Stochastic programming with applications to power systems

Trine Krogh Kristoffersen

PhD thesis, April 2007

Stochastic programming with applications to power systems

Trine Krogh Kristoffersen, Department of Operations Research, University of Aarhus, Denmark.

PhD dissertation, April 2007.

Dissertation committee:

Jørgen Aase Nielsen, University of Aarhus, Jørgen Tind, University of Copenhagen, Stein William Wallace, Molde University College.

Advisor:

Kim Allan Andersen, Aarhus School of Business.

Trine Krogh Kristoffersen
Department of Operations Research
University of Aarhus
Ny Munkegade
Building 1530
DK-8000 Aarhus C
Denmark
trinek@imf.au.dk

Subject classification:
MSC2000 : Primary: 90C15, 90C90; secondary: 37M10.
OR/MS : Stochastic programming, Applications, Time series.

Preface

This thesis presents the results of my work in stochastic programming during my time as a Ph.D. student at Department of Operations Research at the University of Aarhus. The primary focus of my studies has been applications to power systems, including the development and analysis of electricity models to approach the problems that emerged with the incipient restructuring of the power sector in the last decade.

In the course of a three months visit at the University in Duisburg-Essen, I began working on a theoretical problem under the supervision of Prof. Dr. Rüdiger Schultz. The problem was motivated by the needs of incorporating risk management into stochastic programming and considered the inclusion of certain risk measures that were shown to preserve a number of properties and allow for algorithmic treatment in two-stage stochastic linear programming. The work, entitled *Deviation measures in linear two-stage stochastic programming*, was subsequently published in Mathematical Methods of Operations Research, Vol. 62, No. 2, 2005.

By Prof. Schultz I was briefly introduced to the potential of stochastic programming in energy. Later, I started a collaboration with Associate Professor Stein-Erik Fleten, who is working in the field, and became the very interested in electricity applications, which therefore provided a basis for the rest of my work. As a natural consequence of Stein-Erik Fleten's location in Tronheim, my work with him involved the major electricity source of Norway, hydro-power.

The starting point of our work was a planning problem presented by the hydropower company, TrønderEnergi, near Trondheim. In the liberalized power market, a power producer faces the problem of bidding into the day-ahead market with only limited information on the market price. Thus, we proposed a stochastic programming model that sought to reflect the Nordic market conditions as closely as possible, including market price uncertainty and, contrary to the existing literature on the subject, both so-called hourly bids and block bids. The computational results offered valuable insight into the advantages of using a stochastic approach for optimizing bidding strategies. A presentation can be found in the paper *Stochastic programming for optimizing bidding strategies of a Nordic hydro-power producer* published in European Journal of Operational Research, Vol. 181, 2007.

The short-term hydro rescheduling problem, that has not previously been addressed in the literature, came about from same collaboration with Stein-Erik Fleten and TrønderEnergi. Following the completion of the day-ahead bidding, the model establishes a daily production plan that complies with the day-ahead commitments. Uncertainties in both reservoir inflows and market prices were investigated and special effort was made to generate the scenarios that serve as input to the stochastic programming problem. The paper *Short-term hydro-power production planning by stochastic programming* is in press for Computers and Operations Research.

In contrast to the planning problems of a small power producer, the Danish power system operator, Energinet.dk, introduced a problem of a price-taker. In order for the system operator to determine the amount of reserves necessary for balancing supply and demand, we established a stochastic programming model that was able to include the price determination process and account for supply and demand uncertainty. The resulting paper *Power reserve management by two-stage stochastic programming* is joint work with Camilla Schaumbug-Müller and is submitted to an operations research journal.

Finally, the survey *The development in stochastic programming models for power production and trading* must be considered work in progress.

Århus, April 2007
Trine Krogh Kristoffersen

Acknowledgments

Several people have contributed to the progress of this work.

Foremost, I wish to thank my adviser Kim Allan Andersen for his encouraging way of guiding me through the process. His insightful comments and suggestions has been a significant motivation.

During my Ph.D. program I had the pleasure of staying three months by Prof. Dr. Rüdiger Schultz at the University of Duisburg-Essen. I am deeply indebted for his supreme supervision and the privilege to draw from his superior knowledge in the field of stochastic programming. Also, I wish to express my gratitude for the hospitality and the socially enjoyable environment both to him and the other people at the Department of Mathematics.

I owe a large debt of thanks to Associate Professor Stein-Erik Fleten at the Norwegian University of Science and Technology. I highly appreciate our work together within the stochastic programming application area of electricity. Through visits in Trondheim and constant mail correspondence, I was able to benefit greatly from our discussions and gained from him a deeper understanding not only on energy modeling but also in other relevant subject matters. At the same time I wish

to thank Nina Detleflefsen for establishing both this and other contacts and in general for her willingness to help when possible.

I am thankful to Camilla Schaumburg-Müller from the Technical University of Copenhagen for our fruitful work together and for her everlasting enthusiasm. Moreover, I would like to thank her supervisor Hans Ravn for his indispensable advice during our work.

Special thanks goes to my former colleagues at the University of Aarhus, Rasmus Vinther Rasmussen and Christian Roed, who have provided an ideal working atmosphere in always being ready to help with new ideas and in sharing our everyday stories (and sweets) to cheer up life in the office. At the Department of Mathematical Sciences, I would like to thank Randi Mosegaard for linguistic support. Furthermore, I have benefited from very helpful statisticians, among those Preben Blæsild and Anders Holst Andersen.

The main work of my Ph.D. program has focused on applications of stochastic programming to power systems. However, this would never have been possible without Energinet.dk and TrønderEnergi who both generously shared their data and extensive knowledge. I would like to give my thanks to Peter Børre, Henning Parbo and Jens Petersen as well as Berhard Kvaal, Gunnar Aronsen and Lars Olav Hoset.

Last but not least, I want to thank my family and friends for their continuous and outstanding support and Wouter in particular for his understanding and endless time to listen (in spite of the reflection on the phone bill).

Contents

Summary

The field of stochastic programming is concerned with the study of mathematical programming problems subjected to uncertainty. During the past 50 years, stochastic programming problems have encouraged a significant amount of research into structural properties such as continuity, differentiability, convexity and stability, mainly to facilitate the development of efficient solution approaches. In their linear versions, stochastic programming problems have proved particularly suited for decomposition. Mixed-integer formulations, however, are generally found to be computational challenging and the contributions within the literature are fewer. Instead, the area of applications have appeared to attract an increasing attention.

At the same time, with the decentralization of the electricity generation and the deregulation of the power markets, many previous procedures have changed and new planning and operating problems have emerged, making advances in power optimization highly relevant. Furthermore, the presence of uncertainty have been widely acknowledged, motivating applications of stochastic optimization.

Backed up by several studies, applications to power systems successfully illustrate the strengths of stochastic programming. By virtue of the sequential decision process, the stochastic programming models incorporate the information flow of prices, costs, resources etc. and the ability of production and trading decisions to "hedge" against future uncertainty, the result being increases in profit or decreases in risk.

In Chapter 1 of the thesis, we begin with an introduction to the most basic and widely applied stochastic programming problems, two-stage and multi-stage stochastic recourse problems, and state the approximations that allow for computational tractability. We present some of the general solution approaches for linear and mixed-integer two-stage and multi-stage programs, which are applicable to the problems of this thesis and from which some of the suggested solution methods originate.

Chapter 2 continues with two-stage linear recourse problems and in particular the inclusion of the risk measures known as deviation measures. In line with the presentation in Kristoffersen (2005), we show that the resulting mean-risk models

inherit the properties of continuity, differentiability, convexity and stability from the traditional model and can be handled algorithmically by a modification of the L-shaped method. The capability of deviation measures to comply with a linear formulation is utilized in the power optimization models of the following chapters, in which the downside measures, semideviation and expected excess of target, come into play.

The remainder of the thesis is dedicated to power optimization problems within the area of stochastic programming applications. Chapter 3 provides an overview of stochastic programming models in short-term power production and trading with special emphasis on the development prompted by the restructuring of the power sector. The contents of the chapter is work in progress, which at its current stage can be found in Kristoffersen (2007). The subsequent Chapters 4-6 each present a power optimization problem within the most important short-term activities, day-ahead bidding, rescheduling and intra-day balancing, that has become relevant with the restructuring.

Chapter 4 concerns the problem of bidding into the day-ahead electricity market from the perspective of a price-taking Nordic hydro-power producer that is subjected to market price uncertainty. With a time horizon of an operation day, market prices are revealed at once, and we therefore present a two-stage stochastic programming model. The model includes the main features of the Nordic power market by including both so-called hourly bids and block bids, which allows us to analyze the impact of uncertainty on the structure of the bids. The work is a slightly modified version of that by Fleten and Kristoffersen (2007).

