb{E}_P(Y_{x_n}) \\ \text{subject to} \\ P \in \mathcal{P} \end{array} \right. \qquad (9)$$

   and call the solution $P_n^{(1)}$.

4. Solve the second inner problem

$$\left\| \begin{array}{l} \text{Minimize (in } P) : \; \mathcal{A}_P(Y_{x_n}) \\ \text{subject to} \\ P \in \mathcal{P} \end{array} \right. \qquad (10)$$

   call the solution $P_n^{(2)}$ and let $\mathcal{P}_{n+1} = \mathcal{P}_n \cup \{P_n^{(1)}\} \cup \{P_n^{(2)}\}$.

5. If

   (a) $\mathcal{P}_{n+1} = \mathcal{P}_n$ or

(b) the optimal value of (9) equals t_n and the solution of (10) is equal to $\min_{P \in \mathcal{P}_n} \mathcal{A}_P(Y_{x_n})$

then a saddle point is found and the algorithm stops. Otherwise set $n := n + 1$ and goto 2.

The two inner problems (9) and (10) may yield non unique optimizers. In this case we simply choose one of the optimizing worst case measures and proceed with the algorithm.

To see that the second stopping criterion actually yields a saddle point, consider a situation where the condition 5(2) is fulfilled in the $n - th$ run. Note that since (10) does not yield a lower acceptability, then the minimum of the measures in $\mathcal{P}_n$ the point x_n is feasible for the original problem. Since (9) gives the optimal value $\mathbb{E}_{P'}(Y_{x_n})$, there is no measure P giving smaller expectation than P'. Furthermore it is clear from the optimality in (8), that there can be no x, such that $\mathbb{E}_{P'}(Y_x)$. This establishes that (x_n, P') is indeed a saddle point of the problem.

Note that if the measures $P_n^{(1)}$ or $P_n^{(2)}$ are in the convex hull of the measures in $\mathcal{P}_n$ they don't have to be added to the set $\mathcal{P}_{n+1}$ (since the functions E_p and $\mathcal{A}$ are concave and therefore take their minima on the extreme points of convex sets).[1] Whether or not the a discrete probability measure on $\mathbb{R}^n$ is a convex combination of other such measures can be easily checked by linear programming techniques.

Notice that the set t_n is a decreasing sequence of numbers and $\mathcal{P}_n$ is an increasing sequence of sets. The convergence of this algorithm is stated below. Since one cannot exclude that there are several saddle points (in this case the set of saddle points is closed and convex), only a weak limit result is available in general.

Proposition. Assume that $\mathcal{P}$ is compact and convex and that $(P, x) \mapsto \mathbb{E}_P[Y_x]$ as well as $(P, x) \mapsto \mathcal{A}_P[Y_x]$ are jointly continuous. Then every cluster point of (x_n) is a solution of (6). If the saddle point is unique, then the algorithm converges to the optimal solution.

Proof. Since both $\mathbb{X}$ and $\mathcal{P}$ are compact sets, the sequence x_n has a cluster point x^*. W.l.o.g we may in fact assume that this is a limit point. Let

$$\mathbb{X}_n = \{x : x^\top \mathbb{1} = 1, x \geq 0, \mathcal{A}_P(Y_x) \geq q \text{ for all } P \in \mathcal{P}_n)\}.$$

Then $\mathbb{X}_n$ is a decreasing sequence of compact convex sets, which all contain $\mathbb{X}$. Let $\mathbb{X}^+ = \bigcap_n \mathbb{X}_n$. On the other hand, $\mathcal{P}_n$ is an increasing sequence of sets with upper limit $\mathcal{P}^+$, which is the closure of $\bigcup_n \mathcal{P}_n$.

By construction, x_n is a solution of

$$\max_{x \in \mathbb{X}_n} \min_{P \in \mathcal{P}_n} \mathbb{E}_P[Y_x] \tag{11}$$

[1]We are indebted to a referee for pointing this out to us and also led us to discover the second condition for having found an saddle point in step 5 of the algorithm.

i.e. there is a probability P_n^* such that (P_n^*, x_n) is a saddle point. W.l.o.g. we may assume that the sequence (P_n^*) has a limit P^*. Therefore, (P^*, x^*) is a saddle point for

$$\max_{x \in \mathbb{X}^+} \min_{P \in \mathcal{P}^+} \mathbb{E}_P[Y_x]. \tag{12}$$

Since $\min_{P \in \mathcal{P}} \mathbb{E}_P[Y_{x_n}] = \min_{P \in \mathcal{P}_{n+1}} \mathbb{E}_P[Y_{x_n}]$ one sees by continuity that (x^*, P^*) is a saddle point for

$$\max_{x \in \mathbb{X}^+} \min_{P \in \mathcal{P}} \mathbb{E}_P[Y_x]. \tag{13}$$

Finally, notice that $x^* \in \mathbb{X}$. If not, $\inf_{P \in \mathcal{P}} \mathcal{A}_P(Y_{x^*}) < q$. But then there is an x_n such that $\inf_{P \in \mathcal{P}} \mathcal{A}_P(Y_{x_n}) < q$ and in the next step this x_n together with an open neighborhood will be excluded from $\mathbb{X}_{n+1}$, a contradiction.

