Stochastic Optimization of Electricity Portfolios: Scenario Tree Modeling and Risk Management

Andreas Eichhorn, Holger Heitsch, and Werner Römisch

Abstract We present recent developments in the field of stochastic programming with regard to application in power management. In particular we discuss issues of scenario tree modeling, i.e., appropriate discrete approximations of the underlying stochastic parameters. Moreover, we suggest risk avoidance strategies via the incooperation of so-called polyhedral risk functionals into stochastic programs. This approach, motivated through tractability of the resulting problems, is a constructive framework providing particular flexibility with respect to the dynamic aspects of risk.

1 Introduction

In medium term planning of electricity production and trading one is typically faced with uncertain parameters (such as energy demands and market prices in the future) that can be described reasonably by stochastic processes in discrete time. When time passes, additional information about the uncertain parameters may arrive (e.g., actual energy demands may be observed). Planning decisions can be made at different time stages based on the information available by then and on probabilistic information about the future (non-anticipativity), respectively. In terms of optimization, this situation is modeled by the framework of *multistage stochastic programming*; cf. Section 2. This framework allows to anticipate this dynamic decision structure appropriately. We refer to [6, 22, 23, 31, 49, 51, 56, 67, 70, 73] for exemplary

Andreas Eichhorn
Humboldt University, 10099 Berlin, Germany, e-mail: eichhorn@math.hu-berlin.de

Holger Heitsch
Humboldt University, 10099 Berlin, Germany, e-mail: heitsch@math.hu-berlin.de

Werner Römisch
Humboldt University, 10099 Berlin, Germany, e-mail: romisch@math.hu-berlin.de

case studies of stochastic programming in power planning. For a broad overview on stochastic programming models in energy we refer to [71].

However, a stochastic program incorporating a (discrete-time) stochastic process having infinite support (think of probability distributions with densities such as normal distributions) is an *infinite dimensional* optimization problem. For such problems a solution can hardly be found in practice. On the other hand, an a priori limitation to stochastic processes having finite support (think of discrete probability distributions) wouldn't be appropriate to many applications (including power planning). Therefore, for practical problem solving, approximation schemes are required such that general given stochastic processes are replaced by discrete ones with finite support (scenario trees) in such a way that the solutions of a stochastic program incorporating the discrete process are somehow close to the (unknown) solutions of the same program incorporating the original process. Such *scenario tree approximation* schemes will be one major topic in this chapter. Within the methods [36, 32, 34] to be presented, the closeness of the solutions will be ensured by means of suitable stability theorems for stochastic programs [62, 37].

The second major topic of this chapter will be the incorporation of risk management into power production planning and trading based on stochastic programming. In energy risk management, which is typically carried out *ex post* in practice, i.e., after power production planning, derivative products such as futures or options are traded in order to hedge a given production plan. However, decisions about buying and selling derivative products can also be made at different time stages, i.e., the dynamics of the decisions process here is of the same type as in production and (physical) power trading. Moreover, risk management and stochastic optimization rest upon the same type of stochastic framework. Hence, it is suggesting to integrate these two decision processes, i.e., to carry out simultaneously production planning, power trading, and trading of derivative products. E.g., in [3, 4] it has been demonstrated that such an integrated approach based on stochastic programming (electricity portfolio optimization) yields additional overall efficiency.

If risk avoidance is an objective of a stochastic optimization model, risk has to be quantified in a definite way. To this end, a suitable risk functional has to be chosen according to the economic requirements of a given application model. While in short term optimization simple *risk functionals* (risk measures) such as expected utility or Average-Value-at-Risk might be appropriate, the dynamic nature of risk has to be taken into account if medium or long term time horizons are considered. In this case, intermediate cash flows as well as the partial information that is revealed gradually at different time stages may have a significant impact on the risk. Therefore, *multi-period* risk functionals are required [2, 58]. Another important aspect of choosing a risk functional for the use in a stochastic programming model is a technical one: How much does a certain risk functional complicate the numerical resolution of a stochastic program? We argue that *polyhedral risk functionals* are a favorable choice with respect to the tractability of stochastic programs [18]. Also the stability theorems known for stochastic programs without any risk functional remain valid [17, 20] and, hence, there is a justification for scenario tree approximation schemes.

In addition, the class of polyhedral risk functionals provides flexibility, particularly in the multi-period situation.

This paper is organized as follows: after brief reviews on multistage stochastic programming in Section 2, we present scenario tree approximation algorithms in Section 3. After that, in Section 4, we discuss risk functionals with regard to their employment in electricity portfolio optimization. In particular, our concept of polyhedral risk functionals is presented in Section 4.2. Finally, we illustrate the effect of different polyhedral risk functionals with optimal cash flow curves from a medium term portfolio optimization model for a small power utility featuring a combined heat and power plant (CHP).

2 Multistage Stochastic Programming

For a broad presentation of stochastic programming we refer to [63] and [47]. Let the time stages of the planning horizon be denoted by $t = 1, \dots, T$ and let, for each of these time steps, a d-dimensional random vector ξ_t be given. This random vector represents the uncertain planning parameters that become known at stage t, e.g., electricity demands, market prices, inflows or wind power. We assume that ξ_1 is known from the beginning, i.e., a fixed vector in $\mathbb{R}^d$. For $\xi_2, \dots, \xi_T$, one may require the existence of certain statistical moments. The collection $\xi := (\xi_1, \dots, \xi_T)$ can be understood as multivariate discrete time stochastic process. Based on these notations a multistage stochastic program can be written as

$$\min_{x_1,\dots,x_T} \left\{ \mathbb{F}(z_1,\dots,z_T) \;\middle|\; \begin{array}{l} z_t := \sum_{s=1}^{t} b_s(\xi_s)\cdot x_s, \\ x_t = x_t(\xi_1,\dots,\xi_t),\ x_t \in X_t, \quad (t = 1,\dots,T) \\ \sum_{s=0}^{t-1} A_{t,s}(\xi_t) x_{t-s} = h_t(\xi_t) \end{array} \right\} \qquad (1)$$

where x_t is the decision vector for time stage t. The latter may depend and may only depend on the data observed until time t (non-anticipativity), i.e., on $\xi_1, \dots, \xi_t$, respectively. In particular, the components of x_1 are *here and now* decisions since x_1 may only depend on ξ_1 which was assumed to be deterministic. The decisions are subject to constraints: each x_t has to be chosen within a given set X_t. Typically, each X_t is a polyhedron or even a box, potentially further constrained by integer requirements. Moreover, there are dynamic constraints involving matrices $A_{t,s}$ and right-hand sides h_t which may depend on ξ_t in an affinely linear way. For the objective, we introduce wealth values z_t (accumulated revenues) for each time stage defined by a scalar product of x_t and (negative) cost coefficients b_t. The latter may also depend on ξ_t in an affinely linear way. Hence, each z_t is a random variable $(t = 2, \dots, T)$.

The objective functional $\mathbb{F}$ maps the entire stochastic wealth process (cash flow) to a single real number. The classical choice in stochastic optimization is the *expected value* $\mathbb{E}$ (mean) of the overall revenue z_T, i.e.,

$$\mathbb{F}(z_1,...,z_T) = -\mathbb{E}[z_T]$$

which is a *linear* functional. Linearity is a favorable property with respect to theoretical analysis as well as to the numerical resolution of problem (1). However, if risk is a relevant issue in the planning process, then some sort of nonlinearity is required in the objective (or, alternatively, in the constraints). In this presentation, we will discuss *mean-risk* objectives of the form

$$\mathbb{F}(z_1,...,z_T) = \gamma \cdot \rho(z_{t_1},...,z_{t_J}) - (1-\gamma) \cdot \mathbb{E}[z_T]$$

with $\gamma \in [0,1]$ and ρ being a *multi-period* risk functional applied to selected time steps $1 < t_1 < ... < t_J = T$ allowing for dynamic perspectives to risk.

Though the framework (1) considers the dynamics of the decision process, typically only the first stage solution x_1 is used in practice since it is scenario independent whereas x_t is scenario dependent for $t \geq 2$. When the second time stage $t = 2$ is reached in reality one may solve a new problem instance of (1) such that the time stages are shifted one step ahead (rolling horizon). However, x_1 is a good decision in the sense that it anticipates future decisions and uncertainty.

3 Scenario Tree Approximation

If the stochastic input process ξ has infinite support (infinitely many scenarios), the stochastic program (1) is an infinite dimensional optimization problem. For such problems a solution can hardly be found in practice. Therefore, ξ has to be approximated by another process having finite support [36, 34]. Such an approximation must exhibit tree structure in order to reflect the monotone information structure of ξ. It is desirable that *scenario tree* approximation schemes rely on approximation or stability results for (1) (cf., e.g., [37, 20, 52, 54]) that guarantee that the results of the approximate optimization problem are related to the (unknown) results of the original problem.

The recent stability result in [37] reveals that the multistage stochastic program (1) essentially depends on the probability distribution of the stochastic input process and on the implied information structure. Whereas the probability information is based on the characteristics of the individual scenarios and their probabilities, the information structure says something about the availability of information at different time stages within the optimization horizon. The scenario tree construction approach to be presented next consists of both approximation of the probability information and recovering the information structure [32].

Presently, there exist several approaches to generate scenario trees for multistage stochastic programs (see [14] for a survey). They are based on several different principles. We mention here (i) bound-based constructions [7, 16, 26, 50], (ii) Monte Carlo-based schemes [8, 68, 69] or Quasi-Monte Carlo-based methods [54, 55], (iii) (EVPI-based) sampling within decomposition schemes [10, 11, 46], (iv) the target/moment-matching principle [44, 45, 48], and (v) probability metric based approximations [30, 36, 41, 42, 57].

We propose a technique that belongs to the group (v) and is based on probability distances that are associated with the stability of the underlying stochastic program. The input of the method consists of a finite number of scenarios that are provided by the user and, say, are obtained from historical data by data analysis and resampling techniques or from statistical models calibrated to the relevant historical data. Sampling from historical time series or from statistical models (e.g., time series or regression models) is the most popular method for generating data scenarios. Statistical models for the data processes entering power operation and planning models have been proposed, e.g., in [5, 9, 21, 43, 65, 66, 67, 72].

