DISS. ETH Nr. 17244

Hydro-Electric Power Plant Dispatch-Planning – Multi-Stage Stochastic Programming with Time-Consistent Constraints on Risk

A dissertation submitted to the

ETH ZURICH

for the degree of

Doctor of Sciences

presented by

MARTIN DENSING

dipl. phys. ETH

born 27.7.1970

citizen of Ermatingen, Thurgau

accepted on the recommendation of

Prof. Dr. Hans-Jakob Lüthi, examiner
Prof. Dr. Peter Kall, co-examiner

2007

Zusammenfassung

Thema der Arbeit ist die optimale Steuerung eines Pumpspeicher-Elektrizitätskraftwerks unter Unsicherheit. Die Unsicherheit stammt vom Wasserzufluss und von Preisfluktuationen auf einem Spotmarkt, auf dem die Elektrizität gehandelt wird. Das Kraftwerksmodell ist als mehrstufiges stochastisches lineares Programm formuliert. Ein kohärenter und zeitkonsistenter risiko-adjustierter Wert ist in einer mehrperiodigen Risiko-Nebenbedingung berücksichtigt.

Die Arbeit besteht aus zwei Hauptteilen. Der erste Teil behandelt die kohärente Risikomessung –soweit für das spätere Modell von Relevanz–, der zweite die optimale Kraftwerkssteuerung.

Im ersten Teil wird ein einfach zu handhabender, rekursiver risiko-adjustierter Wert definiert, für den in einem Spezialfall eine untere Schranke hergeleitet wird.

Im zweiten Teil werden Optimierungsprobleme von einfachen Kraftwerksmodellen explizit gelöst; die Probleme weisen diesselbe Struktur auf wie jene, die im Zusammenhang mit kohärenter Risikomessung auftreten. Das hochfrequente Handeln am Spotmarkt wird modelliert, und dies trotz der beschränkten Anzahl numerisch bewältigbarer Zeitstufen. Der Szenariobaum wird mittels Aufenthaltszeiten des Spotpreises erzeugt, für welche eine Hauptkomponentenanalyse eine charakteristische Struktur ergibt. Schlussendlich wird das allgemeine Kraftwerksmodell mit der mehrperiodigen Risiko-Nebenbedingung numerisch gelöst.

Abstract

The principal topic is the optimal operation of a hydro-electric pumped storage plant under uncertainty. The uncertainty stems from the water inflow and from the fluctuations of prices at a spot market, on which the electricity is traded.

The model of the plant is formulated as a multi-stage stochastic linear programming problem. A coherent and time-consistent risk-adjusted value is incorporated in a multi-period constraint on risk.

The thesis is made up of two parts. The first part considers coherent risk-adjusted values, the second part the optimal control of the plant.

The first part defines a simple recursive risk-adjusted value, for which in a special case a lower bound can be derived.

In the second part, some optimization problems of simple dispatch models are explicitly solved; the problems are structured similarly to those in coherent risk measurement. The short-term trading on the spot market is modeled in defiance of the limited number of numerically tractable time stages. The scenario tree is generated by the occupation times of the spot price. The principal component analysis of the occupation times exhibit their characteristic pattern. Finally, the general dispatch model subject to the multi-period constraint on risk is numerically solved.

electron *(Greek)*: amber, acquires an electric charge by friction.

risk unclear origin:

- *rixari (Latin)*: to dispute [76];
- *risco (Spanish)*: dangerous cliff [76];
- *rizq (Arabic)*: to seek your reward from God [65].

Acknowledgement

The search for the topic and the subsequent quest for the feasibility of the thesis was an arduous process. The accomplishment of the final version was never secured. This endeavor could not have been realized without the patience of my supporters and their faith in my work. I owe a lot of thanks to the support of Prof. Dr. Peter Kall. He gave me sufficient time to delve into the mathematics of operations research, and his supervision as well as his clear and discerning expertise improved the work considerably. I am very grateful to Prof. Dr. Hans-Jakob Lüthi, who accepted me to write the thesis at the ETH Zürich, introduced to me the world of energy optimization, initiated the original problem setup. The fruitful discussions with Prof. Dr. János Mayer, his expert knowledge, and his essential suggestions were a valuable source for large improvements.

Dr. Jörg Döge (IFOR ETH), who worked on a similar subject, continuously assisted me with information and data. The meeting with Prof. Dr. Paolo Burlando gave me insight into the state-of-the-art modeling of water inflow. I thank Dr. Walter Tobler for proofreading a preliminary draft. I thank Saskia Edskes for giving me the initial impulse and for personal coaching in the looming hour before the defence. Finally, I cordially thank my beloved parents for their enduring confidence in me.

Zürich, October 2007 Martin Densing

Contents

Chapter 1

Introduction

The business environment of power producers has changed with the advent of electricity markets. In Europe, the first large electricity market was Nord Pool in Scandinavia, founded in 1993. An example of a major market place in Central Europe is the European Energy Exchange (EEX) in Leipzig, which opened in 2002. Based on the directive of the European Union concerning a common electricity market in 1996 [31], the ongoing harmonization of national laws facilitates the access of electricity producers to these deregulated markets.

The electricity markets increase the flexibility of production. In the past, the main concern of an electricity producer was to cover the demand with own production; the demand originates from delivery contracts between the producer and costumers. The new markets allow to cover demand by buying electricity on the market. On the other side, own production capacity can be sold on the market.

The markets do not only increase the flexibility of demand coverage, but allow the hedging of positions by derivative instruments. For example, the production of energy can be immunized against falling electricity prices by selling a futures contract[1].

In the environment of electricity markets, a producer of electricity should coordinate power generation with trading activity. Any optimization of a coordinated strategy has to consider the uncertainty in market prices and demand, as well as the uncertainty in production-related quantities like water inflow (in the case of hydro-energy plants). If the uncertain quantities can be described by probability distributions, then stochastic optimization methods can be applied.

The main subject of the thesis is the optimal operation of a hydro-electric pumped storage plant. Because only a single plant is considered, the emphasis is on the trading of electricity on a spot and futures market, and not on the demand[2]; usually, the demand is satisfied by a large power portfolio.

[1]A futures contract is a contract for the purchase or sale of an underlying financial product at a specified price at a future settlement date. In our case, the underlying is 1 MW of electricity over a certain delivery period.

[2]The demand is only considered in an extended version of the model (Appendix A.6).

We present two modeling views of the electricity plant: The first is a *stochastic control* model, which singles out the control from other variables, and the second is a *multi-stage stochastic programming* model, which is suitable for the numerical tests.

The control of the plant is *non-anticipative* (as all human decisions) and desirably *adapted* to movements in exogenous variables, like market prices and water inflow. This leads to a formulation on a *scenario tree*. Numerical tractability restricts the number of stages in the tree, such that it is difficult to incorporate the short-term (hourly) trading activity on the spot market in mid or long-term models. This problem is tackled by a suitable formulation of the dynamics of the spot price. In addition, to generate a moderately sized scenario tree, the dynamics will be reduced to a small number of *factors*. The cases in which a model formulation on a scenario tree is actually needed (and not a simple *path-wise* formulation) are also numerically identified.

To find the optimal operation of the plant, the model uses two financial selection criteria: The expected final value of the plant and a so-called *risk-adjusted value*; they are monetary, financial values, and, optimally, they should be as large as possible. The particular definition of these values is to some extend up to the decision maker of the plant; for the risk-adjusted value, we use a multi-period *coherent* definition, which takes the intertemporal values of the plant into account. Coherency means that the risk-adjusted value has to fulfill a set of axioms that try to capture the notion of financial risk.

The decisions in a multi-period model are commonly adapted to different states and times. Hence, it is reasonable to measure the risk from the viewpoint of different states and times, too. The connection between the different risk measurements is achieved by the property of so-called *time consistency* (exact definition see Sec. 3.5). A *recursive* definition of the risk-adjusted value will ensure time consistency.

Hence, at first glance, it seems that the role of risk-adjusted values is restricted to constrain the risk in the model of the electricity plant. The connection is deeper: Some *electricity dispatch* models have the same structure as optimization problems that arise from the calculation of coherent risk-adjusted values.

The thesis is built up in two parts: The first part is about financial risk measurement, and the second about the model of the electricity plant.

The next chapter contains a review of single-period coherent risk-adjusted values; in particular, we present the duality relation of the risk-adjusted value *Conditional-Value-at-Risk* (CVaR)[1]. Then, we introduce the notations for multiple periods and define a specific form of a risk-adjusted value for processes as well as of a risk-adjusted value for random variables at final time. The definitions are recursive,

[1]Our definition differs slightly from the literature; hence, we use the qualifier 'risk-adjusted value'.

which ensure time consistency (Ch. 3). A particular recursive risk-adjusted value is where a single step of the recursive calculation is similar to CVaR (Ch. 4).

In the second part, we motivate the modeling of the electricity plant, and we introduce the notion of stochastic control as far as needed (Ch. 5). Before we proceed to the general model, we solve some simple dispatch problems analytically (Ch. 6). The general model of the plant is presented in two formulations: As a stochastic control problem, and second, as a multi-stage stochastic linear program on a finite scenario tree, which is numerically solvable (Ch. 7). The scenario tree generation method is explained in detail (Ch. 8), and the model is numerically solved in a case study (Ch. 9).

Eventually, we conclude (Ch. 10) and provide directions for extensions and further research (Ch. 11).

Chapter 2

Single-Period Risk Measurement

In this chapter, we review coherent single-period risk-adjusted values to the extent their theory is used in the further analysis.

2.1 Coherent Risk-Adjusted Value

Let a fixed probability space $(\Omega, \mathcal{F}, \mathbb{P})$ be given. The set Ω consists of the possible outcomes. $\mathcal{F}$ is a σ-algebra of sets in Ω. A set in $\mathcal{F}$ is called an event. The σ-algebra $\mathcal{F}$ is allowed to have infinitely many elements. The expectation of a random variable $X\colon \Omega \to \mathbb{R}$ with respect to $\mathbb{P}$ is denoted by $\mathbb{E}[X]$. Concerning the notation: Random variables are upper case, and arguments of functionals are denoted in square brackets.

In our context, a single-period risk-adjusted value will be defined for bounded random variables $X \in L^\infty := L^\infty(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R})$. The random variable X is interpreted as the value (net worth) of a financial position at the end of the period, where the value is as seen from today. The financial position is typically a portfolio of assets; in this case X is typically the discounted sum of the cash flows[1] that have occurred till the end of the period for these assets. More generally, the value of a position at a specific time can be considered as a sum of two components: Retrospective (occurred cash flows) and prospective (future cash flows). In the current single-period setup, only the cumulative occurred cash flows at final time are relevant, whereas the forthcoming multi-period setup will always consider the past and the future.

In a portfolio of assets, the amount of each asset is adjustable by a decision maker. In the particular case of an electricity plant, a sequence of cash flows over time is generated by dispatching water to trade electricity. Hence, the portfolio of an electricity plant can be considered to consist of the capacity to dispatch water and of positions in other, purely financial contracts, like futures. In fact, because the

[1]cash flow := (uncertain) stream of money at a specific time

decision maker can adjust (subject to some constraints) the amount of dispatched water at each time, the individual dispatch capacities over time can be considered to constitute a portfolio of its own.

According to Artzner et al. [5], a coherent *risk-adjusted value* of a given financial position is interpreted as a deterministic, monetary amount, such that the following axioms are fulfilled.

Definition 1 (Coherency, [5]). A *coherent single-period risk-adjusted value* is the value of a mappping $\pi\colon L^\infty \to \mathbb{R}$ such that the following axioms are fulfilled:

(i) $\pi[X+Y] \geq \pi[X] + \pi[Y] \quad \text{for all } X, Y \in L^\infty,$

(ii) $\pi[\lambda X] = \lambda\pi[X] \quad \text{for all } X \in L^\infty,\ \lambda \geq 0, \lambda \in \mathbb{R},$

(iii) $X \geq Y \ a.s. \Rightarrow \pi[X] \geq \pi[Y] \quad \text{for all } X, Y \in L^\infty,$

(iv) $\pi[X+c] = \pi[X] + c \quad \text{for all } X \in L^\infty,\ c \in \mathbb{R}.$

Some remarks:

- The symbol c in axiom (iv) has two meanings: On the left, it is a constant random variable $c : \Omega \to \mathbb{R}$, and on the right, it is a scalar $c \in \mathbb{R}$. The identification is $c(\Omega) = c \in \mathbb{R}$.
- The mapping π is superadditive and positive homogenous. Therefore π is concave. In the special case where Ω is finite, the continuity of π follows [4, Proof Prop. 2.2].
- In the original definition in [4], $\rho[X] = -\pi[X]$ is called the *risk measure*. In our notation: The greater the value of X the worthier is X, and so the greater is $\pi[X]$.

A position X is called *acceptable* if $\pi[X] \geq 0$ [5]. Let as assume that $X \geq 0$ *a.s.*; specializing property (iii) to $Y = 0$ *a.s.* and property (ii) to $\lambda = 0$ we get $\pi[X] \geq 0$. Hence, all non-negative random variables are acceptable. The risk-adjusted value can be interpreted as the maximal amount of money that can be withdrawn from the position such that it stays still acceptable:

$$\sup\{c \in \mathbb{R} \mid \pi[X-c] \geq 0\} \stackrel{\text{(iv)}}{=} \sup\{c \in \mathbb{R} \mid \pi[X] \geq c\} = \pi[X], \tag{2.1}$$

where we used the *translation invariance* (iv). The maximal amount is negative for positions that are not accepted. In this case, the negative of the risk-adjusted value is interpreted as the minimal amount of money to be added to render the position acceptable.

In the particular case of an electricity plant, the value of the plant will be defined without considering capital costs, depreciation costs, maintenance costs and water charges; only the value that originates directly from the exploitation of cost-free water in the associated trading activity is considered. This can even lead to a positive value X in almost all events: $X \geq 0$ *a.s.* As we have just seen, this implies $\pi[X] \geq 0$ and X is always 'accepted'. But, acceptance can also be viewed with respect to a target as follows. A *total value* can be defined: $X - c$, where c consists of (usually deterministic) net losses that are generated by the mentioned costs and additional, firm-specific accounting amounts. The total value may become negative. Hence, we can define that X is *accepted with respect to a target* c if $\pi[X] \geq c$, or equivalently, by translation invariance, if $\pi[X - c] \geq 0$.

A profound discussion of the axioms in Def. 1 can be found in Artzner et al. [4]. A short interpretation is (X and Y are positions):

(i) The amount of money that can be withdrawn from the portfolio $X + Y$ such that the portfolio stays acceptable is greater than the sum that can be withdrawn from the individual positions X and Y ('diversification reduces risk').

(ii) The amount of money that can be withdrawn from a position is proportional to the size of the position.

(iii) The allowed amount of withdrawn money is larger if the value is larger in every state.

(iv) if an amount c of money is added to the position, then we can withdraw additionally this amount.

The theory of coherent risk measurement can be extended by relaxing the properties of homogeneity and superadditivity and replacing them by concavity (see e.g. [62]). The advantage of relaxing homogeneity is that the risk of a large financial position can be defined to be disproportionately high with respect to a small position, owing to the observation that large positions are difficult to liquidate in limited financial markets. The disadvantage is that the definition of the risk-adjusted value is no longer invariant with respect to different currencies (exchange rate = λ).

2.2 Definition by Probability Measures

Coherent risk-adjusted values (Def. 1) can be constructed with help of the following representation.

Lemma 1 (Definition by probability measures [4]). *Let $\mathcal{P}$ be a set of probability measures on $(\Omega, \mathcal{F})$. The mapping $\pi \colon L^\infty \to \mathbb{R}$, defined by*

$$\pi[X] := \inf_{\mathbb{Q} \in \mathcal{P}} \mathbb{E}_{\mathbb{Q}}[X], \tag{2.2}$$

is a coherent single-period risk-adjusted value, where $\mathbb{E}_{\mathbb{Q}}[\cdot]$ denotes the integration with respect to measure $\mathbb{Q}$.

The proof is easy (see e.g. [4]). In fact, for every given coherent risk-adjusted value π there exists a set $\mathcal{P}$ such that (2.2) holds (see [19]).

The representation (2.2) says that the expectation of the uncertain value X is evaluated with respect to different probability measures $\mathbb{Q} \in \mathcal{P}$. These *test-probability measures* may be considered as stress-scenarios. The worst expectation over the set of test-probability measures is considered to be the risk-adjusted value of X.

A specific coherent risk-adjusted value with representation (2.2) is given by a set $\mathcal{P}$ of probability measures. For the forthcoming examples of coherent risk-adjusted values, only $\mathcal{P}$s that contain probability measures $\mathbb{Q}$ that are absolutely continuous with respect to $\mathbb{P}$ are taken into account:

$$\text{For all } F \in \mathcal{F},\ \mathbb{P}[F] = 0 \implies \mathbb{Q}[F] = 0.$$

This is reasonable because impossible events with respect to $\mathbb{P}$ may be assumed to be impossible in all test-scenarios $\mathbb{Q}$, too. If $\mathbb{Q} \in \mathcal{P}$ is absolutely continuous, then an integral with respect to $\mathbb{Q}$ can be expressed by an integral with respect to $\mathbb{P}$:

Theorem 1 (Radon-Nikodym). *Let $\mathbb{Q}$ be absolutely continuous with respect to $\mathbb{P}$. There exists an $H \in L^1_+(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R})$ such that*

$$\mathbb{E}_{\mathbb{Q}}[Y] = \mathbb{E}[HY] \text{ for all bounded random variables } Y. \tag{2.3}$$

The non-negative random variable H is called the Radon-Nikodym (probability) density *of $\mathbb{Q}$ with respect to $\mathbb{P}$.*

For a proof, see [6, Prop. 17.3, Prop 17.10]. On the other hand, let $H \in L^1_+ := L^1_+(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R})$ with $\mathbb{E}[H] = 1$ be given. H defines a measure $\mathbb{Q}$ by taking the indicator function $Y := \chi_F$ in (2.3): $\mathbb{Q}[F] := \mathbb{E}[H\chi_F]$ for all $F \in \mathcal{F}$ [6, Prop. 17.1]. Hence, if $\mathcal{P}$ contains only absolutely continuous probability measures, then $\mathcal{P}$ can be identified with a set of Radon-Nikodym densities, and the representation (2.2) of coherent risk-adjusted values becomes

$$\pi[X] = \inf_{H \in \widetilde{\mathcal{P}}} \mathbb{E}[HX], \tag{2.4}$$

where $\widetilde{\mathcal{P}} \subseteq \{H \in L^1_+ \mid \mathbb{E}[H] = 1\}$.

The optimization problem (2.4) is linear. Therefore, relaxing the problem by taking the convex hull of $\widetilde{\mathcal{P}}$ does not yield a lower risk-adjusted value. Indeed, let $0 < \lambda < 1$ and $H_1, H_2 \in \widetilde{\mathcal{P}}$. Clearly, $\lambda H_1 + (1-\lambda)H_2 \in L^1_+$, $\mathbb{E}[\lambda H_1 + (1-\lambda)H_2] = 1$, and

$$\begin{aligned} \mathbb{E}\big[\big(\lambda H_1 + (1-\lambda)H_2\big)X\big] &= \lambda\mathbb{E}[H_1X] + (1-\lambda)\mathbb{E}[H_2X] \\ &\geq \min\big(\mathbb{E}[H_1X],\ \mathbb{E}[H_2X]\big). \end{aligned}$$

Therefore, the set $\widetilde{\mathcal{P}}$ can be considered to be convex.

2.3 Conditional-Value-at-Risk

Next, we choose a specific set of Radon-Nikodym densities (i.e. test-probability measures). The at first unmotivated choice leads to a risk-adjusted value that can be interpreted as an expected shortfall.

Let us consider the following risk-adjusted value with uniformly bounded Radon-Nikodym densities for $X \in L^1$:

$$\min_{H \in L^\infty} \mathbb{E}[HX], \qquad \text{s.t.} \begin{cases} \mathbb{E}[H] = 1, & & |q \\ H \leq \dfrac{1}{\alpha} & a.s., & |Z \\ H \geq 0 & a.s., & \end{cases} \tag{2.5}$$

where $\alpha \in (0,1)$, and where we have indicated Lagrange multipliers for the first two constraints: $q \in \mathbb{R}$ and the random variable $Z \in L^1$.

With the help of the Lagrange multipliers, we can formulate the algebraic dual problem:

$$\max_{q \in \mathbb{R},\, Z \in L^1} q - \frac{1}{\alpha}\mathbb{E}[Z], \qquad \text{s.t.} \begin{cases} Z \geq q - X & a.s., \\ Z \geq 0 & a.s.. \end{cases} \tag{2.6}$$

Clearly, if Z^* is optimal, then $Z^*(\omega)$ is as small as allowed by the constraints for every $\omega \in \Omega$. Thus, we have either $Z^*(\omega) = q - X(\omega)$ or $Z^*(\omega) = 0$. Therefore, we can write the problem concisely as

$$\max_{q \in \mathbb{R}} q - \frac{1}{\alpha}\mathbb{E}[(q - X)^+], \tag{2.7}$$

where $(\cdot)^+ := \max(\cdot, 0)$ selects the positive part. For the stochastic programming formulation of problem (2.7), see Kall and Mayer [49, p. 159]. The optimal solutions of problems (2.5) and (2.7) are well-known; the next proposition emphasises their strong duality.

Proposition 1 (CVaR Duality, [1, 80]). *Let X be a bounded random variable.*

(i) Problem (2.5) *is strongly dual to problem* (2.7)*: The problems have optimal solutions, and the optimal objective values are the same. Every α-quantile of X is an optimal solution of* (2.7).

(ii) Let $q_\alpha(X)$ be an α-quantile of X such that the distribution of X has no atom at $q_\alpha(X)$: $\mathbb{P}[X = q_\alpha(X)] = 0$. Then the optimal objective value is

$$\mathbb{E}[X \mid X \leq q_\alpha(X)]. \tag{2.8}$$

Definition 2. Let $X \in L^1$ and $\alpha \in (0,1)$. The *risk-adjusted value Conditional Value-at-Risk* CVaR$[X]$ *at level* α is the optimal objective value of problems (2.5) and (2.7).

Note. The name CVaR originates from equation (2.8); the expectation of X is taken conditionally on values less or equal an α-quantile, where the quantile may be interpreted as a so-called *value-at-risk*. In applications, α is chosen to be small, e.g. $\alpha = 0.05$ or 0.01. The *risk measure* β-CVaR of Uryasev and Rockafellar [80] is defined on losses, which are considered to be positive, and β is large, e.g. $\beta = 0.95$ or 0.99. The correspondence is for a (positive) loss Y

$$\beta\text{-}CVaR = -\mathrm{CVaR}[-Y] \text{ at level } \alpha = 1 - \beta.$$

Acerbi and Tasche [1] use the notion of *Tail Mean*, whereas Artzner et al. [5, p. 16] use *Tail-VaR*:

$$TM_\alpha(X) = TVaR^\alpha(X) = \mathrm{CVaR}[X] \text{ at level } \alpha.$$

The optimal solutions are formulated in terms of quantiles; an α-quantile of a random variable X is every number q_α such that

$$\mathbb{P}[X < q_\alpha] \leq \alpha \leq \mathbb{P}[X \leq q_\alpha]. \tag{2.9}$$

Note that q_α is not always unique; some useful facts about quantiles are provided in Appendix A.1 (p. 112). Let q_α be an arbitrary α-quantile of X. The proposition already states that q_α is an optimal solutions of (2.7) (see [80]); the optimal solution of (2.5) is given by (a combination of) indicator functions $\chi\colon \Omega \to \{0,1\}$:

$$H^* := \begin{cases} \frac{1}{\alpha}\chi_{\{X<q_\alpha\}} + \dfrac{1 - \frac{1}{\alpha}\mathbb{P}[X < q_\alpha]}{\mathbb{P}[X = q_\alpha]}\chi_{\{X=q_\alpha\}}, & \text{if } \mathbb{P}[X = q_\alpha] > 0, \\ \frac{1}{\alpha}\chi_{\{X \leq q_\alpha\}}, & \text{if } \mathbb{P}[X = q_\alpha] = 0, \end{cases} \tag{2.10}$$

where the definition is in the almost surely sense (see [1]).

Because the optimal solutions are known, Proposition 1 can be proved by strong duality.

Proof. We have to show:

(i) Weak duality: For every feasible solutions, the objective value of (2.5) is smaller or equal to the objective value of (2.6).

(ii) q_α is feasible for (2.6), H^* is feasible for (2.5), and the objective values are the same; therefore, they are optimal solutions.

(iii) The optimal solution does not depend on the choice of the α-quantile.

(iv) If $\mathbb{P}[X = q_\alpha] = 0$, then $\text{CVaR}[X] = \mathbb{E}[X | X \le q_\alpha]$.

(i): Let H, Z and q be feasible.

$$\begin{aligned}\mathbb{E}[HX] \ge \mathbb{E}[H(q-Z)] &= qE[H] - \mathbb{E}[HZ] \\ &= q - \mathbb{E}[HZ] \overset{(Z\ge 0)}{\ge} q - \frac{1}{\alpha}\mathbb{E}[Z].\end{aligned}$$

(ii): $q_\alpha \in \mathbb{R}$ and $Z := (q_\alpha - X)^+$ are clearly feasible in (2.6). For (2.5) we have to consider two cases: First, the discontinuous case. The first constraint in (2.5) is

$$\mathbb{E}[H^*] = \frac{1}{\alpha}\mathbb{P}[X < q_\alpha] + \frac{1 - \frac{1}{\alpha}\mathbb{P}[X < q_\alpha]}{\mathbb{P}[X = q_\alpha]}\mathbb{P}[X = q_\alpha] = 1,$$

and therefore fulfilled. The second constraints in (2.5) demands that

$$0 \le \frac{1}{\alpha} \le \frac{1}{\alpha} \quad a.s. \quad \text{on } \{X < q_\alpha\},$$

and

$$0 \le \frac{1 - \frac{1}{\alpha}\mathbb{P}[X < q_\alpha]}{\mathbb{P}[X = q_\alpha]} \le \frac{1}{\alpha} \quad a.s. \quad \text{on } \{X = q_\alpha\}.$$

The first chain of inequalities is trivially fulfilled. The second chain of inequalities is transformed equivalently by multiplying with $(\alpha\mathbb{P}[X = q_\alpha])$. The resulting chain of inequalities is the definition (2.9) of an α-quantile q_α . In the continuous case, definition (2.9) becomes $\mathbb{P}[X < q_\alpha] = \alpha = \mathbb{P}[X \le q_\alpha]$. Hence, the first constraint becomes

$$\mathbb{E}[H^*] = \frac{1}{\alpha}\mathbb{P}[X \le q_\alpha] = 1,$$

and the second is

$$0 \le \frac{1}{\alpha}\chi_{\{X \le q_\alpha\}} \le \frac{1}{\alpha}.$$

Therefore, H^* is feasible. In the continuous case, the objective value of (2.5) with solution H^* is

$$\begin{aligned}\mathbb{E}[H^*X] &= \mathbb{E}[\frac{1}{\alpha}\chi_{\{X\le q_\alpha\}}X] \\ &= \frac{1}{\alpha}\mathbb{E}[X\chi_{\{X\le q_\alpha\}} - q_\alpha\chi_{\{X\le q_\alpha\}}] + \frac{1}{\alpha}\mathbb{E}[q_\alpha\chi_{\{X\le q_\alpha\}}] \\ &= -\frac{1}{\alpha}\mathbb{E}[(q_\alpha - X)^+] + \frac{1}{\alpha}q_\alpha\mathbb{P}[X \le q_\alpha] \\ &= q_\alpha - \frac{1}{\alpha}\mathbb{E}[(q_\alpha - X)^+],\end{aligned}$$

whereas in the discontinuous case it is

$$\begin{aligned}
\mathbb{E}[H^*X] &= \mathbb{E}[\Big(\frac{1}{\alpha}\chi_{\{X<q_\alpha\}} + \frac{1-\frac{1}{\alpha}\mathbb{P}[X<q_\alpha]}{\mathbb{P}[X=q_\alpha]}\chi_{\{X=q_\alpha\}}\Big)X] \\
&= \frac{1}{\alpha}\mathbb{E}[X\chi_{\{X<q_\alpha\}}] + \frac{1-\frac{1}{\alpha}\mathbb{P}[X<q_\alpha]}{\mathbb{P}[X=q_\alpha]}\mathbb{E}[X\chi_{\{X=q_\alpha\}}] \\
&= \Big(\frac{1}{\alpha}\mathbb{E}[X\chi_{\{X<q_\alpha\}} - q_\alpha\chi_{\{X<q_\alpha\}}] + \frac{1}{\alpha}\mathbb{E}[q_\alpha\chi_{\{X<q_\alpha\}}]\Big) + \big(1-\frac{1}{\alpha}\mathbb{P}[X<q_\alpha]\big)q_\alpha \\
&= -\frac{1}{\alpha}\mathbb{E}[(q_\alpha - X)^+] + \frac{1}{\alpha}q_\alpha\mathbb{P}[X<q_\alpha] + q_\alpha\big(1-\frac{1}{\alpha}\mathbb{P}[X<q_\alpha]\big) \\
&= q_\alpha - \frac{1}{\alpha}\mathbb{E}[(q_\alpha - X)^+].
\end{aligned}$$

Therefore, we have strong duality.
(iii): We made no specific choice of the α-quantile q_α. Therefore, the strong duality holds for every α-quantile. Hence, every α-quantile is optimal in (2.6).
(iv): Let us write the expectation in the optimal objective value of (2.7) conditionally:

$$q_\alpha - \frac{1}{\alpha}\mathbb{E}\big[(q_\alpha - X)^+\big] = q_\alpha - \frac{1}{\alpha}\mathbb{P}\big[q_\alpha - X \geq 0\big]\mathbb{E}\big[(q_\alpha - X)^+\big|q_\alpha - X \geq 0\big],$$

because $\Big(\mathbb{P}[X = q_\alpha] = 0 \overset{(2.9)}{\Longrightarrow} \mathbb{P}[q_\alpha - X \geq 0] = \alpha\Big)$, this is

$$= \mathbb{E}\big[q_\alpha - (q_\alpha - X)^+\big|X \leq q_\alpha\big] = \mathbb{E}\big[X\big|X \leq q_\alpha\big]. \qquad \blacksquare$$

Chapter 3

Recursive Risk Measurement

In this chapter, we introduce the multi-period setting. First, the gain of information over time is modeled by a *filtration* of the probability space. Second, we define the recursive risk-adjusted values for a process (called *Case I*) and for a random variable (*Case II*). In the next chapters we will incorporate the risk-adjusted value into a numerically solvable optimal-decision model. The complexity of this model demands finiteness, hence, we consider –as in Artzner et al. [4]– a finite setup, that is, a finite filtration (scenario tree).

3.1 The Scenario Tree

Let a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ be given. As motivated above, we assume the following finite setup.

Finiteness Assumption:

The forthcoming definitions of multi-period risk-adjusted values are considered only for a finite number of time steps $t = 0, 1, \ldots, T$ *and on a finitely generated* σ*-algebra* $\mathcal{F}_T \subseteq \mathcal{F}$*.*

The finite σ-algebra $\mathcal{F}_T \subseteq \mathcal{F}$ of *final* events represents the possible outcomes. Generally, a finite σ-algebra is generated by a finite partition of Ω, and a set in the partition is called an *atom*.

The available information at time t is represented by a sub-σ-algebra $\mathcal{F}_t \subseteq \mathcal{F}_T$, $t = 0, \ldots, T$ with $\mathcal{F}_t \subseteq \mathcal{F}_{t+1}$, $t = 0, \ldots, T-1$. In other words, the gain of information is represented by a *filtration* $(\mathcal{F}_t)_{t=0,\ldots,T}$. The σ-algebra at time zero is chosen to contain no information: $\mathcal{F}_0 := \{\emptyset, \Omega\}$.

A finite filtration $(\mathcal{F}_t)_{t=0,\ldots,T}$ corresponds to a tree: Each atom of $\mathcal{F}_t$ is identified with a node of the tree at step t. In particular, the set Ω is identified with the *root node* (Fig. 3.1), and the *probability of a node* is the probability of the corresponding atom; if n denotes a node, then we write $\mathbb{P}[n]$ for its probability.

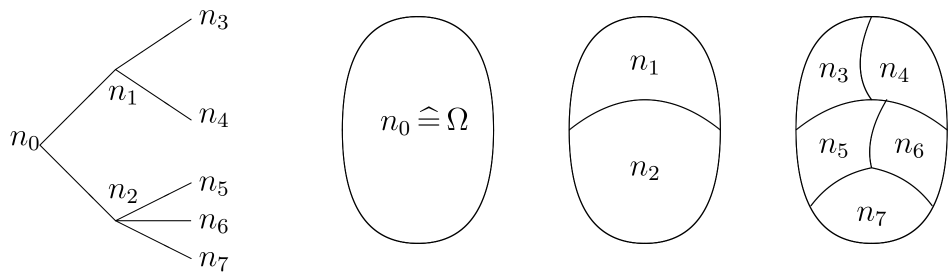


Figure 3.1: A scenario tree generated by a finite filtration $(\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2)$. The filtration corresponds to a sequence of refined partitions of Ω. $\mathcal{F}_0$ has the set $\{n_0\}$ of atoms, $\mathcal{F}_1$ has atoms $\{n_1, n_2\}$, and $\mathcal{F}_2$ has atoms $\{n_3, n_4, n_5, n_6, n_7\}$. For all t, the σ-algebra $\mathcal{F}_t$ is the powerset of the set of atoms.

Definition 3 (Scenario tree of a finite filtration). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space. Let a finite filtration of $\mathcal{F}$ be given: $(\mathcal{F}_t)_{t=0,\dots,T}$. The representing tree together with the probabilities of the nodes is called the *scenario tree* of the filtration with respect to $\mathbb{P}$.

Without loss of generality, we always assume that all nodes of the scenario tree have strictly positive probability under $\mathbb{P}$ *(otherwise the subtree originating from such a node can be neglected).*

Because we will consider different test-probability measures $\mathbb{Q}$ on $(\mathcal{F}, \Omega)$, the above definition can be extended to different probability measures $\mathbb{Q}$ by replacing $\mathbb{P}$ with $\mathbb{Q}$. If the scenario tree is with respect to the (original) probability measure $\mathbb{P}$, we just speak of a *scenario tree for a given (finite) filtration*, and we do not mention that the probabilities of the nodes are according to $\mathbb{P}$. In addition, given a scenario tree with probabilities according to $\mathbb{P}$, we can consider a probability measures $\mathbb{Q}$ on $(\mathcal{F}_T, \Omega)$ (i.e. on the terminal nodes), and we say that $\mathbb{Q}$ is *defined on the tree.*

The set of nodes at time t is denoted by $\mathcal{N}_t$. A *scenario* is defined as the unique path from the root node to a terminal node, $(n_0, n_1, \dots, n_T)$, where $n_t \in \mathcal{N}_t$ for all t. The *history* of node n_t is defined as the unique partial path $(n_0, n_1, \dots, n_{t-1})$, where n_{s-1} is the parent node of n_s for $s = 1, \dots, t$.

A stochastic process is in our case a sequence of random variables over time: $(X_t)_{t=0,\dots,T}$. Let a finite filtration be given: $(\mathcal{F}_t)_{t=0,\dots,T}$. We will consider only $(\mathcal{F}_t)$-*adapted* processes, that is, X_t is $\mathcal{F}_t$-measurable for all t. A standard result of measure theory says that X_t is $\mathcal{F}_t$-measurable if and only if X_t is constant on each atom of the finite σ-algebra $\mathcal{F}_t$, or, equivalently, X_t takes values on the nodes $\mathcal{N}_t$. Hence, by the finiteness assumption of the filtration: *In the forthcoming definitions of multi-period risk-adjusted values, all stochastic processes are defined on a given scenario tree.*

For the rest of the thesis, if not stated otherwise, all equations or inequalities between random variables are to hold almost surely. In our discrete case, this means that an (in-)equality between $\mathcal{F}_t$-measurable random variables holds in all nodes $n \in \mathcal{N}_t$. Nevertheless, we still write '*a.s.*' in some expressions to avoid any ambiguities.

Note. The foregoing definition of a scenario tree uses atomic sets in $\mathcal{F}_t$ (i.e. sets in the abstract space Ω) to define a node at time t (see also [68, Sec. 9]). In applications, these sets (events) are usually generated by the values of a specific stochastic process $(X_t)_{t=0,\ldots,T}$, where the X_ts have finitely-discrete distributions. Thus, we can *define* the filtration by considering the evolution of the process over time:

$$\mathcal{F}_t := \sigma(X_0, \ldots, X_t), \qquad t = 0, \ldots, T.$$

Hence, a node is given by a value of the process at time t conditional on the past values, and a sequence of realizations of the process from time 0 till time T determines a path (scenario) in the tree. Therefore, two different connotations can be associated with scenarios: The event-view and the value-view. Because we will consider several processes, and because in the first part of the thesis the nature of the generating process is irrelevant, we stick to the first view and consider a scenario tree (*event tree*) as a finite filtration on a probability space.

In the following, let a scenario tree of a finite filtration $(\mathcal{F}_t)_{t=0,\ldots,T}$ be given.

3.2 Multi-Period Financial Risk Measurement

Let us consider a financial investment that generates an uncertain series of cash flows. Different stakeholders[1] may require the *acceptability* of the investment:

- If a *creditor* does not accept the investment, the creditor does not grant any credit, or curtails existing credits, and the investment has to be sold or even liquidated.
- If a *supervision authority* does not accept the investment, the authority may impose a fee, a capital increase, or even disallow further operation.
- If the investment becomes too risky, a risk-averse *shareholder* or the *management* requests to change the business strategy.

In every case, the stakeholder's acceptance may be based on the perceived 'value' of the investment. In our multi-period setting, the value at a future time t can depend on the events at time t (i.e. the nodes $n \in \mathcal{N}_t$). We assume a value that is prospective and retrospective: The value in event (node) $n \in \mathcal{N}_t$ is a sum of the cash flows that have realized till time t in event n, and of an assessment of the future cash flows conditional on n (all quantities discounted to $t = 0$). The uncertain

[1] a person with an interest or concern in the investment

sequence of the future financial values is described by an $(\mathcal{F}_t)$-adapted stochastic process $(X_t)_{t=0,\dots,T}$.

As in the single-period case, the acceptability of the value process is assumed to be defined by a lower bound $\rho_{\min} \in \mathbb{R}$ of a risk-adjusted value. In a multi-period setting, two cases of risk-adjusted values can be considered [5]:

Case I: The risk-adjusted value is a number that depends on the whole process $(X_t)_{t=0,\dots,T}$. The number is measured from the viewpoint of today, or it may be measured as viewed from a state (event) at a future time.

Case II: The risk-adjusted value is a number that depends on the single *random variable* X_T at final time T. As in Case I, the risk can be measured at different times (prior to final time) and states.

Case II is appropriate if the stakeholder is interested in the final value only. Case I is appropriate if the stakeholder cares about intertemporal values. The decision whether to use Case I or Case II depends on the specific application. For the forthcoming model of the electricity plant, it is assumed that the intertemporal values are important for the stakeholders, thus, a risk-adjusted value of Case I is applied.

Sometimes, the notion of *liquidity risk* is associated with intertemporal risk. Liquidity risk is the possibility of an intertemporal shortage of cash. Creditors can help by lending money when such a shortage happens. Therefore, a temporary shortage of cash is not always a problem per se. Rather, if the creditor perceives (at any specific point in time) the *value* of the investment to be no longer acceptable, then the creditor grants no longer money.

The forthcoming definition of multi-period risk-adjusted value is based on Artzner et al. [5]. At the time of writing, multi-period financial risk measurement is an active research area. In the following, a short overview of related work is given.

The applied approach is value-based. Instead, the risk measurement could be based on (uncertain) cash flows [38]. The risk measurements from the viewpoint of different states and time can be related to each other by the requirement of *time consistency* (the exact definition in our setting is in Section 3.5). In discrete time, time consistency leads to a *recursive* representation of risk-adjusted values [5,58,81]. The general continuous time case is treated in [20]. Similar approaches specifically for Case II (final random variable) are given in [69,73,74,82]. Similar results in the language of Markov chains were obtained in [12]. Another direction of research is the use of the so-called expected-value-of-perfect-information (EVPI) as a criterion for the hedging[1] of a multi-period income stream; the value of the hedged income stream turns out to have coherency properties [68]. If the time consistency is not

[1]Hedging is an action to reduce risk.

demanded, then general risk-adjusted values on the product space of state and time can be considered, where time is either continuous [17, 18] or discrete [28].

Next, we define the coherent recursive risk-adjusted value in Case I and II, and then the notion of time consistency.

3.3 Case I: Recursive Risk-Adjusted Value for Processes

Let an $(\mathcal{F}_t)$-adapted stochastic process of values be given: $(X_t)_{t=0,\ldots,T}$. The riskiness of the process shall not be measured at time $t = 0$ only, but also in different events at future times $t = 1, \ldots, T$ (i.e. in different nodes of the scenario tree, which is given by the filtration). Hence, for each node of the scenario tree, we measure a risk-adjusted value. The measurement in a node is assumed to be based on the available information given by the node. In other words, the sequence of risk-adjusted values over time is an $(\mathcal{F}_t)$-adapted stochastic process. This *risk-adjusted value process* is denoted by $(R_t^{(X)})_{t=0,\ldots,T}$ for the process of values $(X_t)_{t=0,\ldots,T}$. Occasionally, we will denote the dependence on the stochastic process more explicitly by $(R_t^{(X_0,\ldots,X_T)})_{t=0,\ldots,T}$.

The risk-adjusted value process $(R_t^{(X)})_{t=0,\ldots,T}$ is assumed to fulfill the following assumptions [5].