Therefore (P^*, x^*) is a saddle point for

$$\max_{x \in \mathbb{X}} \min_{P \in \mathcal{P}} \mathbb{E}_P[Y_x], \tag{14}$$

i.e. x^* is a solution of the original problem. □

4 Ambiguity Sets

Typically, ambiguity sets are in some sense neighborhoods of basic models. Basic models are found by estimation from historic data, consistency considerations as the no-arbitrage rule, expert choice or a combination thereof. In all these cases, the found basic model is the most likely one, but model ambiguity is present. To express this ambiguity, one may allow some variations of the basic models in such a way that they differ in distance from the basic model not more than some ϵ.

In particular, consider ambiguity sets of the form

$$\mathcal{P} = \{P : d(P, \hat{P}) \le \epsilon\} \tag{15}$$

where d is some distance for probability measures.

The choice of the distance d is crucial for the final result. In this paper we use the Kantorovich distance (also called the $\mathcal{L}_1$??? distance - see Vallander [1973] and Dall'Aglio [1972]) defined as

$$\begin{aligned} d(P_1, P_2) \;=\; & \sup\{\int f(u)\, dP_1(u) - \int f(u)\, dP_2(u) : \text{ where } f \text{ has Lipschitz} \\ & \text{constant 1, i.e. } f(u) - f(v) \le \|u - v\|_1 \text{ for all } u, v\}. \end{aligned}$$

Here $\|u - v\|_1 = \sum_i |u_i - v_i|$. To ensure that the Kantorovich distance is finite, we restrict ourselves to the space of measures with finite first absolute moment, i.e. $\mathcal{P} = \{P : \int_{\mathbb{R}^n} ||x|| dx < \infty\}$.

The choice is motivated by the fact that the expected return has Lipschitz constant 1 under this distance, i.e.

$$|\mathbb{E}_{P_1}(Y_x) - \mathbb{E}_{P_2}(Y_x)| \leq d(P_1, P_2)$$

for all portfolios x. Hence, the distance of probability models provides a bound for the difference in expectations and therefore a bound in the optimal values of the considered problems.

Furthermore if the chosen acceptability functional is the $\mathbb{A}$V@R, then one has also Lipschitz continuity of the risk functional with respect to the Kantorovich distance, i.e.

$$|\mathbb{A}\mathrm{V@R}_{P_1,\alpha}(Y_x) - \mathbb{A}\mathrm{V@R}_{P_2,\alpha}(Y_x)| \leq \frac{1}{\alpha} d(P_1, P_2)$$

for all α and all portfolio compositions x. If the acceptability functional involves higher moments, then the Fortet-Mourier extension (see Fortet and Mourier [1953]) of the Kantorovich distance appears more appropriate.

By the well known theorem of Kantorovich-Rubinstein (see Rachev [1991]), the Kantorovich ambiguity set (15) can be represented as

$$\{P : d(P, \hat{P}) \leq \epsilon\} = \{P : \text{ there is a bivariate probability } K(\cdot, \cdot) \text{ such that}$$
$$\int_v K(u, dv) = P(u); \int_u K(du, v) = \hat{P}(v); \int_u \int_v \|u - v\|_1 K(du, dv) \leq \epsilon\}.$$

If the probability space Ω is finite, i.e. if it consists of S scenarios $\xi^{(1)}, \ldots, \xi^{(S)}$, with $\xi^{(s)} \in \mathbb{R}^M$ then a probability model is just a S-vector $(P_1, \ldots P_S)$ and the ambiguity set is a polyhedral set

$$\begin{aligned} \{P : d(P, \hat{P}) \leq \epsilon\} &= \{P = (P_1, \ldots, P_S) : P_j = \sum_i K_{i,j}; \sum_j K_{i,j} = \hat{P}_i; \\ &\qquad K_{i,j} \geq 0; \sum_{i,j} \|x^{(i)} - x^{(j)}\|_1 K_{i,j} \leq \epsilon\}. \end{aligned}$$

The bivariate probability K has the interpretation as the solution of *Monge's mass transportation problem.* The Kantorovich distance describes the minimal effort (in terms of expected transportation distances), to change the mass distribution $\hat{P}$ into the new mass distribution P (see Rachev and Rüschendorf [1998]).