The actual scenario tree construction method starts with a finite set of typically individual scenarios where we assume that these scenarios serve as approximation for the original probability information. Although such individual scenarios are convenient to represent a very good approximation of the underlying probability distribution the approximation with respect to the information structure could be poor. In particular, if sampling is performed from non-discrete random variables (e.g., random variables having a density function such as normal distributions), the information structure gets lost in general. But, fortunately, it can be reconstructed approximately by applying techniques of *optimal scenario reduction* successively.

3.1 Scenario reduction

The basis of our scenario tree generation methods is the reduction of scenarios modeling the stochastic data process in stochastic programs. We briefly describe this universal and general concept developed in [15, 33]. More recently, it was improved in [35] and extended to mixed-integer models in [40]. It was originally intended for non-dynamic (two-stage) stochastic programs and, hence, doesn't take into account the information structure when applied in a multistage framework. There are no special requirements on the stochastic data processes (e.g., on the dependence structure or the dimension of the process) or on the structure of the scenarios (e.g. tree-structured or not).

Scenario reduction may be desirable in some situations when the underlying optimization models already happen to be large scale and the incorporation of a large number of scenarios leads to huge programs and, hence, to high computation times. The idea of the scenario reduction framework in [15, 33] is to compute the (nearly) best approximation of the underlying discrete probability distribution by a measure with smaller support in terms of a probability metric which is associated to the stochastic program in a natural way by stability theory [62, 37]. Here, with regard to problem (1), the *norm* $\|\cdot\|_r$ will be used defined by

$$\|\xi\|_r := \left(\sum_{t=1}^{T} \mathbb{E}[|\xi_t|^r] \right)^{\frac{1}{r}}, \tag{2}$$

for a random vector $\xi = (\xi_1, \ldots, \xi_T)$ where $\mathbb{E}[\cdot]$ denotes expectation and $|\cdot|$ denotes some norm in $\mathbb{R}^d$. We aim at finding some $\hat{\xi}$ such that the distance $\|\xi - \hat{\xi}\|_r$ is small. The role of the parameter $r \geq 1$ is to ensure that the stochastic program (1) is well defined provided that $\|\xi\|_r < \infty$. The choice of r depends on the existing moments of the stochastic input process ξ coming across and on whether ξ enters the right-hand side h_t and/or the costs b_t and/or the (technology) matrices $A_{t,s}$. Typical choices are $r = 1$ if either right-hand sides or costs are random and $r = 2$ if both right-hand sides and costs are random. For further details we refer to [36].

The scenario reduction aims at reducing the number of scenarios in an optimal way. If $\xi = (\xi_1, \ldots, \xi_T)$ is a given random vector with finite support, i.e. represented by the scenarios $\xi^i = (\xi_1^i, \ldots, \xi_T^i)$ and probabilities p_i, $i = 1, \ldots, N$, then ones may be interested in deleting of a certain number of scenarios for computational reasons. So the main issue here is to find a suitable index subset $J \subset \{1, \ldots, N\}$. Moreover, if J is given, the question arises, what is the best approximation $\tilde{\xi} = (\tilde{\xi}_1, \ldots, \tilde{\xi}_T)$ of ξ supported only by the scenarios $\xi^j = (\xi_1^j, \ldots, \xi_T^j)$, $j \notin J$. The answer to the latter question, however, can be given directly: In our notation using (2) the problem reads

$$\min \left\{ \sum_{t=1}^{T} \sum_{i=1}^{N} p_i |\xi_t^i - \tilde{\xi}_t^i|^r \;\middle|\; (\tilde{\xi}_1^i, \ldots, \tilde{\xi}_T^i) \in \{(\xi_1^j, \ldots, \xi_T^j)\}_{j \notin J} \right\} \tag{3}$$

and if we define a mapping $j(i)$ such that

$$j(i) \in \arg\min_{j \notin J} \sum_{t=1}^{T} |\xi_t^i - \xi_t^j|^r, \qquad i \in J,$$

the minimum of (3) is attained for scenarios

$$(\hat{\xi}_1^i, \ldots, \hat{\xi}_T^i) = \begin{cases} (\xi_1^{j(i)}, \ldots, \xi_T^{j(i)}), & \text{if } i \in J, \\ (\xi_1^i, \ldots, \xi_T^i), & \text{if } i \notin J. \end{cases} \tag{4}$$

Hence, the best approximation of ξ is obtained for the random vector $\hat{\xi}$ supported by the scenarios $\hat{\xi}^j = (\hat{\xi}_1^j, \ldots, \hat{\xi}_T^j)$ and probabilities q_j, $j \notin J$, where we have

$$\|\xi - \hat{\xi}\|_r^r = \sum_{i \in J} p_i \min_{j \notin J} \sum_{t=1}^{T} |\xi_t^i - \xi_t^j|^r, \tag{5}$$

$$q_j = p_j + \sum_{\substack{i \in J \\ j(i) = j}} p_i. \tag{6}$$

In other words, the *redistribution rule* (6) consists in assigning the new probability to a preserved scenario to be equal to the sum of its former probability and of all probabilities of deleted scenarios that are closest to it.

More complicated is the actual problem of optimal scenario reduction, i.e., finding an optimal choice for the index set J with, say, prescribed cardinality. This problem represents a metric k-median problem which is known to be NP-hard, hence, (polynomial-time) approximation algorithms and heuristics become important. Simple heuristics may be derived from formula (5) for the approximation error. The result are two heuristic algorithms to compute nearly optimal index sets J with given cardinality n.

Algorithm 3.1 *(Forward selection)*

[*Initialization*]
Set $J := \{1, \dots, N\}$.

[*Index Selection*]
Determine an index $l \in J$ such that

$$l \in \arg\min_{u \in J} \sum_{k \in J \setminus \{u\}} p_k \min_{j \notin J \setminus \{u\}} \sum_{t=1}^{T} |\xi_t^k - \xi_t^j|^r$$

and set $J := J \setminus \{l\}$. If the cardinality of J equals n go to the termination step. Otherwise continue with a further index selection step.

[*Termination*]
Determine scenarios according to (4) and apply the redistribution rule (6) for the final index set J.

Algorithm 3.2 *(Backward reduction)*

[*Initialization*]
Set $J := \emptyset$.

[*Index Selection*]
Determine an index $u \notin J$ such that

$$u \in \arg\min_{l \notin J} \sum_{k \in J \cup \{l\}} p_k \min_{j \notin J \cup \{u\}} \sum_{t=1}^{T} |\xi_t^k - \xi_t^j|^r$$

and set $J := J \cup \{l\}$. If the cardinality of J equals n go to the termination step. Otherwise continue with a further index selection step.

[*Termination*]
Determine scenarios according to (4) and apply the redistribution rule (6) for the final index set J.

3.2 Scenario tree construction

Now we turn to the scenario tree construction, where we assume to have a sufficient large set of original or sample scenarios available. Let the (individual) scenarios and probabilities be denoted again by $\xi^i = (\xi_1^i, \ldots, \xi_T^i)$ and p_i, $i = 1, \ldots, N$, respectively, and we assume that $\xi_1^1 = \xi_1^2 = \ldots = \xi_1^N =: \xi_1^*$ (deterministic first stage). The random process with scenarios ξ^i and probabilities p_i, $i = 1, \ldots, N$, is denoted by ξ.

The idea of our tree construction method is to apply the above scenario reduction techniques successively in a specific way. In fact, by the approach of a *recursive scenario reduction* for increasing and decreasing time, respectively, both a forward and backward in time performing method can be derived.

The recursive scenario reduction acts as recovering the original information structure approximately. In the next two subsections we present a detailed description for two variants of our method, the forward and the backward approach. In the following let $I := \{1, \ldots, N\}$.

Forward tree construction

The forward tree construction is based on recursive scenario reduction applied to time horizons $\{1, \ldots, t\}$ with successively increasing time parameter t. It successively computes partitions of I of the form

$$C_t := \{C_t^1, \ldots, C_t^{k_t}\}, \qquad k_t \in \mathbb{N},$$

such that for every t the partitions satisfy the conditions

$$C_t^k \cap C_t^{k'} = \emptyset \quad \text{for} \quad k \neq k', \quad \text{and} \quad \bigcup_{k=1}^{k_t} C_t^k = I.$$

The elements of a partition C_t are called (scenario) clusters. The following forward algorithm allows to generate different scenario tree processes depending on the parameter settings for the reductions in each step.

Algorithm 3.3 *(Forward construction)*

[*Initialization*]
Define $C_1 = \{I\}$ and set $t := 2$.

[*Cluster computation*]
Let be $C_{t-1} = \{C_{t-1}^1, \ldots, C_{t-1}^{k_{t-1}}\}$. For every $k \in \{1, \ldots, k_{t-1}\}$ subject the scenario subsets $\{\xi_t^i\}_{i \in C_{t-1}^k}$ to a scenario reduction with respect to the t-th components only. This yields disjoint subsets of remaining and deleted scenarios I_t^k and J_t^k, respectively. Next, obtain the mappings $j_t^k : J_t^k \to I_t^k$ such that

$$j_t^k(i) \in \arg\min_{j \in I_t^k} |\xi_t^i - \xi_t^j|, \quad i \in J_t^k,$$

according to the reduction procedure (cf. Section 3.1). Finally, define an overall mapping $\alpha_t : I \to I$ by

$$\alpha_t(i) = \begin{cases} j_t^k(i), & i \in J_t^k \text{ for some } k = 1, \ldots, k_{t-1}, \\ i, & \text{otherwise.} \end{cases} \tag{7}$$

A new partition at t is defined now by

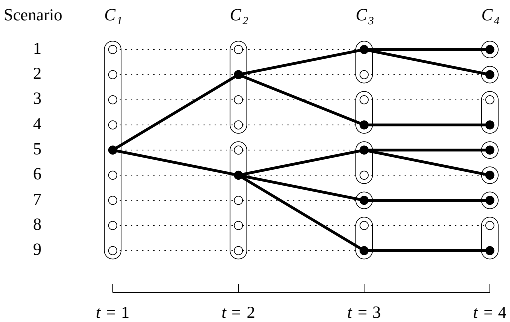
$$C_t := \left\{ \alpha_t^{-1}(i) \,\middle|\, i \in I_t^k, k = 1, \ldots, k_{t-1} \right\}$$

which is in fact a refinement of the partion C_{t-1}. If $t < T$ set $t := t+1$ and continue with a further cluster computation step, otherwise go to the termination step.