(i) The risk-*adjusted* value $R_t^{(X)}$ should be smaller than the value X_t for $t = 0, \ldots, T$:

$$R_t^{(X)} \leq X_t \quad \text{for all } t. \tag{3.1}$$

(ii) As in the single-period theory, a set $\mathcal{P}$ of test-probability measures is considered. Because information is gained over time (for every possible test-probability measure), the risk-adjusted value should increase in average for every 'test-scenario':

$$R_t^{(X)} \leq \mathbb{E}_{\mathbb{Q}}[R_{t+1}^{(X)}|\mathcal{F}_t] \quad \text{for all } \mathbb{Q} \in \mathcal{P}, \text{ for all } t. \tag{3.2}$$

In other words, the risk-adjusted value process is a submartingale for all $\mathbb{Q}$.

(iii) Given a set of test-probability measures, the decision maker is allowed to choose freely among the risk-adjusted value processes that fulfil (3.1) and (3.2). The most favorable, i.e. the largest possible is chosen.

The assumptions (i)-(iii) are equivalent to the following definition:

Definition 4 (Recursive risk-adjusted value, Case I, [5]). Let a set $\mathcal{P}$ of probability measures on the scenario tree be given. The *risk-adjusted value process*

$(R_t^{(X)})_{t=0,\dots,T}$ *of an* $(\mathcal{F}_t)$*-adapted stochastic process* $(X_t)_{t=0,\dots,T}$ *with set* $\mathcal{P}$ *of test-probability measures* is

$$R_t^{(X)} := \begin{cases} X_T, & \text{if } t = T, \\ \min\Big(X_t, \ \min_{\mathbb{Q}\in\mathcal{P}} \mathbb{E}_{\mathbb{Q}}[R_{t+1}^{(X)}|\mathcal{F}_t]\Big), & \text{if } t = 0,\dots,T-1, \end{cases} \tag{3.3}$$

where the minimization of the conditional expectation is pointwise, that is, for every node $n \in \mathcal{N}_t$ separately. $R_0^{(X)}$ is called the *risk-adjusted value* (at time 0) of the process.

We recall that we use sometimes the explicit notation $R_0^{(X_0,\dots,X_T)}$.

For simplicity, we assume in (3.3) that the set $\mathcal{P}$ is such that the pointwise minimum of the conditional expectation is attained. Later on, the attainability is shown for a specific choice of $\mathcal{P}$.

The pointwise minimum of the $\mathcal{F}_t$-measurable conditional expectations takes values on the nodes of $\mathcal{N}_t$ (i.e. is again $\mathcal{F}_t$-measurable). This is obvious since the values of the conditional expectation form a vector in $\mathbb{R}^{|\mathcal{N}_t|}$, and a component-wise minimization does not change this fact. Because the minimum of two random variables in the outer minimization of (3.3) preserves measurability, it follows that the risk-adjusted value process is $(\mathcal{F}_t)$-adapted. By contrast, for infinite $\mathcal{F}_t$ the formulation of (3.3) must be changed: The pointwise infimum of an uncountable set of $\mathcal{F}_t$-measurable random variables may no longer be measurable with respect to $\mathcal{F}_t$ (see e.g. [34, Ch. A.4]). To prevent such cases, 'inf' must be replaced by 'essential inf' [5].

The risk-adjusted value $R_0^{(X)}$ is $\mathcal{F}_0$-measurable and therefore a scalar. It can be considered as a mapping $\bar{\Psi}_0$ from the space of adapted stochastic processes into the reals:

$$\bar{\Psi}_0[X_0,\dots,X_T] := R_0^{(X)},$$

and generally,

$$\bar{\Psi}_t[X_0,\dots,X_T] := R_t^{(X)}, \qquad t = 0,\dots,T,$$

where $\bar{\Psi}_t$ maps into the space of $\mathcal{F}_t$-measurable random variables (cf. the notation in [5]). Note that by (3.3) only the random variables $X_t,\dots,X_T$ are relevant for the calculation of $R_t^{(X)}$.

The axioms of coherency (Def. 1, p. 5) can be translated to the multi-period recursive risk-adjusted value as follows. We use the notation $R_0^{(\lambda X+Y)} := R_0^{(\lambda X_0+Y_0,\dots,\lambda X_T+Y_T)}$, where $\lambda \in \mathbb{R}$ and $(X_t)_{t=0\dots T}$, $(Y_t)_{t=0\dots T}$ are processes.

Lemma 2. *Let* $(X_t)_{t=0\dots T}$ *and* $(Y_t)_{t=0\dots T}$ *be* $(\mathcal{F}_t)$*-adapted stochastic processes. Then the recursive risk-adjusted value is coherent in the sense that*

(i) $R_0^{(X+Y)} \geq R_0^{(X)} + R_0^{(Y)}$,

(ii) $R_0^{(\lambda X)} = \lambda R_0^{(X)}$ *for all* $\lambda \geq 0,\ \lambda \in \mathbb{R}$,

(iii) $X_0 \leq Y_0, \ldots, X_T \leq Y_T \implies R_0^{(X)} \leq R_0^{(Y)}$,

(iv) $R_0^{(X+c)} = R_0^{(X)} + c$ *for all* $c \in \mathbb{R}$,

where in (iv) the left-hand-side c *denotes the constant process* $(c)_{t=0,\ldots,T}$.

Proof. Only (i) is not obvious. We use backward-induction over time. At final time T:

$$R_T^{(X+Y)} \overset{(3.3)}{=} X_T + Y_T \overset{(3.3)}{=} R_T^{(X)} + R_T^{(Y)}.$$

Assume $R_t^{(X+Y)} \geq R_t^{(X)} + R_t^{(Y)}$. Then

$$\begin{aligned} R_{t-1}^{(X+Y)} &\overset{(3.3)}{=} \min\Big(X_{t-1} + Y_{t-1}, \min_{\mathbb{Q}} \mathbb{E}_{\mathbb{Q}}\big[R_t^{(X+Y)} \big| \mathcal{F}_{t-1}\big]\Big) \\ &\geq \min\Big(X_{t-1} + Y_{t-1}, \min_{\mathbb{Q}} \mathbb{E}_{\mathbb{Q}}\big[R_t^{(X)} + R_t^{(Y)} \big| \mathcal{F}_{t-1}\big]\Big), \end{aligned}$$

separate minimization gives smaller values:

$$\geq \min\Big(X_{t-1} + Y_{t-1}, \min_{\mathbb{Q}} \mathbb{E}_{\mathbb{Q}}\big[R_t^{(X)} \big| \mathcal{F}_{t-1}\big] + \min_{\mathbb{Q}} \mathbb{E}_{\mathbb{Q}}\big[R_t^{(Y)} \big| \mathcal{F}_{t-1}\big]\Big),$$

$\min(a+b, c+d) \geq \min\big(a + \min(b,d), c + \min(b,d)\big) = \min(a,c) + \min(b,d)$:

$$\begin{aligned} &\geq \min\big(X_{t-1}, \min_{\mathbb{Q}} \mathbb{E}_{\mathbb{Q}}\big[R_t^{(X)} \big| \mathcal{F}_{t-1}\big]\big) + \min\big(Y_{t-1}, \min_{\mathbb{Q}} \mathbb{E}_{\mathbb{Q}}\big[R_t^{(Y)} \big| \mathcal{F}_{t-1}\big]\big) \\ &\overset{(3.3)}{=} R_{t-1}^{(X)} + R_{t-1}^{(Y)}. \end{aligned}$$

■

Because of the translation invariance *(iv)*, the single-period concept of acceptability (2.1) (p. 5) can be taken over unchanged: $(X_t)_{t=0\ldots T}$ is accepted if $R_0^{(X)} \geq 0$, and $R_0^{(X)}$ is the maximal amount of money that can be withdrawn each time from the value process such the value process stays acceptable.

3.4 Case II: Recursive Risk-Adjusted Value for Final Values

If intertemporal values are not important for a decision maker, then the measurement of the risk of the final value is sufficient. In this case, the recursive calculation (3.3) can be simplified: Only the random variable X_T at final time T enters the recursive calculation.

Definition 5 (Recursive Risk-adjusted Value, Case II). Let a set $\mathcal{P}$ of probability measures be given. The *risk-adjusted value process* $(R_t^{X_T})_{t=0,\dots,T}$ *of the random variable* X_T is

$$R_t^{X_T} := \begin{cases} X_T, & \text{if } t = T, \\ \min\limits_{\mathbb{Q}\in\mathcal{P}} \mathbb{E}_{\mathbb{Q}}[R_{t+1}^{X_T}|\mathcal{F}_t], & \text{if } t = 0,\dots,T-1. \end{cases} \tag{3.4}$$

$R_0^{X_T}$ is called the *risk-adjusted value* (at time 0) of a final value.

The definition is formally a special case of (3.3) if the random variables $X_0,\dots,X_{T-1}$ are so large that they do not enter the recursive calculation (3.3). The risk-adjusted value $R_0^{X_T}$ represents the view of a decision maker who is only interested in the risk of values at final time T.

As in the case for processes, the recursive risk-adjusted value of a random variable X_T at time $t = 0,\dots,T$ can be viewed as the value of a mapping $\bar{\Phi}_t$ from the space of $\mathcal{F}_T$-measurable random variables into the $\mathcal{F}_t$-measurable random variables: $\bar{\Phi}_t[X_T] := R_t^{X_T}$ (cf. the notation in [5]).

3.5 Time Consistency of Risk-Adjusted Value Processes

In this section it is shown that the recursive calculation of the risk-adjusted value ensures time consistency.

The notion of time consistency appears in the literature in different contexts. The forthcoming definition is based on Artzner et al. [5]. A related notion appears in Kydland and Prescott [59] (nobel laureates 2004), where time consistency of economic policies is considered. In their setting, although not explicitly mentioned, time consistency is essentially Bellman's *principle of optimality* [8]. Bellman's principle applies to specially structured multi-period optimal control models. It says that an optimal policy has the property that, whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision [8]. Bellman's principle implies a backward-recursive calculation of an optimal policy (*Dynamic Programming*).

In our setting, time consistency is a property of a decision criterion. The criterion is applied at different time and states of a probability space equipped with a filtration[1]. The decision criterion is applied to financial investment opportunities, such that a decision maker can decide which opportunity to prefer. In particular, in our case, the decision criterion over time is given by the risk-adjusted value process. Hence, in our setting, time consistency is not directly defined in the context of an

[1] the filtration is in our case finite $\widehat{=}$ scenario tree

optimization problem. Nevertheless, time-consistent decision criteria may be used in the objective function or in constraints of optimization problems.

First, we give a definition in words: The decision criterion is *time consistent* if it has the property that

> if an investment opportunity is preferred to another *at a future time* in *all* possible events at that time, then it is preferred *as of today*, too.

Artzner et al. [5] use an extended definition where the future point in time is replaced by a random time (*stopping time*). They show (among other results) that time consistency is equivalent to a recursive calculation of the risk-adjusted value (as implied by Bellman's principle). In our approach, the recursiveness of the risk-adjusted value is the starting definition. Thus, time consistency will be easily verified.

First, we will consider the more accessible Case II: The time consistency of a risk-adjusted value process that measures the risk of a final random variable.

3.5.1 Time Consistency of Risk-Adjusted Value Processes in Case II (Final Values)

In Case II, only the final outcome of the investment opportunity matters. The possible outcomes are the values of an $\mathcal{F}_T$-measurable random variable X_T. The preference of a decision maker over time is represented by a real-valued decision criterion that is calculated for every node on the scenario tree. Thus, at time t, the decision criterion maps X_T into the space of $\mathcal{F}_t$-measurable random variables:

$$X_T \mapsto U_t^{X_T}, \qquad t = 0, \dots, T,$$

where $U_t^{X_T}$ is a $\mathcal{F}_t$-measurable random variable. In fact, in our finite setting, X_T takes values on the terminal nodes of the tree and $U_t^{X_T}$ takes values on the nodes at time t; hence, the mapping can be identified with a mapping from $\mathbb{R}^{|\mathcal{N}_T|}$ to $\mathbb{R}^{|\mathcal{N}_t|}$.

Definition 6 (Time consistency (Case II), [5]). Consider a sequence of mappings from the space of $\mathcal{F}_T$-measurable random variables into the space of $\mathcal{F}_t$-measurable random variables: $X_T \mapsto U_t^{X_T}$, $t = 0, \dots, T$, for all $\mathcal{F}_T$-measurable X_T. The sequence is *time consistent* if for every $t \in \{1, \dots, T\}$ and for every $\mathcal{F}_T$-measurable random variables Y_T and Z_T such that

$$U_t^{Y_T} \geq U_t^{Z_T} \quad a.s.^1,$$

it holds that

$$U_0^{Y_T} \geq U_0^{Z_T}.$$

[1] in our finite setting: on every node $n \in \mathcal{N}_t$

In our finite setting, Y_T, Z_T take values on the terminal nodes of the scenario tree, and the decision criterion $U_t^{Y_T}$, $U_t^{Z_T}$ takes values on the nodes at time t. In other words, the decision criterion is evaluated at the nodes at time t by the decision maker.

Lemma 3. *The recursive risk-adjusted value process for final random variables (Def. 5) is time consistent.*

Proof. Let us consider two random variables Y_T and Z_T that are ordered at a time $0 < t \leq T$ in the sense that $R_t^{Y_T} \geq R_t^{Z_T}$ *a.s.*. Because $X \mapsto \min_{\mathbb{Q}} \mathbb{E}_{\mathbb{Q}}[X|\mathcal{F}_{t-1}]$ is for every random variable X a monotone mapping, the order is preserved:

$$\min_{\mathbb{Q}} \mathbb{E}_{\mathbb{Q}}\left[R_t^{Y_T}\middle|\mathcal{F}_{t-1}\right] \geq \min_{\mathbb{Q}} \mathbb{E}_{\mathbb{Q}}\left[R_t^{Z_T}\middle|\mathcal{F}_{t-1}\right] \quad a.s.,$$

which is by Definition 5 of risk-adjusted values equivalent to

$$R_{t-1}^{Y_T} \geq R_{t-1}^{Z_T} \quad a.s..$$

Repeating the argument for the time steps from $t-1$ to 0 gives $R_0^{Y_T} \geq R_0^{Z_T}$. ■

If a risk-adjusted value is not defined recursively, then time consistency can be violated [5]:

Example. *Let us define the value of a* non-recursive *decision criterion at time $t = 0, \ldots, T-1$ in node $n \in \mathcal{N}_t$ for final a random variable X_T as*

$$U_t^{X_T}\Big|_n := \mathbb{E}_{\mathbb{P}_n}[X_T | X_T \leq q_\alpha^-(X_T)], \tag{3.5}$$

where the probability measure $\mathbb{P}_n$ is the conditional probability measure on the subtree with root n:

$$\mathbb{P}_n[m] := \frac{\mathbb{P}[m \cap n]}{\mathbb{P}[n]} \qquad \text{for every node } m,$$

and $q_\alpha^-(X_T)$ is the lower α-quantile of X_T with respect to $\mathbb{P}_n$ (see facts about quantiles in Appendix A.1, p. 112). We can always divide because $\mathbb{P}[n] > 0$ (general assumption, p. 12). The decision criterion can be interpreted as a variant of the risk-adjusted value CVaR from the viewpoint of different nodes in the scenario tree.

To keep the example simple, a binary, two-step ($T = 2$) scenario tree with equiprobable terminal nodes is considered (Fig. 3.2). If the quantile is taken at $\alpha = 1/2$, one can resort to the simple rule for calculating the criterion $U_t^{X_T}|_n$ at a node n: 'Take the lower half of the values of the conditional distribution of X_T and average these values'.

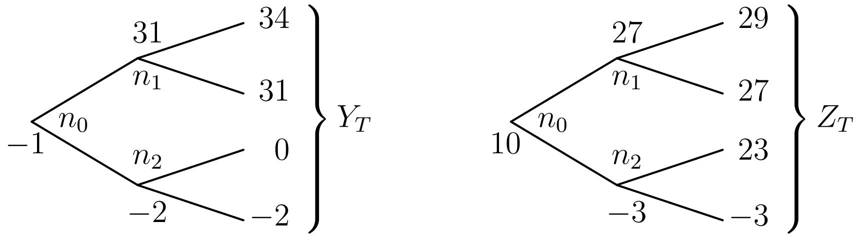


Figure 3.2: A binary scenario tree with random variables Y_T and Z_T which take values in the terminal nodes ($T = 2$). The tree is duplicated for the two random variables such that numbers can be read-off conveniently. The terminal nodes are assumed to be equiprobable. The value of the decision criterion (3.5) is calculated for the nodes n_0, n_1 and n_2. Parameter $\alpha = 1/2$.

The example considers two final outcomes Y_T and Z_T (Fig. 3.2). The decision criterion is calculated at the root node n_0:

$$U_0^{Y_T} = \underbrace{1/\tfrac{1}{2}}_{\substack{\textit{1/conditional}\\ \textit{probability}}} \cdot \Big(\underbrace{1/4}_{\substack{\textit{probability}\\ \textit{under } \mathbb{P}_{n_0}}} \cdot \underbrace{(-2)}_{\textit{value}} + 1/4 \cdot 0 \Big) = \frac{1}{2}(-2+0) = -1,$$

$U_0^{Z_T} = \frac{1}{2}(-3 + 23) = 10$, and at the nodes of the first time step: $U_1^{Y_T}|_{n_1} = 31$, $U_1^{Y_T}|_{n_2} = -2$, $U_1^{Z_T}|_{n_1} = 27$, and $U_1^{Z_T}|_{n_2} = -3$.

Hence, the order of the values of the criterion between Y_T and Z_T is not consistent over time: At the first time step, the values are $31 > 27$ and $-2 > -3$ at n_1 and n_2, respectively, whereas at the root node n_0 the values are $-1 < 10$. Hence, at time $t = 1$, the value of the criterion is larger for Y_T than for Z_T, whereas at time $t = 0$, Z_T has a larger value.

In applications, the risk-adjusted value process serves as a decision criterion over time. If the decision criterion prefers Y_T always to Z_T at a later time t, then it may be desirable that Y_T is preferred to Z_T at time zero, too. Otherwise, the risk-adjusted value process may not be used as decision criterion in a consistent way over time. The expectations of Y_T and Z_T could be used as an additional criterion, such that time consistency says that if for both criteria two random variables are ordered at later time in all events, then these variables should be ordered today in the same way. This two-criteria case is important in mean-risk optimization problems, where the optimal solution is selected with two critera: mean of the value, and risk of the value (risk-adjusted value). We were able to give counter-examples of time-consistency in this case, too. Finally, we may quote Wang [81]:

> "It is well understood that in a dynamic optimization problem, if either the objective function or any of the constraints is not time consistent,

a strategy chosen today may be regretted later on. In other words, if given the opportunity, the strategy, chosen earlier will be abandoned in favor of another one, causing inconsistency in choices over time."

3.5.2 Time Consistency of Risk-Adjusted Value Processes in Case I (Processes)

Time consistency preserves the order of uncertain values, where the order is calculated from the viewpoint of different times: from time t and from time 0. Thus, time consistency applies to values that realize in the future of time 0 and time t. To compare processes over time, the definition of time consistency has to take into account in some way the intermediate random variables of the processes from time 0 to time $t-1$, which lay in the past from the viewpoint of time t. A particular choice is that time consistency is demanded only for processes that are equal from time 0 till time $t-1$. This prevents an influence of the values prior to t on the order.

Definition 7 (Time consistency (Case I), [5]). Let a filtration $(\mathcal{F}_t)_{t=0,\dots,T}$ be given. Consider a sequence of mappings from the space of adapted processes into the space of $\mathcal{F}_t$-measurable random variables: $(X_s)_{s=0,\dots,T} \mapsto U_t^{(X)}$, $t = 0,\dots,T$, for all adapted processes $(X_s)_{s=0,\dots,T}$. The sequence is time consistent if for every $t \in \{1,\dots,T\}$ and for every $(\mathcal{F}_s)$-adapted processes $(Y_s)_{s=0,\dots,T}$ and $(Z_s)_{s=0,\dots,T}$ such that $U_t^{(Y)} \geq U_t^{(Z)}$ *a.s.*[1] and such that

$$Y_s = Z_s \quad a.s. \quad \text{for every } s = 0,\dots,t-1,$$

it holds that

$$U_0^{(Y)} \geq U_0^{(Z)}.$$

Proposition 2. *The recursive risk-adjusted value process that is defined for stochastic processes (Def. 4) is time consistent.*

For the proof, we need the following Lemma, which can be related to Bellman's principle [5].

Lemma 4. *A recursive risk-adjusted value fulfils for every adapted process* $(X_s)_{s=0,\dots,T}$ *and every* $0 < t \leq T$ *that*

$$R_0^{(X)} = R_0^{(X_0,\dots,X_{t-1},R_t^{(X)},\dots,R_t^{(X)})}, \tag{3.6}$$

where we recall the explicit notation $R_t^{(Y_0,\dots,Y_T)} := R_t^{(Y)}$ *for a process* $(Y_s)_{s=0,\dots,T}$.

[1]in our finite setting: on every node $n \in \mathcal{N}_t$

Proof. The proof applies the recursive definition (3.3) to the right-hand-side of (3.6), starting from time T going to time zero:

$$R_T^{(X_0,\ldots,X_{t-1},R_t^{(X)},\ldots,R_t^{(X)})} \overset{(3.3)}{=} R_t^{(X)},$$

$$R_{s-1}^{(X_0,\ldots,X_{t-1},R_t^{(X)},\ldots,R_t^{(X)})} \overset{(3.3)}{=} \begin{cases} \min\left(X_{s-1}, \min_{\mathbb{Q}} \mathbb{E}_{\mathbb{Q}}[R_s^{(X)}|\mathcal{F}_{s-1}]\right) \overset{(3.3)}{=} R_{t-1}^{(X)}, & s = t, \\ \min\left(R_t^{(X)}, \min_{\mathbb{Q}} \mathbb{E}_{\mathbb{Q}}[R_t^{(X)}|\mathcal{F}_{s-1}]\right) \overset{(*)}{=} R_t^{(X)}, & s > t, \end{cases}$$

$$\text{for } s = T, \ldots, t+1, \text{ which implies}$$

$$R_t^{(X_0,\ldots,X_{t-1},R_t^{(X)},\ldots,R_t^{(X)})} = R_t^{(X)}, \qquad (**)$$

where in $(*)$ we used that an $\mathcal{F}_t$-measurable variable is for $t < T$ also $\mathcal{F}_{T-1}$-measurable: $\mathbb{E}_{\mathbb{Q}}[R_t^{(X)}|\mathcal{F}_{T-1}] = R_t^{(X)}$. Applying (3.3) on both sides of $(**)$ yields (3.6). ■

Proof (Proposition). Consider two adapted processes $(Y_s)_{s=0,\ldots,T}$ and $(Z_s)_{s=0,\ldots,T}$ and a time $t \in \{1, \ldots, T\}$ such that

$$R_t^{(Y)} \geq R_t^{(Z)} \quad \text{and} \quad Y_s = Z_s, \ s = 0, \ldots, t-1, \qquad a.s..$$

The monotonicity property (iii) (p. 18) implies

$$R_0^{(Y_0,\ldots,Y_{t-1},R_t^{(Y)},\ldots,R_t^{(Y)})} \geq R_0^{(Z_0,\ldots,Z_{t-1},R_t^{(Z)},\ldots,R_t^{(Z)})},$$

which is by (3.6) equivalent to $R_0^{(Y)} \geq R_0^{(Z)}$. ■

Chapter 4

Risk Measurement with Local-CVaR Sets

In the previous chapter, we defined the recursive risk-adjusted value for a stochastic process (Case I) and for a final random variable (Case II). The associated set $\mathcal{P}$ of test-probability measures was not specified further. In this chapter, we choose the set $\mathcal{P}$ to be *stable*; the stability of $\mathcal{P}$ ensures consistency with single-period risk measurement. In particular, we choose $\mathcal{P}$ in such a way that the associated recursive risk-adjusted value reduced for a single-time period to the risk-adjusted value CVaR (Def. 2, p. 9).

Similar notions related to stable sets are the multiplicative-stable (m-stable) sets of risk-adjusted values in continuous time [20], and the so-called rectangular sets in dynamic-consistent utility theory [30]. The forthcoming definition is based on [5], but, the original general formulation (in terms of so-called density processes) is adapted to our application-oriented focus.

As in the previous chapter, let a probability space $(\Omega, \mathbb{P}, \mathcal{F})$ with a finite filtration $(\mathcal{F}_t)_{t=0,\dots,T}$ be given. In other words, all processes are defined on a given scenario tree, and a node n at stage t, $n \in \mathcal{N}_t$, is identified with an atomic event in $\mathcal{F}_t$.

4.1 Stable Sets of Probability Measures

A probability measure $\mathbb{Q}$ on $\mathcal{F}_T$ (i.e. on the terminal nodes $\mathcal{N}_T$) can be written as a product of *transition probabilities* as follows.

Let $n_T \in \mathcal{N}_T$ be a terminal node such that $\mathbb{Q}[n_T] > 0$[1]. Then the value of the measure $\mathbb{Q}$ of n_T can be written as a product:

$$\mathbb{Q}[n_T] = \frac{\mathbb{Q}[n_T]}{\mathbb{Q}[n_{T-1}]} \frac{\mathbb{Q}[n_{T-1}]}{\mathbb{Q}[n_{T-2}]} \cdots \frac{\mathbb{Q}[n_2]}{\mathbb{Q}[n_1]} \frac{\mathbb{Q}[n_1]}{\mathbb{Q}[n_0]}, \tag{4.1}$$

[1]There may be n_Ts with $\mathbb{Q}[n_T] = 0$; $\mathbb{P}[n_T] > 0 \, \forall n_T$ is assured by general assumption (p. 12).

where $(n_0, n_1, \dots, n_T)$ denotes a root-to-leaf path. The formula holds because n_0 is the root node with $\mathbb{Q}[n_0] = 1$. The fractions are well-defined: If $\mathbb{Q}[n_T] > 0$, then $\mathbb{Q}[n_t] > 0$ for all n_t in the history of n_T.

Definition 8 (Transition probability). Let a scenario tree and a probability measure $\mathbb{Q}$ on the terminal nodes be given. Consider a node n with $\mathbb{Q}[n] > 0$. The *transition probability from node n to a different node m of the probability measure* $\mathbb{Q}$ is

$$q_{nm} := \begin{cases} \dfrac{\mathbb{Q}[m]}{\mathbb{Q}[n]}, & \text{if } n \text{ is in the history of } m, \\ 0 & \text{else.} \end{cases} \tag{4.2}$$

If the measure $\mathbb{Q}$ is the original probability measure $\mathbb{P}$, then we just speak of the *transition probabilities from node n to node m* (without mentioning $\mathbb{P}$). The transition probability q_{nm} is the conditional probability of node n given that node (event) m happens.

Note. In our definition, given a probability measure $\mathbb{Q}$, the transition probability is defined between every pair of different nodes in the scenario tree, whenever the conditioning node n has strictly positive probability. By contrast, assume a non-terminal node n with $\mathbb{Q}[n] = 0$. The successor nodes (events) of n, which are subsets of n, have a probability of zero, too. The transition probabilities from such a node n are not uniquely defined for a given $\mathbb{Q}$.

Given a non-terminal node, the transition probabilities to its immediate successor nodes can be combined into a vector. Such a vector can be defined without a direct reference to a measure $\mathbb{Q}$:

Definition 9 (Vector of single-step transition probabilities). Let a scenario tree be given. Let n be a non-terminal node and denote the immediate successor nodes of n by $m^{(1)}, \dots, m^{(N)}$. A *vector of single-step transition probabilities* associated with node n is

$$\boldsymbol{q}_n = (q_{nm^{(1)}}, \dots, q_{nm^{(N)}})^\top,$$

where $q_{nm^{(i)}} \geq 0$, $i = 1, \dots, N$, and $q_{nm^{(1)}} + \dots + q_{nm^{(N)}} = 1$.

The components of $\boldsymbol{q}_n$ can be identified as the values of a probability measure on a finite sample space of cardinality N.

Given a vector $\boldsymbol{q}_n$ of single-step transition probabilities in each non-terminal node n, a probability measure on the terminal nodes can be defined by building products:

$$\mathbb{Q}[n_T] := q_{n_0 n_1} \cdots q_{n_{T-1} n_T}, \ n_T \in \mathcal{N}_T, \ q_{n_t n_{t+1}} \in \boldsymbol{q}_{n_t}, \ n_t \in \mathcal{N}_t, \ t = 0, \dots, T-1,$$

where $(n_0, n_1, \dots, n_{T-1}, n_T)$ denotes a root-to-leaf-path in the tree and the inclusion '$q_{n_t n_{t+1}} \in \boldsymbol{q}_{n_t}$' means that $q_{n_t n_{t+1}}$ is the component in $\boldsymbol{q}_{n_t}$ that corresponds to the transition from n_t to n_{t+1}.

A set $\mathcal{P}$ of probability measures can be specified as follows. Assume a family of sets $\mathcal{P}_n$ of vectors of single-step transition probabilities associated with every non-terminal node n: $\{\mathcal{P}_n\}_{n \in \mathcal{N}_t,\ t=0,\dots,T-1}$. A probability measure can be constructed as above by selecting a vector from each of the $\mathcal{P}_n$s.

Definition 10 (Stable set of probability measures). Let a scenario tree be given. Consider a family $\{\mathcal{P}_n\}_{n \in \mathcal{N}_t,\ t=0,\dots,T-1}$ of sets of vectors of single-step transition probabilities, where n goes over all non-terminal nodes. The corresponding *stable set* $\mathcal{P}$ of probability measures defined on the scenario tree is

$$\mathcal{P} := \left\{ \left(q_{n_0 n_1} \cdots q_{n_{T-1} n_T} \right)_{n_T \in \mathcal{N}_T} \;\middle|\; \begin{array}{r} \boldsymbol{q}_{n_t} \in \mathcal{P}_{n_t},\ n_t \in \mathcal{N}_t, \\ t = 0, \dots, T-1 \end{array} \right\}, \tag{4.3}$$

where the probability measure on the scenario tree is represented by its values on the set $\mathcal{N}_T$ of terminal nodes, $(n_0, \dots, n_t, n_{t+1} \dots, n_T)$ denotes a root-to-leaf-path in the tree, and where the single-step transition probabilities $q_{n_t n_{t+1}}$ have to be chosen such that if $\widetilde{n}_{t+1}$ is another immediate successor node of n_t then $q_{n_t n_{t+1}} \in \boldsymbol{q}_{n_t} \iff q_{n_t \widetilde{n}_{t+1}} \in \boldsymbol{q}_{n_t}$.

The set $\mathcal{P}$ is 'stable' in the following sense. Consider two elements of $\mathcal{P}$: $\mathbb{Q}$ and $\mathbb{Q}'$. If we replace in some non-terminal node n the vector $\boldsymbol{q}_n$ of $\mathbb{Q}$ by the corresponding vector $\boldsymbol{q}'_n$ of $\mathbb{Q}'$, we get another measure $\mathbb{Q}''$. Because every concatenation of vectors of transition probabilities is a feasible element of $\mathcal{P}$, the measure $\mathbb{Q}''$ is again an element of $\mathcal{P}$.

The stability is desirable in the following context. In Case II, the recursive risk-adjusted value considers a random variable at a single future time T. We may consider the period from time 0 to T as a single period and apply also single-period coherent risk measurement. In particular, let a set $\mathcal{P}$ of test-probability measures on a scenario tree be given. The recursive risk-adjusted value of a random variable X_T should equal the coherent single-period risk-adjusted value with the same set $\mathcal{P}$:

$$R_0^{X_T} \stackrel{!}{=} \min_{\mathbb{Q} \in \mathcal{P}} \mathbb{E}_{\mathbb{Q}}[X_T]. \tag{4.4}$$

If the set $\mathcal{P}$ is stable, this *consistency with single-period coherent risk measurement* is assured (see Appendix A.2, p. 114).

Note. We can only touch the surface of coherent multi-period risk measurement with stable sets; the interested reader is referred to Artzner et al. [5]. We just give a rough outline of a generalization of (4.4) for Case I (stochastic processes). The process is considered at random times (*stopping time*). This gives a random variable that has values of the stopped process. The recursive risk-adjusted value can be written as a minimization of the expected value of this random variable. The minimization goes over a stable set of test-probability measures (as in (4.4)) as well as over all possible stopping times. Hence, we may say that the risk-adjusted value of the stochastic process is the expected value in the most adverse test-scenario, and at the most adverse stop.

4.2 Local-CVaR Sets

By definition, a recursive risk-adjusted value is specified by its associated set $\mathcal{P}$ of test-probability measures. As we have seen, a stable set $\mathcal{P}$ ensures consistency with single-period risk measurement. In addition, the stability will facilitate interpretation as well as computation. In our context, a stable set $\mathcal{P}$ is given by a family of sets $\mathcal{P}_n$ of vectors of transition probabilities associated with all non-terminal nodes n (Def. 10). As in the single-period theory, a vector $\mathbf{q}_n \in \mathcal{P}_n$ is identified with a vector $\mathbf{h}_n$ of a corresponding set $\widetilde{\mathcal{P}}_n$ of Radon-Nikodym probability densities with respect to the vector $\mathbf{p}_n$ of the original probability measure $\mathbb{P}$ (Def. 1, p. 7),

$$\begin{aligned} \mathbf{p}_n &= (p_{nm_1}, \ldots, p_{nm_N})^\top, \qquad && \mathbf{q}_n = (q_{nm_1}, \ldots, q_{nm_N})^\top, \\ \mathbf{h}_n &= (h_{nm_1}, \ldots, h_{nm_N})^\top, && \text{with } q_{nm_i} = h_{nm_i} p_{nm_i}, \quad i = 1, \ldots, N, \end{aligned}$$

where N is the number of immediate successor nodes of n.

A simple choice of $\mathcal{P}_n$ is the set of test-probability measures of the single-period risk-adjusted value CVaR (cf. problem (2.5), p. 8): The density is uniformly bounded from above by a constant $\frac{1}{\alpha}$ with $\alpha \in (0, 1)$. In particular:

$$\begin{aligned} \widetilde{\mathcal{P}}_n^\alpha &:= \left\{ \boldsymbol{h} \in \mathbb{R}^N \middle| \boldsymbol{h} \geq 0,\ \boldsymbol{h} \leq \frac{1}{\alpha}\mathbf{1},\ \boldsymbol{p}_n^\top \mathbf{h} = 1 \right\}, \\ \mathcal{P}_n^\alpha &:= \left\{ \boldsymbol{q} \in \mathbb{R}^N \middle| \frac{q_{nm_i}}{p_{nm_i}} \leq \frac{1}{\alpha},\ i = 1, \ldots, N,\ \boldsymbol{q}_n^\top \mathbf{1} = 1,\ q_{nm_i} \geq 0 \right\}, \end{aligned} \tag{4.5}$$

where $\mathbf{1} = (1, \ldots, 1)^\top \in \mathbb{R}^N$. The corresponding stable set $\mathcal{P}^\alpha$ is given by Definition 10.

Definition 11 (Local-CVaR set of probability measures). Let a scenario tree be given. Let $0 < \alpha < 1$, and let a family $\{\mathcal{P}_n^\alpha\}_{n \in \mathcal{N}_t, t=0,\ldots,T-1}$ of sets of single-step transition probabilities defined by (4.5) be given. The corresponding stable set of probability measures is called the *local-CVaR set* $\mathcal{P}^\alpha$.

Note. For simplicity, the constant α is chosen to be the same for all sets $\mathcal{P}_n^\alpha$. This restriction is not used by most of the subsequent analysis; α could be chosen differently for different nodes n.

The recursive calculation (3.3) of the risk-adjusted value can be written more explicitly by choosing $\mathcal{P} := \mathcal{P}^\alpha$. Let $t = 0, \ldots, T-1$. Then

$$\begin{aligned}
\min_{\mathbb{Q}\in\mathcal{P}^\alpha} \mathbb{E}_\mathbb{Q}[R_{t+1}^{(X)}|\mathcal{F}_t] &\overset{(i)}{=} \min_{\mathbb{Q}\in\mathcal{P}^\alpha} \left(\sum_{n\in\mathcal{N}_t} \Big(\sum_{m\in\mathcal{N}_{t+1}} q_{nm} R_{(t+1)m}^{(X)} \Big) \chi_n \right) \\
&\overset{(ii)}{=} \sum_{n\in\mathcal{N}_t} \Big(\min_{\mathbf{q}_n\in\mathcal{P}_n^\alpha} \sum_{m\in\mathcal{N}_{t+1}} q_{nm} R_{(t+1)m}^{(X)} \Big) \chi_n \\
&\overset{(iii)}{=} \sum_{n\in\mathcal{N}_t} \max_{q\in\mathbb{R}} \Big(q - \frac{1}{\alpha} \sum_{i=1}^{N} p_{nm_i} (q - R_{(t+1)m_i}^{(X)})^+ \Big) \chi_n \\
&\overset{(iv)}{=} \max_{q\in\mathbb{R}} \Big(q - \frac{1}{\alpha} \mathbb{E}[(q - R_{t+1}^{(X)})^+|\mathcal{F}_t] \Big),
\end{aligned} \tag{4.6}$$

with remarks:

(i) The conditional expectation is an $\mathcal{F}_t$-measurable random variable. Since the discrete filtration is represented by a scenario tree, the conditional expectation is constant on each node $n \in \mathcal{N}_t$ and can therefore be represented as a linear combination of indicator-functions χ_n for each node (event) n. The value of $R_{t+1}^{(X)}$ in node $m \in \mathcal{N}_{t+1}$ is denoted by $R_{(t+1)m}^{(X)}$,and q_{nm} is the transition probability (4.2) of $\mathbb{Q}$ from n to m.

(ii) Recall Definition 10 (p. 27) of a stable set: A stable set $\mathcal{P}$ is defined by all possible concatenations of vectors of single-step transition probabilities taken from given sets $\mathcal{P}_n$, where n goes over all non-terminal nodes. Because $\mathcal{P}^\alpha$ is stable the minimization can be performed node-wise over the local sets $\mathcal{P}_n^\alpha$.

(iii) The minimization is a discrete version of the minimization problem of the risk-adjusted value CVaR (2.5). Therefore, we can replace the minimization by the maximization of the dual problem (2.7).

(iv) The maximization is understood pointwise on the atoms of $\mathcal{F}_t$, or, in the language of nodes, for every $n \in \mathcal{N}_t$ separately.

The pointwise optimization problems (4.6) attain finite optimal values (Prop. 1, p. 8).

Similarly to the single-period case in Uryasev and Rockafellar [80], the change from 'min' to 'max' permits us to evaluate the risk-adjusted value by a linear optimization problem as follows.

4.3 Linear Formulation (Case I)

We show that the recursive risk-adjusted value can be calculated by solving a linear optimization problem. Despite we are in a finite setting, the problems are formulated in terms of random variables. The reason is threefold: First, cluttered notation is prevented; second, the structure of the problem becomes apparent; and third, a possible extension to the infinite-dimensional case is prepared.

We consider only Case I (recursive risk-adjusted value for a process). Case II can be treated analogously.

4.3.1 Linear Optimization Formulation (Case I)

The recursive risk-adjusted value with a local-CVaR set of test-probability measures can be formulated as a linear optimization problem:

Lemma 5. *Let a finite filtration $(\mathcal{F}_t)_{t=0,\dots,T}$ be given (scenario tree). Let $(X_t)_{t=0,\dots,T}$ be an $(\mathcal{F}_t)$-adapted stochastic process, and let $\mathcal{P}^\alpha$ be the local-CVaR-set of test-probability measures (Def. 11). Then the recursive risk-adjusted value $R_0^{(X)}$ with set $\mathcal{P}^\alpha$ is given by the optimal objective value of the following stochastic linear optimization problem:*

$$\max R_0,$$
$$\text{s.t.} \begin{cases} R_t \le X_t, & t = 0, \dots, T, \\ R_t \le Q_t - \dfrac{1}{\alpha}\mathbb{E}[Z_{t+1}|\mathcal{F}_t], & t = 0, \dots, T-1, \\ Z_t \ge Q_{t-1} - R_t, & t = 1, \dots, T, \\ Z_t \ge 0, & t = 1, \dots, T, \end{cases} \tag{4.7}$$

where R_t, Q_t and Z_t are $\mathcal{F}_t$-measurable random variables. In addition, we have for feasible $(R_t)_{t=0,\dots,T}$ and for the risk-adjusted value process $(R_t^{(X)})_{t=0,\dots,T}$ (Def. 4) that $R_t \le R_t^{(X)}$ a.s. for all t.

Proof. First, we show that the recursive risk-adjusted value $R_0^{(X)}$ equals the optimal objective value of the following optimization problem:

$$\max_{(R_t)} R_0,$$
$$\text{s.t.} \begin{cases} R_t \le X_t, & t = 0, \dots, T, \\ R_t \le \max\limits_{q \in \mathbb{R}} \left(q - \dfrac{1}{\alpha}\mathbb{E}[(q - R_{t+1})^+|\mathcal{F}_t] \right), & t = 0, \dots, T-1, \end{cases} \tag{4.8}$$

where the 'max' over q is taken pointwise (on the atoms of $\mathcal{F}_t$); by the remarks after (4.6) the maximum is indeed attained. By (3.3) and by transformation (4.6), the risk-adjusted value process $(R_t^{(X)})_{t=0,\dots,T}$ is a feasible solution in problem (4.8). Hence:

$$R_0^{(X)} \le R_0^*, \tag{4.9}$$

where R_0^* is part of an optimal solution in (4.8). For the inverse inequality, we argue by backward-induction over time. Let $(R_t)_{t=0,\dots,T}$ be a feasible solution. At final time, $R_T \leq X_T \overset{(3.3)}{=} R_T^{(X)}$. Let $t = 0, \dots, T-1$. Assume $R_{t+1} \leq R_{t+1}^{(X)}$. Then,

$$\begin{aligned}
R_t &\overset{(4.8)}{\leq} \min\Big\{X_t,\ \max_{q\in\mathbb{R}} \big(q - \frac{1}{\alpha}\mathbb{E}[(q - R_{t+1})^+|\mathcal{F}_t]\big)\Big\},\\
&\overset{(4.6)}{=} \min\big\{X_t,\ \min_{\mathbb{Q}\in\mathcal{P}^\alpha} \mathbb{E}_\mathbb{Q}[R_{t+1}|\mathcal{F}_t]\big\},\\
&\leq \min\big\{X_t,\ \min_{\mathbb{Q}\in\mathcal{P}^\alpha} \mathbb{E}_\mathbb{Q}[R_{t+1}^{(X)}|\mathcal{F}_t]\big\},\\
&\overset{(3.3)}{=} R_t^{(X)}, \qquad t = 0, \dots, T-1.
\end{aligned} \tag{4.10}$$

Combining (4.10) with inequality (4.9) gives $R_0^* = R_0^{(X)}$.

