Ambiguity in portfolio optimization

Georg Ch. Pflug

May/June 2006
Introduction: Risk and Ambiguity

Frank Knight ”Risk, Uncertainty and Profit” (1920)

Risk: the decision-maker can assign mathematical probabilities to random phenomena

Uncertainty: randomness cannot be expressed in terms of specific mathematical models.

Ellsberg (1961)

Uncertainty: the probabilistic model is known, but the realizations of the random variables are unknown (”aleatoric uncertainty”)

Ambiguity: the probability model itself is unknown (”epistemic uncertainty”).
It has frequently been observed that the solutions of portfolio optimization problems are much sensitive w.r.t. the parameters of the distributions (Klein and Bawa (1976), Chopra and Ziemba (1993)). However, the distributions are estimated from data and therefore contain some estimation error and model ambiguity. Therefore we need to combine

*Statistics* ("how to estimate the probability model and its parameters and to assess confidence") and

*Optimization* ("how to find the best decision") in one approach.
Problem formulation

\((\xi_1, \ldots, \xi_M)\) random returns for \(M\) asset categories

\((x_1, \ldots, x_M)\) portfolio weights

\(Y_x = \sum_{m=1}^{M} x_m \xi_m\) portfolio return

\(A(Y_x)\) acceptability functional

Let \(\hat{P}\) be a baseline probability measure for \((\xi_1, \ldots, \xi_M)\). If we fully trust in the validity of this model, we solve the non-ambiguous portfolio optimization problem

Maximize (in \(x\)) : \(\mathbb{E}_{\hat{P}}(Y_x)\)

subject to

\(A_{\hat{P}}(Y_x) \geq q\)

\(x^\top 1 = 1\)

\(x \geq 0\)
Acceptability functionals

Required properties:

(i) translation equivariance

\[ A(Y + c) = A(Y) + c \]

(ii) concavity

\[ A(\lambda Y_1 + (1 - \lambda) Y_2) \geq \lambda A(Y_1) + (1 - \lambda) A(Y_2). \]

(iii) continuity \( Y \mapsto A(Y) \) continuous

(iv) version independence: If \( Y_1 \) and \( Y_2 \) have the same distribution, then \( A(Y_1) = A(Y_2) \)
Examples for such acceptability functionals $\mathcal{A}$ are

- the expectation
  \[ \mathcal{A}(Y) = \mathbb{E}(Y) \]

- the average-value-at risk (conditional value-at-risk, expected shortfall)
  \[ \mathcal{A}(Y) = \text{AV@R}_\alpha(Y) = \frac{1}{\alpha} \int_0^\alpha G_Y^{-1}(p) \, dp, \]
  where $G_Y^{-1}(p) = \inf\{v : P\{Y \leq v\} \geq p\}$

- the (lower) standard deviation corrected expectation (0 < $\lambda$ < 1)
  \[ \mathcal{A}(Y) = \mathbb{E}(Y) - \lambda \text{Std}(Y) \text{ or } \mathbb{E}(Y) - \lambda \text{Std}^{-}(Y), \]
  \[ \text{Std}^{-}(Y) = \sqrt{\mathbb{E}[(Y - \mathbb{E}Y)^-)^2} \]
Stress testing

We solve the portfolio problem for the baseline probability model, but check its performance under new probability models. Example: 500 historic weakly returns from NYSE

<table>
<thead>
<tr>
<th>model</th>
<th>composition optimal for</th>
<th>weakly return</th>
<th>acceptability AV@R_{0.1}</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{P})</td>
<td>(\hat{P})</td>
<td>0.8%</td>
<td>0.92</td>
</tr>
<tr>
<td>(P^-)</td>
<td>(\hat{P})</td>
<td>-0.13%</td>
<td>0.90</td>
</tr>
<tr>
<td>(P^+)</td>
<td>(\hat{P})</td>
<td>1.17%</td>
<td>0.93</td>
</tr>
<tr>
<td>(P^-)</td>
<td>(P^-)</td>
<td>-0.07%</td>
<td>0.92</td>
</tr>
<tr>
<td>(P^+)</td>
<td>(P^+)</td>
<td>1.9%</td>
<td>0.92</td>
</tr>
</tbody>
</table>

(\hat{P}) : historic data
(P^-) : 5% higher prob. for bad scenarios
(P^+) : 5% lower prob. for bad scenarios
The ambiguity problem - maximin formulation

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

Maximize (in $x$) : $\min\{\mathbb{E}_P(Y_x) : P \in \mathcal{P}\}$

subject to

$A_P(Y_x) \geq q$ for all $P \in \mathcal{P}$

$x^\top 1 = 1$

$x \geq 0$
Dealing with uncertainty:

- **Deterministic optimization:**

  \[
  \max \{ F(x) : F_i(x) \geq 0 \}
  \]

  *F* is the objective,  
  \(F_i\) are the constraint functions.