In extension of the problem in Chapter 4, Chapter 5 addresses the problem of determining a daily hydro-power production plan that complies with the day-ahead commitments of the previous day, which is a way of rescheduling. Basically, the problem becomes a matter of spatial distribution of water between the reservoirs when market prices and reservoir inflows are uncertain. To fully capture the future effects of current water releases from the reservoirs, we propose a multi-stage stochastic programming model. The model was first presented in Fleten and Kristoffersen (2006).

In spite of rescheduling, actual production may not match the day-ahead market commitments completely and intra-day balancing is necessary. To ensure sufficiency of balancing resources, however, reserves must be purchased in advance. Chapter 6 presents an application of stochastic programming to the problem of managing such reserves when the imbalances are uncertain at the time of purchasing the reserves. Since this task is the responsibility of the power system operator, the price determination process was included in the model. Still, the model maintains a structure that allows for a solution approach close to common practice. The chapter is a modification of Kristoffersen and Schaumburg-Müller (2007).

Being an important part of modeling, we have dedicated the remainder of the thesis to scenario generation. Chapter 7 gives a selected overview on scenario

generation and reduction methods potentially suitable for applications to power systems. The chapter serves to justify the approach of Chapters 4 and 5 and explains the method used into details. In short, scenario generation starts from a statistical model, from which sampling is possible, and is combined with a scenario reduction method.

A specific description of the statistical models can be found in Chapter 8. The models that determine the distribution of the uncertain data are derived from time series analysis. The univariate distributions describe market prices and reservoir inflows as autoregressive moving average processes, which is also the case for the multivariate distribution.

The main contributions of this thesis are found in Chapters 2, 4 and 5. The contributions of Chapter 2 are within theoretical aspects of stochastic programming, whereas Chapters 4, 5 and 6 contribute within the area of stochastic programming models and applications. The overview of this topic in Chapter 3 is intended to provide a basis for future work and cannot be considered complete. The same applies to Chapter 8 that is ongoing work in scenario generation.

When going through the chapters, we assume the reader is familiar with the basics of probability and measure theory such as probability spaces, random variables and expectations. We further assume some prior knowledge of convex analysis, linear programming as well as mixed-integer linear programming. Finally, an acquaintance with classical statistics is an advantage.

The intension is to maintain a consistent notation throughout the thesis. A list of notation can be found in the back. We include the most important symbols used. Still, since additional notation is sometimes necessary and for ease of exposition, we have explained the notation when used. In cases of ambiguity, the proper use should therefore be clear from the context.

Part I

Stochastic recourse problems

Chapter 1

An introduction to stochastic programming

In this chapter we give a short introduction to the field of stochastic programming, the most commonly known classes of stochastic programming problems and the corresponding terminology and notation. To keep the exposition in line with the rest of the thesis, we restrict the discussion to stochastic mixed-integer linear programming problems. As the of major part of the thesis is devoted to applications, we will not present the structural properties of the stochastic programming problems except for the those of the most basic class. It is worth noting, though, that many results are basically generalizations and follow in rather similar ways. However, we present the most fundamental solution approaches to such problems and include some of the major findings in the development of algorithms.

The field of stochastic programming is concerned with optimization under uncertainty. As the name suggests, its modeling approaches and algorithmic techniques are inherited from mathematical programming, which separates it from the related fields of decision analysis, stochastic control theory and Markov decision processes. Although, mathematical programming is highly recognized and widely used, uncertainty can only be handled by sensitivity or parametric analysis. Stochastic programming overcomes this drawback by including uncertainty explicitly into mathematical programming. Essentially, a stochastic program is a mathematical program in which uncertain data is represented by random variables and an appropriate optimization criterion is selected.

In the following we will confine ourselves to the class of stochastic programming problems referred to as stochastic recourse problems and present the two-stage and multi-stage versions. For other stochastic programming problems such as chance constrained programs, we refer the reader to Prekopa (1995).

1.1 Random optimization

The starting point of stochastic programming is random optimization. We formalize the analysis by the following random mixed-integer linear programming problem, where uncertainty is reflected in data being represented by random variables.

$$\min\{cx \mid Ax = b, T(\omega)x = h(\omega), x \in X\}. \tag{1.1.1}$$

As is also the case in the remainder of the thesis, transposes have been eliminated. We consider a costs minimization framework and let all components have conformable dimensions. $X \subseteq \mathbb{R}^{n_1}_+$ has the property that its convex hull is polyhedral, which allows for integrality restrictions on some of the variables $x \in \mathbb{R}^{n_1}_+$. In a mixed-integer framework the set can thus, without loss of generality, itself be assumed to be polyhedral. We let $\mathbb{R}^n$ be the space of real n-vectors and $\mathbb{R}^{m\times n}$ the space of real $m \times n$-matrices. $c \in \mathbb{R}^{n_1}$ and $b \in \mathbb{R}^{m_1}$ are known vectors and $A \in \mathbb{R}^{m_1\times n_1}$ is a known matrix. $h : \Omega \to \mathbb{R}^{m_2}$ is a random vector and $T : \Omega \to \mathbb{R}^{m_2\times n_1}$ is a random matrix on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. As ω denotes an element of Ω, realization of h and T are denoted $h(\omega)$ and $T(\omega)$.

On one hand, the problem (1.1.1) may represent a *distribution problem* that serves to determine the distributional characteristics of the optimal solutions and objective function values. In a distribution problem, decisions are made after uncertainty is observed, as is the case in sensitivity and parametric analysis. On the other hand, the problem (1.1.1) may be regarded as a so-called *anticipatory problem*, a category into which stochastic recourse problems fall. The challenge of anticipatory problems is to make decisions without anticipating future realizations of the random variables. With these restrictions, however, problem (1.1.1) is not well-defined. To redefine the problem, it is crucial to select an optimization criterion that values future realizations of the random variables and at the same time reflects the preferences of the decision-maker.

Stochastic recourse problems incorporate corrective actions in response to the non-anticipative decisions and employ an optimization criterion that includes the costs of both decision types. Early attempts to formulate the recourse problems were found already in Danzig (1955). We proceed with the presentation of the two-stage and multi-stage stochastic programs with recourse.

1.2 Two-stage stochastic programs with recourse

The most basic stochastic recourse problem is the *two-stage stochastic program with recourse*. To state the problem, assume that non-anticipative decisions represent the main decisions that have already been made and that a temporary violation of the random constraints is allowed. Feasibility is restored through recourse actions that are deferred until the realization of uncertainty is observed. In

this fashion, the decisions are partitioned into two stages according to the information flow and we therefore refer to them as first-stage and second-stage decisions. It should be remarked that the partitioning of decisions need not actually reflect the separation of main decisions and recourse actions but may simply reflect the timing of the decisions such that first-stage decisions are to be made immediately, whereas second-stage decisions can be deferred. Still, we use the terms recourse actions and second-stage decisions interchangeably. Assume further that the decision-maker seeks to minimize direct and expected future costs. Then the two-stage stochastic recourse problem with recourse can be stated as

$$\min\{cx + \mathbb{E}[q(\omega)y(\omega)] \mid Ax = b, Wy(\omega) + T(\omega)x = h(\omega) \\ x \in X, y(\omega) \in Y \ \ \mathbb{P}.a.a.\omega\}. \quad (1.2.1)$$

We denote the expectation operator by $\mathbb{E}[\cdot]$. $Y \subseteq \mathbb{R}^{n_2}_+$ is a non-empty polyhedron that may contain integrality restrictions on some of the variables $y \in \mathbb{R}^{n_2}_+$. The dependency of y on ω reflects the fact that the decisions differ for different realizations of the random variables. $W \in \mathbb{R}^{m_2 \times n_2}$ is a known matrix and $q : \Omega \to \mathbb{R}^{n_2}$ is a random vector on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We refer to x and y as first-stage and second-stage decisions, respectively. c and q are called first-stage and second-stage costs. The first-stage constraints are defined by A and b and the second-stage constraints by the recourse matrix W, the technology matrix T and the right-hand side h. The second-stage constraints are assumed to hold for $\mathbb{P}$-almost all ω, i.e. for $\omega \in \Omega \backslash \Omega'$, where $\mathbb{P}(\Omega') = 0$. The assumption of a known recourse matrix is referred to as fixed recourse. Occasionally, we use the following notation for the expected value of a random variable or vector ξ

$$\mathbb{E}[\xi(\omega)] = \int_\Omega \xi(\omega)\mathbb{P}(d\omega).$$

Moreover, we let $\xi : \Omega \to \mathbb{R}^N$ be the random vector whose components constitute the uncertain data, i.e. $\xi = (q, h, T_1, \ldots, T_{m_2})$, where $, T_1, \ldots, T_{m_2}$ denote the rows of T and $N = n_2 + m_2 + m_2 \times n_1$. To ease notation, we introduce the image measure $\mu = \mathbb{P} \circ \xi^{-1}$ on $\mathbb{R}^N$ and change variables, so that for instance

$$\int_\Omega \xi(\omega)\mathbb{P}(d\omega) = \int_{\mathbb{R}^N} \xi \ \mu(d\xi).$$