We have shown that the risk-adjusted value is the optimal solution of problem (4.8). It remains to compare problem (4.8) with problem (4.7). In particular, we have to show that the optimal objective value of the following problems (P) and (P') are the same.

$$(P) \qquad \begin{aligned} &\max R_0,\\ &\text{s.t.} \begin{cases} R_t \leq X_t, & t = 0, \dots, T,\\ R_t \leq V_t(R_{t+1}), & t = 0, \dots, T-1, \end{cases}\end{aligned}$$

with sub-problems

$$(P_{\text{sub,t}}) \qquad \begin{aligned} V_t(R_{t+1}) :=\ &\max\big\{Q_t - \frac{1}{\alpha}\mathbb{E}[Z_{t+1}|\mathcal{F}_t]\big\},\\ &\text{s.t.} \begin{cases} Z_{t+1} \geq Q_t - R_{t+1},\\ Z_{t+1} \geq 0, \end{cases}\end{aligned}$$

where $(P_{\text{sub,t}})$ is in fact a family of pointwise optimization problems on the nodes at time t; $(P_{\text{sub,t}})$ subsumes the pointwise optimization in (4.8), and the pointwise variable q is subsumed in the random variable Q_t. The problem (P') is the same as (4.7); only the indices of Z_t are shifted to Z_{t+1} to make the link to $(P_{\text{sub,t}})$:

$$(P') \qquad \begin{aligned} &\max R_0,\\ &\text{s.t.} \begin{cases} R_t \leq X_t, & t = 0, \dots, T,\\ R_t \leq Q_t - \frac{1}{\alpha}\mathbb{E}[Z_{t+1}|\mathcal{F}_t], & t = 0, \dots, T-1,\\ Z_{t+1} \geq Q_t - R_{t+1}, & t = 0, \dots, T-1,\\ Z_{t+1} \geq 0, & t = 0, \dots, T-1, \end{cases}\end{aligned}$$

where the variables R_t, Z_t and Q_t are $\mathcal{F}_t$-measurable random variables for all t (same measurability in problem (P)). We have to compare the optimal objective value of (P) and (P').

Let $(R_t)_{t=0\ldots T}$ be a feasible solution of (P). This implies that the sub-problems $(P_{\text{sub,t}})$ have attained optimal values $(Q_t)_{t=0\ldots T-1}$ and $(Z_t)_{t=1\ldots T}$. Inserting the combined solution $\big((R_t)_{t=0\ldots T}, (Q_t)_{t=0,\ldots T-1}, (Z_t)_{t=1\ldots T}\big)$ in the constraints of (P'), we can verify that this yields a feasible solution of (P'). This implies: Optimal objective value of $(P') \geq$ Optimal objective value of (P).

On the other hand, let $\big((R_t^*)_{t=0\ldots T}, (Q_t^*)_{t=0\ldots T-1}, (Z_t^*)_{t=1\ldots T}\big)$ be an optimal solution of (P'). Let t be fixed. Consider problem $(P_{\text{sub,t}})$ with the random variable R_{t+1}^* as parameter. It can be verified that (Q_t^*, Z_t^*) is feasible in $(P_{\text{sub,t}})$. Therefore, $V_t(R_{t+1}^*) \geq Q_t^* - \frac{1}{\alpha}\mathbb{E}[Z_{t+1}^*|\mathcal{F}_t] \geq R_t^*$, holding for all t. Therefore, $(R_t^*)_{t=0\ldots T}$ is feasible in (P). This implies that: Optimal objective value of $(P) \geq$ Optimal objective value of (P').

Combining the two inequalities for the optimal objective values of (P) and (P'), we conclude that the values are the same.

We have shown the equivalence of problem (4.7) with problem (4.8). This equivalence together with the inequalities (4.10) show the second part of the Lemma. ■

Finite Linear Programming Formulation

The foregoing optimization problem (4.7) is formulated in terms of random variables; this formulation is concise, and it highlights the general structure. In our finite setting, the formulation has to be on a scenario tree; according to Section 3.1, the scenario tree is defined to be the representation of the finite filtration $(\mathcal{F}_t)_{t=0,\ldots,T}$. The set of nodes at time $t = 0, \ldots, T$ is denoted by $\mathcal{N}_t$, and let the transition probabilities p_{nm} from each node n to another node m be given by Def. 8 (p. 26). The stochastic processes in the optimization problem are adapted to the finite filtration. Hence, all random variables at time t can be represented by their finitely many realizations on the scenario tree in the nodes $n \in \mathcal{N}_t$:

$$X_t, R_t, Q_t, Z_t \quad \rightarrow \quad (X_{tn})_{n\in\mathcal{N}_t}, (R_{tn})_{n\in\mathcal{N}_t}, (Q_{tn})_{n\in\mathcal{N}_t}, (Z_{tn})_{n\in\mathcal{N}_t},$$

where $X_{tn}, R_{tn}, Q_{tn}, Z_{tn} \in \mathbb{R}$. The conditional expectation $\mathbb{E}[Z_{t+1}|\mathcal{F}_t]$ in (4.7) corresponds pointwise (for each node n) to a finite sum:

$$\mathbb{E}[Z_{t+1}|\mathcal{F}_t]\Big|_n \quad \rightarrow \quad \sum_{m\in\mathcal{N}_{t+1}} p_{nm} Z_{(t+1)m} \quad \text{for all } n \in \mathcal{N}_t,\ t = 0, \ldots, T-1.$$

Note that the sum goes over all nodes at time $t+1$, but only the transition probabilities from node n to its successor nodes can be strictly positive. The constraint $Z_t \geq Q_{t-1} - R_t$ in (4.7) connects two subsequent time steps. In the finite setting, this corresponds to a connection on an edge in the scenario tree:

$$Z_t \geq Q_{t-1} - R_t\,\Big|_n \quad \rightarrow \quad Z_{tn} \geq Q_{(t-1)n^-} - R_{tn} \quad \text{for all } n \in \mathcal{N}_t,\ t = 1, \ldots, T-1,$$

where n^- is the parent node of n.

Thus, the finite linear programming formulation of (4.7) on the scenario tree reads

$$\max R_{0n_0},$$
$$\text{s.t.} \begin{cases} R_{tn} \le X_{tn}, & t = 0,\dots,T,\ \forall n \in \mathcal{N}_t, \\ R_{tn} \le Q_{tn} + \dfrac{1}{\alpha} \displaystyle\sum_{m \in \mathcal{N}_{t+1}} p_{nm} Z_{(t+1)m}, & t = 0 \dots T-1,\ \forall n \in \mathcal{N}_t, \\ Z_{tn} \ge Q_{(t-1)n^-} - R_{tn}, & t = 1 \dots T,\ \forall n \in \mathcal{N}_t, \\ Z_{tn} \ge 0, & t = 1 \dots T,\ \forall n \in \mathcal{N}_t, \end{cases} \tag{4.11}$$

where n_0 denotes the root node of the scenario tree, $X_{tn} \in \mathbb{R}$ are parameters, and $R_{tn} \in \mathbb{R}$, $Q_{tn} \in \mathbb{R}$ and $Z_{tn} \in \mathbb{R}$ are the variables.

Note. The assumption that the parameter α is a constant was not needed in the derivations so far. Hence we can generalize (4.11) to node-dependent α_n:

$$\frac{1}{\alpha} \to \frac{1}{\alpha_n}.$$

The finite representation (4.11) has been derived from the general recursive definition (3.3) (p. 17), where the set of probability measures $\mathcal{P}$ was specialized to the local-CVaR set $\mathcal{P}^\alpha$, and where the inner minimization was replaced by a maximization (see (4.6), p. 29). Let us assume that in the finite setting that $\alpha \le p_{nm}$ for all strictly positive transition probabilities p_{nm}, $n \in \mathcal{N}_t$, $m \in \mathcal{N}_{t+1}$, $t = 0,\dots,T-1$, then it can be verified from (3.3) that

$$R_0^{(X)} = \min_{\substack{n \in \mathcal{N}_t \\ t=0,\dots,T}} X_{tn}.$$

Hence, if α is sufficiently small, then the risk-adjusted value is the global minimum over all scenarios and all time. To avoid such a worst-case risk measurement, a sufficient condition is that all transition probabilities p_{nm} are strictly smaller than α. Hence, each non-terminal node must have a sufficiently large number of immediate successor nodes. Unfortunately, such a tree grows exponentially over time. For example, if each non-terminal node has m immediate successor nodes, then the number of nodes, as well as the number of variables and the number of constraints in problem (4.11) is each a polynomial in m of order T.

4.3.2 Risk-Mean Optimization (Case I)

The recursive risk-adjusted value may be used in a constraint of an optimization problem. The considered risk-mean multi-period stochastic optimization problem maximizes the expected profit under a constraint on the risk-adjusted value (Def. 4, p. 16):

$$\max \mathbb{E}[g(X_0,\dots,X_T)], \\ \text{s.t.} \begin{cases} R_0^{(X)} \geq \rho_{\min}, \\ (X_t)_{t=0,\dots,T} \in \mathcal{X}, \end{cases} \tag{4.12}$$

where $g\colon \mathbb{R}^{T+1} \to \mathbb{R}$ is a measurable function. The recursive risk-adjusted value $R_0^{(X)}$ of the process $(X_t)_{t=0\dots,T}$ is bounded from below by a constant $\rho_{\min} \in \mathbb{R}$. The set $\mathcal{X}$ denotes the feasible set of stochastic processes. Usually, $X_t \in \mathcal{X}_t$, where $\mathcal{X}_t$ is a subset of $\mathcal{F}_t$-measurable random variables. We assume that g and $\mathcal{X}$ are such that the optimal objective value exists and is finite.

In the finite setting on a scenario tree, the foregoing Lemma 5 permits us to write the constraint on risk as a system of linear inequalities:

Proposition 3. *Let $\mathcal{P}^\alpha$ be a local-CVaR set of probability measures on a scenario tree. An optimal solution of problem* (4.12) *on a scenario tree is given by optimal values of the variables $(X_t)_{t=0,\dots,T}$ of the following stochastic linear optimization problem:*

$$\max \mathbb{E}[g(X_0,\dots,X_T)], \\ \text{s.t.} \begin{cases} R_0 \geq \rho_{\min}, & \\ R_t \leq X_t, & t=0,\dots,T, \\ R_t \leq Q_t - \dfrac{1}{\alpha}\mathbb{E}[Z_{t+1}|\mathcal{F}_t], & t=0,\dots,T-1, \\ Z_t \geq Q_{t-1} - R_t, & t=1,\dots,T, \\ Z_t \geq 0, & t=1,\dots,T, \\ (X_t)_{t=0,\dots,T} \in \mathcal{X}, & \end{cases} \tag{4.13}$$

where X_t, R_t, Q_t and Z_t are $\mathcal{F}_t$-measurable random variables.

Alternatively to (4.12), the risk could be minimized under a constraint on the expected profit; the foregoing proposition undergoes only a slight modification.

The proof goes along the same lines as the previous Lemma.

Proof. The optimization problem (4.12) is of the form of problem (P) below, where the abbreviation $X := (X_t)_{t=0,\dots,T}$ is used. By Lemma 5, the risk-adjusted value can be calculated by a linear stochastic program, and the optimal value is attained; this sub-problem is given by (P_{sub}) in condensed form: $v(Y)$ is the objective function with variables $Y := (\mathbf{Y}_t)_{t=0,\dots,T}$. The matrix and the right-hand-side of the linear constraints are denoted by A and b, respectively.

The linear programming formulation of the optimization problem in (4.13) is then of the form (P'):

$$(P) \qquad \begin{aligned} &\max\ \mathbb{E}[g(X)], \\ &\text{s.t.} \begin{cases} \pi[X] \geq \rho_{\min}, \\ X \in \mathcal{X}. \end{cases} \end{aligned} \qquad\qquad (P') \qquad \begin{aligned} &\max\ \mathbb{E}[g(X)], \\ &\text{s.t.} \begin{cases} v(Y) \geq \rho_{\min}, \\ A(X,Y) \leq b, \\ X \in \mathcal{X}. \end{cases} \end{aligned}$$

$$(P_{\text{sub}}) \qquad \begin{aligned} \pi[X] = &\max\ v(Y), \\ &\text{s.t.}\ A(X,Y) \leq b. \end{aligned}$$

We have to compare the optimal objective value of (P) and (P').

Let X be a feasible solution of (P), and let Y be the corresponding optimal value of the sub-problem (P_{sub}). If we insert the combined solution (X, Y) in the constraints of (P'), we see that this yields a feasible solution of (P'). Hence: Optimal objective value of $(P') \geq$ Optimal objective value of (P).

On the other hand, let (X^*, Y^*) be an optimal solution of (P'). Consider problem (P_{sub}) with X^*: Y^* is feasible in (P_{sub}). Therefore, $\pi[X^*] \geq v(Y^*) \geq \rho_{\min}$, and X^* is feasible in (P). This implies: Optimal objective value $(P) \geq$ Optimal objective value (P').

Combining the two inequalities for the optimal objective values of (P) and (P'), we conclude that the values are the same. ■

4.4 A Lower Bound (Case II)

In the forthcoming energy model, the risk will be constrained by a recursive risk-adjusted value for processes (Case I) with a local-CVaR set of test-probability measures. Despite the convenient properties of this risk-adjusted value, a simple economic interpretation can so far not be given. Nevertheless, the related and simpler Case II (random variable) is more tractable, and a lower bound in terms of the single-period risk-adjusted value CVaR can be given. The bound provides a hint how to choose the α parameter for the local-CVaR set.

Proposition 4. *Let $\mathcal{P}^\alpha$ be a local-CVaR set of probability measures on a scenario tree to a given level α. Let X_T be an $\mathcal{F}_T$-measurable random variable. Then the recursive risk-adjusted value $R_0^{X_T}$ with set $\mathcal{P}^\alpha$ (Def. 5, p. 18) is bounded from below by*

$$\underset{(\textit{at level } \alpha^T)}{\text{CVaR}[X_T]} \leq \underset{(\textit{at level } \alpha)}{R_0^{X_T}}, \tag{4.14}$$

where the single-period risk-adjusted value CVaR *is taken at level α^T.*

Proof. By a complete analogue to Lemma 5 (p. 30), we can write the calculation of the recursive risk-adjusted value $R_0^{X_T}$ as a linear program:

$$\max R_0$$
$$\text{s.t.} \begin{cases} R_T \le X_T, & & |H \ge 0 \\ R_t \le Q_t - \frac{1}{\alpha}\mathbb{E}[Z_{t+1}|\mathcal{F}_t], & t = 0, \dots, T-1, & |K_t \ge 0 \\ Z_t \ge Q_{t-1} - R_t, & t = 1, \dots, T, & |L_t \ge 0 \\ Z_t \ge 0, & t = 1, \dots, T, & \end{cases}$$

where we have indicated the random variables H, K_t and L_t as Lagrange-multipliers for the dualization in the next step of the proof; H is $\mathcal{F}_T$-measurable, and K_t, L_t are $\mathcal{F}_t$-measurable. In our finite setup, the straightforward way to obtain the dual linear program involves writing the constraints in matrix-form and transposing. For multi-stage problems, establishing the matrix-form can be cumbersome. An equivalent approach is Lagrangian dualization, which also suits a possible infinite-dimensional extension. The Lagrangian is

$$L\big[(R_t)_{t=0,\dots,T}, (Q_t)_{t=0,\dots,T-1}, (Z_t)_{t=1,\dots,T};\ H, (K_t)_{t=0,\dots,T-1}, (L_t)_{t=1,\dots,T}\big] =$$
$$= R_0 + \mathbb{E}\big[H(X_T - R_T)\big] + \sum_{t=0}^{T-1} \mathbb{E}\big[K_t(Q_t - R_t - \frac{1}{\alpha}\mathbb{E}[Z_{t+1}|\mathcal{F}_t])\big] +$$
$$+ \sum_{t=1}^{T} \mathbb{E}\big[L_t(Z_t - Q_{t-1} + R_t)\big]$$
$$= R_0(1 - K_0) + \sum_{t=1}^{T-1} \mathbb{E}\big[R_t(L_t - K_t)\big] + \mathbb{E}\big[R_T(L_T - H)\big] +$$
$$+ \sum_{t=0}^{T-1} \mathbb{E}\big[Q_t(K_t - \mathbb{E}[L_{t+1}|\mathcal{F}_t])\big] + \sum_{t=0}^{T-1} \mathbb{E}\big[Z_{t+1}(L_{t+1} - \frac{1}{\alpha}K_t)\big] -$$
$$+ \mathbb{E}[HX_T].$$

$$\sup_{\left\{\substack{R_t, Q_t, \\ Z_t \ge 0}\right\}} L < \infty \iff \begin{cases} K_0 = 1, & \\ K_t = L_t, & t = 1, \dots, T-1, \\ H = L_T, & \\ K_t = \mathbb{E}[L_{t+1}|\mathcal{F}_t], & t = 0, \dots, T-1, \\ L_{t+1} \le \frac{1}{\alpha}K_t, & t = 0, \dots, T-1. \end{cases} \tag{4.15}$$

This system of restrictions is equivalent to

$$\mathbb{E}[H] = 1 \quad \wedge \quad \mathbb{E}[H|\mathcal{F}_{t+1}] \le \frac{1}{\alpha}\mathbb{E}[H|\mathcal{F}_t], \qquad t = 0, \dots, T-1.$$

Thus, the dual linear program is

$$
\begin{aligned}
&\min_H \ \mathbb{E}[HX_T],\\
&\text{s.t.} \begin{cases} \mathbb{E}[H] = 1, \\ H \geq 0, \\ \mathbb{E}[H|\mathcal{F}_{t+1}] \leq \dfrac{1}{\alpha}\mathbb{E}[H|\mathcal{F}_t], \quad t = 0, \ldots, T-1, \qquad (*) \end{cases}
\end{aligned}
\tag{4.16}
$$

where H is an $\mathcal{F}_T$-measurable random variable.

We show that this dual is a relaxation of the optimization problem that calculates $\mathrm{CVaR}[X_T]$; this will complete the proof. First, we note that the constraints $\mathbb{E}[H] = 1$ and $H \geq 0$ in problem (4.16) characterize a Radon-Nikodym probability density. Notice that by general assumption, all nodes have strictly positive probability under $\mathbb{P}$. This implies for $H \geq 0$ that

$$
\{\mathbb{E}[H|\mathcal{F}_t] = 0\} \subseteq \{\mathbb{E}[H|\mathcal{F}_{t+1}] = 0\}, \qquad t = 0, \ldots, T-1, \tag{4.17}
$$

with abbreviation $\{Y = 0\} := \{\omega \in \Omega \mid Y(\omega) = 0\}$ for a random variable Y. We will relax the set $(*)$ of constraints:

$$
\mathbb{E}[H|\mathcal{F}_{t+1}] \leq \frac{1}{\alpha}\mathbb{E}[H|\mathcal{F}_t], \qquad t = 0, \ldots, T-1, \qquad (*).
$$

As an implication of $(*)$, we can consider the multiplied constraints (separately on the left- and right-hand sides); the implication holds because the terms are non-negative, and because by inclusion (4.17), we can multiply even on events where the right-hand-side is zero. Hence, $(*)$ implies that

$$
\prod_{t=0}^{T-1} \mathbb{E}[H|\mathcal{F}_{t+1}] \leq \frac{1}{\alpha^T} \prod_{t=0}^{T-1} \mathbb{E}[H|\mathcal{F}_t], \text{ or } H \prod_{t=0}^{T-1} \mathbb{E}[H|\mathcal{F}_t] \leq \frac{1}{\alpha^T} \prod_{t=0}^{T-1} \mathbb{E}[H|\mathcal{F}_t],
$$

where we have used $\mathbb{E}[H|\mathcal{F}_T] = H$ and $\mathbb{E}[H|\mathcal{F}_0] = 1$. By dividing, we get

$$
H \leq \frac{1}{\alpha^T} \qquad \text{on } \Big\{ \prod_{t=0}^{T-1} \mathbb{E}[H|\mathcal{F}_t] > 0 \Big\}.
$$

On the complementary event, $\{\prod_{t=0}^{T-1} \mathbb{E}[H|\mathcal{F}_t] = 0\}$, we have by (4.17) that $\mathbb{E}[H|\mathcal{F}_T] = 0$ on that event, which is $H = 0$ on that event. Hence, the set $(*)$ of constraints implies that $H \leq 1/\alpha^T$ on every event, which is the same constraint as in the optimization problem (2.5) of CVaR at level α^T. Because the other constraints of (2.5) and (4.16) are the same (they just characterize a Radon-Nikodym density), the optimization problem of $\mathrm{CVaR}[X_T]$ is a relaxation of the problem of $R_0^{X_T}$. ■

The definition of $\mathrm{CVaR}[X_T]$ involves only the distribution of X_T (cf. (2.7)), and not the random variable as seen as a mapping $X_T\colon \Omega \to \mathbb{R}$. Hence, the right-hand-side of (4.14) can be additionally minimized over all random variables that have the same distribution-law as X_T:

Corollary 1. *Assume the same setting as in Proposition 4. Then*

$$\underset{(at\ level\ \alpha^T)}{\mathrm{CVaR}[X_T]} \;\; \le \underset{(at\ level\ \alpha)}{\min_{Y_T \in \mathcal{Y}} R_0^{Y_T}}, \tag{4.18}$$

where $\mathcal{Y}$ is the set of $\mathcal{F}_T$-measurable random variables that have the same distribution-law as X_T.

4.5 Summary

Artzner et al. [5] extended the definition of coherent risk measurement to multiple time periods. In the multi-period setting, a recursive definition over time ensures time consistency. Our aim was to show that a special instance of such a recursive risk-adjusted value is –in principle– numerically tractable, and that it can be applied in linear mean-risk optimization problems.

In particular, we considered a finite setup (scenario tree), and a coherent risk-adjusted value that is defined by local-CVaR sets on the scenario tree. A local-CVaR set is a set of test-probability measures that is locally (nodewise) defined by the same box-constraints as the single-period risk measure CVaR. The similarity to CVaR gave a lower bound of the risk-adjusted value in terms of CVaR. By applying the duality of CVaR, the recursive calculation could be written as a multi-stage stochastic linear program, and a lower bound of the risk-adjusted value could be incorporated by a set of linear constraints in multi-stage mean-risk optimization problems.

To prevent the risk-adjusted value to be equal to the worst-case scenario value, we have argued that it suffices that each non-terminal node of the scenario tree has a sufficiently large number of immediate successor nodes. This results in inevitably large sizes of the linear programs, such that the problems are numerically demanding.

Chapter 5

Dispatch and Control

This chapter fixes the general setup of the control of an electricity plant. In the first part of the chapter, the dispatch decision of the electricity plant is explained in general terms. The second part reviews shortly the notion of stochastic control (as far as it is needed for the forthcoming models of the electricity plant).

5.1 General Setting of the Electricity Plant

The hydro-storage plant consists of an upper reservoir of water and of a facility to produce electricity from the potential energy of the water. The electricity is sold on the spot market. The conversion factor of the amount of stored water into the amount of produced electricity is assumed to be constant with respect to different water levels. Hence, for convenience, the amount of water is measured in units of producible electricity (MWh).

We consider a hydro *pumped*-storage plant that has the additional flexibility to pump water from a lower reservoir into the upper reservoir. The lower reservoir is assumed to be large enough to allow maximal pumping at all times. The electricity to operate the pumps is bought on the spot market. Because the water level in the upper reservoir is measured in units of producible electricity, only the back-conversion of electricity to water in the upper reservoir is accompanied by an efficiency-loss factor.

The decision for a specific rate of production (selling) energy or for a specific rate of pumping (buying) energy at a specific point in time is the *dispatch decision* at this specific point in time. In practice, at the EEX market, the spot price of electricity changes hourly. Preferably, the decision should change on that short-term time scale, too. The decision may be based not only on the current spot price, but also on other exogenous quantities, or on the current state of the plant, like the water level.

The market is assumed to have unlimited liquidity, and the plant is assumed to be price-taker on the market; a price-taker cannot influence prices with own trading operations.

The reservoir is a renewable resource that has a natural inflow of water. The dispatch decision must ensure that the water level stays between an upper and a lower bound. In fact, the upper bound is never violated because an arbitrary spill-over is allowed. Because the spill is an energy loss and therefore reduces the value of the plant, optimal dispatch decisions will cause a relatively small spill. If the reservoir would have the additional purpose of a flood detention basin, an excessive spill could be penalized.

The modeling of the plant in terms of *stochastic optimal control* as well as *multi-stage stochastic linear programming* will be given in Chapter 7. Before we proceed to this general model, the next section explains the notion of stochastic optimal control, and Chapter 6 presents some simplified, exactly solvable models. The current section ends with a short literature review of related models:

For short-term planning of a power system, single-period models are sufficient [64]. Mid and long-term planning (several months up to several years) require a dynamic setting. A first category of multi-stage stochastic programming models addresses the generation of power under stochastic demand; the electricity market is no yet considered [3, 40, 41, 63, 66]. A second category uses the flexibility of a spot and futures market. Numerically tractable multi-stage stochastic programming models have usually a finite state space, which leads to a formulation on a scenario tree with a limited number of stages. Therefore, it is difficult to incorporate the hourly trading activity on the spot market in mid and long-term models. One way is to consider a single price during the stages [33, 36].

The use of a scenario tree can be circumvented if the model allows a path-wise formulation over time (no conditional expectations etc. are involved). In such models, the scenario tree can be given by a simple *fan*[1]. Such models can be solved by path-wise Monte-Carlo sampling. Path-wise models with hourly time steps and a comparably large time horizon of several months are numerically tractable [79]. Even large power portfolios (hydro, thermal, wind), together with market activities in spot and futures, have been proven to be numerically solvable by Döge and Lüthi [23, 24]. As a drawback, the decisions are not path-dependent on a scenario tree.

An alternative approach is to view the profit of a hydro-storage power plant as the payoff of a series of contingent option contracts; exercising an option means to produce electricity at a specific point in time [79]. Unfortunately, financial derivative pricing theory cannot be easily carried over to electricity markets, because, in general, the payoff of an electricity plant cannot be replicated by a portfolio of exchange-traded contracts, for which prices are known. A related, novel approach uses financial interest rate theory to valuate a production utility [44].

[1]The scenarios have only the root node in common.

5.2 Stochastic Control with Exogenous Observables

Stochastic control theory considers the optimal control of a stochastic process. In many applied problems, the control cannot influence the observations; hence, the observations are *exogenously* given. By freeing up the probability structure from being influenced by the control, the problem is numerically tractable, and can be formulated as a stochastic programming problem [71].

In the following, we give the general setting of a stochastic control problem with exogenous observables. The time scale is assumed to be discrete with finite horizon: $t = 0, 1, \ldots, T$. In fact, the forthcoming general model of the electricity plant will use a combination of a small and large time scale, but the basic notation remains the same.

Let a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ be given; for the general setting, we do not need any finiteness assumption. The stochastic control model in discrete time consists of the following:

1. The stochastic process of the controlled object: $(\boldsymbol{X}_t)_{t=0,\ldots,T}$. Commonly, the random vector $\boldsymbol{X}_t \colon \Omega \to \mathbb{R}^n$ is called the vector of *state variables* at time t. The state variables have to be in a *feasible* set: $(\boldsymbol{X}_t)_{t=0,\ldots,T} \in \mathcal{X}$ *a.s.*.

2. The stochastic process of the exogenous variables: $(\boldsymbol{E}_t)_{t=0,\ldots,T}$, where $\boldsymbol{E}_t \colon \Omega \to \mathbb{R}^m$. The distribution of this process is fixed; it cannot be influenced by the control. In economic sciences, it is common to refer to them still as 'variables' , because changes in these quantities affect the optimization problem.

3. The stochastic process of the control(-decisions): $(\mathbf{U}_t)_{t=0,\ldots,T-1}$, where $\boldsymbol{U}_t \colon \Omega \to \mathbb{R}^k$. The process is given by a sequence of functions $\mathbf{u}_t \colon (\mathbb{R}^n \times \mathbb{R}^n \times \cdots \times \mathbb{R}^m) \to \mathbb{R}^k$ of the values of the past-and-present state variables and exogenous variables:
$$\boldsymbol{U}_t := \boldsymbol{u}_t(\boldsymbol{X}_0, \ldots, \boldsymbol{X}_t, \boldsymbol{E}_0, \ldots, \boldsymbol{E}_t). \tag{5.1}$$
The vector-valued control-function $\mathbf{u}_t$ is assumed to be measurable[1]. The functions are in an *admissible* set: $(\boldsymbol{u}_t)_{t=0,\ldots,T-1} \in \mathcal{U}$.

4. The state equations: The state variables at time t are a function of the past values of the state variables and controls, and of the past-and-present values of the exogenous variables:
$$\boldsymbol{X}_t = \boldsymbol{f}_t(\boldsymbol{X}_0, \ldots, \boldsymbol{X}_{t-1}, \boldsymbol{U}_0, \ldots, \boldsymbol{U}_{t-1}, \boldsymbol{E}_0, \ldots, \boldsymbol{E}_t) \quad a.s., \tag{5.2}$$
where the vector-valued function $\boldsymbol{f}_t$ is assumed to be measurable.

[1] i.e. Borel-measurable: $\mathcal{B}((\mathbb{R}^n)^{t+1} \times (\mathbb{R}^m)^{t+1})$- $\mathcal{B}(\mathbb{R}^k)$-measurable

5. The objective is to maximize the expectation of a measurable function $g\colon \mathbb{R}^n \times \cdots \times \mathbb{R}^k \to \mathbb{R}$:

$$\sup \mathbb{E}\big[g(\mathbf{X}_0,\dots,\mathbf{X}_T,\mathbf{U}_0,\dots,\mathbf{U}_{T-1})\big],$$
$$\text{s.t.} \begin{cases} \boldsymbol{U}_t := \boldsymbol{u}_t(\boldsymbol{X}_0,\dots,\boldsymbol{X}_t,\boldsymbol{E}_0,\dots,\boldsymbol{E}_t), & t=0,\dots,T-1,\\ \boldsymbol{X}_t = \boldsymbol{f}_t(\boldsymbol{X}_0,\dots,\boldsymbol{X}_{t-1},\boldsymbol{U}_0,\dots,\boldsymbol{U}_{t-1},\boldsymbol{E}_0,\dots,\boldsymbol{E}_t) \quad a.s., & \\ & t=0,\dots,T-1,\\ (\mathbf{u}_t)_{t=0,\dots,T-1} \in \mathcal{U}, & \\ (\boldsymbol{X}_t)_{t=0,\dots,T} \in \mathcal{X} \quad a.s.. & \end{cases} \tag{5.3}$$

To ensure that the expectation in the objective function exists (in the sense of e.g. [6, p. 74]), a sufficient condition is that the integrand is bounded from below. A sufficient condition for boundedness is that the feasible state variables are uniformly bounded from below *a.s.*, and that the admissible control-functions are uniformly bounded, and that the function $g\colon \mathbb{R}^n \times \cdots \times \mathbb{R}^k \to \mathbb{R}$ is continuous, and for the first $(T+1)$ arguments (i.e. the state variables) monotonically increasing. These conditions may be met in applications.

Apart from the mere existence of the expectation in (5.3), the existence of an optimal solution of (5.3) can be assured only under several technical assumptions in the infinite state-space setting, see e.g. Bertsekas and Shreve [10]. We do not attempt to prove the existence of an optimal solution for the stochastic optimization problem of the general electricity plant (in the forthcoming Ch. 7). In fact, the optimization problem that will be numerically solvable uses a finite state space. The resulting finite *linear program* is either infeasible, unbounded, or has an optimal solution.

Chapter 6

Exact Solutions

In this chapter, exact solutions of some simple dispatch problems are given. The problems generalize those of Unger [79, Ch. 6.5].

The problems have the structure of optimization problems that are used to calculate coherent risk-adjusted values. This shows a strong relationship between coherent risk measurement and certain production problems.

Throughout the chapter, the probability space is $(\Omega, \mathcal{F}, \mathbb{P})$, which has no finiteness assumption. In fact, to simplify the proofs, the spot price is assumed to have a *continuous* distribution. Nevertheless, the method of proof is extendable to discrete distributions (cf. Section 11.2).

6.1 Single-Period Production Model

The model considers a single period in time. The single-period model can be considered as an aggregation of a multi-period model that has a stationary distribution of the electricity price [79, Ch. 6.5].

The electricity price (Euro/MWh) is a non-negative random variable S with finite expectation: $S \in L^1_+ := L^1_+(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R})$.

The control is assumed to be given by a Borel-measurable function of the electricity price:

$$\begin{aligned} u\colon \mathbb{R}_+ &\to \mathbb{R}_+, \\ S &\mapsto u(S). \end{aligned}$$

The value of the control-function u is the amount of produced energy (MWh). The maximal capacity of energy production for a single period is denoted by $u_{\max} \in \mathbb{R}_+$. The control-function is assumed to be chosen at the beginning of the period, and the objective is to maximize the expected profit at the end of the period.

The produced energy is drawn from a water reservoir. The initial water level of the reservoir is denoted by $l_0 \in \mathbb{R}$, and has units of (MWh) (see Sec. 5.1 concerning this unit). In this thesis, exact solutions can be given only for water-level constraints 'in expectation', and not in the 'almost sure' sense. Hence, the expected final level is subject to stay above a lower bound $l_{\min} \in \mathbb{R}$. An inflow of water is not explicitly modeled. Nevertheless, l_0 can be a sum of an initial water level and an expected inflow.

The production model can be formulated as a stochastic control problem:

$$\max_u \mathbb{E}[S\,u(S)]$$
$$\text{s.t.} \begin{cases} l_0 - \mathbb{E}[u(S)] \geq l_{\min}, & \\ 0 \leq u(s) \leq u_{\max} & \text{for all } s \in \mathbb{R}_+, \\ u\colon \mathbb{R}_+ \to \mathbb{R}_+, & u \text{ measurable.} \end{cases} \tag{6.1}$$

To exclude trivial cases, we assume that $l_{\min} < l_0$; otherwise, there would be no usable water in the reservoir. The expectation in the objective function exists because $S \in L^1$ and u is bounded and measurable.

To relate the production problem (6.1) to coherent risk measurement, we need the following Lemma.

Lemma 6. *Consider problem* (6.1). *If the usable energy is strictly less than the production capacity,*

$$l_0 - l_{\min} < u_{\max},$$

and if the electricity price is almost surely greater than zero,

$$\mathbb{P}[S = 0] = 0,$$

then the constraint on water is binding in the optimum.

In reality, $\mathbb{P}[S = 0] = 0$ is always true because the electricity price is never zero.

Proof. Let u_1 be an optimal solution of (6.1) under the assumption $l_0 - l_{\min} < u_{\max}$, and let u_2 be an optimal solution of the relaxed problem without the constraint on water (first ineq. in (6.1)). Clearly, $u_2(\cdot) \equiv u_{\max}$, which implies $u_1(s) \leq u_2(s)$ $\forall s \in \mathbb{R}_+$. Because of the assumption $l_0 - l_{\min} < u_{\max}$, u_2 violates the constraint on water. Thus, on a set of positive probability, u_1 is strictly smaller than u_2: $\mathbb{P}[u_1(S) < u_2(S)] > 0$. Let us assume that u_1 is such that the constraint on water is not binding. Then there is a $\lambda \in (0, 1)$ such that a convex combination of u_1 and u_2 has a binding constraint on water:

$$l_0 - l_{\min} = \lambda \underbrace{\mathbb{E}[u_1(S)]}_{< l_0 - l_{\min}} + (1 - \lambda) \underbrace{\mathbb{E}[u_2(S)]}_{> l_0 - l_{\min}} = \mathbb{E}[\lambda u_1(S) + (1 - \lambda) u_2(S)].$$

Clearly, the convex combination is feasible. Its objective value is

$$\mathbb{E}\big[S\big(\lambda u_1(S) + (1-\lambda)u_2(S)\big)\big] = \lambda\mathbb{E}[Su_1(S)] + (1-\lambda)\mathbb{E}[Su_2(S)] \leq \mathbb{E}[Su_2(S)].$$

The last inequality can be made strict because if $\mathbb{P}[S=0]=0$, then

$$\mathbb{P}[\{u_1(S) < u_2(S)\} \cap \{S>0\}] = \mathbb{P}[u_1(S) < u_2(S)] > 0. \qquad \blacksquare$$

Proposition 5. *Let the distribution function of S be continuous, and let the assumptions of Lemma 6 be fulfilled. Then there exists an optimal solution of problem* (6.1) *that is an indicator function*

$$u^*(S) = u_{\max}\chi_{\{S \geq q_{(1-\beta)}(S)\}},$$

where $q_{(1-\beta)}(S)$ is a $(1-\beta)$-quantile of S, with

$$\beta := \frac{l_0 - l_{\min}}{u_{\max}},$$

and $\chi_A\colon \Omega \to \{0,1\}$ denotes the indicator function for sets $A \subseteq \Omega$.

Proof. Let us make the substitution $H := \frac{u(S)}{l}$ *a.s.*, where $l := l_0 - l_{\min}$ and $H \in L^\infty(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R})$. By Lemma 6, we can assume that the constraint on water is an equality. Therefore, a relaxation of problem (6.1) is

$$-\min_H \mathbb{E}[-lHS],$$
$$\text{s.t.} \begin{cases} \mathbb{E}[H] = 1, \\ 0 \leq H \leq \dfrac{1}{\beta} \quad \textit{a.s.}. \end{cases}$$

Except for the sign in front of the 'min', this is the optimization problem (2.5) (p. 8) of the risk-adjusted value $\mathrm{CVaR}[-lS]$ with solution (2.10, second line):

$$H^* = \frac{1}{\beta}\chi_{\{(-lS) \leq q_\beta(-lS)\}} \qquad \textit{a.s.}. \tag{6.2}$$

Let us substitute u^* back and use the definition of β:

$$u^*(S) = u_{\max}\chi_{\{lS \geq -q_\beta(-lS)\}}.$$

Finally, we use the fact that quantiles are positively homogeneous and that $q_\beta(\cdot) = -q_{1-\beta}(-\cdot)$ (Appendix A.1, Lemma 8, p. 113). It can be checked that the solution u^* of the relaxed problem is also feasible in the original problem (6.1). $\blacksquare$

By the foregoing proof, we can associate to problem (6.1) the dual of the optimization problem of the risk-adjusted value CVaR (2.7):

$$\min_{q\in\mathbb{R}} \; lq + u_{\max}\mathbb{E}[(S-q)^+] = -\max_{\widetilde{q}\in\mathbb{R}} \; \widetilde{q} - \frac{1}{\beta}\mathbb{E}[(\widetilde{q}-(-lS))^+], \tag{6.3}$$

where $\widetilde{q} := -lq$, $l := l_0 - l_{\min}$. By Prop. 1 (p. 8), an optimal $\widetilde{q}^*$ is a β-quantile of $-lS$. Therefore, the optimal q^* is a $1-\beta$-quantile of $S \in L^1_+$. Hence, by the non-negativity of S, we can consider on the left-hand side the restricted problem '$\min_{q\geq 0}$'. The minimization problem has an economical interpretation: Let us assume that the owner of the electricity plant wants to sell the plant with the following contract:

- The water surplus $l = l_0 - l_{\min}$ in the reservoir is sold at a fixed price q, which can be chosen freely by the buyer;
- If the (future) electricity price S is greater than q, then the buyer has to pay the amount $u_{\max}(S-q)$ to the seller at the end of the period. This is the buyer's surplus because of the flexibility of producing at full capacity if prices are high.

The buyer minimizes the expected total cost of the contract.

Because the proof gives a correspondence between problem (6.1) and problem (2.5), q can be considered as the Lagrange multiplier of the water constraint. Therefore, applying the theory of parametric linear optimization, we can interpret the optimal quantile q^* as the marginal price of water.

Note. The optimal solution has a zero-one-behavior: If the electricity price surpasses a certain quantile, then the decision maker produces at full capacity. If the price is less, there is no production at all. In general, if the admissible controls take values in a compact subset of $\mathbb{R}^n$, then an optimal control that takes only values on the boundary is called a *bang-bang* solution [60]. It can be shown that, for certain problem classes, bang-bang solutions are extreme points of the feasible set, where the feasible set lays in a space of functions. For further details, see e.g. Bauer's Maximum Principle [83, Example VIII.6.34], Lyapunov's theorem and its proof [83, Prop. VIII.4.8], and the pioneering work of LaSalle [60].

6.2 Single-Period Dispatch Model

Let us extend the foregoing production model by a control-function for *pumping*: $u^- : \mathbb{R}_+ \to \mathbb{R}_+$. The value of u^- is the amount of water that is pumped-up. The control-function for producing is now denoted with a plus sign: u^+. Both functions are assumed to be Borel-measurable.

The efficiency of pumping[1] is $0 < c < 1$. As in the previous model, we exclude the possibility of production at full capacity:

$$u^+_{\max} > l_0 - l_{\min} > 0. \tag{6.4}$$

The stochastic control problem of producing and pumping is

$$\max_{u^+,u^-} \mathbb{E}\Big[Su^+(S) - \frac{1}{c}Su^-(S)\Big],$$
$$\text{s.t.} \begin{cases} l_0 - \mathbb{E}\big[u^+(S) - u^-(S)\big] \geq l_{\min}, \\ 0 \leq u^+(s) \leq u^+_{\max} & \text{for all } s \in \mathbb{R}_+, \\ 0 \leq u^-(s) \leq u^-_{\max} & \text{for all } s \in \mathbb{R}_+, \\ u^\pm \colon \mathbb{R}_+ \to \mathbb{R}_+, & \text{measurable.} \end{cases} \tag{6.5}$$

The expectation in the objective function exists because $S \in L^1_+$ and $u^\pm$ are bounded measurable functions.