[*Termination*]
According to the partition set C_T and the mappings (7) define a scenario tree process ξ_{tr} supported by the scenarios

$$\xi_{\text{tr}}^k = \left(\xi_1^*, \xi_2^{\alpha_2(i)}, \ldots, \xi_t^{\alpha_t(i)}, \ldots, \xi_T^{\alpha_T(i)} \right) \quad \text{for any} \quad i \in C_T^k,$$

and probabilities $q_k := \sum_{i \in C_T^k} p_i$, for each $k = 1, \ldots, k_T$.



Scenario	α_2	α_3	α_4
1	2	1	1
2	2	1	2
3	2	4	4
4	2	4	4
5	6	5	5
6	6	5	6
7	6	7	7
8	6	9	9
9	6	9	9

Fig. 1 Illustration of the clustering by the forward scenario tree construction algorithm 3.3 (left) and the mappings α_t (right) for an example.

We want to conclude this subsection with two remarks regarding algorithm 3.3. Firstly, both heuristic algorithms from Section 3.1 may be used to compute the scenario reduction within the cluster computation step. Secondly, according to (5) the error of the cluster computation step t is

$$\mathrm{err}_t := \sum_{k=1}^{k_{t-1}} \sum_{i \in J_t^k} p_i \min_{j \in I_t^k} |\xi_t^i - \xi_t^j|^r .$$

Furthermore, as shown in [32, Proposition 6.6], the estimate

$$\|\xi - \xi_{\mathrm{tr}}\|_r \le \left(\sum_{t=2}^{T} \mathrm{err}_t \right)^{\frac{1}{r}}$$

holds for the total approximation error. The latter estimate allows to control the construction process by prescribing tolerances for err_t for every $t = 2, \ldots, T$.

Backward tree construction

The idea of the backward scenario tree construction consists in recursive scenario reduction on $\{1, \ldots, t\}$ for decreasing t, $t = T, \ldots, 2$. That results in a chain of index sets

$$I_1 := \{i_*\} \subseteq I_2 \subseteq \cdots \subseteq I_{t-1} \subseteq I_t \subseteq \cdots \subseteq I_T \subseteq I = \{1, \ldots, N\}$$

representing an increasing number of scenario realizations over the time horizon. The following backward algorithm is the counterpart of the forward algorithm 3.3 and allows again to generate different scenario tree processes depending on the parameters for the reduction steps.

Algorithm 3.4 *(Backward construction)*

[*Initialization*]
Define $I_{T+1} := \{1, \ldots, N\}$ and $p_{T+1}^i := p_i$ for all $i \in I_{T+1}$. Further, let be α_{T+1} the identity on I_{T+1} and set $t := T$.

[*Reduction*]
Subject the scenario subset $\{(\xi_1^i, \ldots, \xi_t^i)\}_{i \in I_{t+1}}$ with probabilities p_{t+1}^i ($i \in I_{t+1}$) to a scenario reduction which results in a index set I_t of remaining scenarios with $I_t \subseteq I_{t+1}$. Let be $J_t := I_{t+1} \setminus I_t$. According to the reduction procedure (cf. Section 3.1) obtain a mapping $j_t : J_t \to I_t$ such that

$$j_t(i) \in \arg\min_{j \in I_t} \sum_{k=1}^{t} |\xi_k^i - \xi_k^j|^r , \quad i \in J_t .$$

Define a mapping $\alpha_t : I \to I_t$ by

$$\alpha_t(i) = \begin{cases} j_t(\alpha_{t+1}(i)), & \alpha_{t+1}(i) \in J_t , \\ \alpha_{t+1}(i), & \text{otherwise}, \end{cases} \tag{8}$$

for all $i \in I$. Finally, set probabilities with respect to the redistribution (6), i.e.,

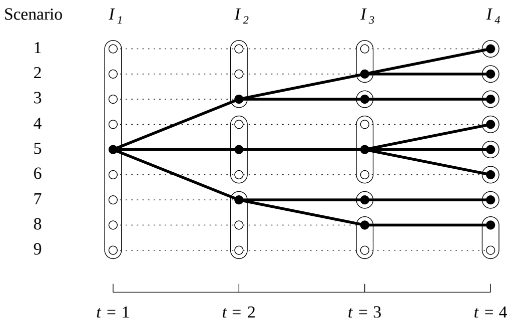
$$p_t^j := p_{t+1}^j + \sum_{\substack{i \in J_t \\ j_t(i)=j}} p_{t+1}^i.$$

If $t > 2$ set $t := t-1$ and continue with performing a further reduction step, otherwise go to the termination step.

[*Termination*]
According to the obtained index set I_T and the mappings (8) define a scenario tree process ξ_{tr} supported by the scenarios

$$\xi_{\text{tr}}^i = \left(\xi_1^*, \xi_2^{\alpha_2(i)}, \ldots, \xi_t^{\alpha_t(i)}, \ldots, \xi_T^{\alpha_T(i)}\right)$$

and probabilities $q_i := p_T^i$, for all $i \in I_T$.



Scenario	α_2	α_3	α_4
1	3	2	1
2	3	2	2
3	3	3	3
4	5	5	4
5	5	5	5
6	5	5	6
7	7	7	7
8	7	8	8
9	7	8	8

Fig. 2 Illustration of the recursive backward scenario tree construction algorithm 3.4 (left) and the mappings α_t (right) for an example. Note that the backward construction yields a clustering similar to the forward variant. Black circles correspond to scenarios contained in the index sets I_t.

We note again that the specific scenario reduction can be performed with both heuristic algorithms of Section 3.1. A similar estimate for the total approximation error $\|\xi - \xi_{\text{tr}}\|_r$ holds as for the forward variant. For details we refer to [36, Section 4.1]. Finally, we mention that all algorithms discussed in this section are implemented and available in GAMS-SCENRED (see `www.gams.com`).

4 Risk Avoidance via Risk Functionals

Risk avoidance requirements in optimization are typically achieved by the employment of a certain *risk functional*. Alternatively, *risk probabilistic constraints* or *risk stochastic dominance constraints* with respect to a given acceptable strategy may be incorporated, i.e., (1) may adopt constraints of the form

$$\mathbb{P}(z_T \leq z_{\text{ref}}) \geq \alpha \qquad \text{or} \qquad z_T \preceq z_{\text{ref}}$$

with (high) probability $\alpha \in (0,1]$ and some acceptable reference level z_{ref} or some acceptable reference distribution z_{ref} and a suitable stochastic ordering relation "$\preceq$". For the relevant background of probabilistic constraints we refer to the survey [59] and to [38, 39]. For a systematic introduction into stochastic order relations we refer to [53] and for recent work on incorporating stochastic dominance constraints into optimization models to [12, 13].

In this section, we focus on risk functionals ρ with regard to their utilization in the objective $\mathbb{F}$ of (1) as suggested, e.g., in [64]; cf. Section 2. Clearly, the choice of ρ is a very critical issue. On the one hand, the output of a stochastic program is highly sensitive to this choice. One is interested in a functional that makes sense from an economic point of view for a given situation. On the other hand, the choice of the risk functional has a significant impact on the numerical tractability of (1) (where ξ may be approximated by a finite scenario tree according to Section 3). Note that reasonable risk functionals are never linear (like the expectation functional), but some of them may be reformulated as infimal value of a linear stochastic program (see Section 4.2).

4.1 Axiomatic Frameworks for Risk Functionals

Basically, a risk functional in a probabilistic framework ought to measure the danger of ending up at low wealth in the future and/or the degree of uncertainty one is faced with in this regard. However, the question what is a good or what is the best risk functional from the viewpoint of economic reasoning cannot be answered in general. The answer depends strongly on the application context. However, various axioms have been postulated by various authors in the last decade that can be interpreted as minimum requirements.

A distinction can be drawn between single-period risk functionals evaluating a stochastic wealth value z_T at one single point in time T and multi-period risk functionals evaluating ones wealth at different time stages, say, $t_1 < t_2 \ldots < t_J$. The latter are typically required for medium or long term models. Of course, from a technical point of view single-period risk measurement can be understood as a special case of multi-period risk measurement. However, with regard to single-period risk functionals there is a relatively high degree of agreement about their preferable properties [1, 25, 58], whereas the multi-period case raises a lot more questions. In the following we pass directly to multi-period risk measurement having single-period risk measurement as a special case in mind.