In the case of a finite probability space Ω it is not difficult to find a solution for the inner problems, i.e. to determine $\inf\{\mathbb{E}_P(Y_x) : d(P, \hat{P}) \leq \epsilon\}$ and $\inf\{\mathcal{A}_P(Y_x) : d(P, \hat{P}) \leq \epsilon\}$.

For a given portfolio composition x, let $y = (y_i)$ with $y_i = x^\top \xi^{(i)}; i = 1, \ldots, S$. Denote by $\hat{P}_i$ the probability mass sitting on y_i under $\hat{P}$. To find the worst case probability $P \in \mathcal{P}$, one has to consider mass transportation from scenarios i to other scenarios j, which in total do need more than ϵ as expected transportation distance.

To this end, for minimizing $\mathbb{E}_P$, let for every pair (i, j), $w_{i,j} = y_i - y_j$ and $v_{i,j} = \|\xi^{(i)} - \xi^{(j)}\|_1$. The needed worst case is found by transferring masses from i to j in a stepwise manner: Starting with the pair i, j for which $w_{i,j}/v_{i,j}$ is maximal, the new masses are assigned in descending order of $w_{i,j}/v_{i,j}$, but only if $w_{i,j} > 0$ until the maximal allowed distance ϵ is reached.

In a similar manner, for minimizing $\mathcal{A}_P$, one sets $w_{i,j} = \frac{\partial \mathcal{A}_P(Y_x)}{\partial P_i} - \frac{\partial \mathcal{A}_P(Y_x)}{\partial P_j}$ and proceeds as before. Therefore the two inner problems are in fact directly solvable and do not need an optimization run.

If the acceptability functional is the average-value at risk, then constraint sets are polyhedral and the outer problems are linear. Therefore, in this case, the whole maximin algorithm is a sequential linear program, as studied in Byrd et al. [2005].

5 Statistical Confidence Sets

Scenario models for asset returns are typically based on statistical data. If the portfolio decision follows a parametric model, as for instance the Markowitz model, then these parameters, as the mean return, the volatility and the correlations are estimated from the given data material and the estimation error may be quantified by estimating the standard errors or by determining confidence sets, see for instance Goldfarb and Iyengar [2003].

Since our approach uses a nonparametric setup, nonparametric confidence sets have to be found. Starting with a basic estimate $\hat{P}_n$ for the probability model for asset returns, a nonparametric confidence set has to be found: The basic estimate may be either the empirical distribution, i.e. the historical model or some variants of it, for instance models with parametric tail estimates to better accommodate extremal events. The second step is the choice of the size a such that a certain confidence level ϵ is reached, i.e.

$$P\{d(P, \hat{P}_n) \geq a\} \leq \epsilon \tag{16}$$

for a large class of models P. The parameter n refers to the number of observations.

As argued before, we use the Kantorovich distance here. Moreover, we consider only model variants which differ only in probabilities and not in values. The argument for doing so is that all model variants must be discrete for the numerical treatment and that variations in values would need some parametric modeling, which we want to avoid.

For getting a confidence set of the form (16), one has to assume that the true probability model P has no mass outside a ball in $\mathbb{R}^M$. For the empirical estimate, which puts mass $1/n$ on each of the n historical observations, one has that

$$\mathbb{E}_P[d(P, \hat{P}_n)] \leq Cn^{-1/M}$$

for some constant C as was shown by Dudley [1968]. Consequently, using Markov's inequality, it follows that

$$P\left\{d(P,\hat{P}_n) \geq \frac{Cn^{-\frac{1}{M}}}{\epsilon}\right\} \leq \frac{\mathbb{E}[d(P,\hat{P}_n)]\epsilon}{n^{-\frac{1}{M}}C} \leq \epsilon.$$

If some smoothness properties of the unknown model P are known, the confidence sets may improved, see for instance Kersting [1978]. In any case, increasing the size n of the data set reduces the confidence set and leads to smaller costs for ambiguity. It should also be noted that in general the shrinking of the confidence sets can be arbitrary slow if the tails of the involved probability measures are heavy enough (see the results in Kersting [1978]).

6 A Numerical Example

The following analysis is intended to demonstrate the impact of the size of the confidence sets (i.e. the robustness) on the optimal solutions of the outlined portfolio selection problem. The presented analysis furthermore makes it possible to asses the value of information in the model by comparing the optimal expected values for different levels of robustness.