To fully illustrate the dynamics of the two-stage decision process, consider the following scheme

$$\text{decide on x} \to \text{observe } q, h, T \to \text{decide on y}.$$

As mentioned above, first-stage decisions must be made with limited information on the future realization of the random data and such as to minimize direct first-stage costs and expected second-stage costs. As the realization of the random

data is revealed, the second-stage decisions can be based on the actual realization and the second stage costs are determined. The dynamics can be demonstrated by formulating of the stochastic recourse program (1.2.1) in terms of dynamic programming. The two-stage stochastic program with recourse is

$$\min\{Q(x) \mid Ax = b, x \in X\}, \tag{1.2.2}$$

with the recourse function

$$Q(x) := cx + \mathbb{E}[\Phi(q, h - Tx)] \tag{1.2.3}$$

and the second-stage value function

$$\Phi(q, h - Tx) := \min\{qy \mid Wy = h - Tx, y \in Y\}. \tag{1.2.4}$$

The dynamic programming formulation (1.2.2)-(1.2.4) illustrates the difficulties in solving the two-stage stochastic program with recourse. Due to the recourse function (1.2.3), (1.2.2) is a non-linear programming problem that involves the evaluation of an integral. Even for an absolutely continuous distribution of the random variables, the problem is non-convex. Most solution approaches therefore rely on an approximation by a discrete distribution with finite support. We assume the approximation of $\xi = (q, h, T)$ is given by a set of scenarios $\{1, \ldots, S\}$ that corresponds to the realizations $\xi^s = (q^s, h^s, T^s), s = 1, \ldots, S$ and probabilities $\pi^s, s = 1 \ldots, S$. The resulting two-stage stochastic program is often referred to as the *deterministic equivalent*.

$$z = \min\ cx + \sum_{s=1}^{S} \pi^s q^s y^s \tag{1.2.5}$$

$$\text{s.t. } Wy^s + T^s x = h^s, Ax = b, x \in X, y^s \in Y. \tag{1.2.6}$$

For an illustration of two-stage stochastic programming scenarios, see Fig. 1.1. The nodes represent decisions points; the node to right first-stage decisions and those to the left scenario-dependent second-stage decisions.

Remark 1.2.1 *Stochastic programming is founded on the assumption of a known probability distribution of the random data, which may seem as a rather strong assumption. Mostly, the distribution is approximated by some discrete distribution with finite support. Nevertheless, mathematical programming assumes the data to be known and specified in advance, which can be seen as specifying a distribution of only one mass point and hence is most likely outperformed by a distribution with a number of mass points. The finite number of mass points that determines the random stochastic programming data define the set of scenarios. Throughout the thesis, we refer to the set of scenarios interchangeably as* $\{1, \ldots, S\}$ *or* $\mathcal{S}$*. For different approaches to approximating the probability distribution, see Chapter 7.*

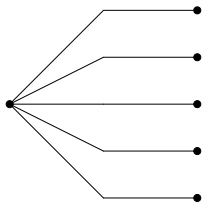


Figure 1.1: Two-stage scenario paths.

So far, we have implicitly assumed a risk neutral decision-maker, who seeks to minimize an expectation-based objective. In the case of other preferences and in particular another attitude towards risk, the objective takes a different form. For simplicity, however, we will in general state the stochastic program as

$$\min\{Q(x) \mid Ax = b, x \in X\}.$$

To further simplify the notation, we will sometimes suppress the representation of the constraints.

Structural properties such as continuity, differentiability, convexity and stability of the two-stage stochastic programs with linear recourse are given in Chapter 2. The results contain the cases of both expectation-based and risk-adjusted objectives. For mixed-integer recourse, see Louveaux and Schultz (2003) for the expectation-based and Märkert and Schultz (2005) for the risk-adjusted case.

For a more general and exhaustive introduction to two-stage stochastic programming, we refer the reader to Birge and Louveaux (1997), Kall and Wallace (1994) and Prekopa (1995).

1.3 Multi-stage stochastic programs with recourse

The *multi-stage stochastic program with recourse* relies on the same ideas as the two-stage version. Decisions are made without anticipating future realizations of uncertain data, which forces a partitioning of decisions into stages according to the information flow. The realization of uncertain data is, however, only gradually revealed and decisions are therefore made dynamically. Since non-anticipativity allows a temporary violation of the random constraints at a stage, feasibility is restored through recourse actions at the following stages at the expense of recourse costs. We assume that the overall aim is to minimize expected future costs.

Initially we formulate the problem by introducing measurability conditions to state the fact that decisions at a stage depend only on the available information

at this point in time.

$$\min\Big\{\mathbb{E}[c(\omega)x_1(\omega)+\cdots+c_T(\omega)x_T(\omega)] \;\Big|\; \sum_{t'\leq t} W_{tt'}(\omega)x_{t'}(\omega)=h_t(\omega),$$
$$t=2,\ldots,T, A_t(\omega)x_t(\omega)=b_t(\omega), x_t(\omega)\in X_t,$$
$$t=1,\ldots,T\ \ \mathbb{P}.a.a.\omega, x_t \text{ measurable w.r.t } \mathcal{F}_t\Big\}. \quad (1.3.1)$$

We consider a finite time horizon indexed by $\{1,\ldots,T\}$. Occasionally, we also refer to the set of time points as $\mathcal{T}$. For now, we let the time points index the stages. The stages represent points in time at which new information arrives and should not be confused with points of decision-making. However, to avoid highly complex notation throughout the rest of the thesis, we use only a single set of indices and leave it to the reader to extract the meaning from the context. X_t are non-empty polyhedra that may contain integrality restrictions on some of the variables $x_t\in\mathbb{R}^{n_t}, t=1,\ldots,T$. We let the variables x_t at time t depend on the realization ω of the random data and set $x^t=(x_1,\ldots,x_t)$. For $t=1,\ldots,T$, $c_t:\Omega\to\mathbb{R}^{n_t}$ and $h_t:\Omega\to\mathbb{R}^{m_t}, b_t:\Omega\to\mathbb{R}^{m_t'}$ are random vectors and $W_{tt'}:\Omega\to\mathbb{R}^{m_t\times n_{t'}}, t'\leq t, A_t:\Omega\to\mathbb{R}^{m_t'\times n_t}$ are random matrices on the probability space $(\Omega,\mathcal{F},\mathbb{P})$. We refer to the constraints determined by $W_{tt'}, t'\leq t, h_t, t=1,\ldots,T$ as coupling constraints and those determined by $X_t, A_t, b_t, t=1,\ldots,T$ as stage-specific constraints. If a solution satisfies both the coupling and stage-specific constraints, it is called admissible. For $t=1,\ldots,T$, we let $\xi_t:\Omega\to\mathbb{R}^{N_t}$ be the random vector $\xi_t=(c_t,h_t,b_t,W_{t1},\ldots,W_{tt},A_t)$ where, as in the remainder of the thesis, the matrices are to be read as listed in rows and where $N_t=n_t+m_t+m_t'+m_t\times n_1+\cdots+m_t\times n_t+m_t'\times n_t$. Information is described by the stochastic process $\{\xi_t\}_{t=1}^T$, and specifically information available at time t by $\xi^t=(\xi_1,\ldots,\xi_t)$. We denote by $\mathcal{F}_t\subseteq\mathcal{F}$ the σ-algebra generated by ξ^t and assume that the σ-algebras form a filtration such that $\mathcal{F}_t\subseteq\mathcal{F}_{t+1}, t=1,\ldots,T-1$ and $\mathcal{F}_1=\{\emptyset,\Omega\}$ and $\mathcal{F}_T=\mathcal{F}$. Non-anticipativity is expressed as measurability of x_t with respect to $\mathcal{F}_t$ which can also be expressed as $x_t=\mathbb{E}[x_t|\mathcal{F}_t]$, where $\mathbb{E}[\cdot|\cdot]$ denotes the condition expectation. Solutions that comply with the non-anticipativity are called implementable. Letting $\mathbb{P}_{\mathcal{F}_{t-1}}$ be a regular conditional probability measure on $\mathcal{F}_{t-1}\times\Omega$, we introduce the image measure $\mu_t=\mathbb{P}_{\mathcal{F}_{t-1}}\circ(\xi^t)^{-1}$ on $\mathbb{R}^{N_t}, t=1,\ldots,T$ and change variables.

