Stochastic Programming:

Models, Approximations, Applications

Werner Römisch

Humboldt-University Berlin
Institute of Mathematics
http://www.mathematik.hu-berlin.de/~romisch

3rd Alumni Summer School on Applied Mathematics

Cairo, Egypt, November 26, 2005

Introduction

What is Stochastic Programming ?
- Mathematics for Decision Making under Uncertainty
- subfield of Mathematical Programming (MSC 90C15)

Stochastic programs are **optimization models**
- having special properties and structures,
- depending on the underlying probability distribution,
- requiring specific approximation and numerical approaches,
- having close relations to practical applications.

Selected recent monographs:

P. Kall/S.W. Wallace 1994, A. Prekopa 1995,
J.R. Birge/F. Louveaux 1997, J. Mayer/P. Kall 2005
A. Ruszczynski/A. Shapiro (eds.), Stochastic Programming, Handbook, Elsevier, 2003
S.W. Wallace/W.T. Ziemba (eds.), Applications of Stochastic Programming, MPS-SIAM Series on Optimization, 2005.

Application: Electricity Portfolio Management

We consider the yearly electricity portfolio management of a municipal German power utility. Its portfolio consists of the following positions:

- power production (based on utility-owned thermal units),
- (mid-term) contracts (provided by large utilities),
- (physical) spot market trading and
- (financial) trading of futures.

The yearly time horizon is discretized into hourly intervals. The underlying stochasticity consists in a bivariate stochastic load and price process that is approximately represented by a finite number of scenarios. The objective is to maximize the total expected revenue. The portfolio management model is a large scale (mixed-integer) multistage stochastic program.

Should the expected revenue be maximized exclusively or should the risk of its production and trading decisions simultaneously be bounded or even minimized ?

Stochastic Programming Model

Let $\{\xi_t\}_{t=1}^T$ be a discrete-time stochastic data process defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and with ξ_t taking values in $\mathbb{R}^d$. The stochastic decision x_t at period t varying in $\mathbb{R}^{m_t}$ is assumed to depend only on $\xi^t := (\xi_1, \ldots, \xi_t)$ (nonanticipativity). Let $\mathcal{F}_t \subseteq \mathcal{F}$ denote the σ-algebra which is generated by ξ^t, i.e., $\mathcal{F}_t = \sigma\{(\xi_1, \ldots, \xi_t)\}$. We have $\mathcal{F}_t \subseteq \mathcal{F}_{t+1}$ for $t = 1, \ldots, T-1$ and we assume that $\mathcal{F}_1 = \{\emptyset, \Omega\}$ (i.e., ξ_1 deterministic) and $\mathcal{F}_T = \mathcal{F}$.

We consider the (linear) stochastic programming model:

$$\min \left\{ \mathbb{E}\left[\sum_{t=1}^{T} \langle b_t(\xi_t), x_t \rangle\right] \left| \begin{array}{l} x_t \in X_t, \\ x_t \text{ is } \mathcal{F}_t - \text{measurable}, t = 1, \ldots, T, \\ A_{t0}x_t + A_{t1}(\xi_t)x_{t-1} = h_t(\xi_t), t = 2, \ldots, T \end{array} \right. \right\}$$

where the sets X_t are nonempty and polyhedral, and $A_{t1}(\cdot)$, $b_t(\cdot)$ and $h_t(\cdot)$ are affinely linear for each $t = 2, \ldots, T$.

To have the model well defined, we assume
$x_t \in L_{r'}(\Omega, \mathcal{F}_t, I\!P; I\!R^{m_t})$ and $\xi_t \in L_r(\Omega, \mathcal{F}_t, I\!P; I\!R^d)$,
where $r \geq 1$ and

$$r' := \begin{cases} \frac{r}{r-1} & , \text{ if only costs are random} \\ r & , \text{ if only right-hand sides are random} \\ r = 2 & , \text{ if only costs and right-hand sides are random} \\ \infty & , \text{ if all technology matrices are random and } r = T. \end{cases}$$

Then nonanticipativity may be expressed as

$$x \in \mathcal{N}_{na}$$

$$\mathcal{N}_{na} = \{x \in \times_{t=1}^{T} L_{r'}(\Omega, \mathcal{F}, I\!P; I\!R^{m_t}) : x_t = I\!E[x_t|\mathcal{F}_t]\,, \forall t\},$$

i.e., as a subspace constraint, by using the conditional expectation $I\!E[\cdot|\mathcal{F}_t]$ with respect to the σ-algebra $\mathcal{F}_t$.

For $T = 2$ we have $\mathcal{N}_{na} = I\!R^{m_1} \times L_{r'}(\Omega, \mathcal{F}, P; I\!R^{m_2})$.

$\longrightarrow$ infinite-dimensional optimization problem

Scenario-based models

Let Ω be finite, i.e., $\Omega = \{\omega_s\}_{s=1}^{S}$, $\mathcal{F}$ power set of Ω.
$p_s := \mathbb{P}(\{\omega_s\})$ (probability of scenario s), $s = 1, \ldots, S$,
$\xi_t^s := \xi_t(\omega_s)$ (data scenario s at stage t) and
x_t^s (decision scenario s at t, $s = 1, \ldots, S$, $t = 1, \ldots, T$.
Let $\mathcal{E}_t$ be a (finite) partition of Ω such that the smallest σ-algebra containing $\mathcal{E}_t$ is just $\mathcal{F}_t$. Then

$$\begin{aligned}
\mathbb{E}[x_t|\mathcal{F}_t] &= \sum_{C\in\mathcal{E}_t} \frac{1}{P(C)} \int_C x_t(\omega)P(d\omega)\chi_C \\
&= \sum_{C\in\mathcal{E}_t} \Big(\sum_{\substack{s=1\\ \omega_s\in C}}^{S} p_s\Big)^{-1} \Big(\sum_{\substack{s=1\\ \omega_s\in C}}^{S} p_s x_t^s\Big)\chi_C
\end{aligned}$$

where χ_C denotes the characteristic function of $C \in \mathcal{E}_t$.