- **Robust optimization (maximin approach):**

  A set \(\Xi\) of possible parameters \(\xi\) is given.

  \[
  \max \{ \min \{ F(x, \xi) : \xi \in \Xi \} : F_i(x, \xi) \geq 0 ; \xi \in \Xi \}
  \]

  Robust optimization produces very conservative portfolios.
Stochastic optimization:

\((\Xi, \mathcal{A}, P)\) is a probability space, typical element is called \(\xi\).

\[
\max \{ \mathbb{E}_P[F(x, \xi)]; \mathbb{E}_P[F_i(x, \xi)] \geq 0 \}
\]

or more generally

\[
\max \{ \mathcal{A}_P[F(x, \xi)]; \mathcal{A}_P^{(i)}[F_i(x, \xi)] \geq 0 \}
\]

where \(\mathcal{A}_P^{(\cdot)}\) are *aggregating probability functionals*, mapping the distribution of \(F(x, \xi)\) under \(P\) to the (extended) real line. Stochastic optimization produces well diversified portfolios.
Ambiguity in portfolio optimization


?  ?
The ambiguity problem - Bayesian formulation

Let $\lambda$ be a probability on the ambiguity set $\mathcal{P}$. $\lambda$ is called the Bayesian prior on the set of probability models. The Bayesian formulation is

\[
\begin{align*}
\text{Maximize (in } x): & \quad \int_{\mathcal{P}} A_P[F(x, \xi)] \ d\lambda(P) \\
\text{subject to} & \quad \int_{\mathcal{P}} A_{P}^{(i)}[F_i(x, \xi)] \ d\lambda(P) \geq 0 \\
& \quad x^\top 1 = 1 \\
& \quad x \geq 0
\end{align*}
\]
The ambiguity problem as robust stochastic optimization

We repeat the maximin formulation of the problem:

\[
\begin{align*}
\text{Maximize (in } x \text{) : } & \min \{ \mathbb{E}_P(Y_x) : P \in \mathcal{P} \} \\
\text{subject to} & \\
A_P(Y_x) & \geq q \text{ for all } P \in \mathcal{P} \\
x^\top \mathbf{1} & = 1 \\
x & \geq 0
\end{align*}
\]
Solution techniques for the maximin portfolio problem

Let

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

then the ambiguity problem reads

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

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 if and only if there is a $P^* \in \mathcal{P}$ such that $(P^*, x^*)$ is a saddle point, i.e.

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

for all $(P, x) \in \mathcal{P} \times \mathbb{X}$.

Our problem is semi-infinite. Direct saddle point methods were used by Rockafellar (1976), Nemirovskii and Yudin (1978).
Successive convex programming (SCP)

1. Set $n = 0$ and $\mathcal{P}_0 = \{\hat{P}\}$ with $\hat{P} \in \mathcal{P}$.
2. Solve the outer problem

\[
\begin{align*}
\text{Maximize (in } x, t): & \quad t \\
\text{subject to:} & \\
& t \leq \mathbb{E}_P(Y_x) \text{ for all } P \in \mathcal{P}_n \\
& \mathcal{A}_P(Y_x) \geq q \text{ for all } P \in \mathcal{P}_n \\
& x^\top \mathbf{1} = 1; x \geq 0
\end{align*}
\]

and call the solution $(x_n, t_n)$.
3. Solve the first inner problem

\[
\begin{align*}
\text{Minimize (in } P): & \quad \mathbb{E}_P(Y_{x_n}) \\
\text{subject to:} & \\
& P \in \mathcal{P}
\end{align*}
\]

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

Georg Ch. Pflug

Ambiguity in portfolio optimization
4. Solve the second inner problem

\[
\begin{align*}
\text{Minimize (in } P): \quad & A_P(Y_{x_n}) \\
\text{subject to:} \quad & P \in \mathcal{P}
\end{align*}
\]

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

5. If \( \mathcal{P}_{n+1} = \mathcal{P}_n \) then stop. Otherwise set \( n := n + 1 \) and goto 2.
Convergence

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

Georg Ch. Pflug
Ambiguity in portfolio optimization
Ambiguity in portfolio optimization
Ambiguity sets as balls - distances of probability measures

We want to use an ambiguity sets of the form \( \mathcal{P} = \{ P : d(P, P^*) \leq \epsilon \} \) as ambiguity sets. But which distance \( d \) to choose?