The alternating decision process of decisions and observations of the random data, can be summarized in the scheme

$$\text{decide on } x_1\to\cdots\to\text{observe } c_t,h_t,b_t,W_{t1},\ldots,W_{tt},A_t\to$$
$$\text{decide on } x_t\to\cdots\to\text{decide on } x_T.$$

The dynamics are made clear in the formulation of the stochastic recourse program by the use of dynamic programming. To write problem (1.3.1) as a dynamic program, the multi-stage stochastic program with recourse is

$$\min\{Q_2(x^1, \xi^1) \mid A_1 x_1 = b_1, x_1 \in X_1\}, \tag{1.3.2}$$

with the recourse function

$$Q_t(x^{t-1}, \xi^{t-1}) := c_{t-1}x_{t-1} + \mathbb{E}[\Phi_t(x^{t-1}, \xi^t)|\mathcal{F}_{t-1}], t = 2, \ldots, T \tag{1.3.3}$$

$$Q_{T+1}(x^T, \xi^T) := c_T x_T \tag{1.3.4}$$

and the value function

$$\Phi_t(x^{t-1}, \xi^t) := \min\Big\{Q_{t+1}(x^t, \xi^t)\Big| \sum_{t'=1}^{t} W_{tt'}x_{t'} = h_t, \\ A_t x_t = b_t, x_t \in X_t\Big\}, t = 2, \ldots, T. \tag{1.3.5}$$

Due to the computational difficulties in (1.3.2)-(1.3.5), the probability distribution is mostly approximated by a discrete distribution with finite support. The approximation may result in a scenario formulation or a scenario tree formulation of the multi-stage stochastic program.

As concerns the *scenario formulation*, we assume the approximate distribution of the stochastic process $\{\xi_t\}_{t=1}^T = \{(c_t, h_t, b_t, W_{t1}, \ldots, W_{tt}, A_t)\}_{t=1}^T$ is given by the scenario paths $\{\xi_t^s\}_{t=1}^T = \{(c_t^s, h_t^s, b_t^s, W_{t1}^s, \ldots, W_{tt}^s, A_t^s)\}_{t=1}^T, s = 1, \ldots, S$ and the scenario probabilities $\pi^s, s = 1, \ldots, S$. Non-anticipativity is explicitly expressed as linear constraints that force decision variables to have the same value if they are based on the same information. This can be formulated by means of so-called scenario bundles. At each stage, non-anticipativity induces a partitioning of the scenarios. Two scenarios are said to be members of the same bundle $\mathcal{B}$ at time t if the scenarios contain the same information up to time t. In this fashion, every scenario s is a member of exactly one bundle $\mathcal{B}(s,t)$ at time t. Based on this, the scenario formulation takes the form

$$z = \min \sum_{s=1}^{S}\sum_{t=1}^{T} \pi^s c_t^s x_t^s \tag{1.3.6}$$

$$\text{s.t.} \sum_{t'=1}^{t} W_{tt'}^s x_{t'}^s = h_t^s, t = 2, \ldots, T, s = 1 \ldots, S \tag{1.3.7}$$

$$A_t^s x_t^s = b_t^s, x_t^s \in X_t, t = 1, \ldots, T, s = 1 \ldots, S \tag{1.3.8}$$

$$\text{if } \mathcal{B}(s_1, t') = \mathcal{B}(s_2, t'), t' \leq t, t', t = 1, \ldots, T, s_1, s_2 = 1 \ldots, S$$

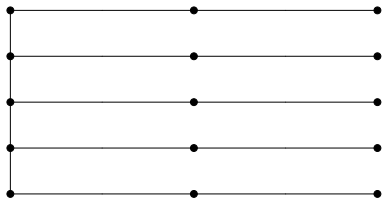


Figure 1.2: Multi-stage scenario paths.

$$\text{then } x_t^{s_1} = x_t^{s_2}, t = 1, \ldots, T, s_1, s_2 = 1 \ldots, S. \tag{1.3.9}$$

Unbundled scenario paths are shown in Fig. 1.2, in which the nodes again represent points of decison-making.

The *scenario tree formulation* arises when clustering the scenario paths to a scenario tree, so that branching occurs with the arrival of new information. In other words, decision variables at a stage are aggregated according to the available information. Decision variables that are based on the same information are replaced by a single variable and thereby automatically have the same values. Since information reveals only gradually, the aggregation of variables induces a tree structure. The non-anticipativity is implicitly given in this tree structure. The scenario tree is built of a set of nodes $\mathcal{N}$. We assume branching occurs at $t = 1, \ldots, T$, although, as already mentioned, the arrival of new information may not occur as often as decision-making. The root node corresponds to time interval $t = 1$. The remaining nodes all have an ascendant node and a set of descendant nodes. For node n, the immediate ascending node is termed n_{-1} with the transition probability $\pi^{n/n_{-1}}$, i.e. the probability that n is the descendant of n_{-1}. The probabilities of the nodes are given recursively by $\pi^1 = 1$ and $\pi^n = \pi^{n/n_{-1}}\pi^{n_{-1}}, n > 1$. The immediate descendants of node n are $\mathcal{N}_{+1}(n)$ and nodes with $\mathcal{N}_{+1}(n) = \emptyset$ are called leaves. Moreover, the path from the root node to node n is denoted by $path(n)$ and $t(n)$ is its length. $\mathcal{N}_t$ is the set $\{n \in \mathcal{N} : t(n) = t\}$ and nodes of $\mathcal{N}_T$ constitute the leaves. Each path from the root node to a leaf represents a scenario and hence the scenario probabilities are $\pi^n, n \in \mathcal{N}_T$. Conversely, given the scenario probabilities, the remaining node and transition probabilities are given by $\pi^n = \sum_{n_+ \in \mathcal{N}_+(n)} \pi^{n_+}$ and $\pi^{n_+/n} = \pi^{n_+}/\pi^n, n_+ \in \mathcal{N}_+(n)$. The ascendant node of node n at $t > 1$ time intervals back in time is n_{-t} with $t(n_{-t}) = t(n) - t$. Finally, for $t = 1, \ldots, T$ the realizations of the uncertain data $\{\xi_t\} = \{(c_t, h_t, b_t, W_{t1}, \ldots, W_{tt}, A_t)\}$ are denoted $\{\xi^n\}_{n \in \mathcal{N}_t} = \{(c^n, h^n, b^n, W^{n1}, \ldots, W^{nn}, A^n)\}_{n \in \mathcal{N}_t}$. Now the scenario tree formulation reads

$$z = \min \sum_{n \in \mathcal{N}} \pi^n c^n x^n \tag{1.3.10}$$

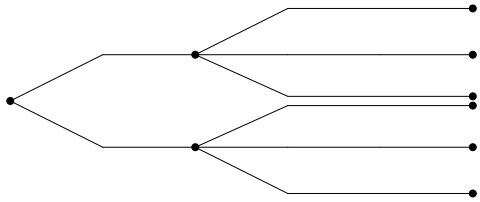


Figure 1.3: Multi-stage scenario tree.

$$\text{s.t.} \sum_{0 \le t' \le t(n)-1} W^{nn_{-t'}} x^{n_{-t'}} = h^n, n \in \mathcal{N}, \tag{1.3.11}$$

$$A^n x^n = b^n, x^n \in X_{t(n)}. \tag{1.3.12}$$

A scenario tree is illustrated in Fig. 1.3. The nodes represent points of decison-making and arrival of new information.

We extend the general formulation of the two-stage stochastic program to the multi-stage version so that the problem

$$\min\{Q(x) \mid Ax = b, x \in X\}$$

in general refers to a stochastic program. As previously, the constraints may not be explicitly displayed.

For an introduction to multi-stage stochastic programming, we again refer to the general textbooks on stochastic programming Birge and Louveaux (1997), Kall and Wallace (1994) and Prekopa (1995) as well as the specific paper on multi-stage stochastic mixed-integer linear programming terminology by Römisch and Schultz (2001). The last paper also provide a number of references on structural properties.

1.4 Solution approaches

Solution approaches to stochastic programming problems often divide into primal and dual decomposition methods. Primal methods aim at decomposing a problem according to stages, whereas dual methods decompose with respect to scenarios. We discuss some major contributions within solution approaches to linear and mixed-integer two-stage and multi-stage stochastic programs.

1.4.1 Two-stage linear programs

The starting point for many stochastic programming solution approaches is the *L-shaped method* introduced by Slyke and Wets (1969) and based on the principles

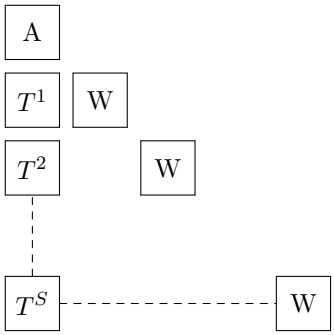


Figure 1.4: Constraint matrix of the deterministic equivalent to the two-stage problem.

of *Benders' decomposition*, cf. Benders (1962). In its basic form it applies to two-stage stochastic programs with linear recourse, i.e. with $X \subseteq \mathbb{R}^{n_1}_+$ and $Y = \mathbb{R}^{n_2}_+$, and assumes a discrete distribution with finite support, the result being a large-scale mathematical program of the form (1.2.5)-(1.2.6). The L-shaped method inherits its name from the structure of this problem or in particular its constraint matrix. Fig. 1.4 illustrates, that for fixed first-stage decisions, the second stage divides into a number of independent subproblems. Due to the stage-wise decomposition, the L-shaped method classifies as a primal decomposition approach.