The nonanticipativity condition (NA) is equivalent to

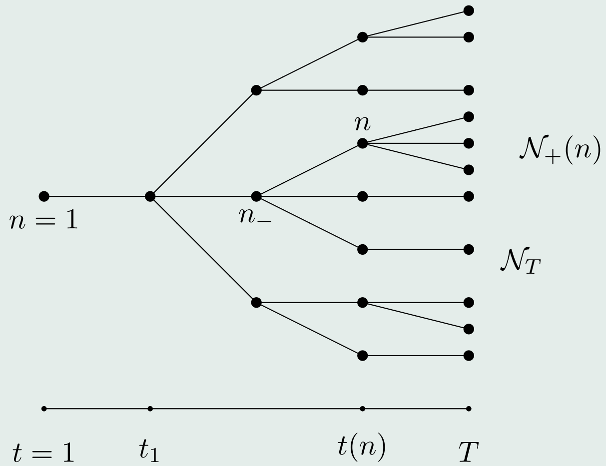
$$x_t^\sigma = (\mathbb{E}[x_t|\mathcal{F}_t])^\sigma = \sum_{\substack{C\in\mathcal{E}_t\\ \omega_\sigma\in C}} \frac{\sum\limits_{\substack{s=1\\ \omega_s\in C}}^{S} p_s x_t^s}{\sum\limits_{\substack{s=1\\ \omega_s\in C}}^{S} p_s}, \forall\sigma = 1,\ldots,S, \forall t.$$

Special case $t = 1$: $\mathcal{E}_1 = \{\Omega\}$ and, hence, (NA) is equivalent to $x_1^\sigma = \sum_{s=1}^S p_s x_1^s$, $\sigma = 1,\ldots,S$, i.e., to $x_1^1 = \ldots = x_1^S$.
Then the stochastic program takes the scenario form:

$$\min \{\sum_{s=1}^{S}\sum_{t=1}^{T} p_s b_t(\xi_t^s) x_t^s : x \text{ satisfies (NA)}, x_t^s \in X_t\,, t = 1,\ldots,T,$$
$$A_{t0}x_t^s + A_{t1}(\xi_t^s)x_{t-1}^s = h_t(\xi_t^s), s = 1,\ldots,S, t = 2,\ldots,T\}$$

Since $\mathcal{F}_t \subseteq \mathcal{F}_{t+1}$, every element of $\mathcal{E}_t$ can be represented as the union of certain elements of $\mathcal{E}_{t+1}$. Representing the elements of $\mathcal{E}_t$ by nodes and the above relations by arcs leads to a tree which is called scenario tree.

A scenario tree is based on a finite set $\mathcal{N} \subset \mathbb{N}$ of nodes where



Scenario tree with $t_1 = 2$, $T = 5$, $|\mathcal{N}| = 23$ and 11 leaves

$n = 1$ stands for the period root node,
n_- is the unique predecessor of node n,
$\text{path}(n) := \{1, \ldots, n_-, n\}$, $t(n) := |\text{path}(n)|$,
$\mathcal{N}_t := \{n : t(n) = t\}$, nodes $n \in \mathcal{N}_T$ are the leaves,
A scenario corresponds to $\text{path}(n)$ for some $n \in \mathcal{N}_T$,
$\mathcal{N}_+(n)$ is the set of successors to node n.
We have $\{\pi_n\}_{n \in \mathcal{N}_T} := \{p_s\}_{s=1}^S$ and $\pi_n := \sum_{n_+ \in \mathcal{N}_+(n)} \pi_{n_+}$, $n \in \mathcal{N}$.

$\{\xi^n\}_{n\in\mathcal{N}_t}$ are the realizations of ξ_t and $\{x^n\}_{n\in\mathcal{N}_t}$ the realizations of x_t.

Then the tree formulation of the model reads:

$$\min \ \{\sum_{n\in\mathcal{N}} \pi_n b_{t(n)}(\xi^n)x^n : x^n \in X_{t(n)}$$
$$A_{t(n)0}(\xi^n)x^n + A_{t(n)1}(\xi^n)x^{n-} = h_{t(n)}(\xi^n), n \in \mathcal{N}\}$$

Note that it holds for the dimensions $|\mathcal{N}| << TS$.

Dynamic programming

Theorem: (Evstigneev 76, Rockafellar/Wets 76)
Under weak assumptions the multistage stochastic program is equivalent to the (first-stage) convex minimization problem

$$\min\left\{\int_{\Xi} f(x_1,\xi)P(d\xi) : x_1 \in X_1\right\},$$

where f is an integrand on $I\!R^{m_1} \times \Xi$ given by

$$f(x_1,\xi) := \langle b_1(\xi_1), x_1\rangle + \Phi_2(x_1,\xi^2),$$

$$\Phi_t(x_1,\ldots,x_{t-1},\xi^t) := \inf\left\{\langle b_t(\xi_t), x_t\rangle + I\!E\left[\Phi_{t+1}(x_1,\ldots,x_t,\xi^{t+1})|\mathcal{F}_t\right] : x_t \in X_t, A_{t,0}x_t + A_{t,1}x_{t-1} = h_t(\xi_t)\right\}$$

for $t = 2,\ldots,T$, where $\Phi_{T+1}(x_1,\ldots,x_T,\xi^{T+1}) := 0$.

$\longrightarrow$The integrand f depends on the probability measure $I\!P$ in a nonlinear way !

Stability

Let us introduce some notations. Let F denote the objective function defined on $L_r(\Omega, \mathcal{F}, I\!P; I\!R^s) \times L_{r'}(\Omega, \mathcal{F}, I\!P; I\!R^m) \to I\!R$ by $F(\xi, x) := I\!E[\sum_{t=1}^{T} \langle b_t(\xi_t), x_t \rangle]$, let

$$\mathcal{X}_t(x_{t-1}; \xi_t) := \{x_t \in X_t | A_{t,0} x_t + A_{t,1}(\xi_t) x_{t-1} = h_t(\xi_t)\}$$

denote the t-th feasibility set for every $t = 2, \ldots, T$ and

$$\mathcal{X}(\xi) := \{x \in \times_{t=1}^{T} L_{r'}(\Omega, \mathcal{F}_t, I\!P; I\!R^{m_t}) | x_1 \in X_1, x_t \in \mathcal{X}_t(x_{t-1}; \xi_t)\}$$

the set of feasible elements with input ξ.
Then the multistage stochastic program may be rewritten as

$$\min\{F(\xi, x) \, : \, x \in \mathcal{X}(\xi)\}.$$

Furthermore, let $v(\xi)$ denote its optimal value and let, for any $\alpha \geq 0$,

$$l_\alpha(F(\xi, \cdot)) := \{x \in \mathcal{X}(\xi) : F(\xi, x) \leq v(\xi) + \alpha\}$$

denote the α-level set of the stochastic program with input ξ.