To perform this analysis the maximin approach was implemented and applied to the following data set, downloaded from finance.yahoo.com: The data consists of monthly returns within the period January 1, 1990 to December 31, 2004 of stocks from six companies, namely

- IBM - International Business Machines Corporation
- PRG - Procter & Gamble Corporation
- ATT - AT&T Corporation
- VER - Verizon Communications Inc
- INT - Intel Corporation
- EXX - Exxon Mobil Corporation

The selection of these six assets was motivated as follows. Among all assets represented in the Dow-Jones index, IBM and PRG show the least correlation and ATT and VER show the highest correlation. INT is the asset with smallest and EXX with largest variance. Under the basic model $\hat{P}$, all observations have the same probability of 1/180.

The 10% average value-at-risk $\mathbb{AV@R}_{0.1}$ was chosen as acceptability functional. In order to translate the acceptance level into a risk level, we used $1 - \mathbb{A}\mathrm{V@R}_{0.1}$ as the risk level. The bounds were set to $\mathbb{A}\mathrm{V@R}_{0,1} \geq 0.9$, i.e. the risk was bounded by 0.1. The ambiguity sets were determined as

$$\mathcal{P} = \{P : d(P, \hat{P}) \leq a\}$$

Since the a in the above formula is hard to interpret, for our analysis we vary the a in dependence of an robustness parameter γ, $\gamma \mapsto a(\gamma)$ and $\mathcal{P} = \{P : d(P, \hat{P}) \leq a(\gamma)\}$ whereby

$$a(\eta) = \max\{\eta : \sup_{d(P,\hat{P}) \leq \eta} \mathbb{E}_P(\xi^{(i)}) \leq (1 + \gamma)\mathbb{E}_{\hat{P}}(\xi^{(i)}) : \text{ for all } i\}.$$

The parameter γ is displayed in all Figures.

The maximin problem was solved by successive linear programming as described above.

Figure 1 shows the change in return, risk and portfolio composition as a function of the robustness parameter γ. As one can see, more robust portfolios are more diversified. The worst case expected return decreases with robustness, but the basic model expected return $\mathbb{E}_{\hat{P}}$ only drops slightly. On the other hand, the worst case risk stays at the bound, because this bound is binding, while the basic model risk $1 - \mathbb{A}\mathrm{V@R}_{0.1,\hat{P}}$ drops significantly. Thus, for the given data set, the price for model robustness is very small while the portfolio composition changes dramatically with the increase in robustness.

The robustness parameter influences the efficient frontier. We have plotted the efficient surface: The optimal expected return as a function of risk and robustness. Figure 2 shows this surface, when risk and return are calculated under the basic probability, while Figure 3 shows the surface when risk and return are calculated under the pertaining worst case. Notice that for each point of the surface, this might be another worst case model. While under $\hat{P}$, the efficient frontier is much deformed by increasing robustness, there is a much smaller influence on the worst cases.

7 Conclusions

We have presented a maximin approach for portfolio selection, which accommodates scenario uncertainty (aleatoric uncertainty) and probability ambiguity (epistemic uncertainty). Ambiguity is modeled by Kantorovich neighborhoods of a basic probability model.

The size of the ambiguity neighborhood may be chosen to correspond to a probabilistic confidence region for the estimated basic model. The more information is collected about the model, the smaller is the ambiguity set and the smaller is the loss in expected return one has to sacrifice for the sake of robustness. Therefore, the value of better statistical information expressed in possible shrinks of the ambiguity set can be assessed by looking at the pertaining losses in expected return for the basic model.

Figure 1: For different robustness parameters γ, the upper figure shows the expected returns, the middle figure shows the risks and the lower figure shows the portfolio composition. The assets are ordered from top to bottom as: EXX, VER, ATT, PRG, INT, IBM.

Figure 2: Efficient frontiers in dependence of the robustness parameter γ. Risk and return are calculated with respect to the basic model $\hat{P}$.

Figure 3: Efficient frontiers in dependence of the robustness parameter γ. Risk and return are calculated with respect to the worst case model.

The chosen example showed a relative small drop in expected return, but a considerable decrease of risk in exchange for a reasonable gain in robustness. It suggests that looking at portfolios which are robust against model ambiguities is very advisable for portfolio managers.

References

Aharon Ben-Tal and Arkadi Nemirovski. Robust optimization—methodology and applications. *Math. Program.*, 92(3, Ser. B):453–480, 2002. ISSN 0025-5610. ISMP 2000, Part 2 (Atlanta, GA).

Richard H. Byrd, Nicholas I. M. Gould, Jorge Nocedal, and Richard A. Waltz. On the convergence of successive linear-quadratic programming algorithms. *SIAM J. Optim.*, 16(2):471–489 (electronic), 2005. ISSN 1052-6234.