Consulting (1.2.2)-(1.2.4), the two-stage stochastic program with recourse can be restated as

$$\min\ cx + \theta \tag{1.4.1}$$

$$\text{s.t. } \Phi(x, \xi^s) < +\infty, s = 1, \ldots, S \tag{1.4.2}$$

$$\sum_{s=1}^{S} \pi^s \Phi(x, \xi^s) \leq \theta \tag{1.4.3}$$

$$Ax = b, x \in X \tag{1.4.4}$$

$$\theta \in \mathbb{R}. \tag{1.4.5}$$

The L-shaped method fixes the first-stage decisions in a master problem that is constructed by relaxing the second-stage feasibility constraints (1.4.2) and the optimality constraint (1.4.3) and restoring the constraints by valid cuts.

Consider some iteration i of the algorithm and solve the master problem. If the problem is infeasible, so is the original problem and the algorithm terminates. A slight modification of the following analysis will suffice if the problem is unbounded. Finally, if neither infeasible nor unbounded, an optimal solution x^i of the master problem is found.

Feasibility cuts are derived from a number of linear subproblems defined for $s = 1, \ldots, S$ and $x \in \mathbb{R}^{n_1}$ by

$$\bar{\Phi}(x, \xi^s) := \min\{es_1 + es_2 \mid Wy + Is_1 - Is_2 = h^s - T^s x, \\ y \in \mathbb{R}^{n_2}_+, s_1, s_2 \in \mathbb{R}^{m_2}_+\}, \quad (1.4.6)$$

where $e = (1, \ldots, 1)$, I is the $m_2 \times m_2$ identity matrix and $s_1, s_2 \in \mathbb{R}^{m_2}_+$ are slack variables. Evidently, the subproblems can be solved separately. If the solution x^i causes the second-stage problem $\Phi(x^i, \xi^s)$ to be infeasible for some $s \in \{1, \ldots, S\}$, then $0 < \bar{\Phi}(x^i, \xi^s) = \sigma^{i,s}(h^s - T^s x^i)$ for a dual optimal solution $\sigma^{i,s}$ to $\bar{\Phi}(x^i, \xi^s)$. Thus, x^i will be cut off by adding the feasibility cut

$$\sigma^{i,s}(h^s - T^s x) \leq 0. \quad (1.4.7)$$

Feasibility cuts of the form (1.4.7) are added until $\Phi(x^i, \xi^s)$ are feasible for all $s = 1, \ldots, S$. By duality, (1.4.7) is a valid inequality for (1.4.1) – (1.4.5) for all $x \in \mathbb{R}^{n_1}$ that do not violate second-stage feasibility.

Having restored second-stage feasibility, the algorithm proceeds by solving the subproblems. If for some $s \in \{1, \ldots, S\}$, $\Phi(x^i, \xi^s)$ is unbounded, so is the original problem and the algorithm terminates. Otherwise, for $s = 1, \ldots, S$ let $\sigma^{i,s}$ be a dual optimal solution to $\Phi(x^i, \xi^s)$. If x^i is such that $\theta^i < \sum_{s=1}^{S} \pi^s \Phi(x^i, \xi^s) = \sum_{s=1}^{S} \pi^s \sigma^{i,s}(h^s - T^s x^i)$ for some $s \in \{1, \ldots, S\}$, then x^i is not optimal in the original problem. Hence, (x^i, θ^i) will be cut off by adding the optimality cut

$$\sum_{s=1}^{S} \pi^s \sigma^{i,s}(h^s - T^s x) \leq \theta \quad (1.4.8)$$

Again, by duality, the optimality cut is a valid inequality to (1.4.1) – (1.4.5). Having restored both feasibility and optimality, the algorithm terminates with an optimal solution to the original problem.

The algorithm can be stated as follows

Algorithm 1.4.1

Step 0 (Initialization). Set $i = 0$ and let the current master problem be

$$\min\ cx + \theta$$
$$s.t.\ \theta \in \mathbb{R}$$

Step 1 (Solve master problem). Set $i = i + 1$. *Solve the current master problem and let* $(x^i, \theta^{i,1}, \ldots, \theta^{i,S})$ *be an optimal solution (If* $\theta^s = -\infty$ *for some* $s \in \{1, \ldots, S\}$ *the variable is ignored in the computation.)*

Step 2 (Add feasibility cuts). For $s = 1, \ldots, S$, *solve problem* (1.4.6) *with* $x = x^i$ *and let* $\sigma^{i,s}$ *be a corresponding dual solution. If* $\sigma^{i,s}(h^s - T^s x^i) > 0$ *for some* $s = 1, \ldots, S$, *add a feasibility cut* (1.4.7) *to the master problem and return to step 2.*

Step 3 (Add optimality cuts). For $s = 1, \ldots, S$, *solve the problem* (1.2.4) *with* $x = x^i$ *and let* $\sigma^{i,s}$ *be a corresponding dual solution. If* $\sum_{s=1}^{S} \pi^s \sigma^{i,s}(h^s - T^s x^i) > \theta^i$, *add an optimality cut* (1.4.8) *to the master problem and return to step 2.*

Step 4 (Termination). Stop. The current solution is optimal.

If it exists, Algorithm 1.4.1 terminates with an optimal solution in a finite number of iterations. Otherwise, the algorithm proves unboundedness or infeasibility of the problem.

There is a different way of considering the use of cutting planes. Duality arguments may prove the recourse function (1.2.3) to be piecewise linear and convex. It is thus possible to build an outer linearization of the function and the optimality cuts can be regarded as supporting hyperplanes in this respect.

In contrast to the optimality cuts (1.4.8) that rely on aggregated information, Wets (1983) and Birge and Louveaux (1988) suggested the use of disaggregated cuts. The idea was to replace the single cut by a number of cuts derived from separate subproblems or so-called bunches of subproblems. Cuts derived from separate subproblems have the form

$$\sigma^{i,s}(h^s - T^s x) \leq \theta^s.$$

Since the disaggregated cuts contain more information, it is expected that the use of the so-called *multi-cut method* involves less iterations and often outperforms the traditional L-shaped method, which is supported by the numerical tests of Birge and Louveaux (1988). Further improvements to the L-shaped method in this direction include *regularized decomposition* proposed by Ruszczyński (1986). The method combines the multi-cut version of L-shaped decomposition with the inclusion of a quadratic regularizing objective function term, the resulting objective being

$$cx + \sum_{s=1}^{S} \pi^s \theta^s + 0.5\alpha \|x - x^{i-1}\|^2,$$

where $\alpha > 0$. This prevents initial solutions from oscillating and allows for cut removal in order to avoid final degeneracy in the master problem. The quadratic objective function term ensures strict convexity which provides for finite convergence of the algorithm.

Among other methods that emanate from the L-shaped method is *stochastic decomposition* by Higle and Sen (1991). The authors use an internal sampling procedure for approximating the probability distribution, and solve the subproblems at only one sample point. The cuts provided by the internal sampling procedure are statistical estimates that converge to the supporting hyperplanes of the original objective function. Finally, the most direct alternative decomposition approach is the method of Dantzig and Wolfe (1960). *Dantzig-Wolfe decomposition* can be regarded as solving the dual to the L-shaped master problem and uses, in contrast to outer linearization and cut generation, inner linearization and column generation. In most cases, the L-shaped method outperforms Dantzig-Wolfe decomposition due to smaller bases of the master problem.

1.4.2 Two-stage mixed-integer programs

We next discuss solution approaches to two-stage stochastic programs with mixed-integer recourse, i.e. problems on the form (1.2.5)-(1.2.6) with $X \subseteq \mathbb{R}^{n_1}_+$ and $Y \subseteq \mathbb{R}^{n_2}_+$ that may contain integrality restrictions on some variables. By virtue of the integrality, the convexity properties that apply to stochastic linear programs are lost, which makes stochastic mixed-integer program challenging from a computational point of view.

Independent of convexity, any mixed-integer linear stochastic program can be stated as its deterministic equivalent, the result being a large-scale problem amenable to *LP-based branch and bound*. The branch and bound may then be conducted by commercial software such as the CPLEX callable library, cf. Cplex Optimization Inc. (2006).

A number of attempts were made at adapting the L-shaped method to two-stage stochastic mixed-integer programs, which lead to a branch and cut procedure known as the *integer L-shaped method*. Laporte and Louveaux (1993) first proposed the derivation of cuts for two-stage programs with purely binary first stage. Based on general duality theory, Carøe and Tind (1998) later provided a full characterization of the integer L-shaped method and derived cuts for two-stage programs with integer first and second stage. Finally, Norkin, Pflug, and Ruszczyński (1998) and Norkin, Ermoliev, and Ruszczyński (1998) suggested the *stochastic branch and bound* principles using statistical estimates of the recourse function instead of ordinary evaluation.