The following conditions are imposed:

(A1) There exists a $\delta > 0$ such that for any $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, I\!P; I\!R^s)$ with $\|\tilde{\xi} - \xi\|_r \le \delta$, any $t = 2, \ldots, T$ and any $x_1 \in X_1$, $x_\tau \in \mathcal{X}_\tau(x_{\tau-1}; \tilde{\xi}_\tau)$, $\tau = 2, \ldots, t-1$, the set $\mathcal{X}_t(x_{t-1}; \tilde{\xi}_t)$ is nonempty (relatively complete recourse locally around ξ).

(A2) The optimal value $v(\xi)$ is finite and the objective function F is level-bounded locally uniformly at ξ, i.e., for some $\alpha > 0$ there exists a $\delta > 0$ and a bounded subset B of $L_{r'}(\Omega, \mathcal{F}, I\!P; I\!R^m)$ such that $l_\alpha(F(\tilde{\xi}, \cdot))$ is nonempty and contained in B for all $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, I\!P; I\!R^s)$ with $\|\tilde{\xi} - \xi\|_r \le \delta$.

(A3) $\xi \in L_r(\Omega, \mathcal{F}, I\!P; I\!R^s)$ for some $r \ge 1$.

Norms in L_r and $L_{r'}$:

$$\|\xi\|_r := \left(\sum_{t=1}^{T} I\!E[\|\xi_t\|^r] \right)^{\frac{1}{r}} \qquad \|x\|_{r'} := \left(\sum_{t=1}^{T} I\!E[\|x_t\|^{r'}] \right)^{\frac{1}{r'}}$$

Theorem:
Let (A1), (A2) and (A3) be satisfied and X_1 be bounded.
Then there exist positive constants L, α and δ such that the estimate

$$|v(\xi) - v(\tilde{\xi})| \leq L(\|\xi - \tilde{\xi}\|_r + D_{\mathrm{f}}(\xi, \tilde{\xi}))$$

holds for all $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, I\!P; I\!R^s)$ with $\|\tilde{\xi} - \xi\|_r \leq \delta$.
Here, $D_{\mathrm{f}}(\xi, \tilde{\xi})$ denotes the filtration distance of ξ and $\tilde{\xi}$ defined by

$$D_{\mathrm{f}}(\xi, \tilde{\xi}) := \sup_{\varepsilon \in (0,\alpha]} \inf_{\substack{x \in l_\varepsilon(F(\xi,\cdot)) \\ \tilde{x} \in l_\varepsilon(F(\tilde{\xi},\cdot))}} \sum_{t=2}^{T-1} \max\{\|x_t - I\!E[x_t|\tilde{\mathcal{F}}_t]\|_{r'}, \|\tilde{x}_t - I\!E[\tilde{x}_t|\mathcal{F}_t]\|_{r'}\},$$

where $\mathcal{F}_t$ and $\tilde{\mathcal{F}}_t$ denote the σ-fields generated by ξ^t and $\tilde{\xi}^t$, $t = 1, \ldots, T$.

The filtration distance of two stochastic processes vanishes if their filtrations coincide, in particular, if the model is two-stage. If solutions exist, the filtration distance is of the simplified form

$$D_{\mathrm{f}}(\xi,\tilde{\xi}) = \inf_{\substack{x\in l_0(F(\xi,\cdot))\\ \tilde{x}\in l_0(F(\tilde{\xi},\cdot))}} \sum_{t=2}^{T-1} \max\{\|x_t-\mathbb{E}[x_t|\tilde{\mathcal{F}}_t]\|_{r'}, \|\tilde{x}_t-\mathbb{E}[\tilde{x}_t|\mathcal{F}_t]\|_{r'}\}.$$

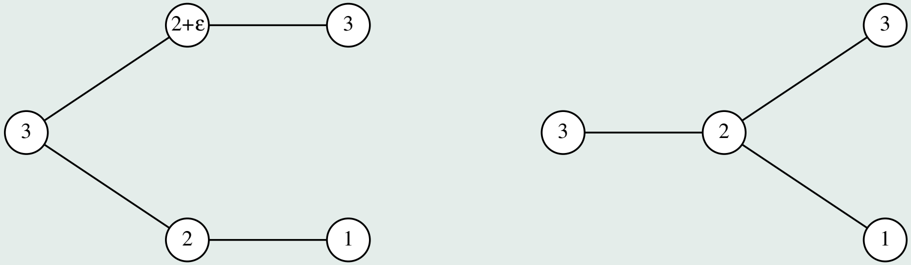
For example, solutions exist if Ω is finite or if $1 < r' < \infty$ implying that the spaces $L_{r'}$ are finite-dimensional or reflexive Banach spaces (hence, the level sets are compact or weakly sequentially compact).

The following example shows that the filtration distance D_{f} is indispensable for the stability result to hold.

Example: (Optimal purchase under uncertainty)
The decisions x_t correspond to the amounts to be purchased at each time period with uncertain prices are ξ_t, $t = 1, \ldots, T$, and such that a prescribed amount a is achieved at the end of a given time horizon. The problem is of the form

$$\min \left\{ \mathbb{E}\left[\sum_{t=1}^{T} \xi_t x_t\right] \;\middle|\; \begin{array}{l} (x_t, s_t) \in X_t = \mathbb{R}_+^2, \\ (x_t, s_t) \text{ is } (\xi_1, \ldots, \xi_t)\text{-measurable}, \\ s_t - s_{t-1} = x_t,\ t = 2, \ldots, T, \\ s_1 = 0,\ s_T = a. \end{array} \right\},$$

where the state variable s_t corresponds to the amount at time t. Let $T := 3$ and ξ_ε denote the stochastic price process having the two scenarios $\xi_\varepsilon^1 = (3, 2+\varepsilon, 3)$ ($\varepsilon \in (0,1)$) and $\xi_\varepsilon^2 = (3, 2, 1)$ each endowed with probability $\frac{1}{2}$. Let $\tilde{\xi}$ denote the approximation of ξ_ε given by the two scenarios $\tilde{\xi}^1 = (3, 2, 3)$ and $\tilde{\xi}^2 = (3, 2, 1)$ with the same probabilities $\frac{1}{2}$.



Scenario trees for ξ_ε (left) and $\tilde{\xi}$

We obtain

$$\begin{aligned} v(\xi_\varepsilon) &= \frac{1}{2}((2+\varepsilon)a + a) = \frac{3+\varepsilon}{2}a \\ v(\tilde{\xi}) &= 2a\,, \qquad \text{but} \\ \|\xi_\varepsilon - \tilde{\xi}\|_1 &\le \frac{1}{2}(0+\varepsilon+0) + \frac{1}{2}(0+0+0) = \frac{\varepsilon}{2}. \end{aligned}$$

Hence, the multistage stochastic purchasing model is not stable with respect to $\|\cdot\|_1$.