Let a certain linear space $\mathscr{Z}$ of discrete-time random processes be given. A random process $z \in \mathscr{Z}$ is basically a collection of random variables $z = (z_{t_1}, \ldots, z_{t_J})$ representing wealth at different time stages. The realization of z_{t_j} is completely known at time t_j, respectively. Moreover, at time stage t_j one may have more information about $(z_{t_{j+1}}, \ldots, z_{t_J})$ than before (at earlier time stages $t_1, \ldots, t_{j-1}$). Therefore,

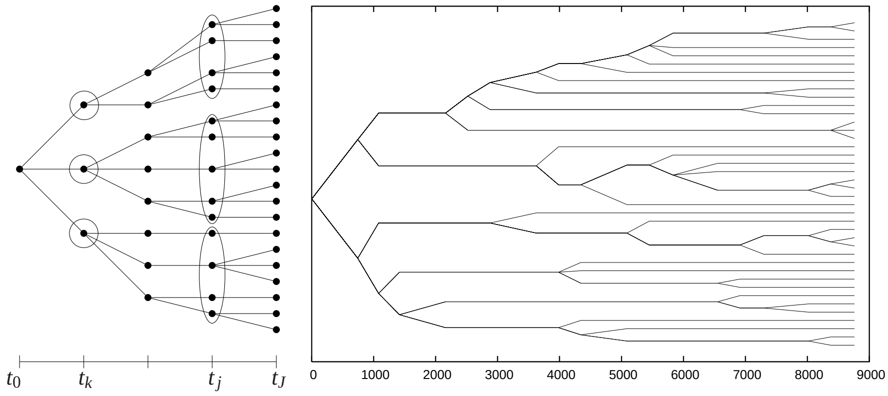


Fig. 3 Left: Illustration of the (discretized) information structure of a stochastic wealth process $z_{t_1},...,z_{t_J}$. At each time stage t_k and in each scenario one can look at subsequent time steps $t_j > t_k$ and consider the discrete (sub-) distribution of z_{t_j} seen from this node. Right: Branching structure of an exemplary scenario tree with 40 scenarios, $T = 8760$ time steps, and approx. $150,000$ nodes used for the simulations in Section 4.3. There is a node at each time step for each scenario.

a multi-period risk functional may also take into account *conditional distributions* with respect to some *underlying information structure*. In the context of the multistage stochastic program (1), the underlying information structure is given in a natural way through the stochastic input process $\xi = (\xi_1,...,\xi_T)$. Namely, it holds that $z_{t_j} = z_{t_j}(\xi_1,...,\xi_{t_j})$, i.e., z is *adapted* to ξ. In particular, if ξ is discrete, i.e., if ξ is given by a finite scenario tree as in Section 3, then also z is discrete, i.e., z is given by the values $z^i_{t_j}$ ($j = 1,...,J$, $i = 1,...,N$) on the scenario tree. However, we will consider general (not necessarily discretely distributed) random processes here and we also write $z^i_{t_j}$ for a realization (outcome) of random variable z_{t_j} even if the number of scenarios (possible outcomes) is infinite.

From a formal point of view, a risk functional ρ is just a mapping

$$z = (z_{t_1},...,z_{t_J}) \in \mathscr{Z} \quad \mapsto \quad \rho(z) \in \mathbb{R}$$

i.e., a real number is assigned to each random wealth process from $\mathscr{Z}$. One may require the existence of certain statistical moments for the random variables z_{t_j} ($j = 1,...,J$), i.e., $\mathbb{E}[|z_{t_j}|^p] < \infty$ for some $p \geq 1$. The J time steps are denoted by $t_1,...,t_J$ to indicate that, with regard to problem (1), they may be only a subset of the time steps $t = 1,...,T$ of the underlying information structure. We assume $1 < t_1 < ... < t_J = T$ and set $t_0 = 1$ for convenience. The special case of *single-period* risk functionals occurs if only one time step is taken into account ($J = 1$, $t_J = T$).

Now, a high number $\rho(z)$ should indicate a high risk of ending up at low wealth values z_{t_j}, a low (negative) number $\rho(z)$ indicates a small risk. In [2] the number $\rho(z)$ is interpreted as the *minimal amount μ of additionally required risk-free capital* such that the process $z_{t_1} + \mu,...,z_{t_J} + \mu$ is acceptable. Such and other intuitions have been formalized by various authors in terms of axioms. As a start, we cite the first two axioms from [2], in addition to convexity as the third axiom. A functional ρ is called a multi-period *convex (capital) risk functional* if the following properties hold for all stochastic wealth processes $z = (z_{t_1}, \dots, z_{t_J})$ and $\tilde{z} = (\tilde{z}_{t_1}, \dots, \tilde{z}_{t_J})$ in $\mathscr{Z}$, and for all (non-random) real numbers μ:

- *Monotonicity*: If $z_{t_j} \le \tilde{z}_{t_j}$ in any case for $j = 1, \dots, J$, then it holds that $\rho(z) \ge \rho(\tilde{z})$.
- *Cash invariance*: It holds that $\rho(z_{t_1} + \mu, \dots, z_{t_J} + \mu) = \rho(z_{t_1}, \dots, z_{t_J}) - \mu$.
- *Convexity*: If $0 \le \mu \le 1$ it holds that $\rho(\mu z + (1-\mu)\tilde{z}) \le \mu\rho(z) + (1-\mu)\rho(\tilde{z})$.

The formulation "$z_{t_j} \le \tilde{z}_{t_j}$ in any case" means that in each scenario i it holds that $z^i_{t_j} \le \tilde{z}^i_{t_j}$. The convexity property is motivated by the idea that *diversification* might decrease risk but does never increase it. Sometimes the following property is also required for all $z \in \mathscr{Z}$:

- *Positive homogeneity*: For each $\mu \ge 0$ it holds that $\rho(\mu z) = \mu\rho(z)$.

Note that, for the single-period case $J = 1$, the first three properties coincide with the classical axioms from [1, 24, 27]. A positively homogeneous convex risk functional is called *coherent* in [1, 2]. We note, however, that other authors do not require positive homogeneity, but claim that risk should rather grow overproportionally, i.e., $\rho(\mu z) > \mu\rho(z)$ for $\mu > 1$; cf. [28, 25]. Clearly, the negative expectation functional $-\mathbb{E}$ is a (single-period) coherent risk functional, whereas the α-Value-at-Risk given by $\mathrm{VaR}_\alpha(z) = -\inf\{\mu \in \mathbb{R} : \mathbb{P}(z \le \mu) > \alpha\}$ is not since it is not convex [1].

For the multi-period case ($J > 1$) the three above axioms are only a basis admitting many degrees of freedom. There are several aspects of risk that could be measured. First of all, one may want to measure the chance of ending up at very low values $z^i_{t_j}$ at each time since very low values can mean bankruptcy (liquidity considerations). In addition, one may want to measure the degree of uncertainty one is faced with at each time step; cf. Fig. 3 (left). A situation where, at some time t_k, one can be sure about the future development of ones wealth z_{t_j} ($j > k$) may be preferred to a situation continuing uncertainty. E.g., low values z_{t_j} may be tolerable if one can be sure that later the wealth is higher again. Hence, one may want to take into account not only the marginal distributions of $z_{t_1}, \dots, z_{t_J}$ but also their chronological order, their interdependence, and the underlying information structure. Therefore, a multi-period risk functional may also take into account the conditional distributions of z_{t_j} given the information $\xi_1, \dots, \xi_s$ with $s = 1, \dots, t_j - 1$ ($j = 1, \dots, J$); cf. Fig. 3 (left). Clearly, there are quite a lot of those conditional distributions and the question arises which ones are relevant and how to weight them reasonably.

The above axioms leave all these questions open. In our opinion, general answers can not be given, the requirements depend strongly on the application context, e.g., on the time horizon, on the size and capital reserves of the respective company, on the broadness of the model, etc. Some stronger versions of cash invariance (translation equivariance) have been suggested, e.g., in [28, 58], tailored to certain situations. However, the framework of polyhedral risk functionals in the next section is particularly flexible with respect to the dynamic aspects.

4.2 Polyhedral Risk Functionals

The basic motivation for polyhedral risk functionals is a technical, but important one. Consider the optimization problem (1). It is basically linear or mixed-integer linear if the objective functional is linear, i.e., $\mathbb{F} = -\mathbb{E}$. In this case it is well tractable by various solution and decomposition methods. However, if $\mathbb{F}$ incorporates a risk functional ρ it is no longer linear since risk functionals are essentially nonlinear by nature. Decomposition structures may get lost and solution methods may take much longer or may even fail. To avoid the worst possible situation one should choose ρ to be at least convex [64]. Then (1) is at least a convex problem (except possible integer constraints contained in X_t), hence, any local optimum is always the global one. As discussed above, convexity is in accordance with economic considerations and axiomatic frameworks.

Now, the framework of polyhedral risk functionals [18, 17] goes one step beyond convexity: polyhedral risk functionals maintain linearity structures even though they are nonlinear functionals. Namely, a polyhedral risk functional ρ is given by

$$\rho(z) = \inf \left\{ \mathbb{E}\left[\sum_{j=0}^{J} c_j \cdot y_j\right] \middle| \begin{array}{l} y_j = y_j(\xi_1, \ldots, \xi_{t_j}) \in Y_j, \\ \sum_{k=0}^{j} V_{j,k} y_{j-k} = r_j \ (j = 0, \ldots, J), \\ \sum_{k=0}^{j} w_{j,k} \cdot y_{j-k} = z_{t_j} \ (j = 1, \ldots, J) \end{array} \right\} \tag{9}$$

where $z = (z_{t_1}, \ldots, z_{t_J})$ denotes a stochastic wealth process being non-anticipative with respect to ξ, i.e., $z_t = z_t(\xi_1, \ldots, \xi_t)$. The notation $\inf\{\,.\,\}$ refers to the infimum. The definition includes fixed polyhedral cones Y_j (e.g., $\mathbb{R}_+ \times \ldots \times \mathbb{R}_+$) in some Euclidean spaces $\mathbb{R}^{k_j}$, fixed vectors c_j, r_j $w_{j,k}$, and matrices $V_{j,k}$, which have to be chosen appropriately. We will give examples for these parameters below. However, functionals ρ defined by (9) are always convex [18, 17].

Observe that problem (9) is more or less of the form (1), i.e., the risk of a stochastic wealth process z is given by the optimal value of a stochastic program. Moreover, if (9) is inserted into the objective of (1) (i.e., $\mathbb{F} = \rho$), one is faced with two nested minimizations which, of course, can be carried out jointly. This yields the equivalent optimization problem

$$\min \left\{ \mathbb{E}\left[\sum_{j=0}^{J} c_j \cdot y_j\right] \middle| \begin{array}{l} x_t = x_t(\xi_1, \ldots, \xi_t) \in X_t, \ \sum_{s=0}^{t-1} A_{t,s}(\xi_t) x_{t-s} = h_t(\xi_t) \\ (t = 1, \ldots, T), \\ y_j = y_j(\xi_1, \ldots, \xi_{t_j}) \in Y_j, \ \sum_{k=0}^{j} V_{j,k} y_{j-k} = r_j, \\ \sum_{k=0}^{j} w_{j,k} \cdot y_{j-k} = \sum_{s=1}^{t_j} b_s(\xi_s) \cdot x_s \ (j = 1, \ldots, J) \end{array} \right\}$$

which is a stochastic program of the form (1) with *linear* objective. In other words: the nonlinearity of the risk functional ρ is transformed into additional variables and additional linear constraints in (1). This means that decomposition schemes and solution algorithms known for linear or mixed-integer linear stochastic programs can also be used for (1) with $\mathbb{F} = \rho$. In particular, as discussed in [18, Section 4.2], dual decomposition schemes (like scenario and geographical decomposition) carry over to the situation with $\mathbb{F} = \rho$. However, the dual problem in Lagrangian relaxation of coupling constraints (also called geographical or component decomposition) contains polyhedral constraints originating from the dual representation of ρ.