Proposition 6. *Let the distribution function of S be continuous. Then an optimal solution of problem* (6.5) *is*

$$\overset{*}{u}{}^+(S) = u^+_{\max}\chi_{\{S \geq q^*\}}, \qquad \overset{*}{u}{}^-(S) = u^-_{\max}\chi_{\{S \leq cq^*\}}, \tag{6.6}$$

where q^*, which is the optimal Lagrange multiplier of the water constraint, is given by a solution of the equation

$$u^+_{\max}\mathbb{P}[S \geq q^*] - u^-_{\max}\mathbb{P}[S \leq cq^*] = l_0 - l_{\min}. \tag{6.7}$$

If the efficiency c of pumping is strictly smaller than 1, then the optimal solution (6.6) exhibits no simultaneous producing and pumping.

We need a short Lemma for the proof.

Lemma 7. *The set of solutions q^* of* (6.7) *is a non-empty, non-negative interval.*

Proof. The left-hand-side of (6.7) is a monotonous (decreasing) function in the variable q. Hence, the set of solutions q^* of (6.7) is an interval. The function tends to $-u^-_{\max}$ for $q \nearrow \infty$, and to $u^+_{\max}$ for $q \searrow 0$. By the continuity of the function, a solution of (6.7) can always be attained under assumption (6.4). Hence, the interval is non-empty. In addition, the interval is non-negative. This holds because $q < 0$ implies $\mathbb{P}[S \geq q] = 1$ and $\mathbb{P}[S \leq cq] = 0$ (since $S \geq 0$). Hence, if $q < 0$ fulfils (6.7), then $u^+_{\max} = l_0 - l_{\min}$, which contradicts assumption (6.4). ■

[1] The production has no efficiency factor (see Section 5.1).

Proof (Proposition 6). Because we postulate the optimal solutions, we just have to show strong duality.

Let $q \geq 0$ be the Lagrange multiplier of the water constraint in (6.5), $Z^+, Z^- \in L^1_+$ those of the upper box-constraints, and set $l := l_0 - l_{\min}$. After straightforward calculation, the Lagrangian dual of problem (6.5) is

$$\min_{q \geq 0} \; lq + u^+_{\max}\mathbb{E}[Z^+] + u^-_{\max}\mathbb{E}[Z^-],$$
$$\text{s.t.} \begin{cases} Z^+ \geq S - q & a.s., \\ Z^- \geq q - \dfrac{1}{c}S & a.s., \\ Z^+ \geq 0, \; Z^- \geq 0 & a.s.. \end{cases} \tag{6.8}$$

Problem (6.8) is equivalent to

$$\min_{q \geq 0} \; lq + u^+_{\max}\mathbb{E}\big[(S - q)^+\big] + u^-_{\max}\mathbb{E}\big[(q - \frac{1}{c}S)^+\big]. \tag{6.9}$$

Problem (6.9) can be explicitly minimized: The derivative of the objective function in (6.9) with respect to q is

$$l + u^+_{\max}\big(\mathbb{P}[S \leq q] - 1\big) - u^-_{\max}\big(\mathbb{P}[\frac{1}{c}S \geq q] - 1\big).$$

For the explicit calculation of this derivative, see Rockafellar and Uryasev [72]; the derivative exists if S has a continuous distribution function. The critical points q^* have to fulfil

$$u^+_{\max}\mathbb{P}[S \geq q^*] - u^-_{\max}\mathbb{P}[S \leq cq^*] = l, \tag{6.10}$$

which is equation (6.7). By the foregoing Lemma, q^* exists and is non-negative. Therefore, the proposed optimal solution q^* is feasible in (6.9). Next we show that the proposed optimal solution (6.6) is feasible in (6.5). Clearly,

$$0 \leq \mathring{u}^+(\cdot) \leq u^+_{\max}, \quad \text{and} \quad 0 \leq \mathring{u}^-(\cdot) \leq u^-_{\max}.$$

It remains to check the constraint on the water level:

$$\begin{aligned} l_0 - \mathbb{E}[\mathring{u}^+(S)] + \mathbb{E}[\mathring{u}^-(S)] &= l_0 - u^+_{\max}\mathbb{P}[S \geq q^*] + u^-_{\max}\mathbb{P}[S \leq cq^*] \\ &\overset{(6.10)}{=} l_{\min}. \end{aligned}$$

Next, we will show weak duality as well as strong duality of the proposed solutions. In fact, the method of Lagrange-dualization implies always weak duality. But because we are in an infinite-dimensional setting, we will show weak duality explicitly.

Let $u^+(\cdot)$ and $u^-(\cdot)$ be a feasible solution of (6.5) and let Z^+, Z^- and q be a feasible solution of (6.8). The objective function of (6.5) is

$$\begin{aligned}
&\mathbb{E}\big[u^+(S)\,S\big] - \frac{1}{c}\mathbb{E}\big[u^-(S)\,S\big] \\
&\overset{(i)}{\le} \mathbb{E}\big[u^+(S)\,(Z^+ + q)\big] - \frac{1}{c}\mathbb{E}\big[u^-(S)\,c\,(-Z^- + q)\big] \\
&= \mathbb{E}\big[u^+(S)\,Z^+\big] + \mathbb{E}\big[u^-(S)\,Z^-\big] + q\Big(\mathbb{E}\big[u^+(S)\big] - \mathbb{E}\big[u^-(S)\big]\Big) \\
&\overset{(ii)}{\le} \mathbb{E}\big[u^+(S)\,Z^+\big] + \mathbb{E}[u^-(S)\,Z^-\big] + ql \\
&\overset{(iii)}{\le} u^+_{\max}\mathbb{E}[Z^+] + u^-_{\max}\mathbb{E}[Z^-] + ql,
\end{aligned}$$

where (i) follows by using the inequalities for Z^+ and Z^- in problem (6.8), (ii) follows by using the constraint on the water level in (6.5), and (iii) follows by using the box-constraints of feasible solutions in (6.5). The final expression is the objective function of (6.9).

Next, we show strong duality. Let q^* be a feasible solution of problem (6.9) that fulfils (6.10), and let $\overset{*}{u}{}^+$ and $\overset{*}{u}{}^-$ be the proposed optimal solutions (6.6). The optimal objective value of problem (6.9) is

$$\begin{aligned}
&u^+_{\max}\mathbb{E}[(S-q^*)^+] + u^-_{\min}\mathbb{E}[(q^* - \frac{1}{c}S)^+] + q^*l \\
&= u^+_{\max}\mathbb{E}[(S-q^*)\chi_{\{S\ge q^*\}}] + u^-_{\min}\mathbb{E}[(q^* - \frac{1}{c}S)\chi_{\{S\le cq^*\}}] + q^*l \\
&\overset{(6.6)}{=} \mathbb{E}[S\overset{*}{u}{}^+(S)] - u^+_{\max}q^*\mathbb{P}[S\ge q^*] - \frac{1}{c}\mathbb{E}[S\overset{*}{u}{}^-(S)] + u^-_{\min}q^*\mathbb{P}[S\le cq^*] + q^*l \\
&= \mathbb{E}[S\overset{*}{u}{}^+(S) - \frac{1}{c}S\overset{*}{u}{}^-(S)] - q^*\big(u^+_{\max}\mathbb{P}[S\ge q^*] - u^-_{\min}\mathbb{P}[S\le cq^*]\big) + q^*l \\
&\overset{(6.10)}{=} \mathbb{E}[S\overset{*}{u}{}^+(S) - \frac{1}{c}S\overset{*}{u}{}^-(S)],
\end{aligned}$$

which is the optimal objective value of problem (6.5). ∎

The algebraic dual problem (6.8) is a *stochastic recourse problem* (see e.g. [49]): The variable q with feasible set $\mathbb{R}_+$ is the so-called first-stage decision. The *recourse* at the second stage is measured in units of Z^+ and Z^-. If the outcome of S is such that $q \in [0, S)$, then Z^+ is strictly positive, and if $c \cdot q \in (S, \infty)$, then Z^- is strictly positive. At the second stage, the total expected cost of recourse is to be minimized: The cost of a (unit) deviation in Z^+ and Z^- is $u^+_{\max}$ and $u^-_{\max}$, respectively.

In the special case $c = 1$, the equivalent formulation (6.9) can be re-ordered to

$$u^+_{\max}\mathbb{E}[S] + \min_{q\ge 0}\Big(lq - u^+_{\max}\mathbb{E}[\min(S,q)] + u^-_{\max}\mathbb{E}[\max(q-S,0)]\Big). \tag{6.11}$$

The minimization can be (formally) identified with a *newsvendor problem* (see e.g. [11]): q is the number of newspapers to buy in advance at price l, S is the stochastic demand for newspapers, $u^+_{\max}$ is the selling price, and $u^-_{\max}$ is the cost to dispose an unsold newspaper. The objective is to minimize the expected total costs.

6.3 Multi-Period Dispatch Model

Let us extend the foregoing model to multiple periods. The electricity price is given by a non-negative stochastic process: $(S_t)_{t=0,\dots,T}$, with $S_t \in L^1_+$ for all t. The control is represented by functions of the electricity prices:

$$u^+_t(S_0, S_1, \dots, S_t) \quad \text{and} \quad u^-_t(S_0, S_1, \dots, S_t), \qquad t = 0, \dots, T. \tag{6.12}$$

The chosen form of the control is non-anticipative: The control-functions do not depend on future electricity prices.

For analytical tractability, constraints on the water level are assumed to hold only in expectation: The expected water level is constrained by a lower bound $l_{\min}$ at each time step. An expected water inflow $w_t \geq 0$, $t = 0, \dots, T$, increases the water level at each time step. The initial water level is assumed to be above the minimal level: $l_0 - l_{\min} > 0$. The stochastic control problem is

$$\begin{aligned} &\max_{(u^\pm_t)} \sum_{t=0}^{T} \mathbb{E}\Big[S_t u^+_t(S_0, \dots, S_t) - \frac{1}{c} S_t u^-_t(S_0, \dots, S_t)\Big], \\ &\text{s.t.} \begin{cases} l_0 + \displaystyle\sum_{t=0}^{s} \Big(\mathbb{E}\Big[u^-_t(S_0, \dots, S_t) - u^+_t(S_0, \dots, S_t)\Big] + w_t\Big) \geq l_{\min}, & \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad s = 0, \dots, T, & \\ 0 \leq u^+_t(\boldsymbol{s}) \leq u^+_{\max} \quad \text{for all } \boldsymbol{s} \in \mathbb{R}^{T+1}_+, \ t = 0, \dots, T, & \\ 0 \leq u^-_t(\boldsymbol{s}) \leq u^-_{\max} \quad \text{for all } \boldsymbol{s} \in \mathbb{R}^{T+1}_+, \ t = 0, \dots, T, & \\ u^\pm_t : \mathbb{R}^{T+1}_+ \to \mathbb{R}_+, \quad \text{measurable}, \qquad t = 0, \dots, T. & \end{cases} \end{aligned} \tag{6.13}$$

The expectation in the objective function exists because $S_t \in L^1_+$ and the $u^\pm_t$ are bounded measurable functions.

Proposition 7. *Let the distribution function of S_t be continuous for all t. Then an optimal solution of problem* (6.13) *is*

$$\begin{aligned} \overset{*}{u}{}^+_t(S_0, \dots, S_t) &= u^+_{\max} \chi_{\{S_t \geq \sum_{s=t}^{T} q^*_s\}}, \quad t = 0, \dots, T, \\ \overset{*}{u}{}^-_t(S_0, \dots, S_t) &= u^-_{\max} \chi_{\{S_t \leq c \sum_{s=t}^{T} q^*_s\}}, \quad t = 0, \dots, T, \end{aligned} \tag{6.14}$$

where $\boldsymbol{q}^* = (q_0^*, \ldots, q_T^*)^\top \in \mathbb{R}^{T+1}$ *is given by a solution of the following system of equations*

$$\begin{cases} \displaystyle\sum_{t=0}^{s} \Big(u_{\max}^- \mathbb{P}\big[S_t \leq c \sum_{s'=t}^{T} q_{s'}^*\big] - u_{\max}^+ \mathbb{P}\big[S_t \geq \sum_{s'=t}^{T} q_{s'}^*\big] + w_t \Big) + l_0 - l_{\min} = v_s, \\ \hfill s = 0, \ldots, T, \\ \boldsymbol{v}^\top \boldsymbol{q}^* = 0, \quad \boldsymbol{v} \geq 0,\ \boldsymbol{q}^* \geq 0, \quad \boldsymbol{v} = (v_0, \ldots, v_T)^\top. \end{cases} \tag{6.15}$$

The existence of $\boldsymbol{q}^*$ is assured in the proof. Its components are the Lagrange multipliers of the water constraints.

Proof. The proof is similar to that for Proposition 6. Because the optimal solutions are already proposed, their optimality is verified by introducing a Lagrangian dual problem: We show feasibility of the solutions, as well as weak and strong duality.

Let $\boldsymbol{q}^* = (q_0^*, \ldots, q_T^*)^\top \geq 0$ be the Lagrange multipliers of the water constraints in problem (6.13). By additionally introducing Lagrange multipliers for the constraints $u_t^\pm \leq u_{\max}^\pm$ for all t in problem (6.13), the (algebraic) Lagrangian dual problem becomes

$$\min_{\boldsymbol{q} \geq 0} \sum_{t=0}^{T} \Big(u_{\max}^+ \mathbb{E}\big[(S_t - \sum_{s=t}^{T} q_s)^+\big] + u_{\max}^- \mathbb{E}\big[(\sum_{s=t}^{T} q_s - \frac{1}{c} S_t)^+\big] + q_t \big(l_0 - l_{\min} + \sum_{s=0}^{t} w_s\big)\Big). \tag{6.16}$$

Because $(\cdot)^+$ is a convex function, the objective function in (6.16) is convex in $\boldsymbol{q}$. The objective function behaves as follows for large $\boldsymbol{q}$. By assumption, we have $l_0 - l_{\min} + \sum_{s=0}^{t} w_s \geq 0$, and

$$\mathbb{E}\Big[\big(\sum_{s=t}^{T} q_s - \frac{1}{c} S_t\big)^+\Big] \geq \mathbb{E}\Big[\sum_{s=t}^{T} q_s - \frac{1}{c} S_t\Big] = \sum_{s=t}^{T} q_s - \frac{1}{c} \mathbb{E}[S_t].$$

Hence, the objective function value gets arbitrarily large if a feasible $\boldsymbol{q}$ gets $\|\boldsymbol{q}\| \nearrow \infty$, where $\|\cdot\|$ is the euclidian norm in $\mathbb{R}^{T+1}$. Hence, a possible minimum of (6.16) is attained in the compact set $\{\boldsymbol{q} \mid \boldsymbol{q} \geq 0,\ \|\boldsymbol{q}\| \leq M\}$, where $M > 0$ is a sufficiently large number. The gradient of the objective function with respect to $\boldsymbol{q}$ is

$$\Big(l_0 - l_{\min} + \sum_{s=0}^{t} w_s + \sum_{t=0}^{T} \Big(u_{\max}^- \mathbb{P}\big[\sum_{s=t}^{T} q_s \geq \frac{1}{c} S_t\big] - u_{\max}^+ \mathbb{P}\big[S_t \geq \sum_{s=t}^{T} q_s\big]\Big)\Big)_{t=0,\ldots,T}.$$

For the explicit calculation of the derivatives, see Rockafellar and Uryasev [72] (the derivative exists if S has a continuous distribution function). Because the objective function (6.16) is convex in $\boldsymbol{q}$, it is continuous in $\boldsymbol{q}$. Because a continuous function on a compact set attains its minimum, the existence of a minimizing $\boldsymbol{q}^*$ is assured. The optimal solutions are Karush-Kuhn-Tucker-points (KKT-points). The corresponding KKT-conditions correspond to the system (6.15).

Next, we show weak duality. Define $U_t^+ := u_t^+(S_0, \dots, S_t)$ and $U_t^- := u_t^-(S_0, \dots, S_t)$ for all t. In particular, U_t^+ and U_t^- are $\mathcal{G}_t := \sigma(S_0, \dots, S_t)$-measurable random variables. We will show a little more than what is needed: We show weak duality for all $\mathcal{G}_t$-measurable random variables U_t^+ and U_t^- that fulfil the constraints implied by problem (6.13) for all t. The objective function of (6.13) is bounded from above (i.e. weak duality) as follows (see notes below):

$$
\begin{aligned}
&\sum_{t=0}^{T} \Big(\mathbb{E}[U_t^+ S_t] - \frac{1}{c}\mathbb{E}[U_t^- S_t] \Big) \\
&\overset{(i)}{=} \sum_{t=0}^{T} \Big(\mathbb{E}\big[U_t^+ (S_t - \sum_{s=t}^{T} q_s)\big] + \mathbb{E}\big[U_t^- (\sum_{s=t}^{T} q_s - \frac{1}{c} S_t)\big] + \sum_{s=t}^{T} q_s \big(\mathbb{E}[U_t^+] - \mathbb{E}[U_t^-]\big)\Big), \\
&\overset{(ii)}{\leq} \sum_{t=0}^{T} \Big(u_{\max}^+ \mathbb{E}\big[(S_t - \sum_{s=t}^{T} q_s)^+\big] + u_{\max}^- \mathbb{E}\big[(\sum_{s=t}^{T} q_s - \frac{1}{c} S_t)^+\big]\Big) \\
&\qquad\qquad + \sum_{s=0}^{T} q_s \sum_{t=0}^{s} \big(\mathbb{E}[U_t^+] - \mathbb{E}[U_t^-]\big), \\
&\overset{(iii)}{\leq} \sum_{t=0}^{T} \Big(u_{\max}^+ \mathbb{E}\big[(S_t - \sum_{s=t}^{T} q_s)^+\big] + u_{\max}^- \mathbb{E}\big[(\sum_{s=t}^{T} q_s - \frac{1}{c} S_t)^+\big]\Big) \\
&\qquad\qquad + \sum_{s=0}^{T} q_s \big(l_0 - l_{\min} + \sum_{t'=0}^{s} w_{t'}\big),
\end{aligned}
$$

where (i) the term $0 = \sum_{s=t}^{T} q_s - \sum_{s=t}^{T} q_s$ was added, (ii) the upper bound of feasible control in (6.13) was used, the accompanying factors are made positive $(\cdot) \to (\cdot)^+$, and the sums in the second term were interchanged, and (iii) the constraints of water in (6.13) were applied. The final expression is the objective function of (6.16). Thus, the objective function of (6.16) is for every $\boldsymbol{q} \geq 0$ an upper bound for problem (6.13), and this holds for all $\mathcal{G}$-measurable, feasible random variables U_t^+ and U_t^-.

Next, we show that the proposed solution (6.14) is feasible in problem (6.13). Clearly,

$$
0 \leq \mathring{u}_t^+(\cdot) \leq u_{\max}^+, \quad \text{and} \quad 0 \leq \mathring{u}_t^-(\cdot) \leq u_{\max}^- \quad \forall t.
$$

It remains to check the constraints on the water level. Let $s = 0, \dots, T$:

$$\begin{aligned}
&\sum_{t=0}^{s} \Big(\mathbb{E}[\overset{*}{u}{}_t^-(S_t)] - \mathbb{E}[\overset{*}{u}{}_t^+(S_t)] + w_t \Big) + l_0 - l_{\min} \\
&= \sum_{t=0}^{s} \Big(u_{\max}^- \mathbb{P}[S_t \le c \sum_{s'=t}^{T} q_{s'}^*] - u_{\max}^+ \mathbb{P}[S_t \ge \sum_{s'=t}^{T} q_{s'}^*] + w_t \Big) + l_0 - l_{\min} \\
&\ge 0,
\end{aligned}$$

where the last inequality follows from the system (6.15).

Next, we show strong duality. Let $\boldsymbol{q}^*$ be a solution of (6.16) that fulfils the system (6.15), and let $\overset{*}{u}{}_t^+$ and $\overset{*}{u}{}_t^-$ be the solution (6.14). The optimal objective value of (6.16) is

$$\begin{aligned}
&\sum_{t=0}^{T} \Big(u_{\max}^+ \mathbb{E}\big[(S_t - \sum_{s=t}^{T} q_s^*)^+\big] + u_{\max}^- \mathbb{E}\big[(\sum_{s=t}^{T} q_s^* - \frac{1}{c} S_t)^+\big] + q_t\big(l_0 - l_{\min} + \sum_{s=0}^{t} w_s\big) \Big) \\
&\overset{(i)}{=} \sum_{t=0}^{T} \Bigg(u_{\max}^+ \mathbb{E}\big[(S_t - \sum_{s=t}^{T} q_s^*) \chi_{\{S_t \ge \sum_{s=t}^{T} q_s^*\}}\big] \\
&\qquad + u_{\max}^- \mathbb{E}\big[(\sum_{s=t}^{T} q_s^* - \frac{1}{c} S_t) \chi_{\{S_t \le c \sum_{s=t}^{T} q_s^*\}}\big] + q_t\big(l_0 - l_{\min} + \sum_{s=0}^{t} w_s\big) \Bigg) \\
&\overset{(ii)}{=} \sum_{t=0}^{T} \Bigg(\mathbb{E}\big[S_t \overset{*}{u}{}_t^+(S_t)\big] - u_{\max}^+ \Big(\sum_{s=t}^{T} q_s^* \Big) \mathbb{P}\big[S_t \ge \sum_{s'=t}^{T} q_{s'}^*\big] \\
&\qquad - \frac{1}{c} \mathbb{E}\big[S_t \overset{*}{u}{}_t^-(S_t)\big] + u_{\max}^- \Big(\sum_{s=t}^{T} q_s^* \Big) \mathbb{P}\big[S_t \le c \sum_{s'=t}^{T} q_{s'}^*\big] + q_t\big(l_0 - l_{\min} + \sum_{s=0}^{t} w_s\big) \Bigg) \\
&\overset{(iii)}{=} \sum_{t=0}^{T} \mathbb{E}\big[S_t \overset{*}{u}{}_t^+(S_t) - \frac{1}{c} S_t \overset{*}{u}{}_t^-(S_t)\big] \\
&\qquad + \sum_{s=0}^{T} q_s^* \Bigg(\sum_{t=0}^{s} \Big(u_{\max}^- \mathbb{P}\big[S_t \le c \sum_{s'=t}^{T} q_{s'}^*\big] - u_{\max}^+ \mathbb{P}\big[S_t \ge \sum_{s'=t}^{T} q_{s'}^*\big] + w_t \Big) + l_0 - l_{\min} \Bigg) \\
&\qquad \overset{(iv)}{=} \sum_{t=0}^{T} \mathbb{E}\big[S_t \overset{*}{u}{}_t^+(S_t) - \frac{1}{c} S_t \overset{*}{u}{}_t^-(S_t)\big] + 0,
\end{aligned}$$

where (i) the $(\cdot)^+$-function was converted into an indicator function, (ii) the proposed solution (6.14) was substituted, (iii) sums were interchanged, and (iv) the system (6.15) was used. The final expression is the optimal objective value of problem (6.13). ■

By assumption, the feasible control is non-anticipative. In fact, it can be verified that the non-anticipativity is not used in the proof. Hence, the bang-bang solution (6.14) is optimal even in the case where anticipative controls are allowed. Because $q_t^* \geq 0$ for all t, the optimal threshold for the maximal production and the maximal pumping decreases over time. If q_t^* is considered to be the marginal price of water in interval $[t, t+1]$, then the solution (6.14) can be interpreted such that the operator shall produce whenever the marginal price of remaining water is less than the spot price.

The numerical evaluation of the q_t^* for non-trivial spot-price distributions can be difficult. In addition, the model assumes that the water level is constrained in expectation; obtaining the mere form of exact solutions for models with almost-sure (path-wise) constraints is unknown. Accordingly, the more realistic model of the next chapter resorts to a numerical solvable optimization problem in terms of a discretized multi-stage stochastic linear programming model.

Chapter 7

The General Dispatch Model

In the previous chapter, the models were analytically solvable, but simple. Now, we present a more realistic model of a hydro-electric pumped storage plant.

First, we consider the different time scales of the state variables and of the exogenous quantities. Then, the stochastic control model is introduced. The straightforward formulation of the model needs two different time scales. By approximating the electricity price with a step function, we can remove the finer time scale. The resulting model is a multi-stage stochastic linear program on a single time scale. In fact, a stochastic control model that has exogenous observables can always be viewed as a stochastic linear programming model (see Section 5.2). Nevertheless, the use of control-language helps to clarify the concepts of our engineering application.

The forthcoming stochastic control model does not assume any disretization of the random variables. Nevertheless, to obtain a numerically solvable problem, the resulting stochastic linear program is formulated on a scenario tree, that is, the filtration has to be finite. The scenario tree will be obtained by a suitable discretization of the random variables (Chapter 8).

Let a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ be given.

7.1 Stochastic Control Formulation

In the following, we will argue that the trading-decisions on the electricity market are on a finer time scale than variations in the state of the hydro-electricity plant. Hence, the formulation has two time scales; the fine *(small)* time scale will be modeled as a subdivision of the coarse *(large)* time scale.

7.1.1 The Small and the Large Time Scale

The dispatched electricity is traded on the spot market. On the EEX spot market, the spot price changes every hour considerably. To benefit from these changes, the amount to dispatch should be decided on that hourly time scale, too. Apart from the electricity price, another exogenous quantity is the water inflow. In reality, the inflow is continuously varying over time. For the moment, we assume that sufficient data is available (which is hardly ever the case) to calibrate an inflow model on an hourly time scale. In a typical modeling, the technical state of the plant is determined by the water level, whereas the financial state is characterized by the realized cumulative profit-and-loss and, possibly, by a current outlook on future revenues. We could choose for the state variables an hourly time scale, too. The most general control would be a function of all past-and-present exogenous and state variables. Unfortunately, current methods are not able to numerically solve the resulting control problem over a longer period of time.

At a typical electricity plant, relative changes of water level induced by the dispatch and water inflow during one hour can be neglected. Likewise, relative changes in the financial state of the electricity plant during one hour are small. Accordingly, the applied model uses a small and a large time scale. Because the hourly electricity price-changes have to be taken into account, the small scale has hourly intervals. The small time scale is measured in fractions of the large time scale: The points in time of the large scale are $t = 0, 1, \ldots, T$, and those of the small scale $\tau = \frac{h}{H}$, $h = 1, 2 \ldots$ where H is the number of hours between consecutive large time steps. All intervals are assumed to be equally spaced (see Sec. 11 for a possible extension).

The choice of the large time scale depends on the particular plant and the decision maker. For example, the full reservoir of the typical plant in the case study is emptied in about a month by producing at full capacity without inflow (cf. Table 9.3, p. 90). Hence, the decision maker should check the state variables monthly. Several time scales in related models were tested by Döge [23, p. 161]: Half-a-month was considered to be a sufficient granulation. A numerically solvable multi-stage stochastic programming model is restricted to a few time periods (commonly less than ten periods), and hence the time steps are modeled as large as possible. Accordingly, we choose the large time scale in the magnitude of one month which suits a planning horizon of several months.

7.1.2 The Exogenous Variables and the Control

The control-functions for the rate (MW) of production and pumping are denoted by u_t^+ and u_t^-, respectively. The control in a specific hour starting at time $t + \frac{h}{H}$ is chosen to be a function of exogenous variables:

$$u_t^{\pm}(S_{t+\frac{h}{H}}, \boldsymbol{E}_0, \ldots, \boldsymbol{E}_t), \qquad h = 1, \ldots, H, \quad t = 0, \ldots, T-1, \tag{7.1}$$

where the current (electricity) spot price is denoted by $S_{t+\frac{h}{H}}$ (Euro/MWh) at hour h in time interval $[t, t+1]$, and $(\boldsymbol{E}_t)_{t=0,\ldots,T}$ is a sequence of random vectors of those exogenous variables that vary on the large time scale. The first argument

electr. price	dispatch rate
s_1	$u_t^{\pm}(s_1, \boldsymbol{E}_0 ... \boldsymbol{E}_t)$
$\vdots$	$\vdots$
s_N	$u_t^{\pm}(s_N, \boldsymbol{E}_0 ... \boldsymbol{E}_t)$

valid in $[t, t+1)$

$\curvearrowright$

electr. price	dispatch rate
s_1	$u_{t+1}^{\pm}(s_1, \boldsymbol{E}_0 ... \boldsymbol{E}_{t+1})$
$\vdots$	$\vdots$
s_N	$u_{t+1}^{\pm}(s_N, \boldsymbol{E}_0 ... \boldsymbol{E}_{t+1})$

valid in $[t+1, t+2)$

Figure 7.1: The change of the decision table from time t to time $t+1$

of the control-function captures the high-frequency hourly variation of the spot price, whereas the remaining arguments capture the movements of slowly varying exogenous quantities. The exact definition of the variables $\boldsymbol{E}_t$ is postponed to Section 7.2.3. The probability distribution of the exogenous quantities, that is the spot prices and $(\boldsymbol{E}_t)_{t=0,\dots,T}$, will be discussed in Chapter 8.

Note that the control-function (7.1) depends only on exogenous quantities. Nevertheless, a realization of the sequence $\boldsymbol{E}_0, \dots, \boldsymbol{E}_t$ determines a historical path and therefore a specific state of the plant at time t. Hence, the control depends implicitly on the state of the plant. Freeing the control-function from the (endogenous) state variables has the advantage that the model allows a straightforward linear formulation (cf. Sec. 7.3).

If the control-function takes into account only finitely many levels of the spot prices, $s_1, \dots, s_N$, then the function reduces to a (finite) decision table (Fig. 7.1). The final numerically solvable model will use such a decision table; for the moment, we proceed in general terms.

The rate of producing and pumping is assumed to be monotonically increasing and decreasing in the spot price, respectively. In addition, the rate of dispatch is physically limited: The upper bounds of producing and pumping are denoted by $u^+_{\max}$ and $u^-_{\max}$, respectively. Accordingly, the set of admissible control-functions is

$$\mathcal{U} := \left\{ (u_t^+, u_t^-)_{t=0,\dots,T-1} \;\middle|\; \begin{array}{l} u_t^{\pm} : \mathbb{R} \times \mathbb{R}^{n_0} \times \cdots \times \mathbb{R}^{n_T} \to \mathbb{R}, \\ \qquad \text{measurable}, \\ 0 \le u_t^+ \le u^+_{\max}, \\ 0 \le u_t^- \le u^-_{\max}, \\ u_t^+(\cdot, \boldsymbol{v}) \text{ monotonically increasing}, \\ u_t^-(\cdot, \boldsymbol{v}) \text{ monotonically decreasing} \\ \qquad \text{for all } \boldsymbol{v} \in \mathbb{R}^{n_0} \times \cdots \times \mathbb{R}^{n_T}, \end{array} \right\} \tag{7.2}$$

where n_t is the dimension of the vector $\boldsymbol{E}_t$.

7.1.3 The State Variables

The plant is determined by a technical as well as a financial state. The technical state at time t is the *water level* L_t (MWh). The financial state is chosen to be the *cumulated, discounted profit-and-loss* P_t (Euro), and the *value of the plant* V_t (Euro). As we have argued in Section 7.1.1, the state variables need not to be checked on the hourly time scale of the short-term dispatch decision; the model assumes that it suffices that their bounds are taken into account only on the large time scale $t = 0, 1, \dots, T$.

The feasible water levels are given as follows. The initial level is denoted by l_0. Subsequently, the level has to stay between a lower and upper bound: $l_{\min} \leq L_t \leq l_{\max}$ *a.s.* for all t. At final time T, the level is restricted to be above an additional lower bound: $l_T \leq L_T$ *a.s.*.

The cumulated discounted profit-and-loss at time t consists of the flown discounted cash flows from time 0 to time t. The value of the plant is defined as follows.

The Value of the Plant (State Equation of Value)

The objective of the optimization problem is chosen to maximize the expected value V_T of the plant at final time T:

$$\sup\ \mathbb{E}[V_T],$$
subject to appropriate constraints specified later.

The existence of the expectation will be shown when the complete model has been specified (p. 61). The considered value is related to operation; operation-independent costs, like depreciation and water charges are not taken into account. Different stakeholders can have different definitions for an operation-related value. We assume that the value is retrospective as well as prospective: The value V_t at time $t = 0, \dots, T$ is assumed to be a sum of the realized cumulated (discounted) profit-and-loss and of a value attributed to future profit-and-loss. The uncertain future profit-and-loss can be considered in two ways:

Dependent on operation: The value of future production is calculated in some way from the path-wise future profit-and-losses in each scenario. In this case, the value depends on the operation of the plant (on how to dispatch). This implies that the stakeholder, who values the plant, knows the future operation, or, at least a probability distribution of it.

Independent on operation: The value is a function of the remaining usable water in the reservoir. This assumes that the stakeholder is not willing or is just not able to judge future operation in detail.

We assume the latter; the prospective part of the value is independent of the operation and is considered to be an agreed definition even to those stakeholders who are not willing to become experts in power optimization and therefore cannot judge future operation. Hence, the value at time t is the sum of occurred cash flows and a value of the expected water that is usable for production in the future. By neglecting discounting[1], the value is

$$V_t := P_t + v_t\Big(\mathbb{E}\big[\sum_{k=t+1}^{T} I_k \big| \mathcal{F}_t\big] + L_t - l_{\min}\Big), \qquad t = 0, \dots, T. \tag{7.3}$$

where v_t (Euro/MWh) denotes a deterministic value of the remaining water at time t, $L_t - l_{\min}$ is the usable water in the reservoir, and the stochastic process of water inflow between consecutive time steps is denoted by

$$(I_t)_{t=1,\dots,T} \quad \text{in (MWh)}, \quad I_t \in L^1_+(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}). \tag{7.4}$$

The expectation in (7.3) is conditioned on the σ-algebra that is generated by the exogenous information: $\mathcal{F}_t := \sigma(\boldsymbol{E}_0, \dots, \boldsymbol{E}_t)$. The probability distribution of inflow will be discussed in Section 8.2 (p. 76). The conditional expectation of I_k exists because I_k is integrable by definition.

The future water in (7.3) is valued by a deterministic factor v_t for simplicity, which is chosen by the stakeholder. For example, the choice is based on a historical selling price. Particularly, the remaining usable water in the reservoir at final time T is valuated. If the final water would not contribute to the value of the plant, or the value of the water is sufficiently small, then an optimization tries to spend all usable water till time T. A remedy is to make the lower bound on the reservoir level at final time T sufficiently high (see [79] for such use of a final water restriction instead of valuing the final water level).

The Constraint on Risk of the Value of the Plant

The decision maker is assumed to be risk-averse: The maximization of the expected final value is subject to a constraint on risk. Common risk measures in energy applications are CVaR [79], target shortfall [33], or multi-period VaR [55].

In our model, the constraint on risk is imposed by a stakeholder who cares about the distributions of the intertemporal values. Hence, the calculation of the risk takes the whole value process $(V_t)_{t=0,\dots,T}$ into account. It is assumed that the risk is bounded by a lower bound on a multi-period risk-adjusted value:

$$\pi[V_0, \dots, V_T] \geq \rho_{\min}.$$

[1]Discounting is neglected for ease of presentation. Discounting factors are introduced in the final stochastic programming problem.

Table 7.1: The state variables and the exogenous quantities

state variables ($t = 0, \ldots, T$)		exogenous variables ($t = 0, \ldots, T-1$)	
L_t :	water level	$S_{t+\frac{h}{H}}$:	spot price at hour h in interval $[t, t+1]$
P_t :	cumulated profit-and-loss		
V_t :	value of plant	$\boldsymbol{E}_t$:	vector of other variables

The risk-adjusted value assigns a number to the distribution of the value process. Due to simplicity, we have introduced multi-period risk-adjusted values only for finitely discrete processes (Ch. 3 and 4). Nevertheless, for the formulation of the stochastic control problem, it suffices to assume that π is a mapping from the space of stochastic processes into the reals, and monotone in its arguments: If $V_t \leq W_t$ *a.s.* for all t, then $\pi[V_0, \ldots, V_T] \leq \pi[W_0, \ldots, W_T]$, which agrees with property (iii) of Lemma 2 (p. 17) for recursive risk-adjusted values. The detailed definition of π in terms of recursive risk-adjusted values will be given in the forthcoming formulation of the multi-stage stochastic program.

Summarizing, the feasible set of state variables is

$$\mathcal{X} := \left\{ (L_t, P_t, V_t)_{t=0,\ldots,T} \;\middle|\; \begin{array}{l} (L_t, P_t, V_t) : \Omega \to \mathbb{R}^3, \text{ measurable}, \\ l_{\min} \leq L_t \leq l_{\max} \quad a.s., \quad t = 0, \ldots, T, \\ l_T \leq L_T \quad a.s., \\ P_0 = 0, \; L_0 = l_0, \\ \pi[V_0, \ldots, V_T] \geq \rho_{\min}. \end{array} \right\} \tag{7.5}$$

The state variables and exogenous variables are recapitulated in Table 7.1.

The only constraint that inter-connects the state variables over the time steps in a non-separable way is the lower bound of the risk-adjusted value. This constraint prohibits an additive separability of the optimization problem. Hence, we cannot use backward-recursive *dynamic programming* solution algorithms [8].

The Change of Profit-and-Loss and of Water Level

The change in the water level in the hth hour of time interval $[t, t+1]$ is

$$L_{t+\frac{h}{H}} - L_{t+\frac{h-1}{H}} = u_t^-(S_{t+\frac{h}{H}}, \boldsymbol{E}_0 \ldots \boldsymbol{E}_t) - u_t^+(S_{t+\frac{h}{H}}, \boldsymbol{E}_0 \ldots \boldsymbol{E}_t) + W_{t+\frac{h}{H}},$$

where $W_{t+\frac{h}{H}}$ is the hourly water inflow; because the model of the plant takes the water level only at time $t = 0, \ldots, T$ into account, the relevant quantity will be the cumulated inflow $I_t = \sum_{h=1}^{H} W_{t+\frac{h}{H}}$. For simplicity, a spill over of the reservoir is not yet modeled and will be considered only in the final, numerically tractable model.

The profit-and-loss in hour h is

$$P_{t+\frac{h}{H}} - P_{t+\frac{h-1}{H}} = S_{t+\frac{h}{H}} u_t^+(S_{t+\frac{h}{H}}, \boldsymbol{E}_0 ... \boldsymbol{E}_t) - \frac{1}{c} S_{t+\frac{h}{H}} u_t^-(S_{t+\frac{h}{H}}, \boldsymbol{E}_0 ... \boldsymbol{E}_t),$$

where $0 < c < 1$ is the efficiency of pumping.

7.1.4 The Stochastic Control Problem

Combining the definitions of the previous two sections, we get the following optimization problem of stochastic control:

$$\sup_{(u_t^\pm)_{t=0,\dots,T-1}} \quad \mathbb{E}[V_T], \qquad \text{subject to}$$

$$P_{t+1} - P_t = \sum_{h=1}^{H} S_{t+\frac{h}{H}} \Big(u_t^+(S_{t+\frac{h}{H}}, \boldsymbol{E}_0 ... \boldsymbol{E}_t) - \frac{1}{c} u_t^-(S_{t+\frac{h}{H}}, \boldsymbol{E}_0 ... \boldsymbol{E}_t) \Big) \quad a.s.,\ t = 0, \dots, T-1, \tag{7.6}$$

(SC)

$$L_{t+1} - L_t = \sum_{h=1}^{H} \Big(u_t^-(S_{t+\frac{h}{H}}, \boldsymbol{E}_0 ... \boldsymbol{E}_t) - u_t^+(S_{t+\frac{h}{H}}, \boldsymbol{E}_0 ... \boldsymbol{E}_t) \Big) + I_{t+1} \quad a.s.,\ t = 0, \dots, T-1, \tag{7.7}$$

$$V_t = P_t + v_t\Big(\mathbb{E}\big[\sum_{k=t+1}^{T} \mathbb{E}[I_k | \mathcal{F}_t] + L_t - l_T \Big) \qquad a.s.,\ t = 0, \dots, T, \tag{7.8}$$

$$(L_t, P_t, V_t)_{t=0,\dots,T} \in \mathcal{X}\ a.s., \quad (u_t^+, u_t^-)_{t=0,\dots,T-1} \in \mathcal{U}.$$

The control formulation has two simplifications: First, the discounting is not written explicitly. Second, a spill-over of the reservoir is not yet allowed; a (inconvenient) dummy control would have to be introduced. Both features can be elegantly reintroduced in the final formulation as a stochastic programming problem. The profit-and-loss does not include marginal costs; they could be included in an obvious way. We assume that the expectation of spot price exists: $S_{t+\frac{h}{H}} \in L_+^1 := L_+^1(\Omega, \mathcal{F}, \mathbb{R})$ for all t, h. By definition (7.4), $I_t \in L_+^1$, and by the boundedness of admissible control-functions, $L_t \in L^1$ for all t. In addition, $S_{t+\frac{h}{H}} \in L^1$ implies $P_t \in L^1$ for all t, hence $V_t \in L^1$ for all t. Thus, the expectation in the objective function exists.

The notation of the control-functions u_t^+ and u_t^- suggests that they are the plus and minus part of a function with no sign restriction.

Proposition 8 (Unger [79]). *Consider the stochastic control problem (SC). If the efficiency of pumping is $c < 1$, then simultaneous producing and pumping is not optimal at each time.*

The proof is given in Appendix A.3 (p. 115).

7.2 Transformation of the State Equations

The state equations of profit-and-loss and of water involve a short and and a large time-scale. In this section, we change the summation in these state equations by using as summation index the price levels instead of the points of the short time-scale. The change allows a reformulation of the model solely on the large time-scale.

The transformation is motivated by the following general observation. To calculate an income over a time period we can consider the path over time of the underlying price process. Equivalently, we can also consider the fraction of time the process is below (or above) different price levels. This second view may be advantageous for some multi-period decision models in general.