However, the estimate for $|v(\xi) - v(\tilde{\xi})|$ in the stability theorem is valid with $L = 1$ since $D_f(\xi, \tilde{\xi}) = \frac{a}{2}$.

Scenario tree approximations for ξ

Reference: Dupačová/Consigli/Wallace 2000

All known approaches consist of two steps:
(a) Simulation of (sufficiently many) scenarios of the stochastic data process ξ;
(b) construction of scenario trees from simulation scenarios or probability distribution information.

(a) Methods:

- Identifying and fitting statistical models to historical data (e.g. (multivariate) time series models).
- sampling or resampling historical data as scenarios.

(b) Methods:

(b1) Construction based on distribution information:

- barycentric tree constructions;
- EVPI-based sampling methods;
- Regression fit to given (higher order) moments.

(b2) Construction from simulation scenarios:

<u>Given:</u> N individual scenarios ξ^i with probabilities p_i and fixed starting point ξ_1^*, i.e., forming a fan.

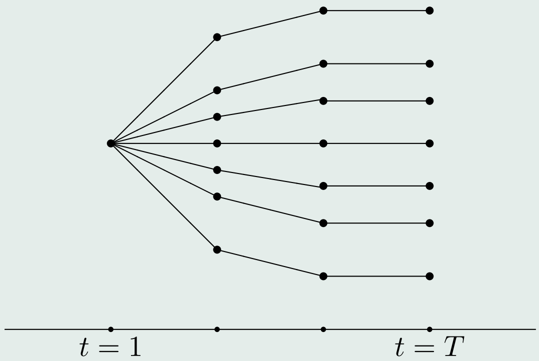


Figure 1: Example of a fan of individual scenarios with $T = 4$ and $N = 7$

Cluster-analysis-based methods:

- Studying the similarity of scenarios for $t = T, \ldots, 2$;
- "Bundling" scenarios in a cluster and definition of succesors and predecessor, respectively, e.g., using the L_r-norm.

Numerical methods for tree construction

Forward and backward algorithms have been developed for constructing a scenario tree $\xi_{\rm tr}$ to approximate a fan ξ of scenarios, i.e., such that $\|\xi - \xi_{\rm tr}\|_r \leq \varepsilon$ and $D_{\rm f}(\xi, \xi_{\rm tr}) \leq \text{Const} \cdot \varepsilon_{\rm f}$.

Algorithm (forward tree construction)
Step 1: Select ε_t such that $\sum_{t=2}^{T} \varepsilon_t \leq \varepsilon$.
Step 2: Choose the stochastic process $\hat{\xi}^2$ with index set I_2 of scenarios and scenario bundles $I_{2,i}$, $i \in I_2$, such that the condition

$$\sum_{i\in I_2} \sum_{j\in I_{2,i}} p_j \|\xi^j - \xi^i\|^{r'} < \min\{\varepsilon_2, \varepsilon_{\rm f}\}^{r'}$$

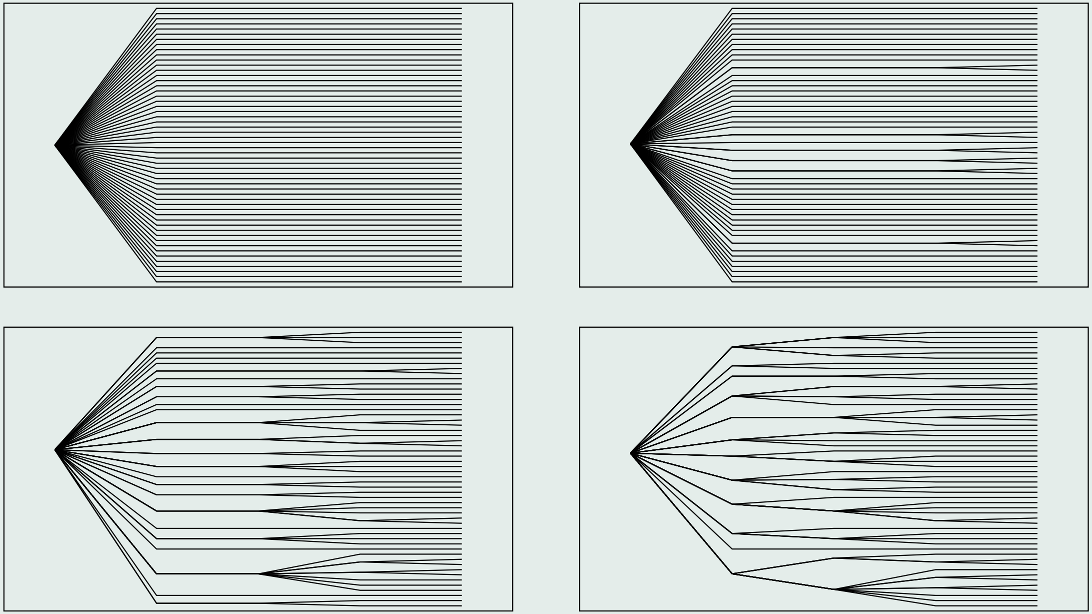
is satified. Hence, I_2 and $I_{2,i}$ are relatively large.
Step t: Determine disjoint index sets I_t^k and J_t^k, where $J_t^k = \bigcup_{i\in I_t^k} J_{t,i}^k$, such that $I_t^k \cup J_t^k = I_{t-1,k}$, and a stochastic process $\hat{\xi}^t$ having N scenarios $\hat{\xi}^{t,i}$ with probabilities p_i and such that

$$\|\hat{\xi}^t - \hat{\xi}^{t-1}\|_{r,t} \leq \varepsilon_t.$$

Set $I_t = \cup_k I_t^k$ and $I_{t,i} = \{i\} \cup J_{t,i}^k$, $i \in I_t^k$, for some k.

Example:

Recursive construction of a bivariate load-price scenario tree starting with $N = 58$ scenarios (illustration, time period: 1 year)



<Start Animation>

Decomposition of convex stochastic programs

Reference: Ruszczynski 03

First idea: Use of standard software for solving the stochastic program in scenario tree form !

But: Models are huge even for small trees and, in addition, special structures are not exploited !

⇒ Decomposition is a successful alternative in many (practical) situations.