As a contrast to the primal approaches, we present a dual solution approach. The approach is due to Carøe and Schultz (1999) and rests on an idea of variable splitting and Lagrangian relaxation that has won great attention. To present the

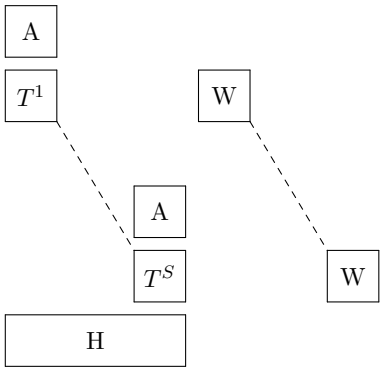


Figure 1.5: Constraint matrix of the two-stage problem with explicit non-anticipativity constraints.

so-called *dual decomposition*, we proceed as follows.

Defining for $s = 1, \ldots, S$ the sets

$$\chi^s = \{(x, y^s) \mid Ax = b, x \in X, Wy^s + T^s x = h^s, y^s \in Y^s\},$$

the deterministic equivalent (1.2.5)-(1.2.6) can be restated as

$$z = \min\Big\{cx + \sum_{s=1}^{S} \pi^s q^s y^s \;\Big|\; (x, y^s) \in \chi^s, s = 1, \ldots, S\Big\}. \tag{1.4.9}$$

We assume the problem is feasible and bounded. The variable splitting applies to the first-stage variables x and consists in the introduction of copies $x^s, s = 1, \ldots, S$. With such copies, non-anticipativity can be explicitly expressed and (1.4.9) is equivalent to

$$z = \min\Big\{\sum_{s=1}^{S} \pi^s (cx^s + q^s y^s)\Big|(x^s, y^s) \in \chi^s, s = 1, \ldots, S, x^1 = \cdots = x^S\Big\}. \tag{1.4.10}$$

We further assume that the non-anticipativity constraints are represented by the equality $\sum_{s=1}^{S} M^s x^s = 0$, where $= (M^1, \ldots, M^S)$ is a suitable $l \times n_1 S$ matrix. It should be remarked that except for the non-anticipativity constraints, the problem (1.4.10) decomposes according to scenarios. For an illustration of the structure of the constraint matrix and its decomposition potential, see Fig. 1.5. We relax the non-anticipativity constraints using Lagrangian relaxation. The Lagrangian

function is

$$L(x, y; \lambda) := \sum_{s=1}^{S} \big(\pi^s (cx^s + q^s y^s) + \lambda M^s x^s\big),$$

with the corresponding dual function

$$D(\lambda) := \min\{L(x, y; \lambda) \mid (x^s, y^s) \in \chi^s, s = 1, \ldots, S\}. \tag{1.4.11}$$

The Lagrangian dual is therefore

$$\max\{D(\lambda) \mid \lambda \in \mathbb{R}^l\}.$$

The Lagrangian relaxation decomposes into scenario subproblems, such that the dual function (1.4.11) is

$$D(\lambda) = \sum_{s=1}^{S} D^s(\lambda),$$

with

$$D^s(\lambda) = \min\{\pi^s (cx^s + q^s y^s) + \lambda M^s x^s \mid (x^s, y^s) \in \chi^s\}.$$

For now, we leave out further details on Lagrangian relaxation and state the branch and cut procedure.

Algorithm 1.4.2

Step 0 (Initialization). Set $\bar{z} = \infty$ and let the $\mathcal{L}$ consist of

$$\begin{aligned} &\min\ Q(x) \\ &s.t.\ Ax = b, x \in X \end{aligned}$$

Step 1 (Termination). If $\mathcal{L} = \emptyset$, then stop. The solution $\bar{x}$ with $\bar{z} = Q(\bar{x})$ is optimal.

Step 2 (Node selection). Select and delete a problem P from $\mathcal{L}$. If P is infeasible, go to step 1. Otherwise, solve the Lagrangian dual to obtain the lower bound $z(P)$ and go step 3.

Step 3 (Bounding). If $z(P) \geq \bar{z}$, then go to step 1. Otherwise,

(i) if the scenario solutions $x^s, s = 1, \ldots, S$ are identical, let $\bar{z} = \min_s\{\bar{z}, Q(x^s)\}$, delete from $\mathcal{L}$ all P' with $z(P') \geq \bar{z}$ and go to step 1.

(ii) if the scenario solutions $x^s, s = 1, \ldots, S$ differ, then compute the average $\bar{x}$ and round it. If $\bar{x}$ is feasible, let $\bar{z} = \min\{\bar{z}, Q(\bar{x})\}$, delete from $\mathcal{L}$ all P' with $z(P') \geq \bar{z}$ and go to step 4.

Step 4 (Branching). Select a component x_j of x and add two new problems to $\mathcal{L}$ obtained from P by adding the constraints $x_j \leq \lfloor \bar{x}_j \rfloor$ and $x_j \geq \lceil \bar{x}_j \rceil$, respectively (integer component) or $x_j \leq \bar{x}_j - \varepsilon$ and $x_j \geq \bar{x}_j + \varepsilon$, respectively (continuous component), where $\bar{x}$ is the average and $\varepsilon > 0$.

1.4.3 Multi-stage linear programs

The primal solution approach to two-stage linear programs, Benders' decomposition or the L-shaped method, extends to multi-stage linear programs. The *nested Benders decomposition* was suggested by Birge (1985) and applies to multi-stage stochastic programs with $X_t = \mathbb{R}^{n_t}_+, t = 1, \ldots, T$ and a discrete distribution with finite support, stated using the scenario tree formulation (1.3.10)-(1.3.12).

The algorithm relies on the dynamic programming formulation (1.3.2)-(1.3.4) and especially the definition of subproblems for every node $n \in \mathcal{N}$ of the scenario tree

$$\min c^n x^n + \theta^n \tag{1.4.12}$$

$$\text{s.t. } \Phi^{n_+}(x^n, \xi^{n_+}) < +\infty, n_+ \in \mathcal{N}_+(n) \tag{1.4.13}$$

$$\sum_{n_+ \in \mathcal{N}_+(n)} \pi^{n_+/n} \Phi^{n_+}(x^n, \xi^{n_+}) \leq \theta^n \tag{1.4.14}$$

$$\sum_{0 \leq t' \leq t(n)-1} W^{nn_{-t'}} x^{n_{-t'}} = h^n \tag{1.4.15}$$

$$A^n x^n = b^n, x^n \in X_{t(n)} \tag{1.4.16}$$

$$\theta^n \in \mathbb{R}. \tag{1.4.17}$$

A master problem is obtained for every node by relaxing the feasibility constraints (1.4.13) and the optimality constraint (1.4.14) and restoring the constraints by valid cuts derived from the descendant nodes. Both feasibility and optimality cuts are derived in the same fashion as for the two-stage case. The extension to multi-stage stochastic programs lies in determining the order of solving the subproblems and deriving the cuts. This order is determined by the directions, forward $DIR = FORE$ and backward $DIR = BACK$, in which the scenario tree is traversed. We state the algorithm as a so-called "fast-forward-fast-back" procedure

Algorithm 1.4.3

Step 0 (Initialization). Set $n = 1$ and $DIR = FORE$. Let the master problem of node 1 be

$$\begin{aligned} \min\ & c^1 x^1 + \theta^1 \\ s.t.\ & A^1 x^1 = b^1, x^1 \in X_{t(1)} \\ & \theta^1 \in \mathbb{R}_+ \end{aligned}$$

Step 1 (Solve master problem and add feasibility cuts). Solve the master problem of node n.

- *(i) If infeasible and $n = 1$, stop. The problem (1.3.2)-(1.3.4) is infeasible.*
- *(ii) If infeasible and $n > 1$, use the current node to derive a feasibility cut that is added to the master problem of the ascendant node n_{-1}. Set $DIR = BACK$, $n = n_{-1}$ and return to step 1.*
- *(iii) If feasible, let (x^n, θ^n) be an optimal solution (if $\theta^n = \infty$, the variable is ignored in the computation). If not all $n \in \mathcal{N}_t$ have been visited, then select a node that has not been visited and return to step 1. If all $n \in \mathcal{N}_t$ have been visited, $DIR = FORE$ and $t(n) < T$, then set $t = t + 1$ and return to step 1. If all $n \in \mathcal{N}_t$ have been visited, $DIR = BACK$ and $t(n) < T$, then go to step 2. If all $n \in \mathcal{N}_t$ have been visited and $t(n) = T$, then $DIR = BACK$ and go to step 2.*

Step 2 (Add feasibility cuts). For all $n \in \mathcal{N}_{t-1}$ do the following. Use the ascendant nodes $\mathcal{N}_+(n)$ to derive an optimality cut.

- *(i) If necessary, add the optimality cut to the master problem of node n, let $t = t - 1$ and go to step 1.*
- *(ii) If unnecessary to add an optimality cut and $t > 1$, then set $t = t - 1$ and select an $n \in \mathcal{N}_t$. If $t = 1$, then $DIR = FORE$. Return to step 1.*
- *(iii) If unnecessary to add an optimality cut and $t = 1$, then stop. The solution x^1 is optimal.*

With only few additional assumptions, the algorithm converges finitely. As for the two-stage L-shaped method, the speed of convergence may improve with the use of multi-cuts and regularization.