7.2.1 Discrete Price Levels and Occupation Times

Consider the right-hand-side of the state equations (7.6) and (7.7). They can formally be written as

$$\sum_{h=1}^{H} f_{t,\omega}\big(S_{t+\frac{h}{H}}(\omega)\big), \tag{7.9}$$

where $f_{t,\omega} : \mathbb{R} \to \mathbb{R}$ is a function of discrete time $t = 0, \dots, T$ and of state $\omega \in \Omega$. To get a numerically solvable, finite multi-stage stochastic programming model, the electricity price is approximated by a step function[1]:

$$S_{t+\frac{h}{H}} \to \widetilde{S}_{t+\frac{h}{H}} := \sum_{i=1}^{N} \bar{s}_i \chi_{\{s_{i-1} < S_{t+\frac{h}{H}} \le s_i\}}, \tag{7.10}$$

where the range of electricity price is discretized by $N+1$ levels

$$0 \le s_0 < s_1 < \cdots < s_N, \tag{7.11}$$

and intermediate prices $\bar{s}_i \in (s_{i-1}, s_i)$ as steps are introduced. Because ω is not altered in the following formula, it will be notationally suppressed. If we use the approximation (7.10) in (7.9), we can interchange the sums:

$$\begin{aligned}
\sum_{h=1}^{H} f_t(\widetilde{S}_{t+\frac{h}{H}}) &= \sum_{h=1}^{H}\sum_{i=1}^{N} f_t(\bar{s}_i)\chi_{\{s_{i-1} < S_{t+\frac{h}{H}} \le s_i\}}, \\
&= \sum_{i=1}^{N} f_t(\bar{s}_i) \sum_{h=1}^{H} \big(\chi_{\{S_{t+\frac{h}{H}} \le s_i\}} - \chi_{\{S_{t+\frac{h}{H}} \le s_{i-1}\}}\big), \\
&= H \sum_{i=1}^{N} f_t(\bar{s}_i)\big(F_{t+1}(s_i) - F_{t+1}(s_{i-1})\big),
\end{aligned} \tag{7.12}$$

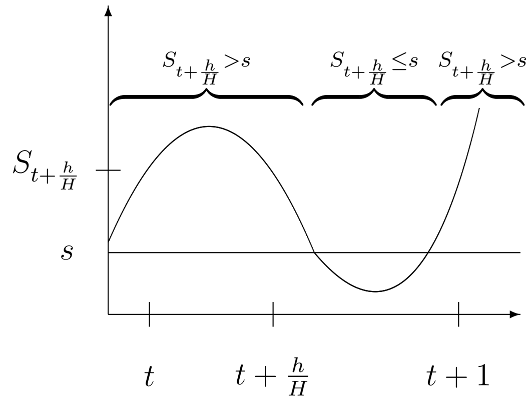
[1]Every measurable positive function is the pointwise limit of step functions [6, Prop. 11.6].

where $F_{t+1}\colon \mathbb{R}_+ \to [0,1]$ is monotonically increasing (ω fixed), and defined as follows.

Definition 12 (Occupation time). The *occupation time* $F_{t+1}(s)\colon \Omega \to [0,1]$ of the electricity price at level s in unit time interval $[t, t+1]$ is the random variable

$$F_{t+1}(s) := \frac{1}{H} \sum_{h=1}^{H} \chi_{\{S_{t+\frac{h}{H}} \leq s\}}, \tag{7.13}$$

where $S_{t+\frac{h}{H}}$ is the electricity price during hour h in the interval.



Note. In a continuous time setting, sums would be replaced by integrals, and the interchange of sums in (7.12) corresponds to the transformation of a time-integral into a Stieltjes integral with measure-inducing function F_{t+1}.

Consider the last line of (7.12), which corresponds to the right-hand-side of the state equation of profit-and-loss or of water. The control-function has to be evaluated only at finitely many electricity prices: $\bar{s}_1, \ldots, \bar{s}_N$. Thus, the control-function reduces to a finite decision table (cf. Fig. 7.1, p. 57). The finiteness allows the following convenient parametrization, which is due to Unger [24].

7.2.2 Parametrization of the Control-Functions

Let $t = 0, \ldots, T-1$. The steps of the control-functions can be parameterized with their normalized differences:

$$x_{it}^{+}(\boldsymbol{E}_0 ... \boldsymbol{E}_t) := \frac{1}{u_{\max}^{+}} \Big(u_t^{+}(\bar{s}_i, \boldsymbol{E}_0 ... \boldsymbol{E}_t) - u_t^{+}(\bar{s}_{i-1}, \boldsymbol{E}_0 ... \boldsymbol{E}_t) \Big), \qquad i = 1, \ldots, N,$$
$$x_{it}^{-}(\boldsymbol{E}_0 ... \boldsymbol{E}_t) := \frac{1}{u_{\max}^{-}} \Big(u_t^{-}(\bar{s}_{i-1}, \boldsymbol{E}_0 ... \boldsymbol{E}_t) - u_t^{-}(\bar{s}_i, \boldsymbol{E}_0 ... \boldsymbol{E}_t) \Big), \qquad i = 1, \ldots, N,$$

where an additional lowest price step $\bar{s}_0 \in [0, s_0)$ had to be defined. Owing to the measurability of $u_t^{\pm}$, the $x_{it}^{\pm}$ are measurable functions, and because $u_t^{+}(\cdot, \boldsymbol{v})$ and

$u_t^-(\cdot, \boldsymbol{v})$ are for all $\boldsymbol{v}$ monotonically increasing and decreasing, respectively, we have $x_{it}^{\pm}(\boldsymbol{E}_0 \ldots \boldsymbol{E}_t) \geq 0$ for all i and t (see properties (7.2) of $u_t^{\pm}$, p. 57). If we assume no production at the lowest price $\bar{s}_0$ and no pumping at the highest price $\bar{s}_N$, then the parametrization is

$$\begin{aligned} u_t^+(\bar{s}_i, \boldsymbol{E}_0 \ldots \boldsymbol{E}_t) &= u_{\max}^+ \sum_{j=1}^{i} x_{jt}^+(\boldsymbol{E}_0 \ldots \boldsymbol{E}_t), \\ u_t^-(\bar{s}_i, \boldsymbol{E}_0 \ldots \boldsymbol{E}_t) &= u_{\max}^- \sum_{j=i+1}^{N} x_{jt}^-(\boldsymbol{E}_0 \ldots \boldsymbol{E}_t). \end{aligned} \tag{7.14}$$

Furthermore, if we assume full-capacity production at $\bar{s}_N$ and full-capacity pumping at $\bar{s}_0$, then

$$\sum_{i=1}^{N} x_{it}^{\pm}(\boldsymbol{E}_0 \ldots \boldsymbol{E}_t) = 1.$$

This is a rather strong assumption, but in accordance with similar models [23, 79] (dropping the assumption gives the relaxed constraint $\sum_{i=1}^{N} x_{it}^{\pm}(\boldsymbol{E}_0 \ldots \boldsymbol{E}_t) \leq 1$). Under the previous assumptions, the set of admissible controls becomes

$$\left\{ \left(x_{it}^{\pm}(\boldsymbol{E}_0 \ldots \boldsymbol{E}_t) \right)_{\substack{i=1,\ldots,N, \\ t=0,\ldots,T-1}} \;\middle|\; x_{it}^{\pm}(\boldsymbol{E}_0 \ldots \boldsymbol{E}_t) \geq 0, \; \sum_{i=1}^{N} x_{it}^{\pm}(\boldsymbol{E}_0 \ldots \boldsymbol{E}_t) = 1 \right\}. \tag{7.15}$$

For fixed t, the values of $(x_{it}^{\pm}(\boldsymbol{E}_0 \ldots \boldsymbol{E}_t))_{i=1,\ldots,N}$ are a decision table: If the spot price is in the interval $(\bar{s}_{i-1}, \bar{s}_i]$, then the additional fraction of production is x_{it}^+ with respect to prices less or equal than $\bar{s}_{i-1}$, and the additional fraction of pumping in interval $[\bar{s}_{i-1}, \bar{s}_i)$ is x_{it}^- with respect to prices greater or equal than $\bar{s}_i$.

7.2.3 Exogenous Variables in the Transformed Model

We have introduced occupation times into the state equations and we have parameterized the control. In particular, the approximation of the state equations for profit-and-loss and for water (7.6)-(7.7) according to the scheme (7.12) together with the parametrization of the control according to (7.14)-(7.15) gives

$$\begin{aligned} P_{t+1} - P_t = u_{\max}^+ H \sum_{i=1}^{N} \Big(\sum_{j=i+1}^{N} x_{jt}^+(\boldsymbol{E}_0 \ldots \boldsymbol{E}_t) \Big) \bar{s}_i \big(F_{t+1}(s_i) - F_{t+1}(s_{i-1}) \big) \\ - u_{\max}^- \frac{1}{C} H \sum_{i=1}^{N} \Big(\sum_{j=1}^{i} x_{jt}^-(\boldsymbol{E}_0 \ldots \boldsymbol{E}_t) \Big) \bar{s}_i \big(F_{t+1}(s_i) - F_{t+1}(s_{i-1}) \big), \\ t = 0, \ldots, T-1, \end{aligned} \tag{7.16}$$

$$L_{t+1} - L_t = l_0 + u^-_{\max} H \sum_{i=1}^{N} \Big(\sum_{j=i+1}^{i} x^-_{jt}(\boldsymbol{E}_0 \ldots \boldsymbol{E}_t) \Big) \big(F_{t+1}(s_i) - F_{t+1}(s_{i-1}) \big)$$
$$- u^+_{\max} H \sum_{i=1}^{N} \Big(\sum_{j=i}^{N} x^+_{jt}(\boldsymbol{E}_0 \ldots \boldsymbol{E}_t) \Big) \big(F_{t+1}(s_i) - F_{t+1}(s_{i-1}) \big)$$
$$+ I_{t+1}, \qquad t = 0, \ldots, T-1. \tag{7.17}$$

The stochastic control problem (SC) (p. 61) with the foregoing approximated state equations (7.16)-(7.17) and with the admissible set of controls (7.15) is formulated entirely on the large time scale; the small time scale is hidden in the occupation times F_t (Def. 12, p. 63). The question (Sec. 7.1.2) about the precise definition of the exogenous random variables $\boldsymbol{E}_t$ can now be satisfactorily answered: To have full *flexibility* in the sense that the control shall have the best ability to react to exogenous events, the $\boldsymbol{E}_t$s consist of all exogenous variables appearing in the transformed model, that is, the water inflow and the occupation times. Accordingly,

$$\boldsymbol{E}_t := \big(I_t, F_{0t}, \ldots, F_{Nt} \big)^\top, \qquad t = 0, \ldots, T, \tag{7.18}$$

where the occupation times are abbreviated by

$$F_{it} := F_t(s_i), \qquad i = 0, \ldots, N, \; t = 0, \ldots, T.$$

In fact, no component of the vector $\boldsymbol{E}_0$ appears explicitly in the stochastic control model. Hence, $\boldsymbol{E}_0$ could be neglected, but is kept for notational coherency.

7.3 The Stochastic Linear Program on a Scenario Tree

Most of the modeling work has been accomplished in terms of the foregoing stochastic control problem. To make this problem numerically tractable, it has to be disretized and linearized. Moreover, we have to specify the constraint on risk in more detail. The resulting problem will be a *multi-stage stochastic program on a scenario tree.*

7.3.1 Linear Formulation on a Scenario Tree

So far, the model of the electricity plant was formulated as a stochastic control problem (problem (SC), p. 61, and its subsequent modifications). We shortly review the model and highlight the advantages of using the control-language. The multiperiod model has time steps $t = 0, \ldots, T$. The problem's objective is to maximize the expected final value of the plant. In the formulation as a control problem, the decision variables (control-functions for producing and pumping: $u_t^{\pm}$) are separated from the state variables. The state variables are the water level L_t, the cumulated profit-and-loss P_t, and the value of the plant V_t. Each state variable is equipped with a state equation, which governs the dynamic behavior of the variable. Because the plant has technical restrictions, the control-functions have to be in an admissible set. Similarly, the state variables have to be in a feasible set, too. Especially, in the case of the plant's value, the feasible set is restricted by a constraint on risk: The risk-adjusted value of the plant's values over time is bounded from below.

The state variables were modeled to vary on the time steps. Apart from this comparatively large time scale, the high-frequency trading activity had to be modeled. Consequently, a (fractional) hourly time scale had to be introduced. The formulation as a control problem allowed to remove the hourly time scale. The exogenous variables of spot prices over time were thereby transformed into a set of occupation times at different price levels (see the modified state equations (7.16)-(7.17)). In view of solving the optimization problem numerically, the number of the price levels was chosen to be finite. The finiteness allowed to parameterize the control in terms of fractions of maximal producing- and pumping-capacity (see admissible set (7.15)).

To solve the problem numerically, the random variables in the stochastic control problem have to be *discretized.* The discretized problem is formulated on a scenario tree (see Sec. 3.1, p. 12), where the tree is generated by a discretized version of the exogenous variables $(\boldsymbol{E}_t)_{t=0,\dots,T}$. The components of the vectors $\boldsymbol{E}_t$ are the occupation times $(F_{1t}, \dots, F_{Nt})$ and the water inflow I_t (definition (7.18)). Consequently, the distributions of the occupations times and of the water inflow must be discretized. A discretization method will be given in Chapter 8; here, for the formulation of the numerically solvable model, we assume that the exogenous variables are already suitably discretized. In terms of σ-algebras, we can say that the discrete exogenous variables generate a finite filtration $(\mathcal{F}_t)_{t=0,\dots,T}$ (cf. Sec. 3.1):

$$\mathcal{F}_t := \sigma(\boldsymbol{E}_0, \dots, \boldsymbol{E}_t), \quad t = 0, \dots, T,$$

where $\mathcal{F}_0 := \{\emptyset, \Omega\}$ because $\boldsymbol{E}_0$ is deterministic. The corresponding scenario tree is defined to be the representation of $(\mathcal{F}_t)_{t=0,\dots,T}$. All state variables, all the variables for the control, and all auxiliary variables take values on specific nodes of the scenario tree; the state equations determine on which node the variables take their values (see details below). The resulting problem is a *multi-stage stochastic program on a scenario tree.*

In the current discrete setting, we can now specify precisely the risk-adjusted value of the value process $(V_t)_{t=0,\dots,T}$ that appears as bounded from below in the set of feasible state variables (7.5) (p. 60):

$$\pi[V_0, \dots, V_T] \geq \rho_{\min}. \tag{7.19}$$

So far, π is a mapping from the space of stochastic processes into the reals. We define

$\pi[V_0, \dots, V_T] := R_0^{(V)}$, where $R_0^{(V)}$ is the recursive risk-adjusted value of the value process (Def. 4, p. 16). The corresponding set of test-probability measures is chosen to be a local-CVaR set (Def. 11, p. 28). The recursiveness ensures time consistency, which rules out regrettable optimal solution (Sec. 3.5, p. 19). In addition, the recursive risk-adjusted value is coherent, which implies the convenient properties (i)-(iv) of Lemma 2 (p. 18), and the stability of the local-CVaR set ensures consistency with single-period coherent risk measurement. Finally, local-CVaR sets enable a linear formulation of the constraint on risk and give a link to the common single-period risk-adjusted value CVaR (Sec. 4.3 and 4.4). The filtration of the recursive risk-adjusted value is chosen to be $(\mathcal{F}_t)_{t=0,\dots,T}$. Indeed, we will see in the next section that $(V_t)_{t=0,\dots,T}$ is adapted to this filtration.

Apart from discretization, the model should desirably be *linear*; the linearity permits to solve large problem instances. Fortunately, all the constraints of the control problem are already linear except the constraint on risk (7.19). By Prop. 3 (p. 34) the constraint on risk can be written as a system of linear constraints with help of auxiliary variables $(R_t)_{t=0,\dots,T}$, $(Q_t)_{t=0,\dots,T-1}$, $(Z_t)_{t=1,\dots,T}$.

7.3.2 The Discretized LP Formulation

In this section, we formulate the model on the scenario tree in detail. The scenario tree is defined by the node set $\mathcal{N}_t$ at time $t = 0, \dots, T$, and by the transition probabilities p_{nm} from each node n to another node m (Def. 8, p. 26). The values of the discretized exogenous variables $\boldsymbol{E}_t = (I_t, F_{1t}, \dots, F_{Nt})$ on node $n \in \mathcal{N}_t$, $t = 0, \dots, T$, is denoted by $(I_{tn}, F_{t1n}, \dots, F_{tNn})$.

The modified state equations for profit-and-loss and inflow (7.16)-(7.17), and the state equation for the value of the plant (7.8) induce that the state variables P_t, L_t and V_t are $\mathcal{F}_t$-measurable for all t. By Prop. 3 (p. 34), the auxiliary variables R_t, Q_t and Z_t of the constraint on risk are also $\mathcal{F}_t$-measurable for all t. Accordingly, for all t, all random variables are represented by their finitely many values on the scenario tree:

$$\begin{aligned} P_t, L_t, V_t &\quad\rightarrow\quad (P_{tn})_{n\in\mathcal{N}_t}, (L_{tn})_{n\in\mathcal{N}_t}, (V_{tn})_{n\in\mathcal{N}_t}, \\ R_t, Q_t, Z_t &\quad\rightarrow\quad (R_{tn})_{n\in\mathcal{N}_t}, (Q_{tn})_{n\in\mathcal{N}_t}, (Z_{tn})_{n\in\mathcal{N}_t}, \\ X_{it}^{\pm} &\quad\rightarrow\quad (X_{itn}^{\pm})_{n\in\mathcal{N}_t} \quad i = 1, \dots, N, \end{aligned}$$

where on the last line the control-functions $x_{it}^{\pm}$ is written as a $\mathcal{F}_t$-measurable random variable:

$$X_{it}^{\pm} = x_{it}^{\pm}(\boldsymbol{E}_0, \dots, \boldsymbol{E}_t), \qquad i = 1, \dots, N,\ t = 0, \dots, T-1. \tag{7.20}$$

On the scenario tree, expectations are finite sums. For the constraint on risk, this was already described in detail by the finite linear programming formulation (4.11)

(p. 33). For the objective function of the stochastic control model (SC) (p. 61), we get

$$\mathbb{E}[V_T] \qquad \to \qquad \sum_{n\in\mathcal{N}_T} p_{n_0 n} V_{Tn},$$

where $p_{n_0 n}$ is the transition probability from root node n_0 to a terminal node n. The conditional expectation in the value of the plant (7.8) corresponds pointwise (for each node n) for $t = 0, \ldots, T-1$, $k = 1, \ldots, T$ to the following sum:

$$\mathbb{E}[I_k|\mathcal{F}_t]\Big|_n \qquad \to \qquad \sum_{m\in\mathcal{N}_k} p_{nm} I_{km}, \qquad \text{for all } n \in \mathcal{N}_t,\ k > t,$$

where the sum goes over all nodes at time k, but where only the transition probabilities from node n to its successor nodes can be strictly positive.

As an additional feature of the stochastic programming model, the discounting is written explicitly with a continuous discounting rate $r \in \mathbb{R}_+$.

To summarize, the stochastic control problem (SC, p. 61) can be reformulated as the following finite stochastic linear program (SLP):

$$\max \sum_{n\in\mathcal{N}_T} p_{n_0 n} V_{Tn} \qquad \text{subject to (I)-(IV):} \tag{SLP}$$

$$\text{(I)} \begin{cases} P_{tn} = \quad u^+_{\max} H \sum_{k=1}^{t} e^{-rk} \sum_{i=1}^{N} \Big(\sum_{j=i+1}^{N} X^+_{j(k-1)n} \Big) \bar{s}_i (F_{ikn} - F_{(i-1)kn}) \\ \qquad - \dfrac{1}{c} u^-_{\max} H \sum_{k=1}^{t} e^{-rk} \sum_{i=1}^{N} \Big(\sum_{j=1}^{i} X^-_{j(k-1)} \Big) \bar{s}_i (F_{ikn} - F_{(i-1)kn}), \\ L_{tn} \le l_0 + u^-_{\max} H \sum_{k=1}^{t} \sum_{i=1}^{N} \Big(\sum_{j=1}^{i} X^-_{j(k-1)} \Big) (F_{ikn} - F_{(i-1)kn}) \\ \qquad - u^+_{\max} H \sum_{k=1}^{t} \sum_{i=1}^{N} \Big(\sum_{j=i+1}^{N} X^-_{j(k-1)} \Big) (F_{ikn} - F_{(i-1)kn}) + \sum_{k=1}^{t} I_{kn}, \\ V_{tn} = P_{tn} + e^{-rt} v_t \Big(L_{tn} - l_T + \sum_{k=t+1}^{T} \sum_{m\in\mathcal{N}_k} p_{nm} I_{km} \Big), \\ \qquad\qquad t = 0, \ldots, T,\ \forall n \in \mathcal{N}_t, \end{cases}$$

$$\text{(II)} \begin{cases} l_{\min} \le L_{tn} \le l_{\max}, & t = 0, \ldots, T,\ \forall n \in \mathcal{N}_t, \\ l_T \le L_{Tn}, & \forall n \in \mathcal{N}_T, \end{cases}$$

$$
\text{(III)}\begin{cases}
R_{0n_0} \geq \rho_{\min}, & \\
R_{tn} \leq V_{tn}, & t = 0, \dots, T,\ \forall n \in \mathcal{N}_t,\\
R_{tn} \leq Q_{tn} + \dfrac{1}{\alpha} \displaystyle\sum_{m \in \mathcal{N}_{t+1}} p_{nm} Z_{(t+1)m}, & t = 0, \dots, T-1,\ \forall n \in \mathcal{N}_t,\\
Z_{tn} \geq Q_{(t-1)n^-} - R_{tn}, & t = 1, \dots, T,\ \forall n \in \mathcal{N}_t,\\
Z_{tn} \geq 0, & t = 1, \dots, T,\ \forall n \in \mathcal{N}_t,
\end{cases}
$$

$$
\text{(IV)}\begin{cases}
\displaystyle\sum_{j=1}^{N} X^+_{jtn} = 1, \quad X^+_{itn} \geq 0, & i = 1, \dots, N,\ t = 0, \dots, T-1, \forall n \in \mathcal{N}_t,\\
\displaystyle\sum_{j=1}^{N} X^-_{jtn} = 1, \quad X^-_{itn} \geq 0, & i = 1, \dots, N,\ t = 0, \dots, T-1, \forall n \in \mathcal{N}_t.
\end{cases}
$$

Parameters

$s_0, \dots, s_N$:	price levels, with intermediate levels $\bar{s}_i \in (s_{i-1}, s_i)$
$l_{0/\min/\max/T}$:	water levels: starting, minimal, maximal, and minimal final
$u^+_{\max}$, $u^-_{\max}$:	maximal production rate, maximal pumping rate
v_t:	value of the remaining water at time t
c:	efficiency of pumping
r:	continuous discounting rate
$\rho_{\min}$:	lower bound of risk-adjusted value
H:	number of hours in time interval $[t-1, t]$
I_{tn}:	water inflow in time interval $[t-1, t]$ in node $n \in \mathcal{N}_t$
F_{itn}:	occupation time of spot price at level s_i in time interval $[t-1, t]$ in node $n \in \mathcal{N}_t$ (Def. 12, p. 63, where $F_{it} := F_t(s_i)$)
p_{nm}:	transition probability from node n to node m

Variables

At time t in node $n \in \mathcal{N}_t$:

P_{tn}:	cumulative profit-and-loss
L_{tn}:	water level
V_{tn}:	value of the plant
R_{0n_0}:	risk-adjusted value in root n_0 (when constraint is binding)
R_{tn}:	upper bound of risk-adjusted value process
Q_{tn}, Z_{tn}:	auxiliary variables for the risk-adjusted value
X^+_{itn}:	production rate (fraction of $u^+_{\max}$) if spot prices are in $(\bar{s}_{i-1}, \bar{s}_i]$
X^-_{itn}:	pumping rate (fraction of $u^-_{\max}$) if spot prices are in $[\bar{s}_{i-1}, \bar{s}_i)$

Constraints (I)-(IV)

We shortly review the constraints (they were thoroughly discussed in the previous two sections):

(I) The constraints correspond to the state equations of the stochastic control model (SC) (p. 61). The state equations for the cumulative profit-and-loss and the water level use the approximated formulation (7.16)-(7.17) (p. 64). Each term has the following form: For each time t and spot price level indexed by i, a fraction $X_{it}^{\pm}$ of dispatch is multiplied with the occupation time of spot prices in the interval $(s_{i-1}, s_i]$. This gives the additional fraction of dispatch in time interval $[t, t+1]$ for price interval $(\bar{s}_{i-1}, \bar{s}_i]$ with respect to prices lower or equal to $\bar{s}_{i-1}$. The water level is constrained by an inequality because an arbitrary spill-over is now allowed (in contrast to the control formulation (SC) where for the sake of simple presentation an equality was used).

(II) The constraints correspond to the feasible water levels in the set of feasible state variables (7.5) (p. 60).

(III) The constraints correspond to the constraint on financial risk in the set of feasible state variables (7.5). Because the applied risk-adjusted value uses a local-CVaR set (Def. 11, p. 28), Proposition 3 (p. 34) allows a formulation as a set of linear constraints (see also the discussion at the beginning of this section). For other useful properties of the risk-adjusted value see the previous discussion on page 60.

(IV) The constraints correspond to the parameterized set (7.15) of admissible controls.

7.4 Extension of the Model: Futures

The foregoing optimization problem (SLP) is a model of the dispatch decision of the hydro-electric pumped storage plant. The assumed decision maker of the plant trades the dispatched electricity only on the spot market; opportunities to invest in other markets do not exist. This restriction diminishes the flexibility to bound financial risk. To increase the flexibility, an extended version of the model considers a decision maker who can invest in futures contracts. In the case study, we will make use of the additional flexibility of futures.

Exchange traded futures-contracts are standardized, and the exchange is the intermediary between the buyer and the seller; this excludes any counterparty risk. Futures on the electricity market are different from those on traditional financial markets: Electricity futures exchange a fixed price against a floating price. Hence, these contracts can be considered to be financial *swaps* [9].

In our case, the floating price is chosen to be the spot price during consecutive time steps. The futures are assumed to be cash settled: The pay-off is the hourly difference between the fixed futures price and the spot price. These assumptions are not artificial: At the EEX (year 2005), Phelix-Base-Month-Futures (Phelix = 'Physical Electricity Index') are monthly futures on the hourly spot price.

The decision maker is assumed to enter all positions in futures at initial time $t = 0$, and the positions are not changed subsequently. This assumption is reasonable if the futures are used to *hedge*[1] the value of the plant over the time horizon, and are not used for short-term trading activities. Accordingly, a future-price process is not modeled.

For each time step, a different position amount p_t (MW) in futures can be chosen. The profit-and-loss of the futures in time period t to $t+1$ is

$$P_{t+1}^{\text{fut}} - P_t^{\text{fut}} = p_t \sum_{h=1}^{H} \big(f_t - S_{t+\frac{h}{H}}\big), \qquad t = 0, \ldots, T-1, \tag{7.21}$$

where f_t is the futures-price (Euro/MWh) as fixed at time zero (discounting is suppressed). Using the approximation (7.10) and the associated transformation into occupation times $F_{it} = F_t(s_i)$, the approximated profit-and-loss is

$$P_{t+1}^{\text{fut}} - P_t^{\text{fut}} = p_t \Big(H f_t - H \sum_{i=1}^{N} \bar{s}_i \big(F_{i(t+1)} - F_{(i-1)(t+1)}\big)\Big), \quad t = 0, \ldots, T-1. \tag{7.22}$$

Because the considered models have a time-horizon within a year, discounting can be neglected.

[1]Hedging is an action to reduce risk

Chapter 8

Scenario Tree Generation

In this chapter, the scenario tree generation method is presented. The scenario tree is generated by a discretized version of the exogenous random variables over time. Let us recall the stochastic process of exogenous variables (7.18):

$$\boldsymbol{E}_t = (I_t, F_{1t}, \ldots, F_{Nt})^\top, \qquad t = 0, \ldots, T,$$

where I_t is the water inflow in time interval $[t-1, t]$, and F_{it} is the occupation time of the spot price at level s_i, $i = 1, \ldots, N$ in time interval $[t-1, t]$ (F_{0t} is neglected; we assume $s_0 = 0$, hence $F_{0t} \equiv 0$ for all t).

We assume that the random variables of water inflow are stochastically independent from the occupation times of spot prices. The reasoning is as follows. The water inflow is a function of several variables: Temperature, precipitation, altitude of reservoir, water branch-off, snowmelt etc. Some of these variables depend on local circumstances. The available historical sample for several water reservoirs was small; a proper correlation estimation would require more data. In addition, the EEX-power market covers a large region of Europe with various electricity sources and heterogenous weather conditions. By contrast, in spatially smaller regions such as in Scandinavia, where hydro-energy is the predominant source for electricity, the local electricity price was reported to be correlated with precipitation [33].

In the following, we give a short overview of the scenario tree generation method. A factor model will reduce the random vector of exogenous variables to a small number of factors. Each factor is modeled by an autoregressive process. Autoregressive means that the value at a specific time is a function of previous values plus an additive, stochastic random quantity, called the *innovation* (*white-noise*). It is assumed that these innovations are stochastically independent with respect to different time and different factors. This assumption will be tested in the case study. The independence allows for a simple discretization scheme of the exogenous variables: The distribution of each innovation will be approximated separately by a *discrete distribution*. The scenario tree is generated *inductively*: The root node corresponds to the starting value of the factors. Given a specific node of the tree at a specific time, the different values of the discretized innovations generate the immediate successor nodes of the scenario tree. In each node, the values of the innovations determine the values of the associated factors. Finally, the values of the factors determine via the factor model the values of the original exogenous variables.

The dimension-reduction and the discretization technique is inspired by Jamishidian and Zhu [48]. We give a short overview of other approaches (for reviews see e.g. [26, 51]):

One approach is to sample complete paths of the exogenous variables over time. Then, the pathes are clustered into a tree by some distance criteria (see e.g. [42]). Conditional random sampling methods can produce directly a tree (see e.g. [56]). Another commonly used method to produce an approximative distribution (conditional on a given node) is by matching moments of the exact distribution. Commonly, the first few moments are matched [46, 47].

Other methods yield directly a bound on the difference in optimal objective value of the approximative optimization problem with respect to the exact optimization problem. A particular kind of such methods reduces the value of a so-called probability metric between the exact and approximative distribution. Under suitable assumptions of the optimization problem, this metric is an upper bound for differences in optimal objective values [27, 39, 43, 67]. Other examples use linear complementarity approaches [77], or barycentric approximations [35].

Some of the aforementioned methods use an iterative disaggregation algorithm: They start with a small tree and insert additional branches into the tree. The node in which to disaggregate is thereby determined by a local or global criteria. For example, the branching can be based on a local information criterion [21], a local infeasibility criterion [54], or a global criteria like complementary relations [77].

Our approach of scenario tree generation is rather unsophisticated; the forthcoming model of the exogenous variables is demanding by itself, such that the tree generation uses the previously outlined, simple discretization method.

8.1 The Model of Occupation Times

In the following, the multivariate vector of occupation times is described by a *statistical factor model*. The factors are related to *principal components* of the sample-covariance-matrix of the occupation times. The proposed factor model is fairly standard and based on Zivot and Wang [84]. The idea to use principal components for scenario generation is from Jamshidian and Zhu [48]. The methodology seems to be applied for the first time to occupation times.

8.1.1 Principal Components

Principal component analysis is a dimension reduction technique. The aim is to explain the sample-covariance-matrix $\widehat{\Sigma}$ of a multivariate random vector by a small number of principal components. Principal components are linear combinations of the components of the random vector. The coefficients $\boldsymbol{b}_k$, $k = 1, \dots, K$, of the principal components are successively chosen according to the following set of rules:

(i) The variance of the sum of the first k principal components is maximal,

(ii) principal components are orthogonal,

(iii) principal components have unit length (Euclidean norm).

The corresponding least-square optimization problem can be solved by Lagrange's method of multipliers (see e.g. [78, p. 42]); it turns out that the principal components are eigenvectors corresponding to the K largest eigenvalues of the matrix $\widehat{\Sigma}$.

In our case, the multivariate quantity is the occupation time at the N different price levels. The $(N \times N)$-sample covariance matrix of occupation times is given by

$$\widehat{\Sigma} = \frac{1}{M} \widehat{F} \widehat{F}^{\top}, \qquad \text{with } \widehat{F} := (\widehat{\boldsymbol{F}}_1, \dots, \widehat{\boldsymbol{F}}_M),$$

where M is the size of the historical sample, and where the mean-adjusted vector of occupation times is given by

$$\widehat{\boldsymbol{F}}_t = \begin{pmatrix} \widehat{F}_{1t} - \overline{F}_1 \\ \vdots \\ \widehat{F}_{Nt} - \overline{F}_N \end{pmatrix}, \qquad \text{with } \overline{F}_i = \frac{1}{M} \sum_{t=1}^{M} \widehat{F}_{it},$$

where $\widehat{F}_{it}$ is the empirical sample at historical time $t = 1, \dots, M$ for level i. The principal components of $\widehat{\Sigma}$ are used for the following factor model.

8.1.2 Statistical Factor Model

The factor model tries to describe the mean-adjusted stochastic process of the multivariate occupation times,

$$\boldsymbol{F}_t = (F_{1t} - \overline{F}_1, \dots, F_{Nt} - \overline{F}_N)^{\top}, \qquad t = 1, 2, \dots,$$

by a related sequence of stochastic *factors* $(\boldsymbol{G}_t)_{t=1,2,\dots}$. Preferably, the dimension K of $\boldsymbol{G}_t$ should be smaller than $\boldsymbol{F}_t$. In our case, the factor model is chosen to be linear and to have no time lags (the use of this rather simple model is justified by the model-parameter estimation in the case study). The linear regression equation of the *statistical factor model* is

$$\boldsymbol{F}_t = \boldsymbol{\mu} + B\boldsymbol{G}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, 2 \ldots, \tag{8.1}$$

where $\boldsymbol{\mu} \in \mathbb{R}^N$ is an intercept, B is an $(N \times K)$-matrix of *factor loadings*, and the residual error is $\boldsymbol{\varepsilon}_t$[1] [84, Ch. 15]. Standard assumptions for the regression are [84]:

- $\boldsymbol{\varepsilon}_t$ has zero mean for all t,
- $(\boldsymbol{\varepsilon}_t)_{t=1,2,\ldots}$ is serially uncorrelated, and the vector $\boldsymbol{\varepsilon}_t$ is uncorrelated for all t,
- $\mathrm{COV}[\boldsymbol{\varepsilon}_t]$ is time invariant,
- factors and errors are not correlated: $\mathrm{COV}[\boldsymbol{G}_t, \boldsymbol{\varepsilon}_s] = 0$ for all t, s.

The input to (8.1) are the estimated factor realizations $\widehat{\boldsymbol{G}}_t := (\widehat{G}_{1t}, \ldots, \widehat{G}_{Kt})^\top$, which are chosen to be given by the foregoing principal component analysis:

$$\widehat{G}_{kt} = \widehat{\boldsymbol{b}}_k^\top \widehat{\boldsymbol{F}}_t, \qquad t = 1, \ldots, M, \ k = 1, \ldots, K, \tag{8.2}$$

where $\widehat{\boldsymbol{b}}_k$ is the kth eigenvector of the sample covariance matrix $\widehat{\Sigma}$ of occupation times. The loadings $\boldsymbol{\beta}_i$ for each price level i, the intercepts μ_i, and the residual variances $\mathrm{VAR}[\varepsilon_{it}] = \sigma_i^2$ are estimated from the regression (8.1):

$$\widehat{F}_{it} = \mu_i + \boldsymbol{\beta}_i^\top \widehat{\boldsymbol{G}}_t + \varepsilon_{it}, \quad t = 1, \ldots, M, \ i = 1, \ldots, N, \tag{8.3}$$

yielding the estimates $\widehat{\mu}_i$, $\widehat{\boldsymbol{\beta}}_i$, and $\widehat{\sigma}_i$. Note that $\widehat{\sigma}_i$ can be interpreted as the residual variance, which cannot be explained by the factors. The tree generation method will choose K as large such that the residual variance is small. If we neglect the residual variance, then the dynamics of the occupation times is described entirely by the dynamics of the factors.

If $\boldsymbol{F}_t$ is covariance stationary, then the sample covariance matrix $\widehat{\Sigma}$ is an estimator of $\mathrm{COV}[\boldsymbol{F}_t]$ (the same for all t), and the columns of B are approximately the first K eigenvectors of $\widehat{\Sigma}$. In this case, the variances and covariances of the factors are

$$\mathrm{COV}[G_{kt}, G_{lt}] = \mathbb{E}[\boldsymbol{b}_k^\top \boldsymbol{F}_t (\boldsymbol{b}_l^\top \boldsymbol{F}_t)^\top] = \boldsymbol{b}_k^\top \mathrm{COV}[\boldsymbol{F}_t] \boldsymbol{b}_l \approx \boldsymbol{b}_k^\top \widehat{\Sigma} \boldsymbol{b}_l = \delta_{kl} \lambda_l, \tag{8.4}$$

where λ_l denotes the eigenvalue of eigenvector $\boldsymbol{b}_l$. Hence, the factors are approximately uncorrelated, and the ratio of the *cumulative variance explained by K factors* is

$$\frac{\sum_{k=1}^K \mathrm{VAR}[G_{kt}]}{\sum_{i=1}^N \mathrm{VAR}[F_{it}]} \approx \frac{\sum_{k=1}^K \lambda_k}{\sum_{i=1}^N \lambda_i}. \tag{8.5}$$

The nearer the ratio is to one, the better is the factor model.

[1] Notational exception: $\boldsymbol{\varepsilon}_t$, though a random vector, is lower case.

8.1.3 Autoregressive Model of the Factors

So far, the factor model has reduced the dynamics of the occupation times to the dynamics of the factors; the estimated time series of the factors was $(\widehat{G}_{kt})_{t=1,\dots,M}$ (8.2). Under the assumption that $(\boldsymbol{F}_t)_{t=1,2,\dots}$ is covariance stationary it follows from (8.4) that the factors are uncorrelated. Indeed, the case study will show that the estimated correlations are small. Accordingly, each factor is modeled separately with a univariate autoregressive process. The case study will validate that it suffices to consider an autoregressive process of order 1 *(AR(1)-process)*. For simplicity, we neglect any trends and seasonality and assume that the factors are covariance stationary. These assumptions are justified by the short time horizon of several months of the model of the electricity plant. The mean-adjusted regression-equation of the $AR(1)$-process for the kth factor is

$$G_{kt} - \overline{G}_k = \phi_k(G_{k(t-1)} - \overline{G}_k) + \epsilon_{kt}, \qquad t = 1, \dots, M, \ k = 1, \dots, K, \tag{8.6}$$

where $\overline{G}_k = \frac{1}{M}\sum_{t=1}^{M} G_{kt}$ denotes the mean, ϕ_k the coupling, and ϵ_{kt}, $t = 1, 2, \dots,$ is the sequence of i.i.d. innovations with vanishing mean and $\mathrm{VAR}[\epsilon_{kt}] = \sigma_k^2$. Plugging the estimated factors $(\widehat{G}_{kt})_{t=1,\dots,M}$ into the regression (8.6) for each k, we can estimate $\widehat{\phi}_k$ and $\widehat{\sigma}_k$.

8.2 The Model of Water Inflow

State-of-the-art models for the daily river throughput in mountainous areas are periodic autoregressive models with external influences like temperature or precipitation [16]. Sufficient amount of data to calibrate such models is rarely available. In addition, local environmental factors influence the water inflow into reservoirs (cf. start of Ch. 8). Hence, different reservoirs may have different inflow dynamics. In our case, to cover a wide class of reservoirs, the process of water inflow considers only the first two moments of the distribution at a specific time: The inflow at time t is the sum of a historical mean, denoted by $\bar{I}_t$, and an i.i.d. perturbation ϵ_{0t} with zero mean and $\mathrm{VAR}[\epsilon_{0t}] = \sigma_0^2$:

$$I_t = \bar{I}_t + \epsilon_{0t}, \qquad t = 1, 2, \dots . \tag{8.7}$$

The i.i.d. assumption is a simplification: If the time steps are sufficiently short, then inflow is clearly autocorrelated. In our model, the time intervals are large (e.g. a month); the autocorrelation of monthly precipitation can be neglected [75]. Thus, by neglecting the storage of water as ice, a monthly inflow model can be assumed to have no autocorrelation. Moreover, the variance of inflow may change during seasons. But, it is questionable if a more detailed model of an idiosyncratic mountainous river adds any value to the analysis (cf. Döge [23, p. 19]).

Model (8.7) gives for the conditional expected future inflow

$$\sum_{s=t+1}^{T} \mathbb{E}[I_s|\mathcal{F}_t] = \sum_{s=t+1}^{T} \bar{I}_s, \tag{8.8}$$

where $\mathcal{F}_t = \sigma(\boldsymbol{E}_0, \ldots, \boldsymbol{E}_t)$. The formula holds because ϵ_{0s} is independent of $\mathcal{F}_t$ for $s > t$, hence $\mathbb{E}[\epsilon_{0s}|\mathcal{F}_t] = \mathbb{E}[\epsilon_{0s}] = 0$. Expression (8.8) is used for the calculation of the value of the plant (7.8) (p. 61).