Direct or primal decomposition approaches:

- starting point: Benders decomposition based on both *feasibility* and *objective* cuts;
- variants: regularization to avoid an explosion of the number of cuts and to delete inactive cuts; nesting when applied to solve the dynamic programming equations on subtrees recursively; stochastic cuts.

Dual decomposition approaches:

(i) Scenario decomposition by Lagrangian dualization of nonanticipativity constraints (solving the dual by bundle subgradient methods, augmented Lagrangian decomposition, variable or operator splitting methods);

(ii) nodal decomposition by Lagrangian dualization of dynamic constraints (same variants as in (i));

(iii) geographical decomposition by Lagrangian relaxation of coupling constraints (same variants as in (i)).

Presently, nested Benders decomposition, stochastic decomposition and scenario decomposition (based on augmented Lagrangians and on operator splitting) are mostly used for convex models !

Expected costs versus risk

Often minimizing expected costs is not the only objective; decisions should also enjoy minimal or bounded risk.
→mean-risk objective

Classical risk measure from financial mathematics:

Value-at-Risk ($p \in (0,1)$):

$$VaR_p(z) := -\min\{r \in \mathbb{R} : \mathbb{P}(z \leq r) \geq p\}$$

$VaR_p(z)$ does not enjoy pleasant properties !

Is there a general concept of risk measures ?

Axiomatic characterization of risk

Let $\mathcal{Z}$ denote a linear space of real random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We assume that $\mathcal{Z}$ contains the constants. A functional $\rho : \mathcal{Z} \to \mathbb{R}$ is called a risk measure if it satisfies the following two conditions for all z, $\tilde{z} \in \mathcal{Z}$:

(i) If $z \leq \tilde{z}$, then $\rho(z) \geq \rho(\tilde{z})$ (monotonicity).

(ii) For each $r \in \mathbb{R}$ we have $\rho(z + r) = \rho(z) - r$ (translation invariance).

A risk measure ρ is called convex if it satisfies the condition

$$\rho(\lambda z + (1 - \lambda)\tilde{z}) \leq \lambda \rho(z) + (1 - \lambda)\rho(\tilde{z})$$

for all z, $\tilde{z} \in \mathcal{Z}$ and $\lambda \in [0, 1]$.
A convex risk measure is called coherent if it is positively homogeneous, i.e., $\rho(\lambda z) = \lambda \rho(z)$ for all $\lambda \geq 0$ and $z \in \mathcal{Z}$.

References: Artzner/Delbaen/Eber/Heath 99, Föllmer/Schied 02

Examples:

(a) No convex risk measure: Value-at-Risk, standard deviation.

(b) **Semideviation of order** p ($\alpha \in (0,1], r \geq 1$):

$$\rho(z) := -\mathbb{E}[z] + \alpha\left(\mathbb{E}[(\max\{0, \mathbb{E}[z] - z\})^r]\right)^{\frac{1}{r}}$$

(c) **Conditional Value-at-Risk** ($p \in (0,1)$):

$$\begin{aligned} CVaR_p(z) &:= \min\{r + \frac{1}{1-p}\mathbb{E}[\max\{0, -z - r\}] : r \in \mathbb{R}\} \\ &= VaR_p(z) + \frac{1}{1-p}\mathbb{E}[\max\{0, -z - VaR_p(z)\}] \end{aligned}$$

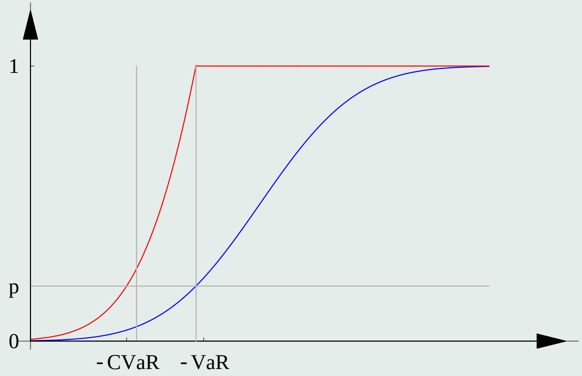
Advantage of $CVaR_p$: linearity properties are preserved.

(Rockafellar/Uryasev 02)

$$CVaR_p(z) := \text{mean of the tail distribution function } F_p$$

where $F_p(t) := \begin{cases} 1 & t \geq -VaR_p(z), \\ \frac{F(t)}{p} & t < -VaR_p(z) \end{cases}$ and

$F(t) := \mathbb{P}(\{z \leq t\})$ is the distribution function of z.



$VaR_p(z)$ and $CVaR_p(z)$ for a continuously distributed z

Home Page
Title Page
Contents
Go Back
Full Screen
Close
Quit

Polyhedral risk measures: One-period case

Definition:
A risk measure ρ on $\mathcal{Z}$ will be called polyhedral if there exist $k, l \in I\!N$, $a, c \in I\!R^k$, $q, w \in I\!R^l$, a polyhedral set $X \subseteq I\!R^k$ and a polyhedral cone $Y \subseteq I\!R^l$ such that

$$\rho(z) = \inf \{ \langle c, x \rangle + I\!E[\langle q, y \rangle] : \langle a, x \rangle + \langle w, y \rangle = z, x \in X, y \in Y \}$$

for each $z \in \mathcal{Z}$. Here, $I\!E$ denotes the expectation on $(\Omega, \mathcal{F}, I\!P)$ and $\langle \cdot, \cdot \rangle$ the scalar product on $I\!R^k$.

The notion *polyhedral* risk measure is motivated by the polyhedrality of $\rho(z)$ as a function of the scenarios of z if z is discrete. Origin: Properties of the Conditional value-at-risk CVaR.

How to generalize this concept to the multiperiod case ?

Multiperiod polyhedral risk measures

When (real) random variables $z_1, ..., z_T$ with $z_t \in L_p(\Omega, \mathcal{F}_t, I\!\!P)$, $1 \leq p \leq +\infty$, are considered that evolve over time and unveil the available information with the passing of time, it may become necessary to use multiperiod risk measures. We assume that a filtration of σ-fields $\mathcal{F}_t$, $t = 1, ..., T$, is given, i.e. $\mathcal{F}_t \subseteq \mathcal{F}_{t+1} \subseteq \mathcal{F}$, and that $\mathcal{F}_1 = \{\emptyset, \Omega\}$, i.e. that z_1 is always deterministic.