Furthermore, the linear combination of two polyhedral risk functionals is again a polyhedral risk functional (cf. [17, Section 3.2.4]). In particular, the case

$$\mathbb{F}(z) = \gamma\rho(z) + \sum_{k=1}^{J} \mu_k \mathbb{E}\left[z_{t_k}\right]$$

with a polyhedral risk functional ρ (with parameters c_j, $w_{j,k}$ etc.) and real numbers γ and μ_k, $k = 1,\ldots,J$, can be fully reduced to the case ρ by setting

$$\hat{c}_j := \gamma c_j + \sum_{k=j}^{J} \mu_k w_{k,k-j} \quad (j = 0,\ldots,J)$$

for the vectors in the objective function of the representation (9) of $\mathbb{F}$ and letting all remaining parameters of ρ unchanged.

Another important advantage of polyhedral risk functionals is that they also behave favorable to stability with respect to (finite) approximations of the stochastic input process ξ [20]. Hence, there is a justification for the employment of the scenario tree approximation schemes from Section 3.

It remains to discuss the issue of choosing the parameters c_j, h_j, $w_{j,k}$, $V_{j,k}$, Y_j in (9) such that the resulting functional ρ is indeed a reasonable risk functional satisfying, e.g., the axioms presented in the previous section. To this end, several criteria for these axioms have been deduced in [18, 17] involving duality theory from convex analysis. However, here we restrict the presentation to examples.

First, we consider the case $J = 1$, i.e., single-period risk functionals evaluating only the distribution of the final value z_T (total revenue). The starting point of the concept of polyhedral risk functionals was the well-known risk functional *Average-Value-at-Risk* AVaR_α at some probability level $\alpha \in (0,1)$. It is also known as Conditional-Value-at-Risk (cf. [61]), but as suggested in [25] we prefer the name Average-Value-at-Risk according to its definition

$$\mathrm{AVaR}_\alpha(z) := \tfrac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_\beta(z)d\beta$$

as an average of Value-at-Risks and avoid any conflict with the use of conditional distributions within VaR and AVaR (see [58] for such constructions). The Average-Value-at-Risk is a (single-period) coherent risk functional which is broadly accepted. $\mathrm{AVaR}_\alpha(z_T)$ can be interpreted as the mean (expectation) of the α-tail distribution of z_T, i.e., the mean of the distribution of z_T below the α-quantile of z_T. It has been observed in [61] that AVaR_α can be represented by

$$\begin{aligned}\mathrm{AVaR}_\alpha(z_T) &= \inf_{y_0\in\mathbb{R}}\left\{y_0 + \tfrac{1}{\alpha}\mathbb{E}[(y_0+z_T)^-]\right\}\\ &= \inf\left\{ y_0 + \tfrac{1}{\alpha}\mathbb{E}[y_{1,2}] \;\middle|\; \begin{array}{l} y_0 \in \mathbb{R}, \\ y_1 = y_1(\xi_1,\ldots,\xi_T) \in \mathbb{R}^2_+, \\ y_0 + z_T = y_{1,1} - y_{1,2} \end{array} \right\}\end{aligned}$$

where $(\,.\,)^-$ denotes the negative part of a real number, i.e., $a^- = \max\{0,-a\}$ for $a \in \mathbb{R}$. The second representation is deduced from the first one by introducing stochastic variables y_1 for the positive and the negative part of $y_0 + z_T$. Hence, AVaR_α is of the form (9) with $J = 1$, $c_0 = 1$, $c_1 = (0, \frac{1}{\alpha})$, $w_{1,0} = (1,-1)$, $w_{1,1} = -1$, $Y_0 = \mathbb{R}$, $Y_1 = \mathbb{R}^2_+ = \mathbb{R}_+ \times \mathbb{R}_+$, and $h_0 = h_1 = V_{0,0} = V_{1,0} = V_{1,1} = 0$. Thus, it is a (single-period) polyhedral risk functional.

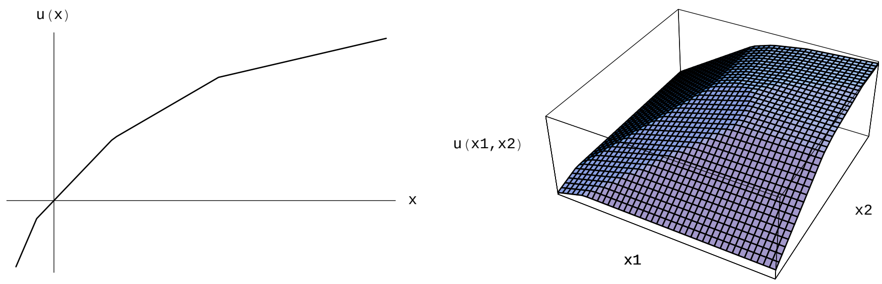


Fig. 4 Monotone and piecewise linear concave utility functions, single-period (left) and two-period ($J = 2$) (right)

Another single-period example for a polyhedral risk functional (satisfying monotonicity and convexity) is expected utility, i.e., $\rho_u(z_T) := -\mathbb{E}[u(z_T)]$ with a non-decreasing concave utility function $u : \mathbb{R} \to \mathbb{R}$; cf. [25]. Typically, nonlinear functions such as $u(x) = 1 - e^{-\beta x}$ with some fixed $\beta > 0$ are used. Of course, in such cases ρ_u is not a polyhedral risk functional. However, in situations where the domain of z_T can be bounded a priori, it makes sense to use piecewise linear functions for u (see Fig. 4, left). Then, according to the infimum representation of piecewise linear convex functions [60, Corollary 19.1.2], it holds that

$$\rho_u(z_T) = \inf\left\{ \mathbb{E}\left[c \cdot y_1\right] \,\middle|\, \begin{array}{l} y_1 = y_1(\xi_1, ..., \xi_T) \in \mathbb{R}^{n+2}_+, \\ w \cdot y_1 = z_T,\ \sum_{i=1}^n y_{1,i} = 1 \end{array} \right\}$$

where n is the number of cusps of u, $w_1, \ldots, w_n$ are the x-coordinates of the cusps, and $c_i = -u(w_i)$ $(i = 1, .., n)$. Thus, ρ_u is a polyhedral risk functional. This approach can also be generalized to the multi-period situation in an obvious way by specifying a (concave) utility function $u : \mathbb{R}^J \to \mathbb{R}$ (see Fig. 4, right). However, specifying an adequate utility function may be difficult in practice, in particular in the multi-period case. Furthermore, expected utility is not cash invariant (cf. Section 4.1), neither in the single-period nor in the multi-period case. Therefore we will focus on generalizations of AVaR_α to the multi-period case.

In the multi-period case $J > 1$, the framework of polyhedral risk functionals allows to model different perspectives to the relations between different time stages. In [18, 19, 17, 58], several examples extending AVaR_α to the multi-period situation in different ways have been constructed via a bottom-up approach using duality

$\mathrm{AVaR}_\alpha(z_0)$	polyhedral representation (9)
$z_0 = \frac{1}{J}\sum_{j=1}^J z_{t_j}$	$\inf\left\{ \frac{1}{J}\left(y_0 + \sum_{j=1}^J \frac{1}{\alpha}\mathbb{E}\left[y_{j,2}\right]\right) \middle\vert \begin{array}{l} y_0 \in \mathbb{R},\, y_j = y_j(\xi_1,\ldots\xi_{t_j}) \in \mathbb{R}\times\mathbb{R}_+ \\ (j=1,\ldots,J-1), \\ y_J = y_J(\xi_1,\ldots\xi_{t_J}) \in \mathbb{R}_+\times\mathbb{R}_+ \\ y_{j,1} - y_{j,2} = z_{t_j} + y_{j-1,1}\ (j=1,\ldots,J) \end{array} \right\}$
$z_0 = \min\{z_{t_1},\ldots,z_{t_J}\}$	$\inf\left\{ y_0 + \frac{1}{\alpha}\mathbb{E}\left[y_{J,2}\right] \middle\vert \begin{array}{l} y_0 \in \mathbb{R},\, y_j = y_j(\xi_1,\ldots\xi_{t_j}) \in \mathbb{R}_+\times\mathbb{R}_+\times\mathbb{R}_+ \\ (j=1,\ldots,J), \\ y_{1,2} - y_{1,3} = 0,\, y_{j,2} - y_{j,3} - y_{j-1,2} = 0\ (j=2,\ldots,J), \\ y_{j,1} - y_{j,2} - y_0 = z_{t_j}\ (j=1,\ldots,J) \end{array} \right\}$

Table 1 Representation (9) of $\mathrm{AVaR}_\alpha(\frac{1}{J}\sum_{j=1}^J z_{t_j})$ (above) and $\mathrm{AVaR}_\alpha(\min\{z_{t_1},\ldots,z_{t_J}\})$ (below)

theory from convex analysis. Here, we restrict the presentation to the most obvious extensions that can be written in the form $\mathrm{AVaR}_\alpha(z_0)$ with a suitable *mixture* z_0 of $(z_{t_1},\ldots,z_{t_J})$. We consider

$$z_0 = \frac{1}{J}\sum_{j=1}^J z_{t_j} \qquad \text{or} \qquad z_0 = \min\{z_{t_1},\ldots,z_{t_J}\}$$

where "$\sum$" and "min" are understood scenariowise, i.e., $z_0^i = \frac{1}{J}\sum_{j=1}^J z_{t_j}^i$ respectively $z_0^i = \min\{z_{t_1}^i,\ldots,z_{t_J}^i\}$ for each scenario i. Hence, in both cases the risk functional $\mathrm{AVaR}_\alpha(z_0)$ depends on the multivariate distribution of $(z_{t_1},\ldots,z_{t_J})$.