8.3 The Generation of Discrete Distributions

Let us recall that the exogenous quantities are the occupation times of the spot price and the water inflow. The random variables in the model of the occupation times are the i.i.d. innovations ϵ_{kt}, $t = 1, 2, \ldots$ of the factors $k = 1, \ldots, K$ (Sec. 8.1.3), and the random variables in the model of the water inflow are the i.i.d. perturbations ϵ_{0t}, $t = 1, 2, \ldots$ (Sec. 8.2). These random variables can be combined into a *random vector of innovations*

$$(\epsilon_{0t}, \epsilon_{1t}, \ldots, \epsilon_{Kt}), \qquad t = 1, \ldots, T. \tag{8.9}$$

By assumption, the water inflow and the factors are stochastically independent (see start of Ch. 8). Hence, the sequence of the random vectors of innovations (8.9) must be serially and contemporarily independent, too. Apart from independence, we assume that *each component* ϵ_{kt} *of the random vector of innovations* (8.9) *is normally distributed.*

The normality of the first component (water inflow) is in accordance to other models on a continuous time scale [23]. The normality of the innovation of the factors is motivated by a (presumed) *asymptotic normality* of the occupation times of the electricity spot price as follows.

The presumption of asymptotic normality is based on the extended Central Limit Theorem in Appendix A.4 (p. 116): If the hourly spot-price is a *weak-dependent* stochastic process, then the vector of occupation times is asymptotically multivariate normally distributed. Asymptotically means that the number of hours for which the occupation time is calculated tends to infinity (Def. 12, p. 63). Hence, if the time scale of the optimization problem of the electricity plant is large in comparison to one hour, for example a month or larger, and if the spot price exhibits only weak dependencies, then the occupation times are approximately normal. Hence, if we choose the innovations ϵ_{kt} in the autoregressive models of the factors to be normal, then the factors are normal, and if the residuals in the factor regression (8.1) are chosen to be normal, then the normality of the factors translates into the normality of the occupation times (the linear transformation in (8.1) from factors into occupation times preserves normality).

The approximative normality of the factors and of the occupation times will be tested in the case study. To summarize: The exogenous variables are modeled such that the random vector of innovations (8.9) has components that are

- normally distributed (conjectured by Central Limit Theorem and verified in the case study),
- stochastically independent from the others (by the factor model and by the independence of water inflow and electricity price),
- constitute an i.i.d. sequence over time (standard assumptions of autoregressive models).

To generate a scenario tree, the normal distributions of the random vector of innovations have to be approximated by finitely-discrete distributions. A suitable choice are *binomial distributions* (cf. Jamshidian and Zhu [48]). For details, see Appendix A.5 (p. 117).

8.4 The Generation of the Scenario Tree

Let us recall again that the scenario tree is generated by the process of the discretized exogenous variables, which are the water inflow and the occupation times (represented by the factors). The combined random quantity in the models of inflow and occupation times is the random vector of innovations (8.9), which in turn is discretized by binomial distributions. Algorithmically, the scenario tree is inductively generated by the discretized innovations over the time steps $t = 0, \ldots, T$:

root node n_0 (time $t = 0$):

The factors are initialized with their historical long-term mean:

$$G_{k0n_0} = \overline{G}_k, \qquad k = 1, \ldots, K.$$

If the end of historical sample time, M, is isochronous with the root node (this work does not assume that), an alternative would be to initialize with the last historically estimated values $\widehat{G}_{kM}$. In any case, the autocorrelation (dependence on initial value) of a mean-reverting AR(1)-process decays exponentially [84, Sec. 3.2.3].

node n^- → successor node n (time $t-1 \to t$):

Let the node n^- be given. The successor node n is determined by a combined value of the discretized random vector of innovations:

$$(\widehat{\sigma}_0 B_{0t}, \widehat{\sigma}_1 B_{1t}, \ldots, \widehat{\sigma}_K B_{Kt}),$$

where the B_{kt} are independent, standardized[1] binomial distributions with parameters J_{kt} and $1/2$, $k = 0, \ldots, K$ (see Appendix A.5, p. 117). The value in node n is denoted by

$$(\epsilon_{0tn}, \epsilon_{1tn}, \ldots, \epsilon_{Ktn}) \in \mathbb{R}^{K+1}. \tag{8.10}$$

The parent node n^- has $\prod_{k=0}^{K}(J_{kt}+1)$ successor nodes. Because the components of the vector are stochastically independent, the transition probability from node n^- to node n is the product of the marginal probability of the binomial distribution probabilities:

$$p_{n^- n} = \prod_{k=0}^{K} \binom{J_{kt}}{j_{kt}^{(n)}} \left(\frac{1}{2}\right)^{J_{kt}},$$

where node n is associated to the vector of realizations $(j_{0t}^{(n)}, \ldots, j_{Kt}^{(n)})$ of the binomial distributions; each vector in the range

$$(0, \ldots, 0) \leq (j_{0t}, \ldots, j_{Kt}) \leq (J_{0t}, \ldots, J_{Kt}).$$

is associated to a different successor node. The values of the exogenous variables in node n at time t are computed as follows:

Factors: Let the value of the kth factor in node n^- be given: $G_{k(t-1)n^-}$. The value of the kth *factor* in node n, G_{ktn}, is given by the autoregressive model (8.6), where the innovation is set to the value ϵ_{ktn}.

Occupation Times: The value of the ith *occupation time* in node n is given by the factor model (8.3):

$$F_{itn} = \widehat{\mu}_i + \widehat{\boldsymbol{\beta}}_i^\top (G_{1tn}, \ldots, G_{Ktn})^\top, \qquad i = 1, \ldots, N. \tag{8.11}$$

Water Inflow: The value of *water inflow* in node n a time t is given by the model (8.7):

$$I_{tn} = \bar{I}_t + \epsilon_{0tn}.$$

Note. Because the price levels are ordered, that is $s_i > s_{i-1}$ for all i, the original random vectors of occupation times are ordered, too: For each time t, we have

$$F_{it} \geq F_{(i-1)t} \quad a.s., \qquad i = 2, \ldots, N.$$

By contrast, the discretized values (8.11) may not be ordered: There might be $F_{itn} < F_{(i-1)tn}$ for a specific triple (i, t, n). In the optimization model of the electricity plant, the occupation times appear in the state equations (and only there). Fortunately,

[1] i.e. unit variance, zero mean

the state equations are in a sense robust against such a 'wrong' order as follows. The values F_{itn} and $F_{(i-1)tn}$ enter the first two state equations (see model (SLP), p. 68) in the form of the term

$$f(\bar{s}_{i-1})\big(F_{(i-1)tn} - F_{(i-2)tn}\big) + f(\bar{s}_i)\big(F_{itn} - F_{(i-1)tn}\big) + f(\bar{s}_{i+1})\big(F_{(i+1)tn} - F_{itn}\big),$$

where f is a monotonically increasing function. By re-ordering the terms, we obtain

$$\begin{aligned} &f(\bar{s}_{i-1})\big(F_{itn} - F_{(i-2)tn}\big) + f(\bar{s}_i)\big(F_{(i-1)tn} - F_{itn}\big) + f(\bar{s}_{i+1})\big(F_{(i+1)tn} - F_{(i-1)tn}\big) \\ &+ 2\big(F_{itn} - F_{(i-1)tn}\big) \underbrace{\Big(f(\bar{s}_i) - \frac{1}{2}\big(f(\bar{s}_{i-1}) + f(\bar{s}_{i+1})\big)\Big)}_{\approx 0}. \end{aligned}$$

Hence, if $F_{itn} < F_{(i-1)tn}$, then we can consider the ordered vector

$$(\ldots, F_{(i-1)tn}, F_{itn}, \ \ldots) \ \rightarrow \ (\ldots, F_{itn}, F_{(i-1)tn}, \ \ldots),$$

which leads only to a slightly changed value in the state equations.

Chapter 9

Case study

The goals of the case study are as follows.

(1) *Operationalization* and *quality-testing* of the scenario tree generation method,

(2) testing how the *constraint on risk* influences the optimal objective value and the optimal decisions,

(3) testing how the *flexibility of the feasible decisions* influences the optimal objective value, where flexibility means the ability to react to exogenous events or to endogenous states.

The structure of the case study is as follows.

First, the parameters of the scenario tree generation method are estimated, and the applied estimation methods are validated. As a by-product, a distinctive pattern of the covariances of the occupation times of the spot prices is provided.

Second, we examine the sensitivity of the optimal objective value and (partially) of the optimal solution with respect to changes in various model properties: Tree size, increased flexibility by futures, stochasticity of water inflow, right-hand-side of risk constraint, multi-period versus single-period risk constraint, relaxation of non-anticipativity, and restriction to state-independent decisions.

Third, we test the quality of the scenario tree generation method with a self-contained benchmark method, which uses Monte-Carlo sampling.

9.1 Estimations for the Scenario Tree Model

In this section, we validate the factor model of the occupation times of the electricity prices: First, we estimate the statistical factor, then we estimate the autoregressive models of each estimated factor. At the end of the section, the estimation of the model of water inflow is discussed shortly.

9.1.1 Principal Components of Occupation Times

First, we give some statistical properties of the underlying historical electricity spot-price data, which consists of an hourly time series from 1 January 2002 to 1 February 2005 taken from the EEX market [32].

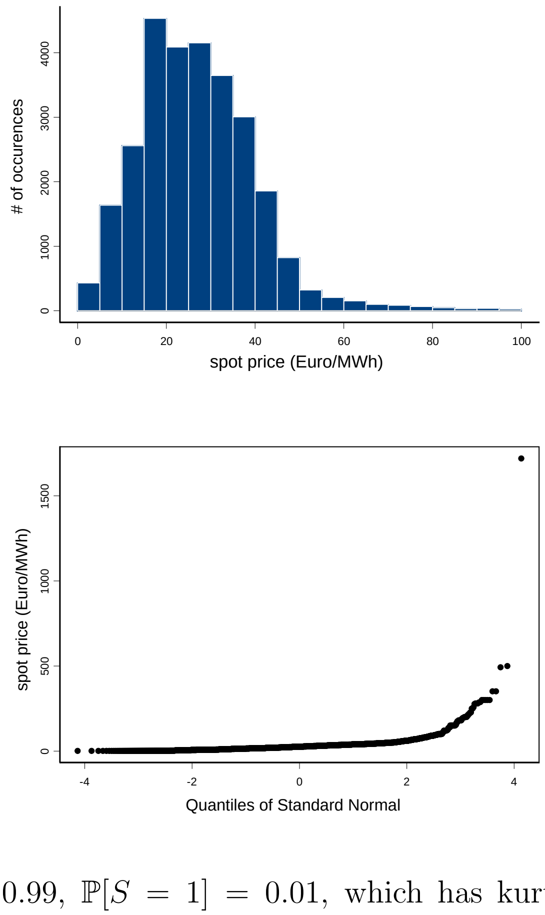
The summary statistics are in Table 9.1. The spot price is extremely leptokurtic (Def. kurtosis: $\text{mean}(Y^4)/\text{mean}(Y^2)^2 - 3$, where $Y = S - \text{mean}(S)$). The kurtosis can be large because higher moments are sensitive to outliers. For example, consider

Table 9.1: Summary statistics of the spot price from 1/1/02 to 1/2/05: The quantiles at thirteen different probability levels; the first four moments; the histogram (clipped at 100 Euro/MWh); and the quantile-quantile plot with respect to the standard normal distribution.

probability	quantile (Euro/MWh)
0.01	3.34
0.05	8.05
0.1	12.00
0.2	16.13
0.3	19.06
0.4	22.68
0.5	25.67
0.6	29.18
0.7	32.89
0.8	36.85
0.9	42.25
0.95	48.60
0.99	79.97

min	0.00 Euro/MWh
max[a]	1719.72 Euro/MWh
mean	27.61 Euro/MWh
std.dev.	19.64 Euro/MWh
skewness	26.58
kurtosis	2021.57

[a]7. Jan. 2003, 6-7 p.m. (cf. [23])



a two-point distribution $\mathbb{P}[S = 0] = 0.99$, $\mathbb{P}[S = 1] = 0.01$, which has kurtosis $= 95.0101$ (cf. [53]). The outliers can be caused by an unforeseen disruption of production capacity in the EEX area, or by an unexpected high demand of electricity. By choosing the level s_N in the set of price levels $s_1, \ldots, s_N$ (7.11) (p. 62) sufficiently high, the model can take such outliers into account to some extend. The spot price is not normally distributed: The null hypothesis that the historical spot price is normally distributed is rejected (0.00% p-value) in a Jarque-Bera test[1]. The logarithm of the spot price gave the same test result. Thus, the distribution of the spot price is far from normal. We will see that the occupation times have a much more pronounced similarity to a normal distribution.

The price levels s_i, $i = 1, \ldots, N$, of the occupation times are chosen to be the empirical quantiles of the spot price at $N = 13$ probability levels (Table 9.1). The intermediate points $\bar{s}_i \in (s_{i-1}, s_i)$ with $s_0 = 0$ are chosen to be the conditional means.

The points in time, $t = 0, 1, \ldots, T$, of the stochastic programming model are chosen to be one month apart. For each month, the (normalized) occupation times at the thresholds s_i are calculated. Thus, we get a monthly, multivariate time series of occupation times of dimension $N = 13$. The estimation of the factor model for this time series goes as described in Section 8.1: The sample covariance matrix, its eigenvalues and eigenvectors $\widehat{\boldsymbol{b}}_k$ (principal components) are calculated from the mean-adjusted time series (Sec. 8.1.1). Then, a sufficient number K of principal components has to be determined with the criterium of variance explained (8.5) (p. 75). We assume that 95% explained variance is sufficient.

The result is that the observed variance explained by the first principal component is 90%, and by the first two principal component it is 96% (Fig. 9.1). Hence,

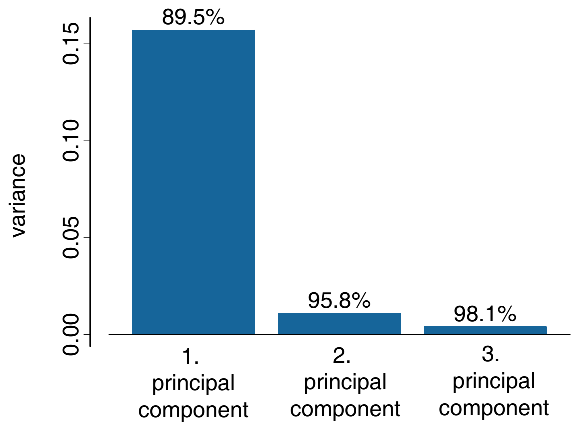


Figure 9.1: The first three principal components of the vector of occupation times. Vertical axis: variance (eigenvalue of sample-covariance matrix). On top of bars: Percentage of cumulated variance.

it suffices to select the first two principal components as the only factors ($K = 2$).

The coefficient-vectors $\widehat{\boldsymbol{b}}_k$ of the principal components $k = 1, 2, 3$ have a distinctive form (see Fig. 9.2). We give a possible, somewhat preliminary interpretation of the first two principal components $\widehat{\boldsymbol{b}}_k$, $k = 1, 2$. Because the ith coefficient

[1] See the end of the current section for more information on this statistical test.

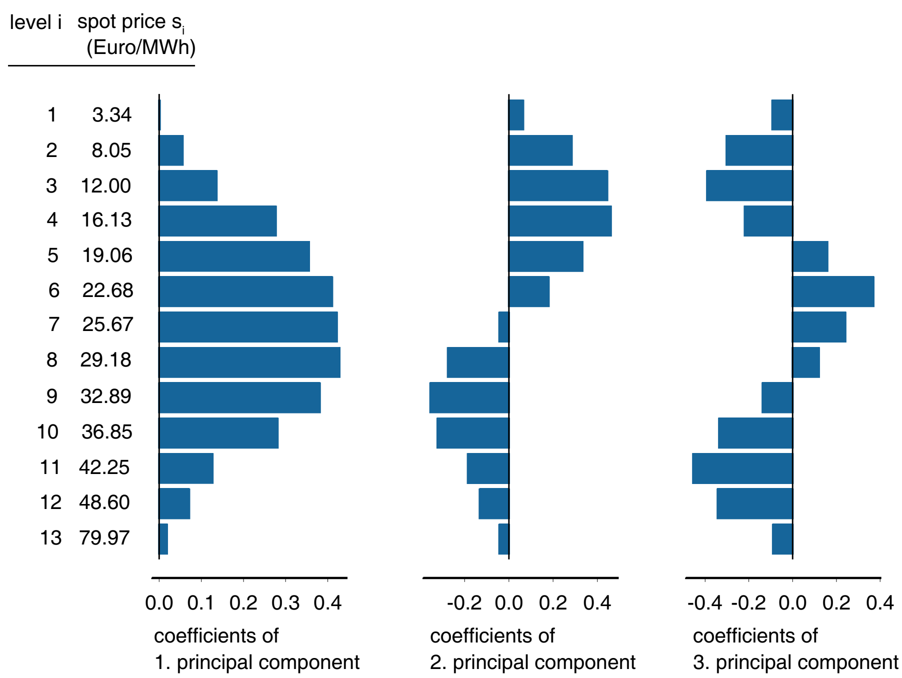


Figure 9.2: The vector $\widehat{\boldsymbol{b}}_k$ of coefficients of the first three principal components $k = 1, 2, 3$. Each coefficient corresponds to a price level $s_1, \ldots, s_{13}$ of occupation times (cf. Table 9.1).

$b_{ki} := (\widehat{\boldsymbol{b}}_k)_i$ corresponds to the occupation time at price level s_i, the differences of the ith and the $(i-1)$th coefficient corresponds to the *occupation time in price interval* $(s_{i-1}, s_i]$, i.e., in terms of the factor model, the occupation time in price interval $(s_{i-1}, s_i]$, $i = 1, \ldots, N$, and time $(t-1, t]$ is

$$F_t(s_i) - F_t(s_{i-1}) = \mu_i - \mu_{i-1} + (b_{1i} - b_{1(i-1)})G_{1t} + (b_{2i} - b_{2(i-1)})G_{2t} + \text{residual}.$$

The observed differences $\hat{b}_{1i} - \hat{b}_{1(i-1)}$, $i = 1, \ldots, 13$, of the first principal component are as follows (Fig. 9.2). The differences for $i \leq 8$ have positive sign, whereas the differences for $i > 8$ have negative sign. Hence, if the first factor G_{1t} increases over time, and if movements in the second factor G_{2t} are neglected, then the frequencies for high prices increase, whereas the frequencies for low prices prices decrease (and vice versa for G_{1t} decreasing). In other words, the first factor can be associated with a *shift* between high and low prices. The differences $\hat{b}_{2i} - \hat{b}_{2(i-1)}$ of the second principal component have for medium-price intervals, $i = 4, \ldots, 8$, a positive sign, whereas for high- and low-price intervals a negative sign. Hence, the second factor either disperses prices (G_{2t} decreases) or concentrates them near the mean (G_{2t} increases). Hence, the second factor can be associated with *volatility*.

Empirically, a similar pattern of principal components is observed for the term structure of interest rates: The first three principal components have the figurative names *shift*, *tilt*, and *hump* (also called *butterfly*) [37, 48]. In addition, the author was able to reproduce semi-analytically the same pattern of principal components for the occupation times of a mean-reverting gaussian markov process (Ornstein-Uhlenbeck process) [22]. Ornstein-Uhlenbeck processes are suitable models for the empirical spot price [15, 61].

The foregoing principal component analysis suggests to condense the original time series of occupation times, which had dimension 13, into a 2-dimensional time series of factors (8.2) (p. 75): $(\widehat{G}_{kt})_{t=0,1,\ldots}$, $k = 1, 2$. The time series of the factors is used to fit the regression equation of the factor model (8.3) (p. 75) by an Ordinary Least-Square method. The variability explained by the regression model can be expressed by the sample multiple correlation coefficient R^2 which gave a median of 0.9986. The regression gives the estimated factor loadings and intercepts of the factor model.

Having estimated the factor model, it remains to estimate the time series model of the factors themselves. Before we proceed, we test the assumption of normality for the occupation times and for the factors (as stated in Sec. 8.3).

Only marginal normality is tested: We use a Jarque-Bera univariate normality test. The Jarque-Bera test statistic is a function of the sample skewness and the sample kurtosis [84, Ch. 3]. The null hypothesis of the test is normality. If the null hypothesis is true, then discrepancies as large or larger than the one observed in the test would take place with probability (of so-called) *p-value*. The test is applied to the occupation time at each of the thirteen price levels, as well as to the two factors. The result is as follows (Table 9.2): With a p-value of 21% or higher, the null hypothesis of normality is not rejected for the occupation times from the 4th price level till the 10th price level. For the two factors, the null hypothesis is not rejected with a p-value of 60% or higher. The relatively large deviation from normality of the occupation times at very small and large levels can be attributed to the scarcity of historical data of these rare events. Generally, in comparison to traditional financial market data, only a small amount of historical electricity price data is available (38 data points on the monthly time scale in this case study).

9.1.2 The AR(1)-Model of the Factors

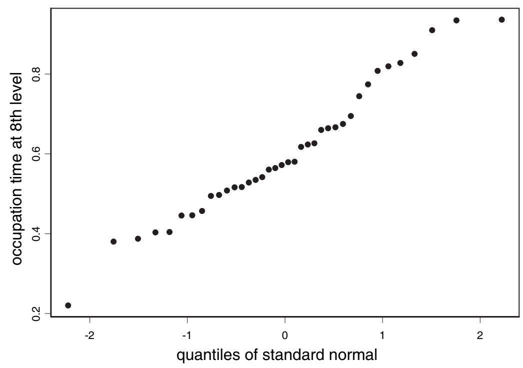
The time series of factors is modeled by AR(1)-processes (Sec. 8.1.3). Next, we will check whether AR(1)-processes are in fact a suitable choice.

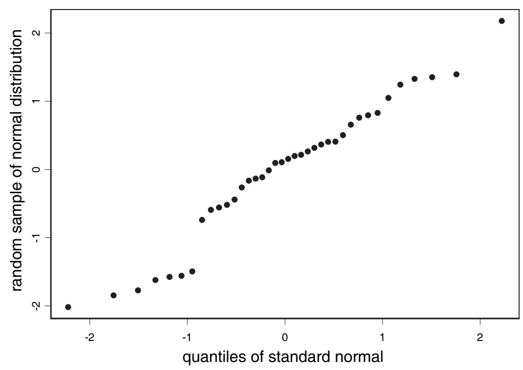
First, the empirical autocorrelation and the empirical partial autocorrelation are calculated. In general, autocorrelations of an AR(1)-process are exponentially decaying, and partial autocorrelations are insignificant after the first time lag [84]. The observed empirical autocorrelation as well as the empirical partial autocorrelation do no reject an AR(1)-model (Fig. 9.3).

Second, the sample covariance matrix of the factors is calculated. If all the assumptions of the factor model of the occupation times would be exactly met, then

Table 9.2: On the left: Jarque-Bera normality test for the historical occupation times $\widehat{F}_i := (\widehat{F}_{it})_{t=1,\dots,M}$, $i = 1, \dots, 13$, and for the factors $(\widehat{G}_{kt})_{t=1,\dots,M}$, $k = 1, 2$. The probability of the quantile s_i of spot price is indicated in parentheses. On the right above: Quantile-quantile plot of $\widehat{F}_8$ with respect to normal distribution. Below: The same plot for a randomly generated normal sample.

occupation time		p-value
$\widehat{F}_1$	(0.01)	0.0000
$\widehat{F}_2$	(0.05)	0.0001
$\widehat{F}_3$	(0.1)	0.0144
$\widehat{F}_4$	(0.2)	0.2146
$\widehat{F}_5$	(0.3)	0.2648
$\widehat{F}_6$	(0.4)	0.4309
$\widehat{F}_7$	(0.5)	0.6929
$\widehat{F}_8$	(0.6)	0.7384
$\widehat{F}_9$	(0.7)	0.8088
$\widehat{F}_{10}$	(0.8)	0.5324
$\widehat{F}_{11}$	(0.9)	0.0145
$\widehat{F}_{12}$	(0.95)	0.0035
$\widehat{F}_{13}$	(0.99)	0.0000
factor		
$\widehat{G}_1$		0.6026
$\widehat{G}_2$		0.6513





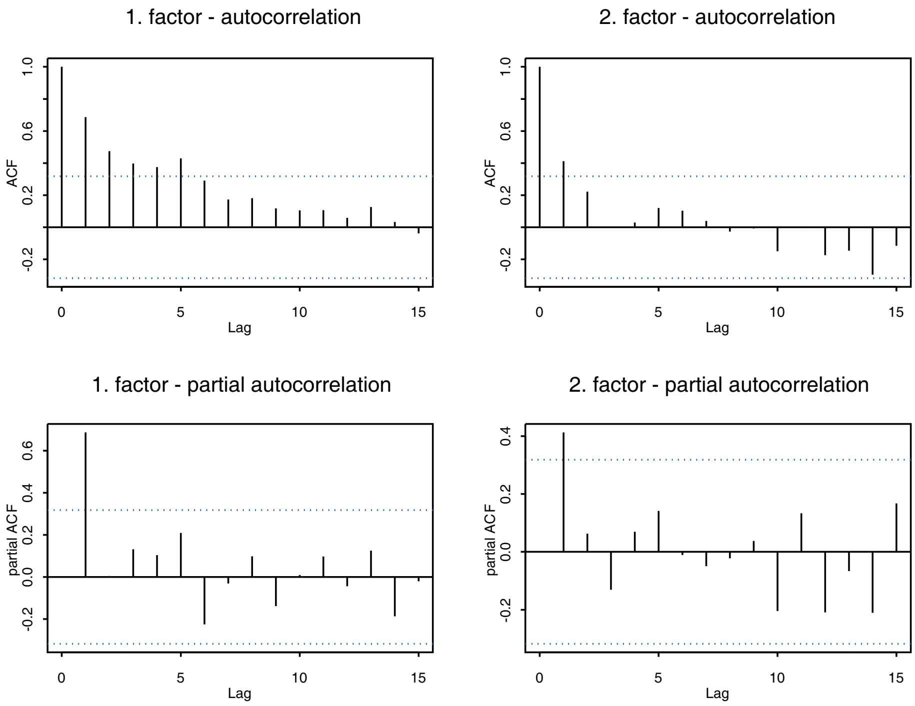


Figure 9.3: The sample autocorrelation and the sample partial-autocorrelation of the two factors. The interval bounded by dotted line is the 95%-probability around-zero range of the sample (partial-)autocorrelations for an i.i.d. normally distributed process.

the factors are identical to principal components and therefore uncorrelated. The estimated covariance matrix of the factors is

$$\widehat{\mathrm{COV}}(\widehat{G}_1, \widehat{G}_2) = \begin{pmatrix} 0.16 & -2.7 \cdot 10^{-17} \\ -2.7 \cdot 10^{-17} & 0.011 \end{pmatrix},$$

which indicates that the factors can be considered to be uncorrelated.

The foregoing empirical auto- and cross-correlations suggest that each factor is modeled separately as an AR(1)-process. The parameters of the AR(1)-model were estimated by maximum likelihood [84, Ch. 3].

The estimated mean-reverting parameters were $\widehat{\phi}_1 = 0.80 \pm 0.10$, $\widehat{\phi}_2 = 0.45 \pm 0.15$ with estimated residual variances $\widehat{\sigma}_1^2 = 0.073$, $\widehat{\sigma}_2^2 = 0.0092$. The residuals had not any significant autocorrelation (5% level), and the Ljung-Box test [84, Ch. 3] did not reject the null of white noise (5% level).

Note. The chosen AR(1)-models of the factors have no drift term or seasonality, whereas in historical data, yearly seasonality exists [15, 61]. In the case study, the optimization model has a time horizon not longer than four months. So we can focus on a single season only. The exclusion of a drift is based on the poor data base: We have to estimate a monthly factor model based on a historical data of length of three years only, which makes it difficult to estimate drifts. In addition, it can be doubted if a historically estimated drift can be used for the future drift. Another reason of the exclusion of drifts and seasonality is the new approach by means of occupation times, which demands to keep the model simple. Nevertheless, it is possible to incorporate seasonality and drift, and we will tackle this in future versions of this work.

9.1.3 The Water Inflow

The available historical data of water inflow is a daily time series over a single hydrological year (starting in October). The data was provided by an industrial partner in an already aggregated form: The originally available daily value is averaged over the last five years. The available data set is additionally aggregated to yield a series of monthly inflows: $(\bar{I}_t)_{t=1,2...}$. Because the original data is already highly aggregated, the simple model (8.7) (p. 76) is considered to be appropriate for a monthly modeling. The estimated standard deviation of the monthly inflow is $\widehat{\sigma}_0 = 1.96\,\mathrm{GWh}$, which is used for the i.i.d. innovations ε_{0t}, $t = 1, 2 \dots$ of the model.

The yearly inflow is $\sum_{t=1}^{12} \bar{I}_t = 54\,\mathrm{GWh}$, which is the same order of magnitude as the storage capacity of the reservoir (cf. Table 9.3).

9.1.4 Notation of Scenario Tree Topology

In the foregoing sections, we discussed the parameter-estimation for the factor model of occupation times and for the model of water inflow. Let us recall that the occupation times and the inflow constitute the exogenous variables, which generate the scenario tree. In particular, the scenario tree is generated by the inflow and by the two factors of the occupation times (Sec. 8.4, p. 78). Hence, there are *three exogenous values in each node of the tree.* For simplicity, in the case study, the granularity of the discretization of the factors and of the inflow is constant over time. In other words, the number of immediate successor nodes of each non-terminal node is chosen to be the same[1]. Thus, the notation of the scenario tree can be kept simple. The following notation is compatible with that of Dupačová et al. [26].

Let a and b be the number of values of the discretized innovation (conditional on every non-terminal node) of the first and second factor of occupation times, respectively, and let c be the respective number for the water inflow. Then the tree is denoted by

$$(a \cdot b \cdot c)^T,$$

where T is the number of stages. Hence, there are $a \cdot b \cdot c$ immediate successor nodes for each non-terminal node, and the tree has $((a \cdot b \cdot c)^{T+1} - 1)/(a \cdot b \cdot c - 1)$ nodes. Dupačová et al. [26] use a more compact notation: The product $a \cdot b \cdot c$ is written as a single number.

9.2 The Parameters of the Electricity Plant

In the foregoing section, we were concerned with the estimation of the parameters of the scenario tree. The optimization model of the electricity plant (SLP) (p. 68) has several additional parameters. The chosen parameter values are given in Table 9.3. In the following, these choices are discussed.

The maximal power of production and pumping, the efficiency of pumping, and the bounds on the water level are chosen from a mid-sized Swiss power plant.

The default value of the lower bound of the water level at final time l_T is set to the overall lower level $l_{\min}$. Hence, the reservoir can be emptied down to the minimum level $l_{\min}$. Thus, in the default setting, the final filling degree is triggered not by l_T but by the perceived price of water v_T at final time T.

The perceived price of water and the parameters of the constraint on risk are chosen with help of computational experiments by solving the model over a range of values as follows.

Let us recall that the value of the plant is that perceived by a stakeholder who supervises the operation of the plant, and the value is both retrospective, counting the occurred cash flows, and prospective, valuing the future usable water (see (7.3), p. 59). For simplicity, the stakeholder is assumed to value the future water by a fixed price, which does not change over the optimization period: $v_t = v = const.$

[1] An alternative is to equip the nodes far in the future with less successor nodes. The reasoning is that the first stage is considered as the most important one.

Table 9.3: Parameters of the electricity plant (default values)

category	parameter		value
technology	maximal power of production	$u^+_{\max}$	60 MW
	maximal power of pumping	$u^-_{\max}$	16 MW
	efficiency of pumping	c	70 %
	upper water level	$l_{\max}$	41 GWh
	lower water level	$l_{\min}$	10 GWh
	initial water level	l_0	40 GWh
decision maker	perceived price of future water	$v_t \ \forall t$	55 Euro/MWh
	level of recursive risk	α	25 %
	lower water level at final time	l_T	10 GWh
market	price of futures	$f_t \ \forall t$	28 Euro/MWh

The value of future water is chosen sufficiently high such that the reservoir is not emptied down to $l_{\min}$ at final time in every scenario (see Figure 9.4 for a typical dependence).

The constraint on risk in model (SLP) (p. 68) has two parameters: The right-hand-side $\rho_{\min}$ and the level α of the local-CVaR set of probability measures. An increase of α relaxes the risk constraint, as does a decrease of $\rho_{\min}$. If the risk-adjusted value would be an ordinary single-period CVaR, then an α-level of 0.05 or 0.01 is usually used in practice [80]. In our case, the risk is measured by a recursive risk-adjusted value for the process of values over time, and the set of test-probability measures is a local-CVaR set (Def. 11, p. 28). The simplified version of this risk-adjusted value is defined for a single random final value, for which the lower bound (4.14) (p. 35) can be given. Based on the lower bound, a judicious choice is that α^T is in the range of ordinary single-period CVaR-levels: For example $(0.25)^3 = 0.02$ or $(0.25)^4 = 0.004$, where $T = 3, 4$ are typical final times of moderately sized models.

In a test for a 4-period model, the level α was varied over three values ($0.25 = \sqrt[4]{0.004}$, $0.32 = \sqrt[4]{0.01}$, $0.47 = \sqrt[4]{0.05}$), and the optimal objective value in dependence of $\rho_{\min}$ was calculated. The results do not qualitatively differ for different α (Fig. 9.5); this was already observed for single-period models with a constraint on risk measured by single-period CVaR [57]. Hence, the severity of the constraint on risk will be investigated by changing the lower bound $\rho_{\min}$; the level α is held fixed. Table 9.4 shows a backtest: We are interested in the specific level of CVaR that makes the value of CVaR$[V_T]$ equal to the value of the recursive risk-adjusted value $R_0^{(V_1,\dots,V_T)}$. The results indicate that the CVaR-levels are in the order of magnitude comparable to α^T.

Note. The (recursive) calculation of CVaR takes only values lower or equal to the α-quantile of the distribution into account. As a function of α, the α-quantile of a

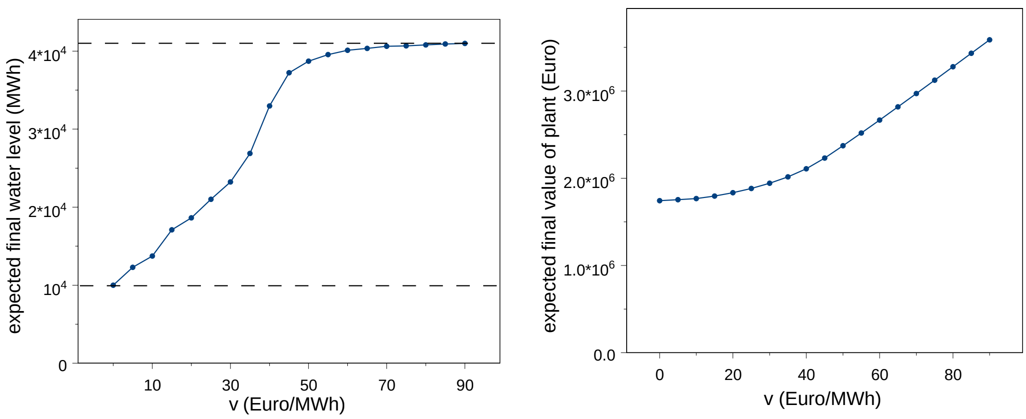


Figure 9.4: Influence of the value of future water ($v = v_t = const.$) on the final water level (left) and on the final value of plant (right). On the left: The physical lower and upper water levels are indicated by dashed lines. – Parameters: Tree topology $(4 \cdot 1 \cdot 2)^4$, $\rho_{\min} = 10^5$ Euro, no futures.

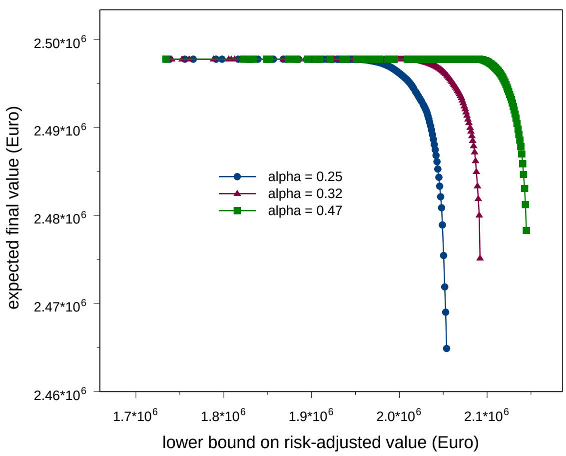


Figure 9.5: The risk-mean diagram for different α-levels. Vertical axis: expected final value of the plant, $\mathbb{E}[V_T]$. Horizontal axis: lower bound $\rho_{\min}$ on risk-adjusted value. The right end-points of the curves are the last feasible solutions that were numerically obtainable; to the left, the curves continue horizontally. Every data point is a change of optimal basis. – Parameters: Tree topology $(4 \cdot 1 \cdot 2)^4$, no futures.

Table 9.4: The CVaR-levels in comparison to multi-period risk levels α. In particular, the entries are $\max\{\beta \in (0,1) \mid \mathrm{CVaR}^{\beta}[V_T] \leq \rho_{\min}\}$, where V_T is the value of the plant at final time T, $\rho_{\min}$ is the lower bound of the recursive constraint on risk, and CVaR^{β} denotes single-period CVaR at level β. The levels are calculated for different multi-period risk parameters α and $\rho_{\min}$. – Parameters as in Fig. 9.5.

α	$\rho_{\min}$ (Euro)			
	$2.04 \cdot 10^6$	$2.05 \cdot 10^6$	$2.08 \cdot 10^6$	$2.15 \cdot 10^6$
0.25	0.024	0.025	-	-
0.32	0.039	0.041	0.042	-
0.47	0.040	0.044	0.061	0.090

finitely discrete distribution is constant for sufficiently low values of α. Hence, in our case of a finite scenario tree, a decrease of α below a certain threshold does no longer tighten the constraint on risk. Thus, especially for sparse scenario trees, α has to be chosen sufficiently large.

Note. In most of the considered scenario and parameter settings, the value of the plant is positive (cf. the discussion after Def. 1, p. 5). Hence, the risk-adjusted value ($= \rho_{\min}$, if the constraint on risk is binding) is positive, too. If additional deterministic costs are taken into account, the value is shifted downwards. The translation invariance of recursive risk-adjusted values ensures that the same shift can be applied to $\rho_{\min}$ (property (iv), p. 18), yielding a negative risk-adjusted value.

The *extended* dispatch problem of the electricity plant allows the additional flexibility to enter positions in futures contracts (Sec. 7.4, p. 70). For each time interval, a different, but fixed position can be chosen. As discussed, the futures are used to hedge the production of the plant over the time horizon, and not for short-term trading. This justifies the assumption that the positions are not dynamically adapted and hold fixed until maturity. For simplicity, the price of the futures is chosen to be the same for all maturities, and is approximately equal to the averaged historical spot price (cf. [23, Sec. 2.2.3.2] for a similar *risk-neutral* assumption).

Because the time horizon is less than a year, the discount rate r is of minor importance and set to zero.

9.3 Implementational Setup

The model (SLP) (p. 68) with the extended profit-and-loss of futures (7.22) was implemented in the modeling language of the software `GAMS` 22.0 [14]. The linear-program solver in `GAMS` was the dual simplex solver `CPLEX` 9.1. The scenario tree generation method was implemented mainly in the language `S` of the statistical software `S-PLUS` 6.2 [7]. The time-critical part of the scenario tree generation was implemented in the language `C` (compiled with Microsoft VisualStudio) [52], and the statistical estimation part used the `S-PLUS` module `S+FinMetrics` [84]. The hardware was an AMD Athlon 64 2 GHz processor with 1 GByte memory, and the operating system was Windows XP.

As an example for the model size, let the tree topology be $(5 \cdot 2 \cdot 6)^3$; therefore, the tree has 219'661 nodes. The model in `GAMS` used 714 MByte of memory, and the matrix had 1'109'287 rows, 1'197'153 columns, with 14'815'953 non-zeroes. The presolver of `CPLEX` reduced this sparse matrix to 457'623 rows, 545'490 columns, with 13'296'627 nonzeros. `GAMS` built the model in 103 seconds, and `CPLEX` solved the model in 3.8 hours by executing 508'852 pivot steps.

9.4 Variation of the Scenario Tree Size

In this section, for a fixed number of time steps, we test the sensitivity of the optimal objective value with respect to variations in the size of the scenario tree, that is, with respect to the granularity of the discretization of the exogenous variables. If the discretization becomes sufficiently fine, then the scenario tree generation method should stabilize itself.

In the test, the tree has three time periods: $T = 3$ (though, the same qualitative behavior was observed for less or more periods). We test the discetization for each exogenous variable separately; a straightforward combined increase surpasses numerical tractability.

First, the discretization of the first factor is tested. The test starts with tree topology $(2 \cdot 1 \cdot 1)^3$, which defines a binary tree where the first factor has a conditional two-point distribution, and where the second factor and the water inflow are deterministic. The tree is made finer until the first factor has fourteen different conditional values: $(14 \cdot 1 \cdot 1)^3$. The result suggests that a five-point distribution of the first factor may be considered to be sufficient (Fig. 9.6).

Second, the discretization of the second factor is varied. The test goes from tree topology $(5 \cdot 1 \cdot 1)^3$ to $(5 \cdot 7 \cdot 1)^3$. The result indicates that the second factor has a minor influence on the optimal objective value (Fig. 9.6): The transition from a deterministic one- to a two-point distribution reduces the optimal objective value by 1.4 %. A further increase in the number of discretizations leaves the objective value almost unchanged. This is in accordance to the fact that the second factor contributes much less to the variance of the occupation times in comparison to the first factor (Fig. 9.1, p. 83).