V. K. Chopra and W. T. Ziemba. The effect of errors in means, variances and covariances on optimal portfolio choice. *J. Portfolio Management*, Winter:6–11, 1993.

Robert T. Clemen and Robert T. Winkler. Combining probability distributions from experts in risk analysis. *Risk Analysis*, 19(2):187–203, 1999.

G. Dall'Aglio. Fréchet classes and compatibility of distribution functions. In *Symposia Mathematica, Vol IX (Convegno di Calcolo delle Probabilità, INDAM, Rome, 1971)*, pages 131–150. Academic Press, London, 1972.

F. Delbaen. Coherent measures of risk on general probability spaces. In K. Sandmann and P.J. Schönbucher, editors, *Advances in Finance and Stochastics. Essays in Honour of Dieter Sondermann*, pages 1–38. Springer Verlag, New York, 2002.

R. M. Dudley. The speed of mean Glivenko-Cantelli convergence. *Ann. Math. Statist*, 40:40–50, 1968.

Daniel Ellsberg. Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics*, 75(4):643–669, nov 1961. ISSN 0033-5533.

H. Föllmer and A. Schied. Convex measures of risk and trading constraints. *Stochastics and Finance*, 6:429–447, 2002.

R. Fortet and E. Mourier. Convergence de la répartition empirique vers la répartition théorique. *Ann. Sci. Ecole Norm. Sup. (3)*, 70:267–285, 1953.

D. Goldfarb and G. Iyengar. Robust portfolio selection problems. *Math. Oper. Res.*, 28(1):1–38, 2003. ISSN 0364-765X.

Götz Dietrich Kersting. Die Geschwindigkeit der Glivenko-Cantelli-Konvergenz gemessen in der Prohorov-Metrik. *Math. Z.*, 163(1):65–102, 1978. ISSN 0025-5874.

Roger W. Klein and Vijay S. Bawa. The effect of estimation risk on optimal portfolio choice. *Journal of Financial Economics*, 3(3):215–231, June 1976.

Frank H. Knight. *Risk, Uncertainty and Profit.* Houghton Mifflin Company, Boston, MA, 1921.

Harry M. Markowitz. *Portfolio selection: Efficient diversification of investments.* Cowles Foundation for Research in Economics at Yale University, Monograph 16. John Wiley & Sons Inc., New York, 1959.

A. Nemirovskii and D.B. Yudin. Cesaro convergence of the gradient method for approximation of saddle points of convex-concave functions. *Doklady AN SSSR*, 239:1056–1059, 1978.

Georg Ch. Pflug. Subdifferential representations of risk measures. *Math. Program.*, 108(2-3, Ser. B):339–354, 2006. ISSN 0025-5610.

Svetlozar T. Rachev. *Probability metrics and the stability of stochastic models.* Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Ltd., Chichester, 1991. ISBN 0-471-92877-1.

Svetlozar T. Rachev and Ludger Rüschendorf. *Mass transportation problems. Vol. I.* Probability and its Applications (New York). Springer-Verlag, New York, 1998. ISBN 0-387-98350-3. Theory.

R. Tyrrell Rockafellar. Monotone operators and the proximal point algorithm. *SIAM J. Control Optimization*, 14(5):877–898, 1976. ISSN 1095-7138.

R. Tyrrell Rockafellar. *Convex analysis.* Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1997. ISBN 0-691-01586-4. Reprint of the 1970 original, Princeton Paperbacks.

R. Tyrrell Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2(3):21–41, 2000.

Alexander Shapiro and Shabbir Ahmed. On a class of minimax stochastic programs. *SIAM J. Optim.*, 14(4):1237–1249 (electronic), 2004. ISSN 1052-6234.

Yesim Tokat, Svetlozar Rachev, and Eduardo Schwartz. The stable non-gaussian asset allocation: A comparison with the classical gaussian approach. *Journal of Economic Dynamics and Conrol*, 27:937–969, 2003.

S.S. Vallander. Calculation of the Wasserstein distance between probability distributions on the line. *Theor. Prob. Appl.*, 18:784 – 786, 1973.

Li-Yong Yu, Xiao-Dong Ji, and Shou-Yang Wang. Stochastic programming models in financial optimization: A survey. *AMO — Advanced Modeling and Optimization*, 5(1), 2003.

J. Zackova. On minimax solutions of stochastic linear programming problems. *Casopis pro Pěstovânî Mathematiky*, 91:423–430, 1966.

William T. Ziemba and Raymond G. Vickson. *Stochastic optimization models in finance*. Academic Press, 1973.