As shown in Table 1, both $\mathrm{AVaR}_\alpha(\frac{1}{J}\sum_{j=1}^J z_{t_j})$ and $\mathrm{AVaR}_\alpha(\min\{z_{t_1},\ldots,z_{t_J}\})$ can be written in the form (9), i.e., they are indeed *multi-period polyhedral risk functionals*. Moreover, they are multi-period coherent risk functionals in the sense of Section 4.1. Clearly, the latter of the two functionals is the most reasonable multi-period extension of AVaR with regard to liquidity considerations, since AVaR is applied to the respectively lowest wealth values in each scenario; this worst case approach has also been suggested in [2, Section 4].

4.3 Illustrative Simulation Results

Finally, we illustrate the effects of different polyhedral risk functionals by presenting some optimal wealth processes from an electricity portfolio optimization model [21, 19]. This model is of the form (1), it considers the one year planning problem of a *municipal power utility*, i.e., a price-taking retailer serving heat and power demands of a certain region; see Fig. 5. It is assumed that the utility features a combined heat and power (CHP) plant that can serve the heat demand completely but the power demand only in part. In addition, the utility can buy power at the day-ahead spot market of some power exchange, e.g., the European Energy Exchange EEX. Morover, the utility can trade monthly (purely financial) futures (e.g., Phelix futures at EEX).

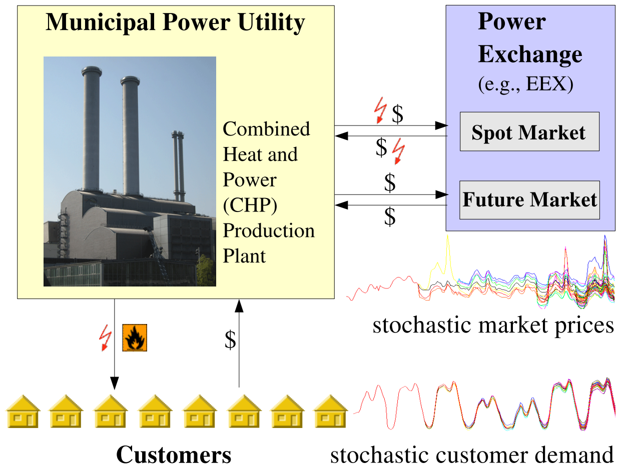


Fig. 5 Schematic diagram for a power production planning and trading model under demand and price uncertainty (portfolio optimization).

The objective $\mathbb{F}$ of this model is a mean-risk objective as discussed in Section 1 incorporating a polyhedral risk functional ρ and the expected total revenue $\mathbb{E}[z_T]$; the weighting factor is set to $\gamma = 0.9$. Time horizon is one year in hourly discretization, i.e., $T = 8760$. Time series models for the uncertain input data (demands and prices) have been set up (see [21] for details) and approximated according to Section 3 by a finite scenario tree consisting of 40 scenarios; see Fig. 3 (right). The scenario tree has been obtained by the forward construction procedure of Algorithm 3.3. It represents the uncertainty well enough on the one hand, and, on the other hand the moderate size of the tree guarantees computational tractability. For the risk time steps t_j we use 11 PM at the last trading day of each week ($j = 1,\ldots,J = 52$). Note that, due to the limited number of branches in the tree, a finer resolution for the risk time steps doesn't make sense here. The resulting optimization problem is very large-scale, however, it is numerically tractable due to the favorable nature of polyhedral risk functionals. In particular, since we modeled the CHP plant without integer variables, it is a linear program (LP) which could be solved by ILOG CPLEX in about one hour.

In Fig. 6 (as well as in Fig. 7) the optimal cash flows are displayed, i.e., the wealth values z_t for each time step $t = 1,\ldots,T$ and each scenario, obtained from optimization runs with different mean-risk objectives. The price parameters have been set such that the effects of the risk functionals may be observed well although these settings yield negative cash values. These families of curves differ in shape due to different policies of future trading induced by the different risk functionals; see Fig. 8. Setting $\gamma = 0$ (no risk functional at all) yields high spread for z_T and there is no future trading at all (since we worked with fair future prices). Using $\mathrm{AVaR}_\alpha(z_T)$ ($\gamma = 0.9$) yields low spread for z_T but low values and high spread at $t < T$. This shows that, for the situation here, single-period risk functionals are not appropriate.

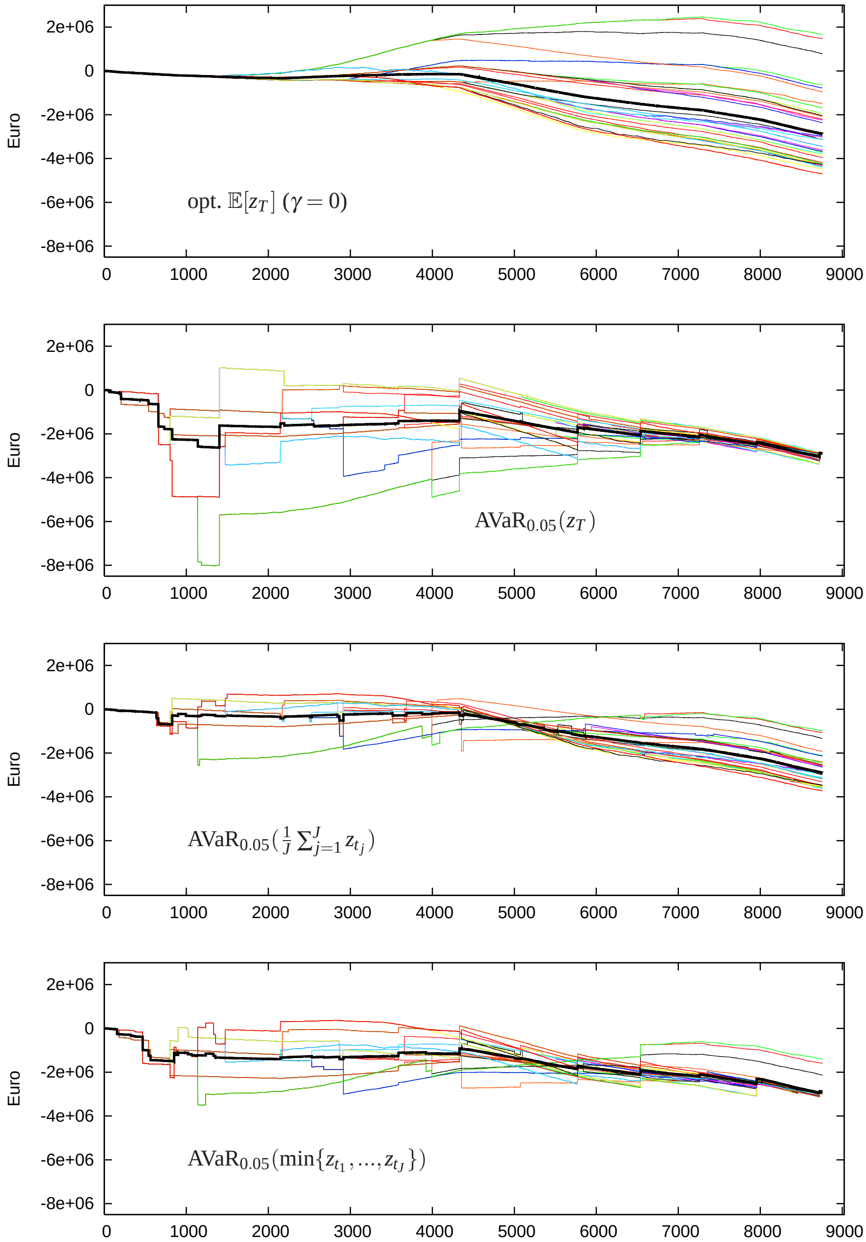


Fig. 6 Optimal cash values z_t (wealth) over time $(t = 1,\ldots,T)$ with respect to different risk functionals. Each curve in a graph represents one of the 40 scenarios. The expected value of the cash flows is displayed in black.

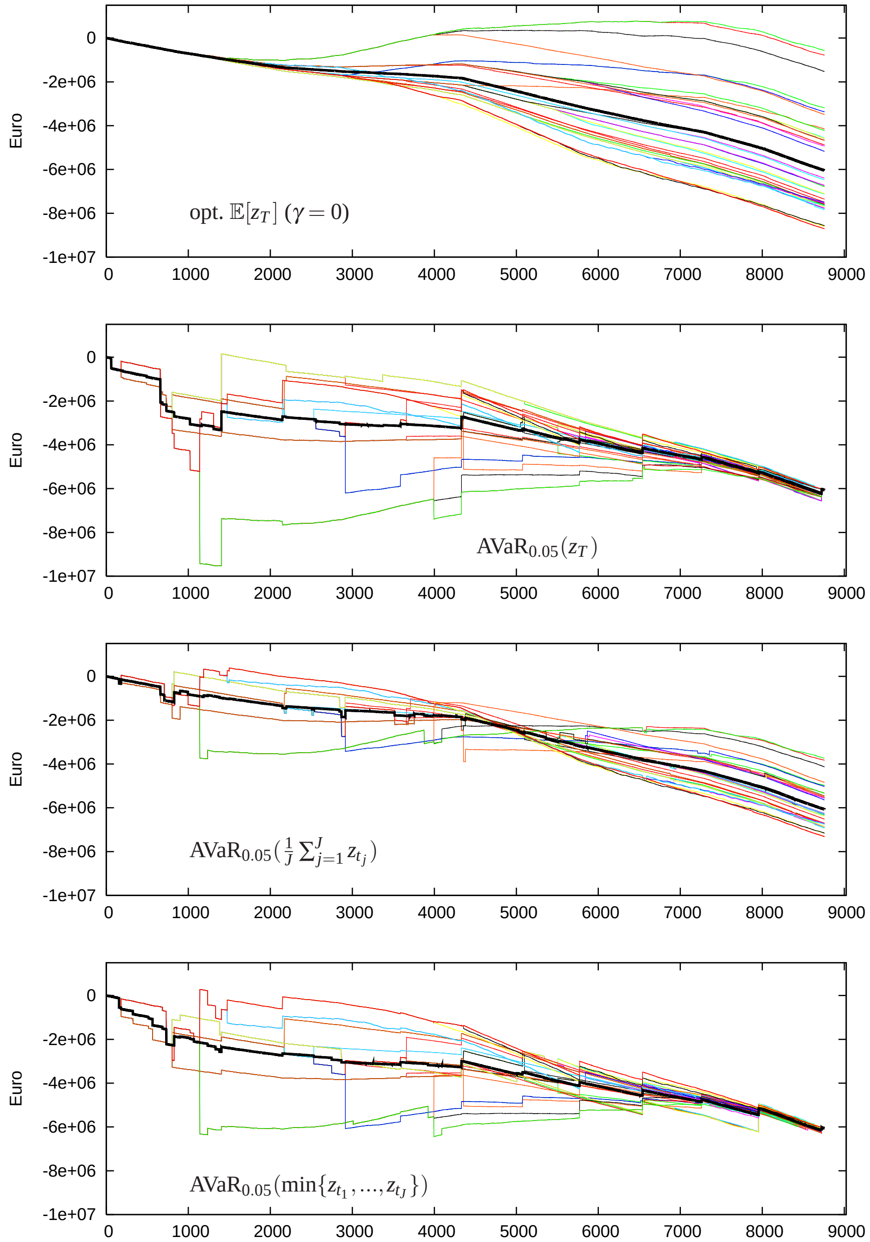


Fig. 7 Optimal cash values z_t (wealth) over time ($t = 1,\ldots,T$) with respect to different risk functionals. For these calculations, slightly higher fuel cost parameters have been used such that the graphs demonstrate the nature of the risk functionals best.