Like as traditional Benders' decomposition, nested Benders' decomposition relies on an outer linearization. Although this is generally preferred in the literature, inner linearizations have also been suggested. Other approaches to multi-stage linear programs are Lagrangian relaxation procedures such as *progressive hedging* and

augmented Lagrangian decomposition. For linear stochastic programs, both algorithms converge. The augmented Lagrangian decomposition approach by Mulvey and Ruszczyński (1995) rests on a diagonal quadratic approximation of the Lagrangian and a proposal to solve the resulting subproblems by an interior point method. In contrast to the progressive hedging algorithm, the non-anticipativity to be relaxed is determined by scenario branching. Since the progressive hedging applies more generally to multi-stage mixed-integer programs, we defer the discussion of this approach to the next section.

1.4.4 Multi-stage mixed-integer programs

To address mixed-integer recourse problems, i.e. problems with $X_t \subseteq \mathbb{R}^{n_t}_+, t = 1, \ldots, T$, where integrality restrictions may apply to some variables, we consider the scenario formulation (1.3.6)-(1.3.9).

To some extent, the dual decomposition approach is similar in spirit to the *progressive hedging* algorithm suggested by Rockafellar and Wets (1991). Both approaches are motivated by relaxation of the non-anticipativity constraints. The dual decomposition approach, however, automatically produces admissible solutions and resolves implementability by branch and bound, whereas progressive hedging iterates between admissible and implementable solutions. We briefly state the components of the progressive hedging approach.

Introduce the copies $(x^s_1, \ldots, x^s_T), s \in \mathcal{S}$ and divide the scenarios into bundles. For every bundle $\mathcal{B}$, set

$$\bar{x}^{\mathcal{B}}_t = \sum_{s:\mathcal{B}(s,t)=\mathcal{B}} \pi^s x^s_t \Big/ \sum_{s:\mathcal{B}(s,t)=\mathcal{B}} \pi^s$$

and let the non-anticipativity constraints be expressed as

$$\text{if } \mathcal{B}(s,t) = \mathcal{B} \text{ then } x^s_t = \bar{x}^{\mathcal{B}}_t.$$

Making non-anticipative decisions may be regarded as a means of "hedging" against uncertainty, thus the name progressive hedging.

Define for $s = 1, \ldots, S$ the sets

$$\chi^s = \Big\{(x^s_1, \ldots, x^s_T) \Big| \sum_{t'=1}^{t} W^s_{tt'} x^s_{t'} = h^s_t, t = 2, \ldots, T, \\ A^s_t x^s_t = b^s_t, x^s_t \in X_t, t = 2, \ldots, T\Big\}$$

and, motivated by the augmented Lagrangian relaxation, let

$$L(x;\lambda) := \sum_{s=1}^{S} \sum_{t=1}^{T} \pi^s \big(c^s_t x^s_t + \lambda^s_t x^s_t + 0.5\alpha(x^s_t - \bar{x}^{\mathcal{B}(s,t)}_t)^2\big).$$

Then the algorithm seeks to solve the quadratic problem

$$\min\{L(x;\lambda) \mid (x_1^s,\ldots,x_T^s) \in \chi^s, s = 1,\ldots,S\}$$

by decomposition it into the scenario subproblems given by

$$\min\Big\{\sum_{t=1}^{T}\big(c_t^s x_t^s + \lambda_t^s x_t^s + 0.5\alpha(x_t^s - \bar{x}_t^{\mathcal{B}(s,t)})^2\big)\Big|(x_1^s,\ldots,x_T^s) \in \chi^s\Big\}.$$

Now the progressive hedging algorithm reads

Algorithm 1.4.4

Step 0 (Initialization). Set $i = 0$. *Let* $(\lambda_1^{0,s},\ldots,\lambda_T^{0,s}) = 0, s = 1,\ldots,S$. *Solve the scenario subproblems without the augmenting term and let* $(x_1^s,\ldots,x_T^s), s = 1,\ldots,S$ *be an optimal solution. Compute the solution* $(\bar{x}_1^{0,\mathcal{B}(s,1)},\ldots,\bar{x}_T^{0,\mathcal{B}(s,T)}), s = 1,\ldots,S$. *Let the current problem be*

$$\begin{aligned}\min\ & \sum_{s=1}^{S}\sum_{t=1}^{T}\pi^s\big(c_t^s x_t^s + \lambda_t^{i-1,s}x_t^s + 0.5\alpha(x_t^s - \bar{x}_t^{i-1,\mathcal{B}(s,t)})^2\big)\\ s.t.\ & (x_1^s,\ldots,x_T^s) \in \chi^s, s = 1,\ldots,S\end{aligned}$$

Step 1 (Admissibility). Set $i = i + 1$. *Solve the current problem and let* $(x_1^{i,s},\ldots,x_T^{i,s}), s = 1,\ldots,S$ *be an optimal solution. The solution is admissible but not necessarily implementable.*

Step 2 (Implementability). Compute the solution $(\bar{x}_1^{i,\mathcal{B}(s,1)},\ldots,\bar{x}_T^{i,\mathcal{B}(s,T)}), s = 1,\ldots,S$. *The solution is implementable but not necessarily admissible.*

Step 3 (Termination). If some termination criteria are met, stop. Otherwise, go step 4.

Step 4 (Multiplier update). Let $\lambda_t^{i,s} = \lambda_t^{i-1,s} + \alpha(x_t^{i,s} - \bar{x}_t^{i,\mathcal{B}(s,t)}), t = 1,\ldots,T, s = 1,\ldots,S$ *and return to step 1.*

The progressive hedging algorithm ensures implementable solutions in all iterations and potentially convergence towards admissibility. As for possible termination criteria, iterations may be continued until a predefined limit is reached, implementable solutions remain nearly unchanged or the integer components of the solutions do not change from iteration to iteration.

Løkketangen and Woodruff (1996) have tested the performance of the progressive hedging algorithm and use tabu search for solving the quadratic scenario

subproblems close to optimality. The results are found to be very encouraging and indicate convergence in practice.

The authors of the related Lagrangian relaxation approach, *dual decomposition*, state the extension from two-stage problems to multi-stage problems, although they admit the extension may suffer from dimensionality problems. Lagrangian relaxation also finds its use in relation to certain coupling constraints. For suitable applications and further discussion, see Chapter 3.

Chapter 2

Deviation measures in two-stage stochastic linear programming

The present chapter addresses the inclusion of risk measures in two-stage stochastic recourse programs and the its impact on structural properties and algorithmic treatment.

As a starting point we consider a two-stage stochastic program with linear recourse. Whereas optimization in the traditional setting is based solely on expectation, we include risk measures that reflect dispersions of the random objective. Presenting the resulting mean-risk models, we aim to extend existing results for the expectation-based model. In particular, we discuss structural properties such as continuity, differentiability and convexity and address stability issues. Furthermore, we propose algorithmic treatment with a slight variation of the L-shaped method.

2.1 Introduction

Stochastic programming deals with optimization under uncertainty. Starting from a random optimization problem, the corresponding stochastic program depends heavily on the criteria for selecting an optimal solution. Traditionally, optimality rests on the expectation of the random objective. In many respects, however, it is appropriate to take risk into consideration. Combining expectation and risk, the model is referred to as a mean-risk model.

Risk measures treated in the literature encompass probabilities, dispersions and conditional expectations. Still, given the variety, no risk measure is unambiguously recommendable. For recent overviews on the topic, see e.g. Schultz (2003) on the probability of exceeding target, semideviation and conditional value at risk and Eichhorn and Römisch (2005) on risk measures within the class of

so-called polyhedral risk measures. A wide range of issues are covered, among these smoothness and convexity properties, compatibility with asymptotic results as well as algorithmic potential.

In this chapter we employ three dispersion measures; central deviation, semideviation and expected excess of target, all referred to as deviation measures. The motivation behind deviation measures is their consistency with stochastic dominance principles, an attractive behavior of the mean-risk models as well as their practical tractability. The above dispersion measures were already investigated in Märkert and Schultz (2005) in the case of mixed-integer linear programs and in Ahmed (2004) in the case of linear programs, the latter with emphasis on computational issues.

The idea is to extend existing results from the expectation-based framework to the mean-risk models considered here. Thus, we will aim at confirming that the models are well-posed, posses a number of useful analytical properties and are indeed in tune with stability results. As the deviation measures are based on piecewise linear operations they enable algorithmic treatment when the distribution of the random variables is discrete. Although the mean-risk models do not immediately possess the usual decomposable structure, computational accessibility by simple modifications of standard solution approaches is possible.