Definition: (Artzner et al. 01, 02)
A functional ρ on $\times_{t=1}^{T} L_p(\Omega, \mathcal{F}_t, I\!\!P)$ is called multiperiod risk measure if
(i) If $z_t \leq \tilde{z}_t$ a.s., $t = 1, ..., T$, then $\rho(z_1, ..., z_T) \geq \rho(\tilde{z}_1, ..., \tilde{z}_T)$ *(monotonicity)*,
(ii) For each $r \in I\!\!R$ we have $\rho(z_1 + r, ..., z_T + r) = \rho(z) - r$ *(translation invariance)*,
are satisfied. It is called a multiperiod coherent risk measure, if ρ is convex and positively homogeneous on $\times_{t=1}^{T} L_p(\Omega, \mathcal{F}_t, I\!\!P)$ in addition.

It is a natural idea to introduce risk measures as optimal values of certain multistage stochastic programs.

Definition: A multiperiod risk measure ρ on $\times_{t=1}^{T} L_p(\Omega, \mathcal{F}_t, I\!P)$ is called multiperiod polyhedral if there are $k_t \in I\!N$, $c_t \in I\!R^{k_t}$, $t = 1, \ldots, T$, $w_{t\tau} \in I\!R^{k_{t-\tau}}$, $t = 1, \ldots, T$, $\tau = 0, ..., t-1$, and polyhedral cones $Y_t \subset I\!R^{k_t}$, $t = 1, \ldots, T$, such that

$$\rho(z) = \inf \left\{ I\!E \left[\sum_{t=1}^{T} \langle c_t, y_t \rangle \right] \middle| \begin{array}{c} y_t \in L_p(\Omega, \mathcal{F}_t, I\!P; I\!R^{k_t}),\ y_t \in Y_t \\ \sum_{\tau=0}^{t-1} \langle w_{t,\tau}, y_{t-\tau} \rangle = z_t,\ \ t = 1, \ldots, T \end{array} \right\}.$$

Remark: A convex combination of (negative) expectation and of a multiperiod polyhedral risk measure is again a multiperiod polyhedral risk measure.
Our original multistage stochastic program then reads

$$\min \{ (1-\gamma) I\!E[F(\xi, x)] - \gamma \rho(F(\xi, x)) : x \in \mathcal{X}(\xi) \}$$

(mean-risk objective)

Theorem:

Let ρ be a functional on $\times_{t=1}^{T} L_p(\Omega, \mathcal{F}_t, I\!\!P)$ having the form in the previous definition. Assume

(i) complete recourse: $\langle w_{t,0}, Y_t \rangle = I\!\!R$ $(t = 1, ..., T)$,

(ii) dual feasibility: $\left\{ u \in I\!\!R^T : c_t + \sum_{\nu=t}^{T} u_\nu w_{\nu,\nu-t} \in -Y_t^* \right\} \neq \emptyset$,

where the sets Y_t^* are the (polyhedral) polar cones of Y_t.

Then ρ is Lipschitz continuous on $\times_{t=1}^{T} L_p(\Omega, \mathcal{F}_t, I\!\!P)$ and the following dual representation holds whenever $p \in (1, +\infty)$ and $\frac{1}{p} + \frac{1}{p'} = 1$:

$$\rho(z) = \sup \left\{ -I\!\!E \left[\sum_{t=1}^{T} \lambda_t z_t \right] \;\middle|\; \begin{array}{l} \lambda_t \in L_{p'}(\Omega, \mathcal{F}_t, I\!\!P), \;\; t = 1, \ldots, T \\ c_t + \sum_{\nu=t}^{T} I\!\!E\left[\lambda_\nu | \mathcal{F}_t\right] w_{\nu,\nu-t} \in -Y_t^* \end{array} \right\}$$

Corollary:

Let ρ be a functional on $\times_{t=1}^{T} L_p(\Omega, \mathcal{F}_t, I\!P)$, $p \in (1, \infty)$, having the form of a polyhedral risk measure. Let the above conditions (i) and (ii) be satisfied and assume that the set

$$\Lambda_\rho := \left\{ \lambda \in \times_{t=1}^{T} L_{p'}(\Omega, \mathcal{F}_t, I\!P) \;\middle|\; c_t + \sum_{\nu=t}^{T} I\!E\left[\lambda_\nu | \mathcal{F}_t\right] w_{\nu,\nu-t} \in -Y_t^* \right\}$$

is contained in

$$\mathcal{D}_T := \left\{ \lambda \in \times_{t=1}^{T} L_1(\Omega, \mathcal{F}_t, I\!P) \;\middle|\; \lambda_t \geq 0, \sum_{t=1}^{T} I\!E\left[\lambda_t\right] = 1 \right\}.$$

Then ρ is a multiperiod polyhedral and coherent risk measure.

Example: (Naive multiperiod extensions of CVaR)

A first idea is to incorporate the Conditional-Value-at-Risk at all time periods and to consider the weighted sum

$$\rho_1(z) := \sum_{t=2}^{T} \gamma_t CVaR_{\alpha_t}(z_t) = \sum_{t=2}^{T} \gamma_t \inf_{r \in I\!R} \left\{ r + \frac{1}{\alpha_t} I\!E\left[(r+z)^-\right] \right\}$$

with some weights $\gamma_t \geq 0$, $\sum_{t=1}^{T} \gamma_t = 1$, and some confidence levels $\alpha_2, \alpha_3, \ldots, \alpha_T \in (0,1)$. Here, $a^- = \max\{0, -a\}$.

Then ρ is a multiperiod polyhedral and coherent risk measure and the corresponding dual feasible set is of the form

$$\Lambda_1 = \left\{ \lambda \in \times_{t=1}^{T} L_{p'}(\Omega, \mathcal{F}_t, I\!P) \;\middle|\; \begin{array}{l} \lambda_1 = 0 \\ 0 \leq \lambda_t \leq \frac{\gamma_t}{\alpha_t} \; (t = 2, \ldots, T) \\ I\!E\left[\lambda_t\right] = \gamma_t \end{array} \right\}.$$

By interchanging sum and minimization one arrives at the variant

$$\rho_2(z) = \inf_{r \in I\!R} \left\{ r + \sum_{t=2}^{T} \beta_t I\!E\left[(z_t + r)^-\right] \right\}$$

of the above risk measure. Its dual representation is

$$\Lambda_2 = \left\{ \lambda \in \times_{t=1}^{T} L_{p'}(\Omega, \mathcal{F}_t, \mathbb{P}) \;\middle|\; \begin{array}{l} \lambda_1 = 0\,,\ \sum_{t=1}^{T} \mathbb{E}[\lambda_t] = 1, \\ 0 \le \lambda_t \le \beta_t \ \ (t = 2, ..., T) \end{array} \right\}.$$

However, both multiperiod coherent risk measures do not depend on the filtration, i.e. on the information flow.