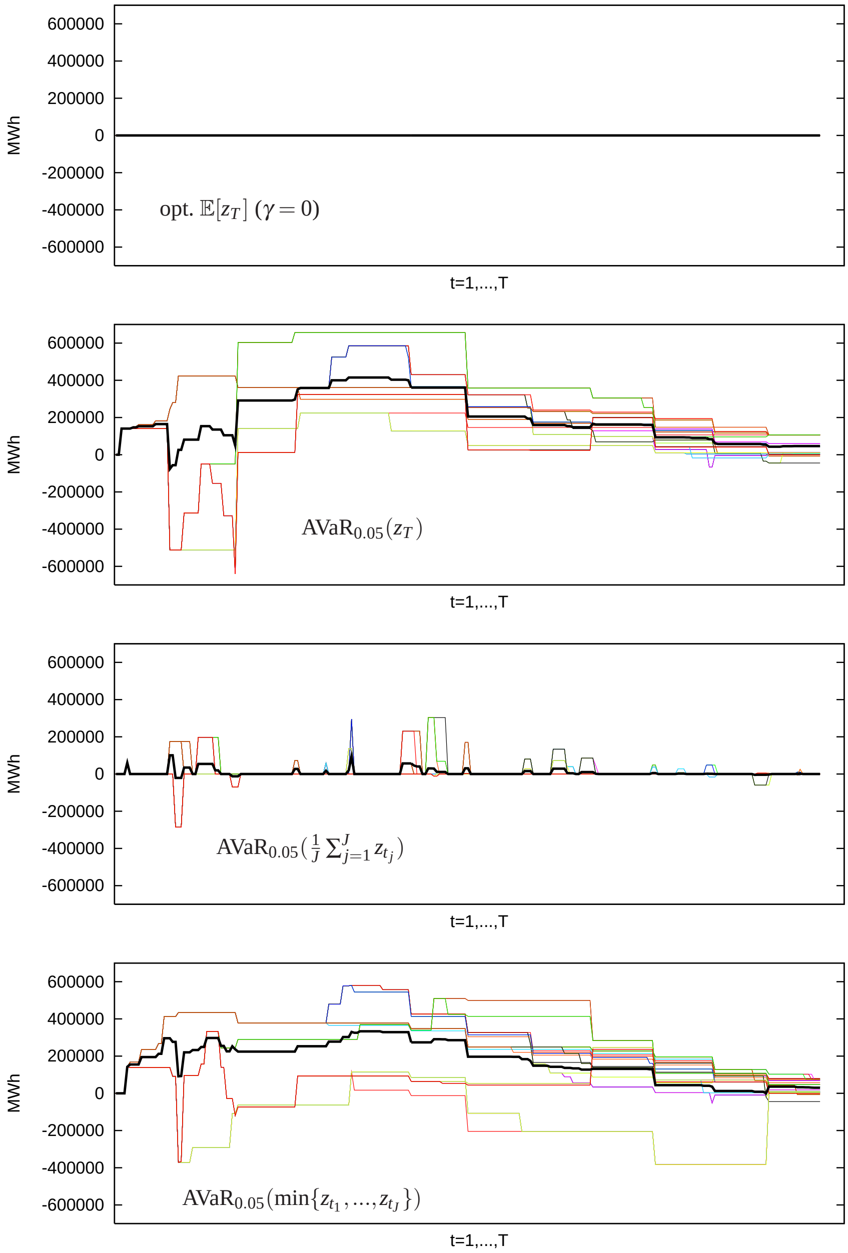


Fig. 8 Optimal future stock over time with respect to different polyhedral risk functionals. The expected value of the future stock over time is displayed in black.

The employment of multi-period polyhedral risk functionals yields spread that is better distributed over time. However, the way how this is achieved is different: The functional $\mathrm{AVaR}_\alpha(\min\{z_{t_1},\dots,z_{t_J}\})$ aims at finding a level y_0 as high as possible such that the curves rarely fall below that level, whereas $\mathrm{AVaR}_\alpha(\frac{1}{J}\sum_{j=1}^{J} z_{t_j})$ aims at equal spread at all times. In the latter case, futures are held only for short time periods, whereas in the other cases futures are held longer. Finally, we note that the effects of the risk functionals cost only less than 1% of the expected overall revenue $\mathbb{E}[z_T]$.

5 Conclusions

Multi-stage stochastic programming models are discussed as mathematical tools for dealing with uncertain (future) parameters in electricity portfolio and risk management. Since statistical information on the parameters (like demands, spot prices, inflows or wind speed) is often available, stochastic models may be set up for them so that scenarios of the future uncertainty are made available. To model the information flow over time, the scenarios need to be tree-structured. For this reason a general methodology is presented in Section 3 that allows to generate scenario trees out of the given set of scenarios. The general method is based on stability argument for multistage stochastic programs and does not require further knowledge on the underlying multivariate probability distribution. The method is flexible and allows to generate scenario trees whose size enables a good approximation of the underlying probability distribution on the one hand and allows for reasonable running times of the optimization software on the other hand. Implementations of these scenario tree generation algorithms are available in GAMS-SCENRED.

A second issue discussed in the paper is risk management via the incorporation of risk functionals into the objective. This allows maximizing expected revenue und minimizing risk simultaneously. Since risk functionals are nonlinear by definition, a natural requirement consists in preserving computational tractability of the (mixed-integer) optimization models and, hence, in reasonable running times of the software. Therefore, a class of risk functionals is presented in Section 4.2 that allow a formulation as linear (stochastic) program. Hence, if the risk functional (measure) belongs to this class, the resulting optimization model does not contain additional nonlinearities. If the expected revenue maximization model is (mixed-integer) linear, the linearity is preserved. A few examples of such *polyhedral* risk functionals are provided for multi-period situations, i.e., if the risk evolves over time and requires to rely on multivariate probability distributions. The simulation study in Section 4.3 for the electricity portfolio management of a price-taking retailer provides some insight into the risk minimization process by trading electricity derivatives. It turns out that the risk can be reduced considerably for less than 1% of the expected overall revenue.

Acknowledgements This work was supported by the DFG Research Center MATHEON *Mathematics for Key Technologies* in Berlin (http://www.matheon.de) and by the German Ministry of Education and Research (BMBF) under the grant 03SF0312E. The authors wish to thank the anonymous referees for their helpful comments.

References

1. Artzner, P., Delbaen, F., Eber, J.M., Heath, D.: Coherent measures of risk. Mathematical Finance **9**, 203–228 (1999)
2. Artzner, P., Delbaen, F., Eber, J.M., Heath, D., Ku, H.: Coherent multiperiod risk adjusted values and Bellman's principle. Annals of Operations Research **152**, 5–22 (2007)
3. Blaesig, B.: Risikomanagement in der Stromerzeugungs- und Handelsplanung, *Aachener Beiträge zur Energieversorgung*, vol. 113. Klinkenberg, Aachen, Germany (2007). PhD Thesis
4. Blaesig, B., Haubrich, H.J.: Methods of risk management in the generation and trading planning. In: IEEE St. Petersburg PowerTech Proceedings (2005)
5. Burger, M., Klar, B., Müller, A., Schindlmayr, G.: A spot market model for pricing derivatives in electricity markets. Quantitative Finance **4**, 109–122 (2004)
6. Cabero, J., Baillo, A., Cerisola, S., Ventosa, M., Garcia-Alcalde, A., Beran, F., Relano, G.: A medium-term integrated risk management model for a hydrothermal generation company. IEEE Transactions on Power Systems **20**, 1379–1388 (2005)
7. Casey, M.S., Sen, S.: The scenario generation algorithm for multistage stochastic linear programming. Mathematics of Operations Research **30**, 615–631 (2005)
8. Chiralaksanakul, A., Morton, D.P.: Assessing policy quality in multi-stage stochastic programming. Preprint, Stochastic Programming E-Print Series, www.speps.org (2004)
9. Clewlow, L., Strickland, C.: Energy Derivatives: Pricing and Risk Management. Lacima, London (2000)
10. Corvera Poiré, X.: Model generation and sampling algorithms for dynamic stochastic programming. Ph.D. thesis, Department of Mathematics, University of Essex, UK (1995)
11. Dempster, M.A.H.: Sequential importance sampling algorithms for dynamic stochastic programming. Zapiski Nauchnykh Seminarov POMI **312**, 94–129 (2004)
12. Dentcheva, D., Ruszczyński, A.: Optimization with stochastic dominance constraints. SIAM Journal on Optimization **14**, 548–566 (2003)
13. Dentcheva, D., Ruszczyński, A.: Portfolio optimization with stochastic dominance constraints. Journal of Banking & Finance **30**, 433–451 (2006)
14. Dupačová, J., Consigli, G., Wallace, S.W.: Scenarios for multistage stochastic programs. Annals of Operations Research **100**, 25–53 (2000)
15. Dupačová, J., Gröwe-Kuska, N., Römisch, W.: Scenario reduction in stochastic programming: An approach using probability metrics. Mathematical Programming **95**, 493–511 (2003)
16. Edirisinghe, N.C.P.: Bound-based approximations in multistage stochastic programming: Nonanticipativity aggregation. Annals of Operations Research **85**, 103–127 (1999)
17. Eichhorn, A.: Stochastic programming recourse models: Approximation, risk aversion, applications in energy. Ph.D. thesis, Department of Mathematics, Humboldt University, Berlin (2007)
18. Eichhorn, A., Römisch, W.: Polyhedral risk measures in stochastic programming. SIAM Journal on Optimization **16**, 69–95 (2005)
19. Eichhorn, A., Römisch, W.: Mean-risk optimization models for electricity portfolio management. In: Proceedings of the 9th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS). Stockholm, Sweden (2006)
20. Eichhorn, A., Römisch, W.: Stability of multistage stochastic programs incorporating polyhedral risk measures. Optimization **57**, 295–318 (2008)