- *Transportation distance, Wasserstein distance*

\[
d_T(P_1, P_2) = \sup \left\{ \int h(u) \, dP_1(u) - \int h(u) \, dP_2(u) : |h(u) - h(v)| \leq |u - v| \right\}
\]
▶ **Kolmogorov–Smirnov distance**

\[
d_{KS}(P_1, P_2) = \sup\{ |P_1((−∞, a]) − P_2((−∞, a])| : a ∈ \mathbb{R} \}
\]

▶ **Variational distance**

\[
d_V(P_1, P_2) = \sup\{ |P_1(A) − P_2(A)| : A \text{ mesurable} \}
\]

If densities exist, this is equivalent to the \( L_1 \) distance:

\[
d_{L_1}(P_1, P_2) = ∫ \left| \frac{dP_1}{dμ} - \frac{dP_2}{dμ} \right| dμ.
\]
Statistical estimation and confidence sets

- The ambiguity set must reflect our current information about $P$
- If our information is based on statistical estimation, the ambiguity set must coincide with a confidence set
- If the distance is too fine, no statistical confidence set can be constructed without further assumptions (e.g. for the variational distance)

If we base our estimate for $P$ on some historical data $\xi_1, \xi_2, \ldots, \xi_n$, we use the empirical distribution

$$\hat{P}_n\{\xi_i\} = \frac{1}{n}$$
Confidence sets

\[ P\{d_{KS}(P, \hat{P}_n) \geq M/\sqrt{n}\} \leq 58 \exp(-2M^2) \]

(Dvoretzky, Kiefer, Wolfowitz inequality)

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

for some constant \( C \) (Dudley (1969)). By Markov’s inequality

\[ P\{d_T(P, \hat{P}_n) \geq \epsilon\} \leq \mathbb{E}[d(P, \hat{P}_n)]/\epsilon \leq n^{-1/M} C/\epsilon. \]

Under smoothness conditions on \( P \), the confidence sets may improved (Kersting (1978)).
Example

6 assets

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

Risk constraint: $\text{AV@R}_{0.1} \geq 0.9$

Ambiguity set: $\mathcal{P} = \{P : d_T(P, \hat{P}) \leq \epsilon\}$.

In order to make the ambiguity sets more interpretable, we define a robustness parameter $\gamma$ as the maximal relative change of the expected returns and relate this to $\epsilon$ by

$$
\epsilon = \max\{\eta : \sup_{d_T(P, \hat{P}) \leq \eta} \mathbb{E}_P(\xi^{(i)}) \leq (1 + \gamma)\mathbb{E}_{\hat{P}}(\xi^{(i)}) : \text{for all } i\}. \quad (1)
$$
Working with Transportation distances

By the well known theorem of Kantorovich-Rubinstein, the Transportation ambiguity set can be represented as

$$\{ P : d_T(P, \hat{P}) \leq \epsilon \} = \{ P : \text{there is a bivariate probability } K(\cdot, \cdot) \text{ s.t.}$$

$$\int_v K(u, dv) = P(u); \int_u K(du, v) = \hat{P}(v); \int_u \int_v \|u - v\|_1 K(du, dv) \leq \epsilon \}. $$

If the probability space $\Omega$ is finite, $\Omega = \{x^{(1)}, \ldots, x^{(n)}\}$, the ambiguity set is a polyhedral set

$$\{ P : d_T(P, \hat{P}) \leq \epsilon \} = \{ P = (P_1, \ldots, P_S) : P_j = \sum_i K_{ij}; \sum_j K_{ij} = \hat{P}_i;$$

$$K_{ij} \geq 0; \sum_{i,j} \|x^{(i)} - x^{(j)}\|_1 K_{ij} \leq \epsilon \}. $$
Illustration of the Transportation distance

\[ d_T(P_1, P_2) = \text{minimal } \mathbb{E} \left( \text{transported mass } \times \text{distance} \right). \]
The bivariate probability $K$ has the interpretation as the solution of *Monge’s mass transportation problem*. The Transportation distance describes the minimal effort (in terms of expected transportation distances), to change the mass distribution $\hat{P}$ into the new mass distribution $P$ (Rachev and Rüschendorf (1998)). In the case of a finite probability space $\Omega$ it is not difficult to find a solution for the inner problems, i.e. to determine

$$\inf\{\mathbb{E}_P(Y_X) : d_T(P, \hat{P}) \leq \epsilon\}$$

and

$$\inf\{A_P(Y_X) : d_T(P, \hat{P}) \leq \epsilon\}$$

by a special algorithm.
The solution of the Example

Expected returns, risks and portfolio composition. The assets from top to bottom are: EXX, VER, ATT, PRG, INT, IBM.
Efficient frontiers in dependence of the robustness parameter $\gamma$.
Risk and return are calculated w.r.t. the basic model $\hat{P}$.
Efficient frontiers in dependence of the robustness parameter $\gamma$. Risk and return are calculated w.r.t. the worst case model.
Types of problems in Finance and Risk management

- Pricing (Equation solving): $F(x) = y$
- Optimization: $\max F(x)$
- Game Theory (Equilibrium): $\max \min F(x, y)$
Conclusions

► In order to capture scenario uncertainty (aleatoric uncertainty) and probability ambiguity (epistemic uncertainty) we use a probabilistic maximin approach.

► The ambiguity neighborhood should be chosen in such a way that it corresponds to a probabilistic confidence regions for which bounds for the covering probability are available.

► The result of model ambiguity is a further diversification of the portfolio

► It turns out that often the "price" to be paid for including ambiguity in the optimization problem is very little.

► Minimax models are also relevant for the pricing of swing options in Electricity markets.