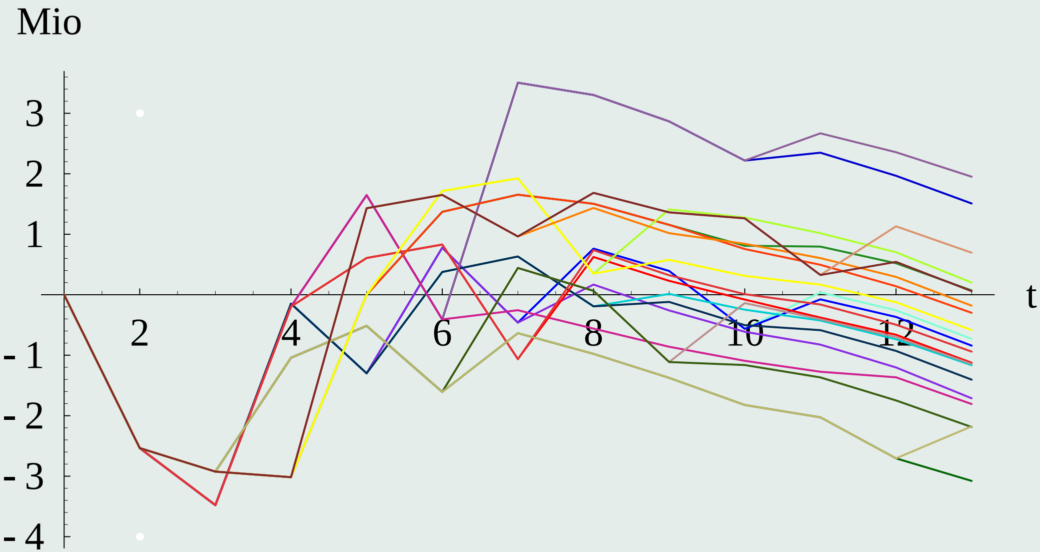
Example:

Multiperiod risk measure ρ_4 depending on the filtration

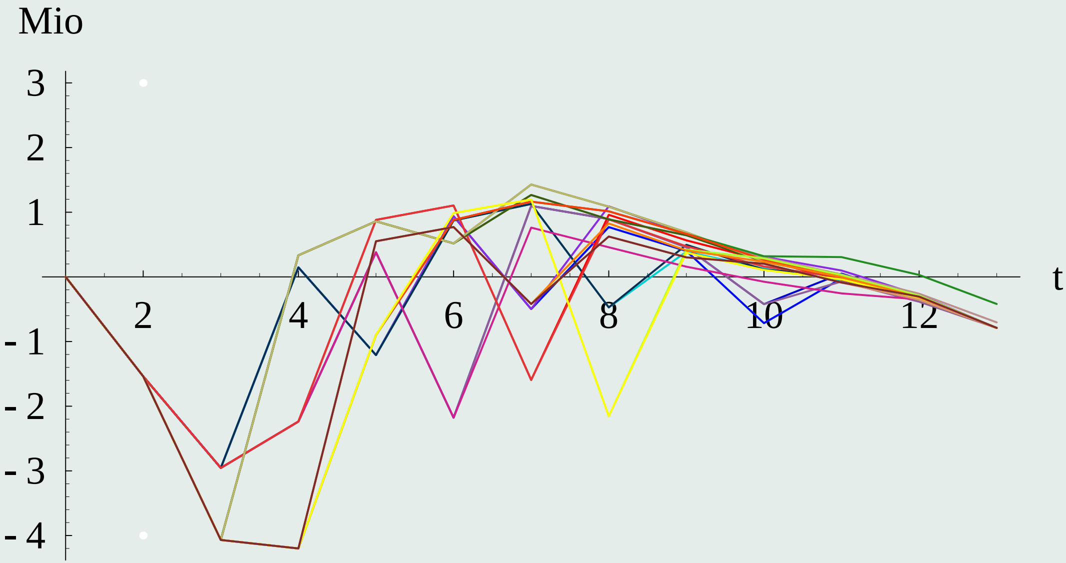
$$\Lambda_4 = \left\{ \lambda \in \times_{t=1}^{T} L_{p'}(\Omega, \mathcal{F}_t, \mathbb{P}) \;\middle|\; \begin{array}{l} \lambda_1 \equiv 0 \\ 0 \le \lambda_t \le \frac{1}{(T-1)\alpha_t} \ (t = 2, ..., T) \\ \lambda_t = \mathbb{E}\left[\lambda_{t+1} | \mathcal{F}_t\right] \ (t = 2, ..., T-1) \\ \mathbb{E}\left[\lambda_2\right] = ... = \mathbb{E}\left[\lambda_T\right] = \frac{1}{T-1} \end{array} \right\}$$

Electricity portfolio management (continued)

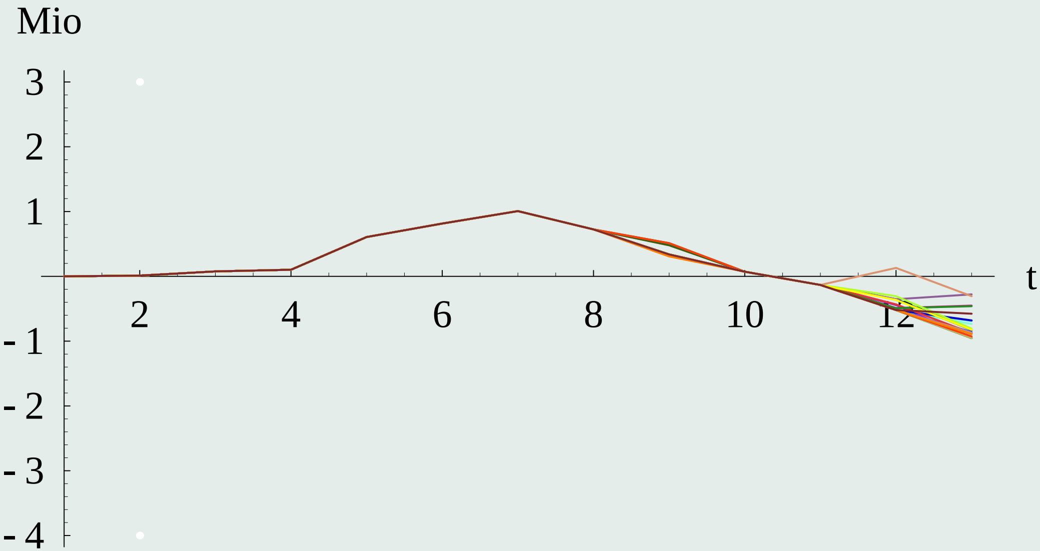
Test runs were performed on real-life data of the utility DREWAG Stadtwerke Dresden GmbH leading to a MIP containing about 2.4 million variables in case of 21 load-price scenarios. The objective function consists in a convex combination of expectation and (multiperiod) risk functional with a coefficient $\gamma \in [0, 1]$, where $\gamma = 0$ corresponds to no risk.