Third, the discretization of the water inflow is tested, where the tree topology goes from $(5 \cdot 2 \cdot 1)^3$ to $(5 \cdot 2 \cdot 6)^3$. The result indicates that the influence is small (Fig. 9.6): The transition from a deterministic one-point distribution to a two-point

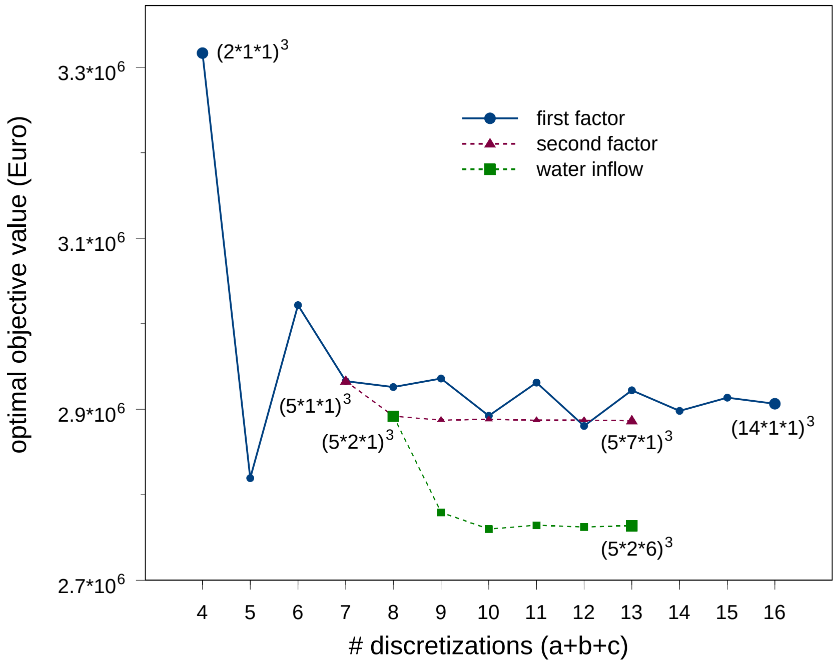


Figure 9.6: The influence of the discretization of exogenous variables (first factor, second factor, water inflow) on the optimal objective value. The tree topology is varied as follows. First factor: $(2 \cdot 1 \cdot 1)^3$ to $(14 \cdot 1 \cdot 1)^3$; second factor: $(5 \cdot 1 \cdot 1)^3$ to $(5 \cdot 7 \cdot 1)^3$; and water inflow: $(5 \cdot 2 \cdot 1)^3$ to $(5 \cdot 2 \cdot 6)^3$. – Other parameters: $\rho_{\min} = 1.6 \cdot 10^6$.

distribution reduces the optimal objective value by 3.7 %. A further increase in the number of discretizations leaves the objective value almost unchanged. This is in accordance with the fact that the reservoir serves as a buffer for the water inflow; thus the influence of higher-moment inflow-variations on the operability of the electricity plant are small. Hence, a two-point distribution seems to be sufficient, too.

A very similar behavior was observed for four stages. Due to limited computational resources, there are less tested tree-topologies. E.g., with the same lower bound $\rho_{\min}$ as in Fig. 9.6 we have

topology	optimal objective value (Euro)	change
$(2,1,1)^4$	3604226	
$(4,1,2)^4$	3071034	-15% to $(2,1,1)^4$
$(5,1,2)^4$	3002917	-2.2% to $(4,1,2)^4$
$(5,1,3)^4$	2977544	-0.8% to $(5,1,2)^4$
$(4,1,4)^4$	3048540	1.5% to $(4,1,2)^4$
$(4,2,1)^4$	3200062	-11% to $(2,1,1)^4$
$(4,4,1)^4$	3192789	-0.2% to $(4,2,1)^4$
$(5,2,1)^4$	3111932	-2.8% to $(4,2,1)^4$

Hence, by these tests for three and four stages, we may conclude that the first factor requires more than a two-point distribution, the second factor has marginal influence, and the water inflow requires at least a two-point distribution. In the following, due to an acceptable computing time, we use four time periods with tree topologies $(4 \cdot 1 \cdot 2)^4$ and $(4 \cdot 2 \cdot 2)^4$.

9.5 The Constraint on Risk

Let us recall that the constraint on risk in the model is given by a lower bound of the recursive risk-adjusted value: $R_0^{(V_0,\ldots,V_T)} \geq \rho_{\min}$, where the whole stochastic process $(V_t)_{t=0,\ldots,T}$ of the value of the plant is taken into account (Def. 4, p. 16). In the node-wise setting of the optimization problem (SLP) (p. 68), this constraint reads $R_{0n_0} \geq \rho_{\min}$, where R_{0n_0} is the corresponding variable for the root node n_0. If the constraint is binding in the optimum, then the optimal $R^*_{0n_0}$ is the risk-adjusted value. Generally, the feasible variables $(R_{tn})_{t=0,\ldots,T,\, n\in\mathcal{N}_t}$ are an upper bound of the risk-adjusted value process (Lemma 5, p. 30), and a lower bound of the value: $R_{tn} \leq V_{tn}$ for all t, $n \in \mathcal{N}_t$ (see (SLP)). The latter constraint can be binding, but the marginal (dual variable) of the constraint can still be zero; for zero marginals, a small change of V_{tn} may not influence the optimal solution of the model. An example over four time intervals ($T = 4$) is shown in Figure 9.7. In the example, at the root node and at $t = 1$, the inequality is strict in every node n: $R^*_{0n} < V^*_{0n}$.

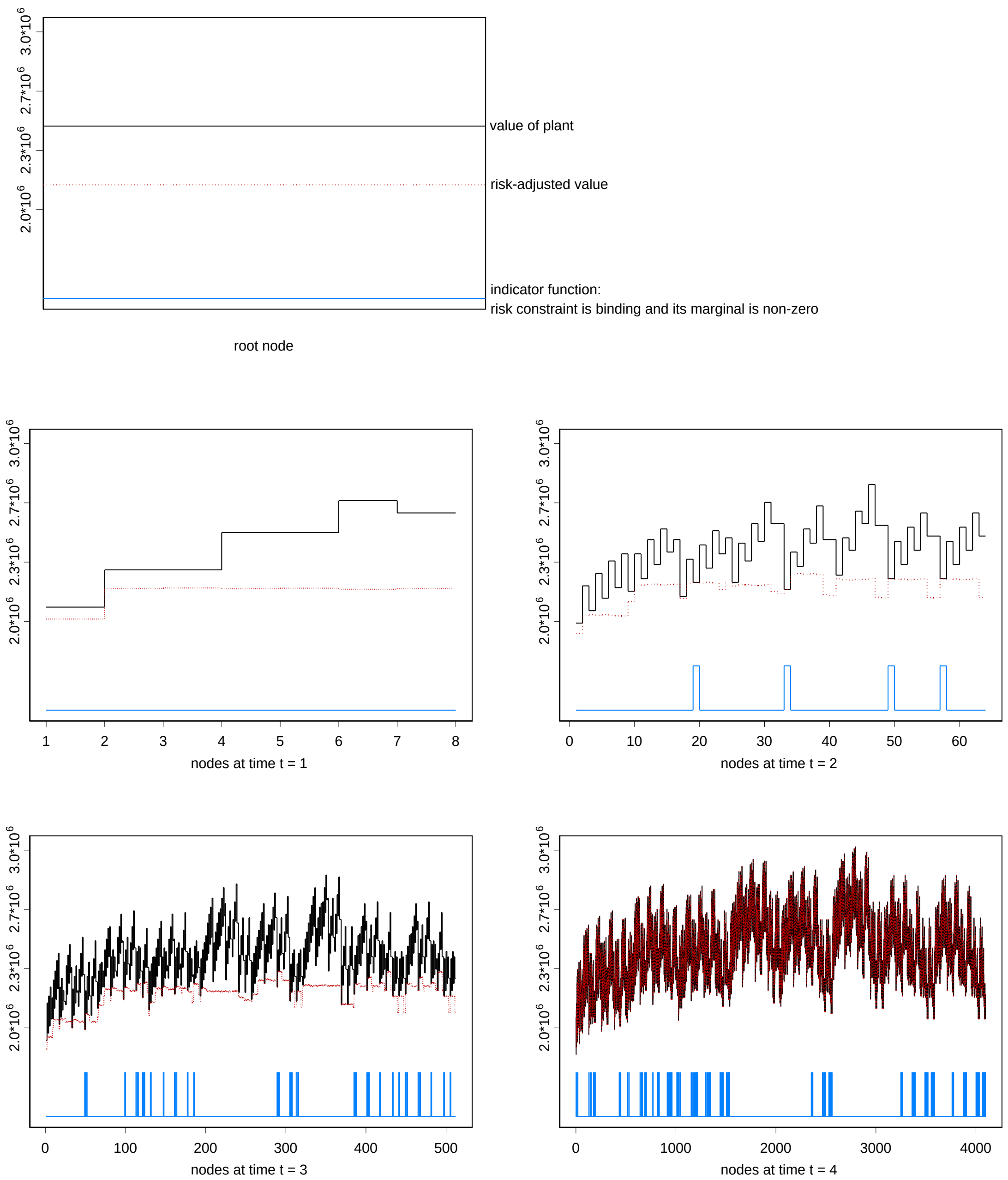


Figure 9.7: The plots correspond to the time steps $t = 0, \ldots, 4$ of the scenario tree. Horizontal axis: The nodes at time t. Every non-terminal node has 8 immediate successor nodes. For better visibility the point-wise values associated to the nodes are extended to staircase-lines. Vertical axis: The optimal value of the plant V^*_{tn} (Euro) and the optimal upper bound R^*_{tn} (Euro) for the risk-adjusted value process which coincides for $t = 0$ with the risk-adjusted value itself. The constraint $R^*_{tn} \leq V^*_{tn}$ has to hold in every node. An indicator function selects those nodes n for which this constraint has a non-zero marginal value. – Parameters: Tree topology: $(4 \cdot 1 \cdot 2)^4$, $\rho_{\min} = 2.14 \cdot 10^6$ Euro.

At $t = 2$, the inequality is binding in four different nodes (e.g. in the 19th node). From the figure, it is apparent that the constraint is binding in the nodes n at $t = 2$ where V_{2n} is small.

At final time $t = 4$, the inequality is always binding, which means that all uncertainty is resolved: The risk-adjusted value equals the actual value of the plant in all nodes. Nevertheless, not all terminal nodes have non-zero marginal.

For a further investigation, we recall from Artzner et al. [5] that the dual formulation of the recursive calculation (3.3), (p. 17) is

$$R_0^{(V)} = \min_{\tau \in \mathcal{T}, \mathbb{Q} \in \mathcal{P}} \mathbb{E}_{\mathbb{Q}}[V_\tau], \tag{9.1}$$

where $\mathcal{T}$ is the set of $(\mathcal{F}_t)$-stopping times with values in $\{0, \ldots, T\}$. Let $(\tau^*, \mathbb{Q}^*)$ be an optimal solution. If we restrict to optimal $\mathbb{Q}^*$ we just have

$$R_0^{(V)} = \min_{\tau \in \mathcal{T}} \mathbb{E}_{\mathbb{Q}^*}[V_\tau], \tag{9.2}$$

which is a classical optimal stopping problem. There is a smallest and largest optimal solution. The smallest solution is

$$\tau^*_{\min} = \min \left\{ t \in \{0, \ldots, T\} \mid R_t^{(V)} = V_t \right\}. \tag{9.3}$$

We have to keep in mind that our particular setup is slightly different, because the recursive calculation (3.3) was transformed into a linear optimization problem. Investigation of the numerical optimal solution showed that for a node n and its successor node m it can happen that $R^*_{tn} = V^*_{tn}$ as well as $R^*_{(t+1)m} = V^*_{(t+1)m}$, where both equalities have non-zero marginals. In fact only the first occurrence (from $t = 0$ onwards) of equality is relevant, because the conditions further in the future are not taken into account for the recursive calculation of $R^*_{0n_0}$.

Next, the constraint on risk is varied in different ways.

9.5.1 Optimal Objective Value in Dependence of $\rho_{\min}$

In this test, we vary the lower bound $\rho_{\min}$ of the constraint on risk. Because $\rho_{\min}$ is part of the right-hand-side of the linear maximization problem (SLP) (p. 68), the risk-mean frontier is a piece-wise affine-linear and concave function.

The impact on the optimal objective value is calculated in four cases: With futures, without futures, with stochastic inflow, and with deterministic inflow.

The tree topology was $(4 \cdot 1 \cdot 1)^4$ for deterministic inflow and $(4 \cdot 1 \cdot 2)^4$ for stochastic inflow; with different topologies we got qualitatively similar results. The findings concerning the inflow are as expected (Fig. 9.8): The uncertainty of the water inflow shifts the risk-mean frontier downward.

Future contracts are incorporated into the model according to Sec. 7.4 (p. 70): For every month of the planning horizon, we can enter a different position in futures.

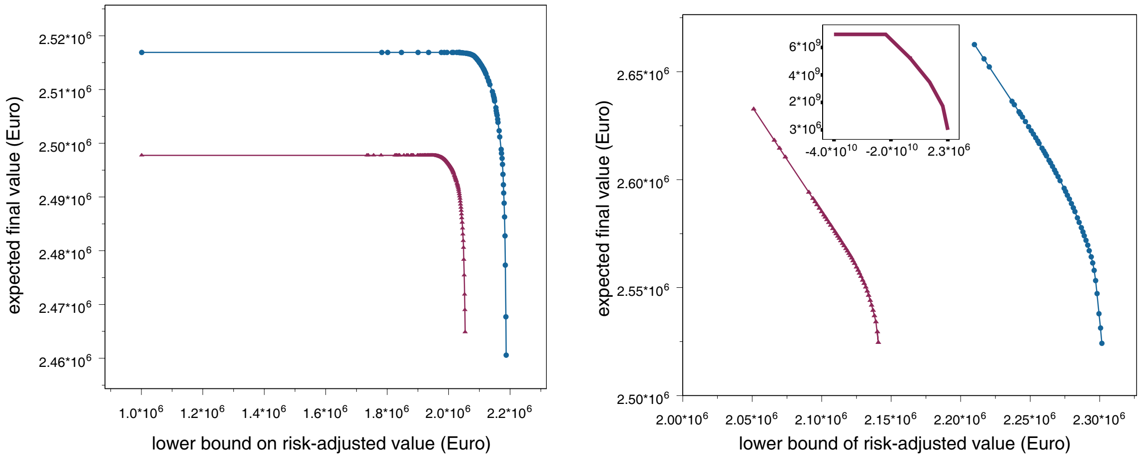


Figure 9.8: The risk-mean diagram. Vertical axis: $\mathbb{E}[V_T]$ (expected final value of the plant); horizontal axis: $\rho_{\min}$ (lower bound on risk-adjusted value). On the left: Investment in futures is prohibited; on the right: Fixed positions in monthly futures are allowed. Upper-right curves: deterministic inflow; lower-left curves: stochastic inflow. The right end-point of the curves are the last feasible solutions that were numerically obtainable. – Tree topology: deterministic inflow $(4 \cdot 1 \cdot 1)^4$, stochastic inflow $(4 \cdot 1 \cdot 2)^4$.

For simplicity, the positions are not changed over time, and no self-financing portfolio process is involved (nevertheless we could make the value process self-financing by introducing a saving account). According to (7.22), p. 71, we assume that the profit-and-loss of the future contributes to the value of the plant only in the delivery month (we assume zero interest rates); in reality, the profit-and-loss is realized daily by so called margin-calls. If the additional flexibility of futures contracts is not allowed, then the constraint on risk is active only in a relatively small interval of the lower bound $\rho_{\min}$ (Fig. 9.8); the optimization problem becomes infeasible by a further increase of $\rho_{\min}$. Hence, the model cannot respond well to different bounds on risk. This behavior does not change if the stochastic water inflow is altered to a deterministic one.

Let us relate this behavior to traditional portfolio optimization in finance. In a typical portfolio optimization problem, an investor has the choice between risky assets that have high expected returns and relatively secure assets that have low expected returns. The risk can be reduced by investing more into the secure assets. By contrast, the dispatch problem of the electricity plant considers only the single asset of electricity. If the electricity plant is restricted to sell or buy on the spot market only, a reduction of risk is limited. The reduction can be larger if the decision maker has the additional flexibility to open positions in futures contracts (Fig. 9.8). The subsequent analysis (if not stated otherwise) includes always futures contracts.

Table 9.5: Optimal first-stage solution for different lower bounds $\rho_{\min}$ of the risk-adjusted value. The optimal dispatch decisions for the thirteen price levels are $X_{i0}^{\pm}$, $i = 1, \ldots, 13$ (+: producing, −: pumping, missing values are zero). The optimal fixed future positions for the four monthly time intervals are p_1, p_2, p_3, p_4 (MW). Sign convention: p_i positive means 'short'-position, i.e. we earn money if the underlying electricity prices are falling. – Tree topology $(4 \cdot 1 \cdot 2)^4$.

solution	$\rho_{\min}$ (Euro)			
	$2.1407 \cdot 10^6$	$2.13 \cdot 10^6$	$2.0 \cdot 10^6$	$1.0 \cdot 10^6$
$X_{1\,0}^+, X_{1\,0}^-$	,	,	,	,
$X_{2\,0}^+, X_{2\,0}^-$	,	,	,	,
$X_{3\,0}^+, X_{3\,0}^-$	,	,	,	,
$X_{4\,0}^+, X_{4\,0}^-$	, 0.882	,	,	,
$X_{5\,0}^+, X_{5\,0}^-$	, 0.118	, 0.805	,	,
$X_{6\,0}^+, X_{6\,0}^-$	,	, 0.195	, 1	, 1
$X_{7\,0}^+, X_{7\,0}^-$	,	,	,	,
$X_{8\,0}^+, X_{8\,0}^-$	,	,	,	,
$X_{9\,0}^+, X_{9\,0}^-$	,	,	,	,
$X_{10\,0}^+, X_{10\,0}^-$	0.195,	0.799,	1,	1,
$X_{11\,0}^+, X_{11\,0}^-$	0.805,	0.201,	,	,
$X_{12\,0}^+, X_{12\,0}^-$	,	,	,	,
$X_{13\,0}^+, X_{13\,0}^-$	,	,	,	,
p_1	$3.531 \cdot 10^4$	$4.213 \cdot 10^4$	$1.161 \cdot 10^5$	$6.753 \cdot 10^5$
p_2	$5.144 \cdot 10^3$	$2.464 \cdot 10^3$	$1.827 \cdot 10^3$	$1.827 \cdot 10^3$
p_3	$-5.008 \cdot 10^2$	$-2.974 \cdot 10^3$	$-2.905 \cdot 10^3$	$-2.905 \cdot 10^3$
p_4	$-2.860 \cdot 10^3$	$-2.860 \cdot 10^3$	$-2.860 \cdot 10^3$	$-2.860 \cdot 10^3$

9.5.2 Optimal Solution in Dependence of $\rho_{\min}$

An example of an optimal solution in dependence of the right-hand-side of the constraint on risk ($\rho_{\min}$) is shown in Table 9.5. In the example, the largest right-hand-side is chosen to be tight ($\rho_{\min} = 2.141 \cdot 10^6$ is already infeasible). The dispatch decision is depicted only for time $t = 0$ (*first-stage decision*). The result indicates that the dispatch decision changes not drastically by varying $\rho_{\min}$. If the right-hand-side is sufficiently lowered, the dispatch decision becomes constant, and a further change in the value distribution of the plant is generated only by the futures position. These findings are in agreement with the aforementioned risk-mean frontiers. Similar results were obtained for different tree topologies.

In addition, in all numerically solved model instances, the optimal dispatch decision exhibited a bang-bang behavior (cf. Table 9.5): There exists a threshold level for production and a threshold level for pumping such that the plant produces (pumps)

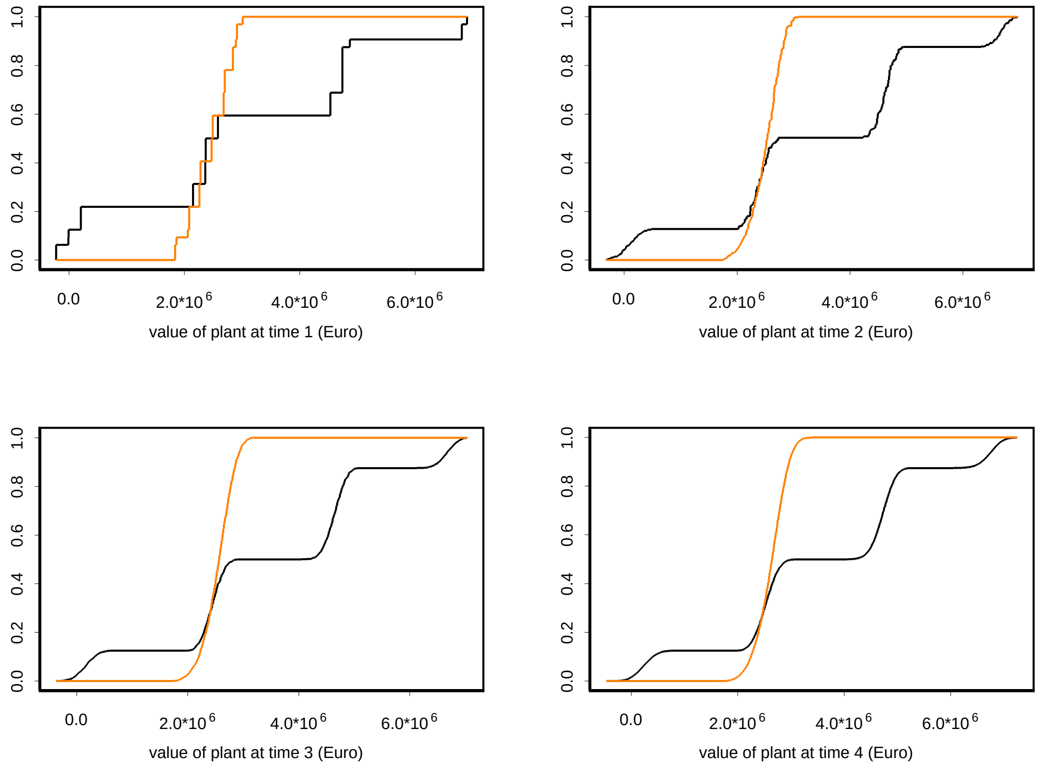


Figure 9.9: The optimal value of the plant for different lower bounds (ρ_{min}) of the risk-adjusted value. The value's distribution function is plotted for each time stage. Gray curve: Tight lower bound ($2 \cdot 10^6$ Euro). Black curve: Loose constraint (10^6 Euro). – Tree topology $(4 \cdot 2 \cdot 2)^4$.

at full capacity if the spot price is larger (smaller) than the respective level. Because $X_{it}^{\pm}$ is the additional fractions of production (pumping) capacity at price level s_i, $X_{it}^{\pm}$ is observed to be equal either to zero or one, or, the sum of adjacent values equals one ('the real threshold is between two adjacent s_is'). We have not yet a proof for that phenomenon, as we were able to give for the problems in Chapter 6.

9.5.3 The Value over Time in Dependence of ρ_{min}

The distribution of the value of the plant *over time* in dependence on the constraint on risk is considered. A tight and a loose lower bound (ρ_{min}) is tested. The result is as follows (Fig. 9.9). If the bound is tight, then the values are concentrated in a relatively small range. If the bound is loose, then the values of the plant can spread widely; even negative values occur. The staircase curve for the loose constraint can be attributed to the position in futures in the first period, which was observed to

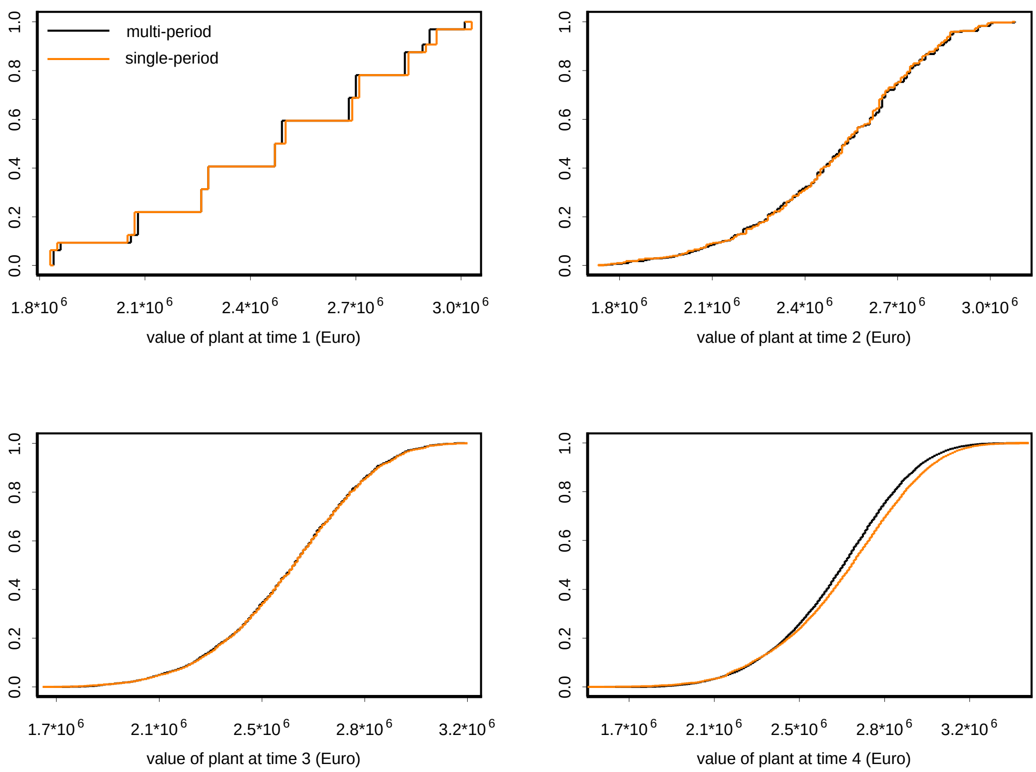


Figure 9.10: The value of the plant for different constraints on risk. The value's distribution function is plotted for each time stage. The two different constraints on risk are the recursive multi-period risk-adjusted value and the single-period CVaR. Optimal objective value ($\mathbb{E}[V_T]$) for CVaR: $2.65 \cdot 10^6$ Euro, and for the multi-period risk: $2.62 \cdot 10^6$ Euro. – Parameters: CVaR-level $\alpha = 0.05$; lower bound of risk-adjusted value $\rho_{\min} = 2 \cdot 10^6$; tree topology $(4 \cdot 2 \cdot 2)^4$.

have a six times larger amount than for the tight constraint.

9.5.4 Comparison with a Single-Period Risk-Adjusted Value

In this test, the multi-period recursive risk-adjusted value is replaced by the single-period risk-adjusted value CVaR (Def. 2, p. 9):

$$R_0^{(V_0,\ldots,V_T)} \geq \rho_{\min} \quad \rightarrow \quad \mathrm{CVaR}[V_T] \geq \rho_{\min}.$$

An example of the corresponding distributions of values of the plant over time is shown in Figure 9.10. The example illustrates that a CVaR-constraint of the final value forces implicitly the intermediate values to have small variations, too;

qualitatively similar results were observed for different tree topologies. Whether this observation depends on the particular parameter setting of the electricity plant or constitutes a general behavior in multi-period models is subject to future research. Nevertheless, the properties of a recursive multi-period risk-adjusted value ensure in every possible problem-instance that the risk of intermediate values is actually bounded.

The aforementioned test contrasts to Eichhorn and Römisch [29]. In their work, the CVaR-constraint of the final value does not prevent large variations of intermediate values. Hence, without considering multi-period risk, their model is inherently different to ours. Consequently, they observe significant differences in intermediate values by applying multi-period risk with respect to CVaR. Their multi-period risk measure seems not to be time-consistent. The scenario tree in their case study has an extraordinary topology: The tree has 8760 time-steps and 40 scenarios (for comparison, a binary scenario tree would have 2^{8760} scenarios). As a remark, they consider a weighted sum of mean and risk in the objective; it can be shown that this approach is equivalent to the mean in the objective and the risk bounded by a constraint.

9.6 Expected-Value-of-Perfect-Information

The decisions on the scenario tree are non-anticipative, that is, the decision at time t can depend only on the information up to time t. This was reflected notationally in the decision variables (7.20) (p. 67):

$$X_{it}^{\pm} = x_{it}^{\pm}(\boldsymbol{E}_0, \ldots, \boldsymbol{E}_t), \qquad i = 1, \ldots, 13,\ t = 0, \ldots, T-1, \tag{9.4}$$

where the decision variables at time t depend only on the past and present exogenous variables: $\boldsymbol{E}_0, \ldots, \boldsymbol{E}_t$, and we recall that the dispatch table $X_{it}^{\pm}$, $i = 1, \ldots, 13$, denotes the fraction of production/pumping at the different electricity spot price levels s_i. The term *here-and-now* solution is sometimes used for such non-anticipative optimal solutions of optimization problems (see e.g. [50, p. 138]). Taking into account the filtration that is generated by the exogenous variables, we can equivalently say that the process $(X_{it}^{\pm})_{t=0,\ldots,T-1}$ is adapted to the filtration for all i.

If the decision maker would know the future evolution of the exogenous variables, then the non-anticipativity of the decisions could be relaxed:

Definition 13 (Anticipative decisions). The decisions are anticipative if they can depend on past *and future* exogenous variables $\boldsymbol{E}_0, \ldots, \boldsymbol{E}_T$:

$$x_{it}^{\pm}(\boldsymbol{E}_0, \ldots, \boldsymbol{E}_T) \qquad \text{for all } i, t. \tag{9.5}$$

The relaxation of non-anticipativity says that all decision variables (even those at intermediate times) can take values for each scenario separately. This implies by the state equations that the state variables P_t, L_t and V_t take different values for each scenario, too. In other words, using the fact that the scenario tree corresponds to a finite filtration $\mathcal{F}_t = \sigma(\boldsymbol{E}_0, \ldots, \boldsymbol{E}_t)$, $t = 0, \ldots, T$, the anticipativity says that the decisions $X_{it}^{\pm}$ and the state variables P_t, L_t and V_t are assumed to be just $\mathcal{F}_T$-measurable for all t, and no longer $\mathcal{F}_t$-measurable.

The optimal solution of the anticipative problem is called a *wait-and-see* solution [50, p. 8]. Because non-anticipativity is relaxed, the objective value of the wait-and-see solution (WS) is always larger (maximization problem) than that of the here-and-now solution (HN). The difference is the so-called *expected-value-of-perfect-information* (EVPI) [50, p. 140]:

$$\text{EVPI} := \text{WS} - \text{HN} \geq 0.$$

The EVPI can be interpreted to be the maximum amount a decision maker would be ready to pay in return for complete (and accurate) information about the future (see e.g. [11, p. 137]). In the case of the model of the electricity plant, the information about the future consists in knowing the future values of the water inflow and of the occupation times of spot prices.

Next, we calculate the EVPI for the optimization problem of the electricity plant. We shall emphasize that the same scenario tree (of exogenous variables) is used in the HN-problem as in the WS-problem, as well as are all parameters; the only difference is that in the WS-problem the variables take independently values for each scenario, whereas in the HN-problem the value of a variable at time t has to be identical in different scenarios whenever the scenarios have a node $n \in \mathcal{N}_t$ of the scenario tree in common.

For the numerical example, we assume that positions in futures are not allowed; if the future is known, futures give an arbitrarily high profit (unless there would be an upper bound on the positions). The absence of futures implies that the model is very insensitive with respect to the constraint on risk (cf. Fig. 9.8). Thus, for simplicity, we consider the model without constraint on risk.

The results are as follows (Table 9.6). The EVPI depends heavily on the lower bound on the final water level: If the reservoir must be as full at final time as at the beginning ($L_T \geq l_0$), then the information about the future is more valuable than if the reservoir can be emptied ($L_T \geq l_{\min}$).

9.7 State-Independent Decisions

The counterpart of the foregoing anticipative decisions (Def. 13) are decisions that do not use any information from the exogenous process $(\boldsymbol{E}_t)_{t=0,\ldots,T}$ at all. Such decisions are identical in all scenarios. Hence, the function in the original non-anticipative dispatch decision (9.4) is replaced by constant values:

Table 9.6: The expected-value-of-perfect-information (EVPI) for different lower bounds of final water level L_T: Initial level l_0 and minimal level $l_{\min}$. The model is without futures contracts and without the constraint on risk. $L_T \geq l_{0/\min}$ means $L_{Tn} \geq l_{0/\min}\ \forall n \in \mathcal{N}_T$. Tree topology: $(4 \cdot 2 \cdot 2)^4$.

	objective value (Euro)		EVPI
water level	here-and-now (HN)	wait-and-see (WS)	(WS − HN)/HN
$L_T \geq l_0$	520'776	937'399	80 %
$L_T \geq l_{\min}$	2'495'064	2'664'400	7 %

Definition 14 (State-independent decisions, [23, p. 57]). The decisions are state-independent if they do not depend on exogenous variables $\boldsymbol{E}_0, \ldots, \boldsymbol{E}_t$:

$$x_{it}^{\pm}(\boldsymbol{E}_0, \ldots, \boldsymbol{E}_t) = x_{it}^{\pm} = \text{const.} \qquad \text{for all } i, t.$$

Note that the dependence on time t and on the exogenous electricity price levels s_i is still given; The dispatch decision at a specific time t is still a dispatch table $x_{it}^{\pm}$, $i = 1, \ldots, 13$, with respect to the thirteen spot prices levels s_i.

If we consider the scenario tree of exogenous variables, a state-independent dispatch decision (dispatch table) at a specific stage is identical for all nodes on that stage.

State-independent decisions (as well as the foregoing wait-and-see decisions) allow to formulate the stochastic optimization model path-wise [23,79]. The path-wise formulation considers just a *fan* of scenarios; a non-trivial scenario tree is not needed. Thus, it is crucial to know whether the original non-anticipative decisions on a scenario tree are better than the state-independent decisions. The non-anticipative decisions are allowed to depend on the history of the exogenous variables; hence they depend implicitly on the state variables (therefore the notion 'state-independent decision' for their counterpart). Thus, non-anticipative decisions are useful if the constraints on the state variables are relatively tight.

In the test, the optimal objective value of the problem that is confined to state-independent decisions is compared with the problem that has the original, non-anticipative decisions. It has to be emphasized that the original optimization problem is altered only in a single aspect: The dispatch table at each time t is no longer allowed to vary over the set of nodes at time t of the scenario tree. The optimal objective value (expected final value of the plant, $\mathbb{E}[V_T]$) is calculated for different bounds on final water level and for different bounds on risk.

The result is as follows (Table 9.7). If the final water level must be large ($L_T \geq l_0$), then an adaptive decision has a considerable advantage, and tighter bounds of risk do still lead to a feasible solution. If the reservoir can be emptied ($L_T \geq l_{\min}$), then the differences are much less pronounced. The test was also executed with different prices of future water giving similar results.

Table 9.7: The optimal objective value for state-independent decisions and for non-anticipative decisions. The lower bounds on the final water level and on the risk-adjusted value are varied. – Tree topology $(4 \cdot 1 \cdot 2)^4$.

constraints		objective value (Euro)	
$L_T \geq$	$R_0 \geq$	state-independent	non-anticipative
$l_{\min}$	0	4'216'743	4'625'948
$l_{\min}$	$1.9 \cdot 10^6$	2'370'202	2'779'487
$l_{\min}$	$2.0 \cdot 10^6$	*infeas*	2682305
$l_{\min}$	$2.1 \cdot 10^6$	*infeas*	2585123
$l_{\min}$	$2.2 \cdot 10^6$	*infeas*	*infeas*
l_0	0	24'030	1'023'124
l_0	$0.1 \cdot 10^5$	14'312	1'013'406
l_0	$0.2 \cdot 10^5$	*infeas*	1'003'688
l_0	$3.6 \cdot 10^5$	*infeas*	670'734
l_0	$3.7 \cdot 10^5$	*infeas*	*infeas*

9.8 Quality-Test of the Scenario Tree Generation Method

The applied quality-test is a heuristic technique to detect rough errors in the scenario tree generation method. The test is based on Kaut and Wallace [51] and compares the original method with a benchmark method.

9.8.1 Definition of Stability and Bias

We need the following definitions. The objective function of the optimization problem to be tested is denoted by $g(\cdot,\cdot)$, where the first argument holds the variables (which have to fulfil certain constraints), and the second argument holds the exogenous values. The optimal objective value of the benchmark-optimization problem is denoted by $g(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{E}})$, where $\widetilde{\boldsymbol{x}}$ denotes the optimal variables, and where $\widetilde{\boldsymbol{E}}$ denotes the exogenous values. The vector $\widetilde{\boldsymbol{E}}$ (which is a notational abbreviation for the stochastic process of exogenous values over time) is assumed to have the exact probability distribution. Let $\boldsymbol{E}^l$, $l = 1, \dots, L$, be approximations of $\widetilde{\boldsymbol{E}}$ of different qualities. Let $g(\boldsymbol{x}^l, \boldsymbol{E}^l)$ be the optimal objective value for the approximation $\boldsymbol{E}^l$ with optimal solution $\boldsymbol{x}^l$. A scenario tree generation method should fulfil the following three requirements [51].

(i) *In-sample stability*: $g(\boldsymbol{x}^l, \boldsymbol{E}^l) \approx g(\boldsymbol{x}^m, \boldsymbol{E}^m), \quad l, m \in \{1, \dots, L\}$. The required approximation quality depends on the particular problem. In words: If scenario trees of different, but sufficient qualities are considered, then the optimal objective value should be (almost) the same.

(ii) *Out-of-sample stability*: $g(\boldsymbol{x}^l, \widetilde{\boldsymbol{E}}) \approx g(\boldsymbol{x}^m, \widetilde{\boldsymbol{E}}), \quad l, m \in \{1, \ldots, L\}$. In words: If the objective value function of the benchmark model is evaluated with the optimal solutions of different approximative models, then the optimal objective value should be (almost) the same.

(iii) *Absence of bias*: $g(\boldsymbol{x}^l, \widetilde{\boldsymbol{E}}) \approx g(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{E}}), \quad l = 1, \ldots, L$. In words: An optimal solution of every approximative model should be an (almost) optimal solution of the benchmark model.

9.8.2 Quality-Test with Monte-Carlo Sampling

The benchmark method is chosen to use Monte-Carlo sampling. Monte-Carlo sampling that creates a scenario tree would usually need conditional sampling. For statistically reasonable sample sizes, the resulting tree would be numerically not tractable. Moreover, the scarcity of available historical data, which is used in the statistical estimation procedures, may prohibit a proper comparison of the original method with the benchmark method. Hence, for simplicity, we assume the following setup:

- The optimization problem is restricted to use state-independent decisions (cf. Sec. 9.7); In this problem formulation, the scenarios of the exogenous variables need not (but can be) in tree-form. Hence, unconditional Monte-Carlo sampling can be used for the benchmark model, resulting in a fan of scenarios.

- The quality-test considers only the generation method for the occupation times, and not for the water inflow (the data and models for inflow are not far developed).

- An hourly time-series model is used to generate random sample-paths of spot prices. Such a (sufficiently long) sample-path is used for the statistical estimation in the original method. The sample-paths also generate sample-paths of monthly occupation times (by counting frequencies) in the benchmark method.

The workflow of the scenario generation methods is shown in Table 9.8. The hourly spot price model is Lucia and Schwartz' model [61]. It is a mean-reverting autoregressive (AR(1)-)model of the logarithm of the spot price where the daily and yearly variations, as well as the different behavior at holidays and week-ends, is taken into account by a deterministic additive function. The model was fitted to hourly data from 1 January 2002 to 10 March 2005 from the EEX [32].

The quality-tests of stability and bias (Sec. 9.8.1) is performed for single and three-period models (Table 9.9). The constraint on risk is neglected in the test; it is

Table 9.8: The workflow of the scenario generation methods.

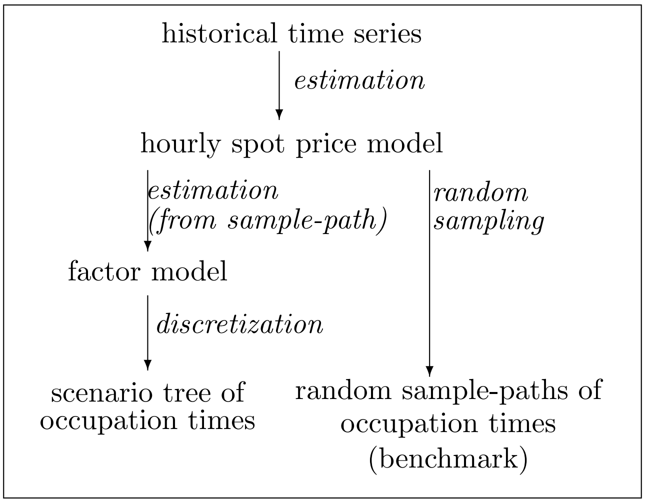


Table 9.9: The optimal objective value for different quality-tests of the scenario-tree generation method. The *in-sample test* calculates an optimal solution $\boldsymbol{x}^l$ with respect to scenario tree $\boldsymbol{E}^l$, $l = 1, \ldots, 5$. The *out-of-sample test* applies the $\boldsymbol{x}^l$s and the scenarios $\widetilde{\boldsymbol{E}}$ from the benchmark model (Monte-Carlo model with 10'000 scenarios). The *bias test* calculates the optimal solution $\widetilde{\boldsymbol{x}}$ of the benchmark model. – Constraint on risk is relaxed.

		quality-tests (deviation from in-sample in %)		
l	tree topology of $\boldsymbol{E}^l$	in-sample $g(\boldsymbol{x}^l, \boldsymbol{E}^l)$	out-of-sample $g(\boldsymbol{x}^l, \widetilde{\boldsymbol{E}})$	bias $g(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{E}})$
1	$(6{\cdot}6{\cdot}1)^1$	864'164	860'283 (0.4%)	871'980 (0.9%)
2	$(12{\cdot}12{\cdot}1)^1$	863'703	865'109 (0.2%)	(1.0%)
3	$(2{\cdot}2{\cdot}1)^3$	2'019'806	2'015'473 (0.2%)	2'029'175 (0.5%)
4	$(4{\cdot}4{\cdot}1)^3$	2'011'987	2'021'970 (0.5%)	(0.9%)
5	$(6{\cdot}6{\cdot}1)^3$	1'981'209	2'011'129 (1.5%)	(2.4%)

assumed that the relaxed model suffices to test the scenario tree generation method. The resulting differences between the original method and the Monte-Carlo method are in the range of several percentages or smaller. Let us recall that the original method uses only the two first principal components of the occupation times, and the used factor model is valid only under several theoretical assumptions. Considering this, the differences are acceptable.