21. Eichhorn, A., Römisch, W., Wegner, I.: Mean-risk optimization of electricity portfolios using multiperiod polyhedral risk measures. In: IEEE St. Petersburg PowerTech Proceedings (2005)
22. Fleten, S.E., Kristoffersen, T.K.: Short-term hydropower production planning by stochastic programming. Computers and Operations Research **35**, 2656–2671 (2008)
23. Fleten, S.E., Wallace, S.W., Ziemba, W.T.: Hedging electricity portfolios via stochastic programming. In: Greengard and Ruszczyński [29], pp. 71–93
24. Föllmer, H., Schied, A.: Convex measures of risk and trading constraints. Finance and Stochastics **6**, 429–447 (2002)
25. Föllmer, H., Schied, A.: Stochastic Finance. An Introduction in Discrete Time, *De Gruyter Studies in Mathematics*, vol. 27, 2nd edn. Walter de Gruyter, Berlin (2004)
26. Frauendorfer, K.: Barycentric scenario trees in convex multistage stochastic programming. Mathematical Programming **75**, 277–293 (1996)
27. Frittelli, M., Rosazza Gianin, E.: Putting order in risk measures. Journal of Banking & Finance **26**, 1473–1486 (2002)
28. Frittelli, M., Scandolo, G.: Risk measures and capital requirements for processes. Mathematical Finance **16**, 589–612 (2005)
29. Greengard, C., Ruszczyński, A. (eds.): Decision Making under Uncertainty: Energy and Power, *IMA Volumes in Mathematics and its Applications*, vol. 128. Springer, New York (2002)
30. Gröwe-Kuska, N., Heitsch, H., Römisch, W.: Scenario reduction and scenario tree construction for power management problems. In: A. Borghetti, C.A. Nucci, M. Paolone (eds.) IEEE Bologna PowerTech Proceedings (2003)
31. Gröwe-Kuska, N., Kiwiel, K.C., Nowak, M.P., Römisch, W., Wegner, I.: Power management in a hydro-thermal system under uncertainty by Lagrangian relaxation. In: Greengard and Ruszczyński [29], pp. 39–70
32. Heitsch, H.: Stabilität und Approximation stochastischer Optimierungsprobleme. Ph.D. thesis, Department of Mathematics, Humboldt University, Berlin (2007)
33. Heitsch, H., Römisch, W.: Scenario reduction algorithms in stochastic programming. Computational Optimization and Applications **24**, 187–206 (2003)
34. Heitsch, H., Römisch, W.: Generation of multivariate scenario trees to model stochasticity in power management. In: IEEE St. Petersburg PowerTech Proceedings (2005)
35. Heitsch, H., Römisch, W.: A note on scenario reduction for two-stage stochastic programs. Operations Research Letters **35**, 731–738 (2007)
36. Heitsch, H., Römisch, W.: Scenario tree modeling for multistage stochastic programs. Mathematical Programming **118**, 371–406 (2009)
37. Heitsch, H., Römisch, W., Strugarek, C.: Stability of multistage stochastic programs. SIAM Journal on Optimization **17**, 511–525 (2006)
38. Henrion, R.: Some remarks on value-at-risk optimization. Management Science and Engineering Management **1**, 111–118 (2006)
39. Henrion, R.: Structural properties of linear probabilistic constraints. Optimization **56**, 425–440 (2007)
40. Henrion, R., Küchler, C., Römisch, W.: Discrepancy distances and scenario reduction in two-stage stochastic mixed-integer programming. Journal of Industrial and Management Optimization **4**, 363–384 (2008)
41. Hochreiter, R.: Computational optimal management decisions - the case of stochastic programming for financial management. Ph.D. thesis, University of Vienna, Austria (2005)
42. Hochreiter, R., Pflug, G.C.: Financial scenario generation for stochastic multi-stage decision processes as facility location problems. Annals of Operations Research **152**, 257–272 (2007)
43. Hochreiter, R., Pflug, G.C., Wozabal, D.: Multi-stage stochastic electricity portfolio optimization in liberalized energy markets. In: System Modeling and Optimization, IFIP International Federation for Information Processing, pp. 219–226. Springer, Boston, MA, USA (2006)
44. Høyland, K., Kaut, M., Wallace, S.W.: A heuristic for moment-matching scenario generation. Computational Optimization and Applications **24**, 169–185 (2003)
45. Høyland, K., Wallace, S.W.: Generating scenario trees for multi-stage decision problems. Management Science **47**, 295–307 (2001)

46. Infanger, G.: Planning under Uncertainty - Solving Large-Scale Stochastic Linear Programs. Scientific Press Series. Boyd & Fraser, Danvers, MA, USA (1994)
47. Kall, P., Mayer, J.: Stochastic Linear Programming. Springer (2005)
48. Kaut, M., Wallace, S.W.: Evaluation of scenario generation methods for stochastic programming. Pacific Journal of Optimization **3**, 257–271 (2007)
49. Krasenbrink, B.: Integrierte Jahresplanung von Stromerzeugung und -handel, *Aachener Beiträge zur Energieversorgung*, vol. 81. Klinkenberg, Aachen, Germany (2002). PhD Thesis
50. Kuhn, D.: Generalized bounds for convex multistage stochastic programs, Lecture Notes in Economics and Mathematical Systems, vol. 548. Springer (2005)
51. Li, T., Shahidehpour, M., Li, Z.: Risk-constrained bidding strategy with stochastic unit commitment. IEEE Transactions on Power Systems **22**, 449–458 (2007)
52. Mirkov, R., Pflug, G.C.: Tree approximations of dynamic stochastic programs. SIAM Journal on Optimization **18**, 1082–1105 (2007)
53. Müller, A., Stoyan, D.: Comparison Methods for Stochastic Models and Risks. Wiley, Chichester, UK (2002)
54. Pennanen, T.: Epi-convergent discretizations of multistage stochastic programs. Mathematics of Operations Research **30**, 245–256 (2005)
55. Pennanen, T.: Epi-convergent discretizations of multistage stochastic programs via integration quadratures. Mathematical Programming **116**, 461–479 (2009)
56. Pereira, M.V.F., Pinto, L.M.V.G.: Multi-stage stochastic optimization applied to energy planning. Mathematical Programming **52**, 359–375 (1991)
57. Pflug, G.C.: Scenario tree generation for multiperiod financial optimization by optimal discretization. Mathematical Programming **89**, 251–271 (2001)
58. Pflug, G.C., Römisch, W.: Modeling, Measuring, and Managing Risk. World Scientific, Singapore (2007)
59. Prékopa, A.: Probabilistic programming. In: Ruszczyński and Shapiro [63], chap. 5, pp. 267–351
60. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton, NJ, USA (1970)
61. Rockafellar, R.T., Uryasev, S.: Conditional value-at-risk for general loss distributions. Journal of Banking & Finance **26**, 1443–1471 (2002)
62. Römisch, W.: Stability of stochastic programming problems. In: Ruszczyński and Shapiro [63], chap. 8, pp. 483–554
63. Ruszczyński, A., Shapiro, A. (eds.): Stochastic Programming, *Handbooks in Operations Research and Management Science*, vol. 10, 1st edn. Elsevier, Amsterdam (2003)
64. Ruszczyński, A., Shapiro, A.: Optimization of convex risk functions. Mathematics of Operations Research **31**, 433–452 (2006)
65. Schmöller, H.K.: Modellierung von Unsicherheiten bei der mittelfristigen Stromerzeugungs- und Handelsplanung, *Aachener Beiträge zur Energieversorgung*, vol. 103. Klinkenberg, Aachen, Germany (2005). PhD Thesis
66. Schmöller, H.K., Hartmann, T., Kruck, I., Haubrich, H.J.: Modeling power price uncertainty for midterm generation planning. In: A. Borghetti, C.A. Nucci, M. Paolone (eds.) IEEE Bologna PowerTech Proceedings (2003)
67. Sen, S., Yu, L., Genc, T.: A stochastic programming approach to power portfolio optimization. Operations Research **54**, 55–72 (2006)
68. Shapiro, A.: Inference of statistical bounds for multistage stochastic programming problems. Mathematical Methods of Operations Research **58**, 57–68 (2003)
69. Shapiro, A.: Stochastic programming approach to optimization under uncertainty. Mathematical Programming **112**, 183–220 (2007)
70. Takriti, S., Krasenbrink, B., Wu, L.S.Y.: Incorporating fuel constraints and electricity spot prices into the stochastic unit commitment problem. Operations Research **48**, 268–280 (2000)
71. Wallace, S.W., Fleten, S.E.: Stochastic programming models in energy. In: Ruszczyński and Shapiro [63], chap. 10, pp. 637–677
72. Weron, R., Bierbrauer, M., Trück, S.: Modeling electricity prices: jump diffusion and regime switching. Physica A **336**, 39–48 (2004)
73. Wu, L., Shahidehpour, M., Li, T.: Stochastic security-constrained unit commitment. IEEE Transactions on Power Systems **22**, 800–811 (2007)