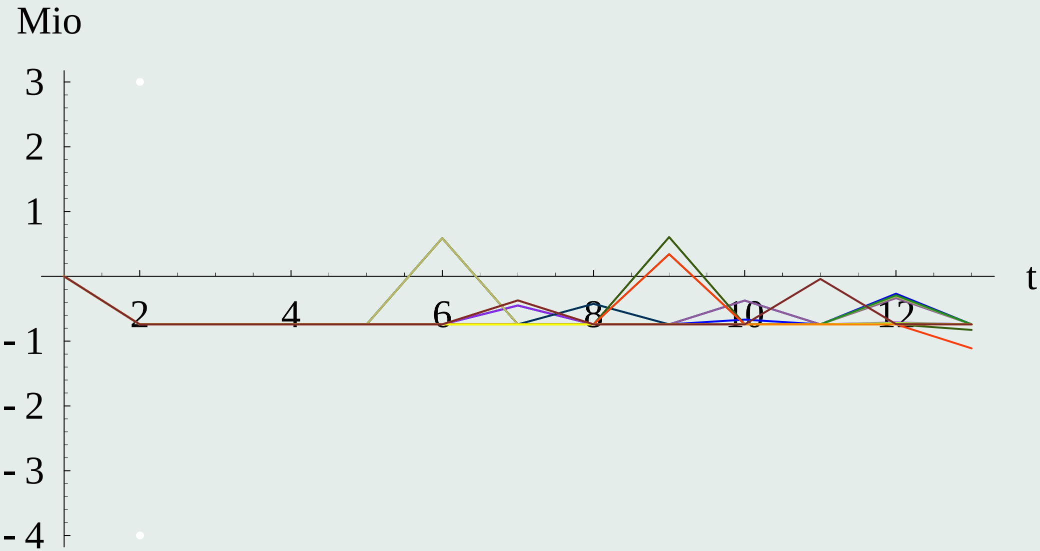
Total revenue and $\gamma = 0$



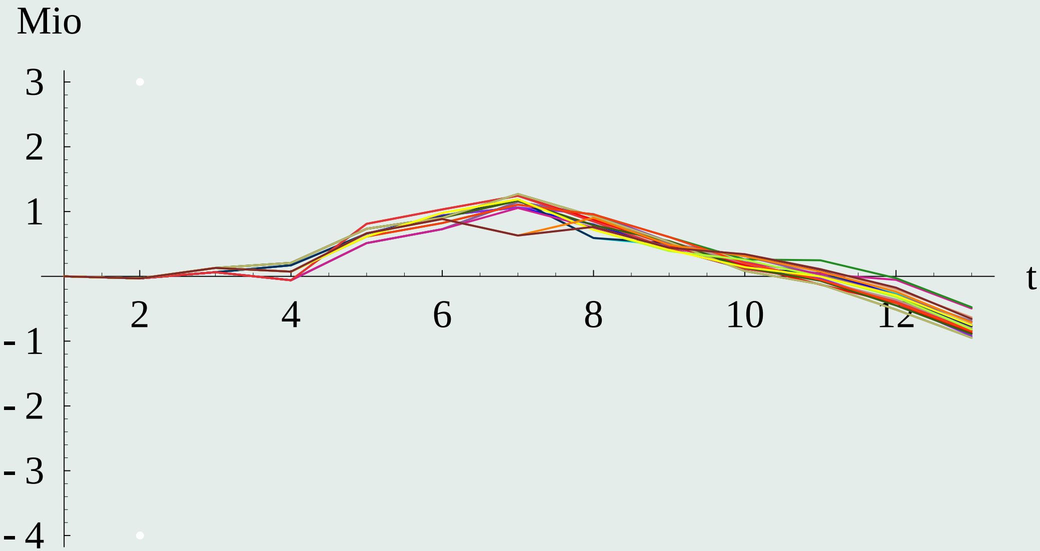
Total revenue with $CVaR_{0.05}$ and $\gamma = 0.25$



Total revenue with ρ_1 and $\gamma = 0.25$



Total revenue with ρ_2 and $\gamma = 0.25$



Total revenue with ρ_4 and $\gamma = 0.25$

References:

Artzner, P., Delbaen, F., Eber, J.-M., Heath, D.: Coherent measures of risk, *Mathematical Finance* 9 (1999), 203–228.

Dupačová, J., Consigli, G., Wallace, S.W.: Scenarios for multistage stochastic programs. *Annals Operations Res.* 100 (2000), 25–53.

Eichhorn, A., Römisch, W.: Polyhedral risk measures in stochastic programming, *SIAM Journal on Optimization* 2005.

Eichhorn, A., Römisch, W., Wegner, I.: Mean-risk optimization of electricity portfolios using multiperiod polyhedral risk measures, *IEEE St. Petersburg Power Tech* 2005.

Evstigneev, I.: Measurable selection and dynamic programming. *Mathematics of Operations Research* 1 (1976) 267–272.

Föllmer, H., Schied, A.: *Stochastic Finance: An Introduction in Discrete Time*, Walter de Gruyter, Berlin 2002 (Second Edition: 2004).

Heitsch, H., Römisch, W., Strugarek, C.: Stability of multistage stochastic programs, Preprint, 2005 and submitted to SIAM Journal on Optimization.

Heitsch, H., Römisch, W.: Scenario tree modeling for multistage stochastic programs, Preprint, 2005.

Rockafellar, R.T., Uryasev, S.: Conditional value at risk for general loss distributions, *Journal of Banking & Finance* 26 (2002), 1443–1471.

Ruszczyński, A., Shapiro, A. (eds.): Stochastic Programming, Handbook of Operations Research and Management Science Vol. 10, Elsevier, Amsterdam, 2003.