Chapter 10

Conclusion

We were able to build a multi-stage stochastic programming model for the optimization of *high-frequency* dispatch-decisions over a *long-term* horizon, and the operability of the concept could be numerically validated. In addition, for simple dispatch models, exact optimal *bang-bang* solutions are derived, and the optimization problems are related to coherent risk measurement.

The applied *factor model* and *principal component analysis* of occupation times of the electricity spot price allowed to consider a small number of relevant factors, which exhibit a characteristic pattern (Figure 9.2, p. 84). The low number of factors enabled the generation of moderately sized scenario trees. In particular, four-period models with sufficient granularity in the discretization of the exogenous variables could be solved (1 period $\widehat{=}$ 1 month).

The presence of a substantial *expected-value-of-perfect-information* indicates that stochasticity is a vital part in the modeling of electricity plants. The superiority of *non-anticipative* decisions (adapted on a scenario tree) over *state-independent* decisions depends on the constraints: If the constraints on the water level are tight, then the adaptiveness increases the expected value of the plant considerably, and the bound on risk can be tighter (Table 9.7, p. 105).

The constraint on risk is given by a lower bound on a recursive *risk-adjusted value* that has the desirable properties of *multi-periodicity*, *coherency* and *time consistency*. Moreover, it is consistent with single-period coherent risk measurement and fits into the optimization model as a set of linear constraints. In the simple case of the risk of a final outcome, a lower bound of the risk-adjusted value can be given. Although a single-period risk-adjusted value gave similar results in terms of intertemporal values of the plant, and the presented counterexample of time consistency for a non-recursive risk-adjusted value (Fig. 3.2, p. 22, cf. Artzner et al. [5]) may rarely occur in applications, only the applied risk-adjusted value can ensure the aforementioned desirable properties in all problem instances.

Chapter 11

Further Work

11.1 Extensions for Risk Measurement

- The linear optimization problem for the recursive risk-adjusted value can be algebraically dualized, which yields a coherent risk-adjusted value on the product space of state and time. Strong duality in the case of infinite state spaces could be investigated.
- The set of test-probability measures can be related to Lagrange-multipliers of a non-anticipativity constraint [68]. This concept is not yet fully extended to multiple periods. The goal would be a unification of the concepts of expected-value-of-perfect-information (EVPI), coherent risk-adjusted values, optimization of production models, and the concept of flexibility [62].
- Time consistency preserves an order over time, and its relation to stochastic dominance over time may be instructive. In addition, time consistency considers the view of future, better informed observers, similarly to EVPI.

11.2 Extensions for the Power Plant Model

- The presented model is for a single electricity plant; it could be extended to a whole power portfolio. In addition, the model formulation allows in principle the incorporation of demand (see Appendix A.6, p. 118).
- An alternative choice of modeling would be to bound the water level only in expectation [79], which is a less severe restriction leading to a greater flexibility in the dispatch decisions.
- The parametric value of future water, v_t, could be calculated by the linear optimization model of the electricity plant as follows. In the model (SLP) (p. 68), the water level is constrained in each node. The node-wise optimal

dual variable can be interpreted as the marginal price of water in that node. In the linear *dual* optimization problem, the parameter v_t is in the technology matrix, in the right-hand-side and in the objective function. If v_t is interpreted as a variable in the dual problem, we get a quadratic optimization problem in v_t and in the remaining dual variables.

- The characteristic *principle component* structure of the occupation times (Fig. 9.2, p. 84) seems to be valid for general mean-reverting Gaussian Markov processes [22]. In addition, in an alternative study, an hourly model for the *logarithmic* spot price may be considered, which ensures positive model prices by definition.

- The exact solutions of the simple dispatch problems could be extended to distributions of the spot price S that are discrete. For example, the production model (6.1) is basically the optimization problem of CVaR, which does not require continuity. Hence, the optimal solution can be extended to the case where S has a general distribution. Specifically, the density (6.2) could be replaced with the expression in the first line in (2.10).

  In addition, a preliminary study of the complementary equations of the general dispatch problem indicates that it has a (rather complicated) bang-bang solution, too.

  The exactly solvable multi-period model could be extended to infinitely many periods or to continuous time.

- The number of hours may be chosen differently for different time steps of the multi-stage problem. For example, later stages could be be separated by longer time intervals, because ignorance prevents a detailed modeling of the far future. – Most of the performed analysis carries over unchanged.

- The value of the plant could depend on future production. For example:

$$\widetilde{V}_t := \sum_{k=1}^{t} e^{-rk} \cdot (P_k - P_{k-1}) + \sum_{k=t+1}^{T} e^{-rk} \cdot \mathbb{E}_{\mathbb{P}^*}[P_k - P_{k-1} | \mathcal{F}_t], \qquad t = 0, \ldots, T,$$

  where the expectation is taken with respect to an appropriate probability measure $\mathbb{P}^*$.

Appendix A

A.1 Quantiles

Let X be a random variable. The distribution function of X is denoted by $F\colon q \mapsto \mathbb{P}[X \leq q]$. The aim is to define inverses of F. Unfortunately, the set of pre-images of a given value[1] $\alpha \in (0,1)$ can have several elements (the distribution function is flat) or it can be empty (the distribution function has a jump). To encompass all pre-images, an 'upper' and 'lower' inverse is defined.

Definition 15 (Upper-, Lower-Quantile). Let $\alpha \in (0,1)$. The *upper-quantile of a random variable* X is

$$\begin{aligned} q_\alpha^+ &:= \sup\{q \mid \mathbb{P}[X \leq q] \leq \alpha\} \\ &= \sup\{q \mid \mathbb{P}[X < q] \leq \alpha\}, \\ &= \inf\,\{q \mid \mathbb{P}[X \leq q] > \alpha\}, \end{aligned} \tag{A.1}$$

and the *lower-quantile* is

$$\begin{aligned} q_\alpha^- &:= \inf\,\{q \mid \mathbb{P}[X \leq q] \geq \alpha\}, \\ &= \inf\,\{q \mid \mathbb{P}[X < q] \geq \alpha\}, \\ &= \sup\{q \mid \mathbb{P}[X \leq q] < \alpha\}. \end{aligned} \tag{A.2}$$

The second and third equalities in the quantile definitions are not that obvious; hence a proof is provided.

Proof. We consider the first equalities for the lower- and upper-quantiles as the definitions. So we have to proof the remaining two.

Let us proof that $q_\alpha^+ = q_\alpha := \sup\{q \mid \mathbb{P}[X < q] \leq \alpha\}$. Because $\mathbb{P}[X < q] \leq \mathbb{P}[X \leq q]\ \forall q$ it holds that $q_\alpha^+ \leq q_\alpha$. Because F is monotone, F has only countably many points of discontinuity, i.e. points q with $\mathbb{P}[X = q] > 0$. By definition,

[1] $\alpha \in (0,1)$ is to ensure finiteness; the whole unit interval $\alpha \in [0,1]$ could be taken into account by taking pre-images $\pm\infty$.

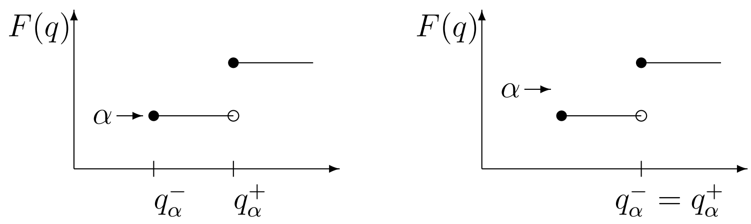


Figure A.1: Upper- and lower-quantiles of a discrete random variable with distribution function F for different levels α. If the pre-image is empty, then the α-quantile is unique (picture on the right).

$\mathbb{P}[X \le q] = \mathbb{P}[X < q] + \mathbb{P}[X = q]$. Let us assume that $q_\alpha^+ < q_\alpha$ would hold. By the countability of discontinuities, there exists a $q \in \mathbb{R}$ with

$$\mathbb{P}[X \le q] = \mathbb{P}[X < q] \text{ and } q_\alpha^+ < q < q_\alpha. \tag{A.3}$$

The inequality $q < q_\alpha$ implies that $\mathbb{P}[X < q] \le \alpha$, and therefore, by (A.3), $\mathbb{P}[X < q] = \mathbb{P}[X \le q] \le \alpha$. But this is a contradiction to $q_\alpha^+ < q$ in (A.3).

The same reasoning can be applied to q_α^-.

The second equalities are yielded by considering the complementary set of constraints with respect to the first equalities. ■

Clearly, the lower- is smaller than the upper-quantile: $q_\alpha^- \le q_\alpha^+$. The $\pm$-notation should be a reminder of the fact that the mapping $\alpha \mapsto q_\alpha^+$ is right-continuous on $\alpha \in (0,1)$, and the mapping $\alpha \mapsto q_\alpha^-$ is left-continuous on $\alpha \in (0,1)$ (see e.g. [34, Lemma 2.72]).

Definition 16 (Quantile). Let q_α^- and q_α^+ be the lower- and upper-quantile of a random variable at a level α. Then $q \in [q_\alpha^-, q_\alpha^+]$ is called an *α-quantile*.

If there are ambiguities, the random variable X and the level α are written explicitly: $q = q_\alpha(X)$.

Lemma 8. *We have*

(i) $q \in [q_\alpha^-, q_\alpha^+] \iff \big\{q \bigm| \mathbb{P}[X < q] \le \alpha \le \mathbb{P}[X \le q]\big\}$,

(ii) $q_\alpha^-(X) = -q_{1-\alpha}^+(-X)$,

(iii) $q(\lambda X) = \lambda q(X)$ *for* $\lambda \ge 0$.

Proof. (i): Because $F\colon q \mapsto \mathbb{P}[X \le q]$ is monotone, the set $\{q \mid \mathbb{P}[X \le q] \ge \alpha\}$ is an interval. Because F is right-continuous, the interval is closed:

$$\big[\inf\{q \mid \mathbb{P}[X \le q] \ge \alpha\}, \infty\big).$$

Because $q \mapsto \mathbb{P}[X < q]$ is monotone, the set $\{q \mid \mathbb{P}[X < q] \leq \alpha\}$ is an interval. Because $q \mapsto \mathbb{P}[X < q]$ is left-continuous, the interval is closed:

$$\big(-\infty, \sup\{q \mid \mathbb{P}[X < q] \leq \alpha\}\big].$$

The intersection of the intervals is therefore

$$\begin{aligned}\big\{q \mid \mathbb{P}[X < q] \leq \alpha \leq \mathbb{P}[X \leq q]\big\} &= \big[\inf\{q \mid \mathbb{P}[X \leq q] \geq \alpha\}, \sup\{q \mid \mathbb{P}[X < q] \leq \alpha\}\big] \\ &\stackrel{\text{def}}{=} [q_\alpha^-, q_\alpha^+].\end{aligned}$$

(ii):
$$\begin{aligned} q_\alpha^-(X) &\stackrel{\text{def}}{=} \inf\{q \mid \mathbb{P}[X \leq q] \geq \alpha\} \\ &= -\sup\{-q \mid \mathbb{P}[X \leq q] \geq \alpha\} \\ &= -\sup\{\,\tilde{q} \mid \mathbb{P}[X \leq -\tilde{q}] \geq \alpha\} \\ &= -\sup\{\,\tilde{q} \mid \mathbb{P}[-X \geq \tilde{q}] \geq \alpha\} \\ &= -\sup\{\,\tilde{q} \mid \mathbb{P}[-X < \tilde{q}] \leq 1-\alpha\}, \end{aligned}$$
which is the form of the second line in (A.1). Therefore
$$\stackrel{\text{def}}{=} -q_{1-\alpha}^+(-X).$$

(iii): By (i), q is an α-quantile of λX if and only if it is in the set

$$\begin{aligned}\{q \mid \mathbb{P}[\lambda X < q] \leq \alpha \leq \mathbb{P}[\lambda X \leq q]\} &= \{\lambda\tilde{q} \mid \mathbb{P}[\lambda X < \lambda\tilde{q}] \leq \alpha \leq \mathbb{P}[\lambda X \leq \lambda\tilde{q}]\} \\ &= \lambda\{\tilde{q} \mid \mathbb{P}[X < \tilde{q}] \leq \alpha \leq \mathbb{P}[X \leq \tilde{q}]\}. \qquad \blacksquare\end{aligned}$$

A.2 Consistency with Single-Period Risk Measurement

Assume that the set $\mathcal{P}$ of test-probability measures is stable. We show (4.4) (p. 27): The recursive risk-adjusted value $R_0^{X_T}$ equals the single-period coherent risk-adjusted value $\min_{\mathbb{Q}\in\mathcal{P}} \mathbb{E}_\mathbb{Q}[X_T]$.

Consider a single step in the recursive calculation (3.4) (p. 19) of $R_0^{X_T}$. It is a minimization of the conditional expectation of $R_{t+1}^{X_T}$. We make use of (4.6) (p. 29) and the stability of $\mathcal{P}$ to write it as

$$\min_{\mathbb{Q}\in\mathcal{P}} \mathbb{E}_\mathbb{Q}[R_{t+1}^{X_T}|\mathcal{F}_t] = \sum_{n\in\mathcal{N}_t} \Big(\min_{\mathbf{q}_n\in\mathcal{P}_n} \sum_{m\in\mathcal{N}_{t+1}} q_{nm} R_{(t+1)m}^{X_T} \Big) \chi_n, \tag{A.4}$$

where the definitions as for (4.6) are used. A minimization over two time-steps at once is

$$\begin{aligned}
\min_{\mathbb{Q}\in\mathcal{P}} \mathbb{E}_{\mathbb{Q}}[R^{X_T}_{t+2}|\mathcal{F}_t] &\overset{\text{def}}{=} \min_{\mathbb{Q}\in\mathcal{P}} \left(\sum_{n\in\mathcal{N}_t} \Big(\sum_{m_1\in\mathcal{N}_{t+1}} q_{nm_1} \sum_{m_2\in\mathcal{N}_{t+2}} q_{m_1m_2} R^{X_T}_{(t+2)m_2}\Big)\chi_n \right) \\
&\overset{\text{stbl.}}{=} \sum_{n\in\mathcal{N}_t} \Big(\min_{\mathbf{q}_n\in\mathcal{P}_n} \sum_{m_1\in\mathcal{N}_{t+1}} q_{nm_1} \min_{\mathbf{q}_{m_1}\in\mathcal{P}_{m_1}} \sum_{m_2\in\mathcal{N}_{t+2}} q_{m_1m_2} R^{X_T}_{(t+2)m_2}\Big)\chi_n \\
&\overset{(A.4)}{=} \sum_{n\in\mathcal{N}_t} \Big(\min_{\mathbf{q}_n\in\mathcal{P}_n} \sum_{m_1\in\mathcal{N}_{t+1}} q_{nm_1} \min_{\mathbb{Q}_2\in\mathcal{P}} \mathbb{E}_{\mathbb{Q}_2}[R^{X_T}_{t+2}|\mathcal{F}_{t+1}]\chi_{m_1}\Big)\chi_n \\
&\overset{(A.4)}{=} \min_{\mathbb{Q}_1\in\mathcal{P}} \mathbb{E}_{\mathbb{Q}_1}\Big[\min_{\mathbb{Q}_2\in\mathcal{P}} \mathbb{E}_{\mathbb{Q}_2}[R^{X_T}_{t+2}|\mathcal{F}_{t+1}]\Big|\mathcal{F}_t\Big].
\end{aligned}$$

Hence, owing to the stability of $\mathcal{P}$, the minimization over two time steps equals the recursive minimization over the single steps. A similar calculation gives for each $m = 1, 2, \ldots, T-t$ that

$$\begin{aligned}
&\min_{\mathbb{Q}\in\mathcal{P}} \mathbb{E}_{\mathbb{Q}}[R^{X_T}_{t+m}|\mathcal{F}_t] \\
&\quad = \min_{\mathbb{Q}_1\in\mathcal{P}} \mathbb{E}_{\mathbb{Q}_1}\Big[\cdots \min_{\mathbb{Q}_{m-1}\in\mathcal{P}} \mathbb{E}_{\mathbb{Q}_{m-1}}\Big[\min_{\mathbb{Q}_m\in\mathcal{P}} \mathbb{E}_{\mathbb{Q}_m}[R^{X_T}_{t+m}|\mathcal{F}_{t+m-1}]\Big|\mathcal{F}_{t+m-2}\Big] \cdots \Big|\mathcal{F}_t\Big].
\end{aligned}$$

Setting $t = 0$ and $m = T$, we find for the conditional expectation on the left-hand-side that

$$\mathbb{E}_{\mathbb{Q}}[R^{X_T}_T|\mathcal{F}_0] = \mathbb{E}_{\mathbb{Q}}[R^{X_T}_T] \overset{(3.4)}{=} \mathbb{E}_{\mathbb{Q}}[X_T],$$

whereas we have on the right-hand-side the recursive definition of $R^{X_T}_0$.

A.3 Segregative Production and Pumping

We prove that an optimal solution of (SC) (p. 61) does not simultaneously produce and pump.

Let $u^+_t(\cdot) := u^+_t(\cdot, \boldsymbol{E}_0 \ldots \boldsymbol{E}_t)$ and $u^-_t(\cdot) := u^-_t(\cdot, \boldsymbol{E}_0 \ldots \boldsymbol{E}_t)$ be the control-functions for production and pumping at time t. Let us assume that for every time t the $u^\pm_t$ are such that the constraints in the optimization problem are fulfilled. The idea of the proof is to define for every feasible control-functions (u^+_t, u^-_t) the new functions

$$\begin{aligned}
\tilde{u}^+_t &:= \max(u^+_t - u^-_t, 0) \qquad \text{for all } t, \\
\tilde{u}^-_t &:= \max(u^-_t - u^+_t, 0) \qquad \text{for all } t.
\end{aligned}$$

By definition, the new functions are not simultaneously strictly positive. It rests to verify that the new functions are admissible, give the same water levels as the old functions, and lead to an increased profit.

For ease of readability, we suppress the time index t: $u^\pm := u^\pm_t$; so the following expressions are supposed to hold for all t.

First, we verify that the functions $\tilde{u}^+$ and $\tilde{u}^-$ are in the set (7.2) (p. 57) of admissible control-functions. Clearly,

$$0 \leq \tilde{u}^+ \leq u^+_{\max} \quad \text{and} \quad 0 \leq \tilde{u}^- \leq u^-_{\max}.$$

The maximization $\max(\cdot, 0)$, as well as the functions u^+ and $-u^-$, are monotonically increasing. Hence, $\tilde{u}^+$ is monotonically increasing. The functions u^- and $-u^+$ are monotonically decreasing. Hence, $\tilde{u}^-$ is monotonically decreasing. Thus, the control-functions are admissible.

The control functions enter the state equation of water level (7.7) (p. 61) in the form of the following expression:

$$\begin{aligned}\tilde{u}^- - \tilde{u}^+ &= \max(u^- - u^+, 0) - \max(u^+ - u^-, 0)\\ &= \max(u^- - u^+, 0) - \max\big(-(u^- - u^+), 0\big)\\ &= u^- - u^+.\end{aligned}$$

Thus, the water level is the same as with the original control-functions.

Next, we consider the differences in the state equation of profit-and-loss (7.6) between the original and the new control-functions. Let $s \in \mathbb{R}_+$ denote spot price in a specific hour. The difference in the respective expressions is

$$\begin{aligned}&\Big(s \cdot \tilde{u}^+(s) - \frac{1}{c} s \cdot \tilde{u}^-(s)\Big) - \Big(s \cdot u^+(s) - \frac{1}{c} s \cdot u^-(s)\Big)\\ &= s\Big(\max\big(u^+(s) - u^-(s), 0\big) - u^+ + \frac{1}{c}\Big(u^-(s) - \max\big(u^-(s) - u^+(s), 0\big)\Big)\Big)\\ &= s\Big(\frac{1}{c} - 1\Big)\Big(u^+(s) + u^-(s) - \max\big(u^+(s), u^-(s)\big)\Big) \geq 0,\end{aligned}$$

where the last inequality holds because $s \geq 0$, $\frac{1}{c} - 1 > 0$ and $u^+, u^- \geq 0$. If there is a spot price s in a specific hour such that $u^+(s) > 0$ and at the same time $u^-(s) > 0$, then the profit-and-loss is strictly increased.

By the state equation of the value of the plant (7.8), the value of the plant is increasing in the profit-and-loss. Thus, if we take the new control, the objective value, which is the expected final value, is increased.

The risk-adjusted value was assumed to be increasing in the value of the plant. Thus, the constraint on risk is still satisfied under the new control.

Overall, we have shown that the new control-functions fulfil every constraint of the optimization problem and that they strictly increase the objective value whenever the old control-functions do simultaneously produce and pump.

A.4 Asymptotics of Occupation Times

Consider the occupation time $F_t(s)$ at time t and price level s (Def. 12, p. 63). If the sequence of random variables of the hourly spot price $(S_{\frac{h}{H}})_{h=1,2,\ldots}$ is i.i.d., then the indicator functions $(\chi_{\{S_{\frac{h}{H}} \leq s\}})_{h=1,2...}$ are also i.i.d., and the classical Central Limit Theorem says that the asymptotic ($H \to \infty$) distribution of $F_t(s)$ is normal, that is, the distribution function of the normalized occupation time converges pointwise and uniformly to the standard normal distribution function. It may be noted that in the i.i.d. case, the occupation time is identical to the empirical distribution function of the spot price.

The Central Limit Theorem is valid under a weaker assumption as follows.

Definition 17. Let $\mathcal{G}$ and $\mathcal{H}$ be sub-σ-algebras of the σ-algebra $\mathcal{F}$ of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The *strong mixing coefficient* is

$$\alpha(\mathcal{G}, \mathcal{H}) := \sup_{G \in \mathcal{G}, H \in \mathcal{H}} \Big| \mathbb{P}[G \cap H] - \mathbb{P}[G]\mathbb{P}[H] \Big|.$$

The strong mixing coefficient measures the stochastic dependence of two σ-algebras. The strong mixing coefficient for a process $(X_t)_{t=1,2,...}$ with gap m is defined as

$$\alpha_m := \sup_{n \in \mathbb{N}} \alpha\Big(\sigma(X_0, X_1, \dots, X_n),\ \sigma(X_{n+m}, X_{n+m+1}, \dots)\Big).$$

Proposition 9 (Central Limit Theorem for mixing processes, [70]). *If the strong mixing coefficients of the electricity price $(S_{\frac{h}{H}})_{h=1,2,...}$ are summable, i.e.*

$$\sum_{m=0}^{\infty} \alpha_m < \infty, \tag{A.5}$$

then the vector of occupation times $\big(F_t(s_0), \dots, F_t(s_N)\big)$ is asymptotically ($H \to \infty$) multivariate normally distributed.

The foregoing proposition is useful because the hourly electricity price can be modeled by autoregressive processes [15, 61]. Many types of autoregressive processes have summable strong mixing coefficients [2, 25]. Therefore, the vector of occupation times may be considered to have asymptotically a multivariate normal distribution. To clarify: The definition of an autoregressive process allows for non-normal innovations, such that the process at a specific time is generally not normal (see e.g. [13, Def. 3.1.2]).

A.5 Binomial Approximation of Normal Distribution

Consider an i.i.d. sequence of Bernoulli random variables $(X_i)_{i=1,2,...,J}$: $\mathbb{P}[X_i = 1] = \mathbb{P}[X_i = 0] = 1/2$. The sum of these variables has a binomial distribution. By the classical Central Limit Theorem, the sum is asymptotically ($J \to \infty$) normally distributed. The binomial distribution with parameters ($p = 1/2, J$) has values $j = 0, 1, \dots, J$ with probabilities

$$p_j = \binom{J}{j} \left(\frac{1}{2}\right)^J, \qquad j = 0, \dots, J. \tag{A.6}$$

To get an approximation of a *standardized* normal distribution, the binomial distribution is mean-adjusted and divided by its standard deviation, giving values

$$\frac{1}{\sqrt{\frac{1}{4}J}}\Big(j - \frac{1}{2}J\Big), \qquad j = 0, \dots, J. \tag{A.7}$$

Empirically, the approximation quality is good for low values of J. A proven error bound in terms of the supremum-norm of the difference between the standard normal and standardized binomial distribution function is $1/\sqrt{2\pi J}$ [45]. Let B_{kt}, $k = 0, \dots, K$, $t = 1, \dots, T$, be stochastically independent, standardized binomial distributions with parameters J_{kt} and $1/2$; the number of attainable values, $J_{kt} + 1$, is allowed to vary over different components k and time t. The random vector of innovations (8.9) on p. 77 is approximated by these standardized binomial distributions:

$$(\epsilon_{0t}, \epsilon_{1t}, \dots, \epsilon_{Kt}) \approx (\widehat{\sigma}_0 B_{0t}, \widehat{\sigma}_1 B_{1t}, \dots, \widehat{\sigma}_K B_{Kt}), \qquad t = 1, \dots, T. \tag{A.8}$$

For a given t, the discrete distribution on the right-hand-side has $\prod_{k=0}^{K}(J_{kt} + 1)$ different values.

A.6 Profit-and-Loss of Demand

Let us assume that the demand of electricity is modeled by an exogenous hourly stochastic process $(D_{\frac{h}{H}})_{h=1,2,\dots}$. The demand is sold at a fixed price c_d to a costumer. The profit-and-loss in time interval t to $t+1$ is

$$P_{t+1}^{\text{dem}} - P_t^{\text{dem}} = \sum_{h=1}^{H} D_{t+\frac{h}{H}}(c_d - S_{t+\frac{h}{H}}),$$

where the discounting factors are suppressed for simplicity. Let us approximate the demand with a step function (similarly to (7.10)):

$$D_{t+\frac{h}{H}} \to \widetilde{D}_{t+\frac{h}{H}} := \sum_{i=1}^{M} \bar{d}_i \chi_{\{d_{i-1} < D_{t+\frac{h}{H}} \le d_i\}},$$

where the range of the demand is discretized in $M+1$ levels:

$$d_0 < d_1 < \cdots < d_M,$$

with intermediate levels $\bar{d}_i \in (d_{i-1}, d_i)$. Then the approximated profit-and-loss can be written as

$$P_{t+1}^{\text{dem}} - P_t^{\text{dem}} = H \sum_{i,j=1}^{N,M} \bar{d}_j (c_d - \bar{s}_i) \Big(F_{t+1}(s_{i+1}, d_{j+1}) - F_{t+1}(s_i, d_j) \Big),$$

where

$$F_{t+1}(s,d) := \frac{1}{H} \sum_{h=1}^{H} \chi_{\{S_{t+\frac{h}{H}} \le s,\ D_{t+\frac{h}{H}} \le d\}}$$

is the joint *price-demand* occupation time in the time interval $[t, t+1]$.

Alternatively, the demand could be considered as a dependent process as follows. Empirically, the demand is strongly correlated with the electricity price [15]. Hence, a simple model of demand is $D_t = f(S_t)$, where $f\colon \mathbb{R} \to \mathbb{R}$ is a measurable function, and the associated profit-and-loss is

$$P_{t+1}^{\text{dem}} - P_t^{\text{dem}} = H \sum_{i=1}^{N} f(\bar{s}_i)(c_d - \bar{s}_i) \Big(F_{t+1}(s_i) - F_{t+1}(s_{i-1}) \Big).$$

In principle, using the foregoing proposals, we can integrate the demand in the existing modeling framework.

Bibliography

[1] C. Acerbi and D. Tasche. On the coherence of expected shortfall. *Journal of Banking & Finance*, 26(7):1487–1503, 2002.

[2] D. W. K. Andrews. First order autoregressive processes and strong mixing. Discussion Paper 664, Cowles Foundation for Research in Economics, Yale University, March 1983.

[3] T. Archibald, C. Buchanan, K. McKinnon, and L. Thomas. Nested benders decomposition and dynamic programming for reservoir optimization. *The Journal of the Operational Research Society*, 50(5):468–479, 1999.

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

[5] P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, and H. Ku. Coherent multiperiod risk measurement and Bellman's principle. Technical report, Department of Mathematics, ETH Zürich, November 2004.

[6] H. Bauer. *Mass- und Integrationstheorie*. Walter de Gruyter, second edition, 1992.

[7] R. A. Becker, J. M. Chambers, and A. R. Wilks. *The New S Language*. Wadsworth & Brooks, 1988.

[8] R. Bellman. *Dynamic Programming*. Princton University Press, 1957.

[9] F. Benth and S. Koekebakker. Stochastic modelling of financial electricity contracts. Technical report, Department of Mathematics, University of Oslo, 2005.

[10] D. P. Bertsekas and S. E. Shreve. *Stochastic optimal control: The discrete-time case*. Athena Scientific, 1996.

[11] J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer, New York, 1997.

[12] K. Boda and J. A. Filar. Time consistent dynamic risk measures. *Mathematical Methods of Operations Research*, 63:169–186, 2006.

[13] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer, New York, second edition, 1991.

[14] A. Brooke, D. Kendrick, and A. Meeraus. *GAMS: A User's Guide (Release 2.25)*. The Scientific Press, 1992. `http://www.gams.com`.

[15] M. Burger, B. Klar, A. Müller, and G. Schindlmayr. A spot market model for pricing derivatives in electricity markets. *Quantitative Finance*, 4:109–122, 2004.

[16] P. Burlando, E. Fraschini, and A. Khne. Impact of climate change on river runoff in mountainous areas. In *Proceedings of the final workshop 1998 Wallingford, UK – River basin modelling, management and flood mitigation*, volume EUR 18287, pages 251–268, 1999.

[17] P. Cheridito, F. Delbaen, and M. Kupper. Coherent and convex monetary risk measures for bounded càdlàg processes. *Stochastic Processes and their Applications*, 112:1–22, 2004.

[18] P. Cheridito, F. Delbaen, and M. Kupper. Coherent and convex monetary risk measures for unbounded càdlàg processes. *Finance and Stochastics*, 10(3):427–448, 2006.

[19] F. Delbaen. Coherent risk measures on general probability spaces. In K. Sandmann and P. Schönbucher, editors, *Advances in Finance and Stochastics*, pages 1–37. Springer, Berlin, 2002.

[20] F. Delbaen. The structure of m-stable sets and particular of the set of risk neutral measures. Technical report, Department of Mathematics, ETH Zürich, May 2003.

[21] M. Dempster and R. Thompson. EVPI-based importance sampling solution procedures for multistage stochastic linear programms on parallel MIMD architectures. *Annals of Operations Research*, 90:161–184, 1999.

[22] M. Densing. The covariance structure of occupation times of an Ornstein-Uhlenbeck process. Technical report, Institute for Operations Research, University of Zürich, January 2006.

[23] J. Döge. *Valuation of Flexibility for Power Portfolios – A Dynamic Risk Engineering Approach*. PhD thesis, ETH Zürich, 2005. Nr. 16438.

[24] J. Döge, H.-J. Lüthi, and P. Schiltknecht. Risk management of power portfolios and valuation of flexibility. *OR Spectrum*, 28(2):267–287, April 2006.

[25] P. Doukhan. *Mixing: Properties and examples*, volume 85 of *Lecture Notes in Statistics*. Springer, New York, 1994.

[26] J. Dupačová, G. Consigli, and S. W. Wallace. Scenarios for multistage stochastic programs. *Annals of Operations Research*, 100:25–53, 2000.

[27] J. Dupačová, N. Gröwe-Kuska, and W. Römisch. Scenario reduction in stochastic programming. *Mathematical Programming (Series A)*, 95:493–511, 2003.

[28] A. Eichhorn and W. Römisch. Polyhedral risk measures in stochastic programming. *SIAM Journal of Optimization*, 16(1):69–95, 2005.

[29] A. Eichhorn and W. Römisch. Mean-risk optimization models for electricity portfolio management. In *Proceedings of the 9th International Conference on Probabilistic Methods Applied to Power Systems*, 2006.

[30] L. G. Epstein and M. Schneider. Recursive multiple priors. *Journal of Economical Theory*, 113:1–31, 2003.

[31] EU European Council. Directive 96/92/EC of the European Parliament and of the Council of 19 December 1996 concerning common rules for the internal market in electricity. *Official Journal of the European Union*, L 27:20–29, 1997.

[32] European Electricity Exchange, Leipzig, Germany. *Historical Market Data*, 2005. `http://www.eex.de`.

[33] S.-E. Fleten, S. W. Wallace, and W. T. Ziemba. Hedging electricity portfolios via stochastic programming. In C. Greengard and A. Ruszczyński, editors, *Decision Making Under Uncertainty: Energy and Power*, volume 128 of *The IMA Volumes in Mathematics and its Applications*, pages 71–94. Springer, New York, 2002.

[34] H. Föllmer and A. Schied. *Stochastic Finance*. de Gruyter Studies in Mathematics, 2002.

[35] K. Frauendorfer. Barycentric scenario trees in convex multistage stochastic programming. *Mathematical Programming (Series B)*, 75:277–294, 1996.

[36] K. Frauendorfer and J. Güssow. Stochastic multistage programming in the operation and management of power system. In K. Marti, editor, *Stochastic Optimization Techniques – Numerical Methods and Technical Applications*, pages 199–222. Springer, 2002.

[37] K. Frauendorfer and M. Schürle. Term structure models in multistage stochastic programming: Estimation and approximation. *Annals of Operations Research*, 100:189–209, 2000.

[38] M. Frittelli and G. Scandolo. Risk measures and capital requirements for processes. *Mathematical Finance*, 16(4):589–613, 2006.

[39] N. Gröwe-Kuska, H. Heitsch, and W. Römisch. Scenario reduction and scenario tree construction for power management problems. In A. Borghetti, C. Nucci, and M. Paolone, editors, *IEEE Bologna Power Tech Proceedings*, 2003.

[40] N. Gröwe-Kuska, K. C. Kiwiel, M. P. Nowak, W. Römisch, and I. Wegner. Power management in a hydro-thermal system under uncertainty by Lagrangian relaxation. In C. Greengard and A. Ruszczyński, editors, *Decision Making under Uncertainty: Energy and Power*, volume 128 of *The IMA Volumes in Mathematics and its Applications*, pages 39–70. Springer, New York, 2002.

[41] N. Gröwe-Kuska and W. Römisch. Stochastic unit commitment in hydrothermal power production planning. In S. Wallace and W. Ziemba, editors, *Applications of Stochastic Programming*, MPS/SIAM Series in Optimization, chapter 30. SIAM, 2005.

[42] N. Gülpınar, B. Rustem, and R. Settergren. Simulation and optimization approaches to scenario tree generation. *Journal of Economic Dynamics and Control*, 28:1291–1315, 2004.

[43] H. Heitsch and W. Römisch. Scenario reduction algorithm in stochastic programming. *Computational Optimization and Application*, 24:187–206, 2003.

[44] J. Hinz. Valuing production capacities on flow commodities. Technical report, Institute for Operations Research, ETH Zürich, 2005. To appear in Mathematical Methods of Operations Research.

[45] C. Hipp and L. Mattner. On the normal approximation to symmetric binomial distributions. Technical report, Institute for Mathematics, University of Karlsruhe and Lübeck, April 2005.

[46] K. Høyland, M. Kaut, and S. W. Wallace. A heuristic for moment-matching scenario generation. *Computational Optimiziation and Applications*, 24:169–185, 2003.

[47] K. Høyland and S. W. Wallace. Generating scenario trees for multistage decision problems. *Management Science*, 47(2):295–307, 2001.

[48] F. Jamshidian and Y. Zhu. Scenario simulation: Theory and methodology. *Finance and Stochastics*, 1:43–67, 1997.

[49] P. Kall and J. Mayer. *Stochastic Linear Programming: Models, Theory, and Computation*. Springer, New York, 2005.

[50] P. Kall and S. W. Wallace. *Stochastic Programming*. John Wiley & Sons, 1994.

[51] M. Kaut and S. W. Wallace. Evaluation of scenario-generation methods for stochastic programming. *Stochastic Programming E-Print Series*, 14, May 2003. `http://www.speps.org`.

[52] B. W. Kernighan and D. M. Ritchie. *The C Programming Language*. Prentice Hall, Inc., second edition, 1988.

[53] T. H. Kim and H. White. On more robust estimation of skewness and kurtosis. *Finance Research Letters*, 1:56–70, 2004.

[54] P. Klaassen. Solving stochastic programming models for asset/liability management using iterative disaggregation. In *Worldwide Asset and Liability Modeling*, pages 427–463. Cambridge University Press, 1998.

[55] P. R. Kleindorfer and L. Li. Multi-period VaR-constrained portfolio optimization with application to the electric power sector. *The Energy Journal*, 26(1):1–26, 2005.

[56] R. Kouwenberg. Scenario generation and stochastic programming models for asset liability management. *European Journal of Operational Research*, 134:279–292, 2001.

[57] P. Krokhmal, J. Palmquist, and S. Uryasev. Portfolio optimization with conditional value-at-risk objective and constraints. *Journal of Risk*, 4(2):43–68, 2002.

[58] M. Kupper, P. Cheridito, and F. Delbaen. Dynamic monetary risk measures for bounded discrete-time processes. *Electronic Journal of Probability*, 11:57–106, 2006.

[59] F. E. Kydland and E. C. Prescott. Rules rather than discretion: The inconsistency of optimal plans. *The Journal of Political Economy*, 85(3):473–492, 1977.

[60] J. P. LaSalle. Time optimal control systems. *Proceedings of the National Academy of Sciences of the United States fo America*, 45(4):573–577, 1959.

[61] J. J. Lucia and E. S. Schwartz. Electricity prices and power derivatives: Evidence from the nordic power exchange. *Review of Derivatives Research*, 5:5–50, 2002.

[62] H.-J. Lüthi and J. Döge. Convex risk measures for portfolio optimization and concepts of flexibility. *Mathematical Programming*, 104:541–559, 2005.

[63] D. P. Morton. An enhanced decomposition algorithm for multistage stochastic hydroelectric scheduling. *Annals of Operations Research*, 64:211–235, 1996.

[64] M. P. Nowak, R. Nürnberg, W. Römisch, R. Schultz, and M. Westphalen. Stochastic programming for power production and trading under uncertainty. In W. Jäger and H.-J. Krebs, editors, *Mathematics – Key Technology for the Future*, pages 632–646. Springer, 2003.

[65] N. Osman. *Kleines Lexikon deutscher Wörter arabischer Herkunft.* Beck, München, sixth edition, 2002.

[66] M. V. Pereira and L. M. Pinto. Multi-stage stochastic optimization applied to energy planning. *Mathematical Programming*, 52:359–375, 1991.

[67] G. C. Pflug. Scenario tree generation for multiperiod financial optimization by optimal discretization. *Mathematical Programming*, 89:251–271, 2001.

[68] G. C. Pflug. A value-of-information approach to measuring risk in multi-period economic activity. *Journal of Banking & Finance*, 30:695–715, 2006.

[69] F. Riedel. Dynamic coherent risk measures. *Stochastic Processes and Applications*, 112:185–200, 2004.

[70] E. Rio. *Théorie asymptotique des processus aléatoires faiblement dépendants*, volume 31 of *Mathématiques & Applications*. Springer, Berlin, 2000.

[71] R. T. Rockafellar. Optimization under uncertainty. Lecture Notes, University of Washington.

[72] R. T. Rockafellar and S. Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26:1443–1471, 2002.

[73] B. Roorda, H. Schumacher, and J. Engwerda. Coherent acceptability measures in multiperiod models. *Mathematical Finance*, 15:589–612, 2005.

[74] A. Ruszczyński and A. Shapiro. Optimization of risk measures. In G. Calafiore and F. Dabbene, editors, *Probabilistic and Randomized Methods for Design under Uncertainty*. Springer, London, 2005.

[75] J. D. Salas, J. A. Ramirez, P. Burlando, and R. A. Pielke. Stochastic simulation of precipitation and streamflow processes. In T. D. Potter and B. R. Colman, editors, *Handbook of Weather, Climate and Water*, chapter 33, pages 607–640. John Wiley & Sons, 2003.

[76] E. Seebold. *Kluge – Etymologisches Wörterbuch der deutschen Sprache*. Walter de Gruyter, 24 edition, 2002.

[77] S. Siegrist. *A Complementary Approach to Multistage Stochastic Linear Programs*. PhD thesis, University of Zürich, 2006. `http://www.dissertationen.unizh.ch/2006/siegrist`.

[78] Y. L. Tong. *The Multivariate Normal Distribution*. Springer Series in Statistics. Springer, New York, 1990.

[79] G. Unger. *Hedging Strategy and Electricity Contract Engineering*. PhD thesis, ETH Zürich, 2002. Nr. 14727.

[80] S. Uryasev and R. Rockafellar. Optimiziation of conditional value-at-risk. *Journal of Risk*, 2:21–41, 2000.

[81] T. Wang. A class of dynamic risk measures. Technical report, Faculty of Commerce and Business, University of British Columbia, 1999.

[82] S. Weber. Distribution-invariant dynamic risk measures, information, and dynamic consistency. *Mathematical Finance*, 16(2):419–442, 2006.

[83] D. Werner. *Funktionalanalysis*. Springer, Berlin, fourth edition, 2002.

[84] E. Zivot and J. Wang. *Modelling Financial Time Series with S-Plus®*. Springer, New York, 2003.

Curriculum Vitae

name	Martin Densing
date of birth	27 July 1970
citizen	Ermatingen (TG)

1977 - 1984	primary school, Ermatingen
1984 - 1985	secondary school, Ermatingen
1985 - 1990	grammar school, Kreuzlingen (TG)
1990	military service (FR)
1991 - 1996	studies at the faculty of physics, ETH Zürich
1996	diploma thesis in theoretical physics, University of Zürich
1996 - 1999	financial consultant, Integrated Risk Management AG, Zürich
2000 - 2005	scientific collaborator at IOR, University of Zürich
2003 - 2006	PhD student by Prof. P. Kall (IOR) and Prof. H.-J. Lüthi (IFOR, ETH Zürich)
2006 -	risk consultant, SunGard Corp., Winterthur

IOR, IFOR = Institute for Operations Research