The chapter is organized as follows. In section 2.2 we extend the traditional linear two-stage stochastic recourse program to mean-risk models and put these into perspective with stochastic dominance. Section 2.3 contains prerequisites known from the expectation-based case and in section 2.4 and 2.5 similar structure and stability properties for the mean-risk models are analyzed. Algorithmic issues are presented in section 2.6.

2.2 The two-stage linear stochastic program

As should be clear from Chapter 1, a stochastic recourse program reflects a way of including uncertainty into optimization, the two-stage version being the most basic one. As the name indicates, decisions are made stage-wise. By non-anticipativity, first-stage decisions are to be taken independently of uncertain data, whereas the second stage allows for recourse actions when uncertainty has been disclosed. The aim of the stochastic program is to select first-stage decisions in an optimal way, optimality depending on which criterion is applied. To formalize this, as in Chapter 1, we are given the random linear program

$$\min\{cx + q(\omega)y \mid T(\omega)x + Wy = h(\omega), Ax = b, x \in X, y \in \mathbb{R}^{n_2}_+\}. \tag{2.2.1}$$

As before, $X \subseteq \mathbb{R}^{n_1}_+$ is a polyhedron, which we further assume is nonempty. Moreover, the costs $q : \Omega \to \mathbb{R}^{n_2}$ and the right-hand side $h : \Omega \to \mathbb{R}^{m_2}$ are random vectors and the technology matrix $T : \Omega \to \mathbb{R}^{m_2 \times n_1}$ is a random matrix on some

probability space $(\Omega, \mathcal{A}, \mathbb{P})$ as previously defined. The value function

$$\Phi(t_1, t_2) = \min\{t_1 y \mid Wy = t_2, y \in \mathbb{R}^{n_2}_+\} \tag{2.2.2}$$

is essential in the formulation of the corresponding stochastic program. According to the two-stage framework, the variables x and y of (2.2.1) are to be fixed before and after observing $h(\omega), q(\omega), T(\omega)$, respectively, and, therefore, the total costs of the sequential decision process compute as $cx + \Phi(q(\omega), h(\omega) - T(\omega)x)$. Finding an optimal $x \in \{x \in X \mid Ax = b\}$ may be understood as selecting the "best" random objective from the indexed family $(cx + \Phi(q(\cdot), h(\cdot) - T(\cdot)x))_{x \in X: Ax = b}$.

Considering the expectation-based criterion

$$Q_{\mathbb{E}}(x) := \int_\Omega (cx + \Phi(q(\omega), h(\omega) - T(\omega)x))\mathbb{P}(d\omega), \tag{2.2.3}$$

the traditional stochastic program is the optimization problem

$$\min\{Q_{\mathbb{E}}(x) \mid Ax = b, x \in X\}. \tag{2.2.4}$$

From a stochastic viewpoint, optimizing an expectation tacitly implies repeating the decision process several times and safety issues are addressed only inadequately. This has lead to the concept of mean-risk models. Here, we measure risk by quantitative deviations of the random objectives from either the mean or some preselected target. We introduce the central deviation

$$Q_{\mathcal{D}}(x) := \int_\Omega \big|cx + \Phi(q(\omega), h(\omega) - T(\omega)x) - Q_{\mathbb{E}}(x)\big|\mathbb{P}(d\omega), \tag{2.2.5}$$

the semideviation

$$Q_{\mathcal{D}^+}(x) := \int_\Omega \max\big\{cx + \Phi(q(\omega), h(\omega) - T(\omega)x) - Q_{\mathbb{E}}(x), 0\big\}\mathbb{P}(d\omega) \tag{2.2.6}$$

and the expected excess of a given target $\eta \in \mathbb{R}$

$$Q_{\mathcal{D}^\eta}(x) := \int_\Omega \max\big\{cx + \Phi(q(\omega), h(\omega) - T(\omega)x) - \eta, 0\big\}\mathbb{P}(d\omega). \tag{2.2.7}$$

Accordingly, problem (2.2.4) extends into the mean-risk model

$$\min\{Q_{\mathbb{E}}(x) + \varrho Q_{\mathcal{R}}(x) \mid Ax = b, x \in X\}, \tag{2.2.8}$$

where $\varrho \in \mathbb{R}_+$ is a suitable weight factor and $Q_{\mathcal{R}}$ is the risk term, i.e. $\mathcal{R} = \mathcal{D}$, $\mathcal{R} = \mathcal{D}^+$ or $\mathcal{R} = \mathcal{D}^\eta$.

In having to select the "best" of a family of random variables, the stochastic dominance approach deserves attention. Although allowing a simple trade-off

analysis, mean-risk models are unable to capture the entire gamut of risk-averse preferences. Nevertheless, for the above deviation measures, the mean-risk approach is consistent with second order stochastic dominance (provided that certain conditions on ϱ are met). For some aspects of stochastic dominance, see Ogryczak and Ruszczyński (1999) and Ogryczak and Ruszczyński (2001).

We briefly outline stochastic dominance results for the deviation measures (2.2.5)–(2.2.7). Considering the random variables $f(x,\cdot) = cx + \Phi(q(\cdot), h(\cdot) - T(\cdot)x), x \in \{x \in X \mid Ax = b\}$, stochastic dominance suggests a partial ordering by point-wise comparisons of performance functions constructed from the distribution functions. Relevant performance functions are

$$F_x^{(1)}(z) := \mathbb{P}(\{\omega \in \Omega \mid f(x,\omega) \leq z\}), \quad z \in \mathbb{R},$$
$$F_x^{(2)}(z) := \int_z^{+\infty} (1 - F_x^{(1)}(z'))dz', \quad z \in \mathbb{R}.$$

As smaller outcomes are preferred over larger, the relation of second degree stochastic dominance is defined as follows

$$f(x_1,\cdot) \succ_{(2)} f(x_2,\cdot) \iff F_{x_1}^{(2)}(z) \leq F_{x_2}^{(2)}(z), \quad \forall z \in \mathbb{R}.$$

Recall that $Q_{\mathbb{E}}(x)$ and $Q_{\mathcal{R}}(x)$ denote the mean and the risk of $f(x,\cdot)$. Now the mean-risk model is said to be consistent with second degree stochastic dominance if

$$f(x_1,\cdot) \succ_{(2)} f(x_2,\cdot) \Rightarrow Q_{\mathbb{E}}(x_1) + \varrho Q_{\mathcal{R}}(x_1) \leq Q_{\mathbb{E}}(x_2) + \varrho Q_{\mathcal{R}}(x_2),$$

i.e. if the mean-risk model inherits a ranking already existing with respect to stochastic dominance. Provided the random variables $f(x,\cdot), x \in \{x \in X \mid Ax = b\}$ have finite first moments, the mean-risk models resulting from $\mathcal{R} = \mathcal{D}$, $\mathcal{R} = \mathcal{D}^+$ and $\mathcal{R} = \mathcal{D}^\eta$ are indeed consistent with second degree stochastic dominance when $\varrho \in [0, 1/2]$, $\varrho \in [0,1]$ and for all $\varrho \geq 0$, respectively, cf. Ogryczak and Ruszczyński (1999), proposition 7.

2.3 Prerequisites

Again, as previously, let $\xi : \Omega \to \mathbb{R}^N$ be the random vector whose components constitute the random data, i.e. $\xi = (q, h, T_1, \ldots, T_{m_2})$, where $T_1, \ldots, T_{m_2}$ denote the rows of T and $N = n_2 + m_2 + m_2 \times n_1$. Moreover, let $\mathcal{P}(\mathbb{R}^N)$ represent the set of all probability measures on $\mathbb{R}^N$ and introduce from this set the image measure $\mu = \mathbb{P} \circ \xi^{-1}$. Changing variables in (2.2.3),

$$Q_{\mathbb{E}}(x,\mu) = \int_{\mathbb{R}^N} f(x,\xi)\mu(d\xi) \tag{2.3.1}$$

and similarly for (2.2.5)–(2.2.7), where $f(x,\xi) = cx + \Phi(q, h - Tx)$ for $x \in \mathbb{R}^{n_1}$.

As the results below concern continuity of the objective function with respect to x alone, μ alone and (x,μ) jointly, dependence of both x and μ is marked explicitly.

We will impose the assumptions

(A1) (Complete recourse) For all $t \in \mathbb{R}^{m_2}$, there exists a $y \in \mathbb{R}^{n_2}_+$ such that $Wy = t$.

(A2) (Dual feasibility) For μ-almost all $\xi \in \mathbb{R}^N$, there exists a $\sigma \in \mathbb{R}^{m_2}$ such that $\sigma W \leq q$.

(A3) (Finite second moment) $\int_{\mathbb{R}^N} \|\xi\|^2 \mu(d\xi) < +\infty$.

The assumptions $(A1)$ and $(A2)$ ensure feasibility and boundedness of problem $\Phi(q, h - Tx)$ for all $x \in \mathbb{R}^{n_1}$ and μ-almost all $\xi \in \mathbb{R}^N$ and are sufficient to establish certain properties of the value function Φ.

The following basis decomposition theorem applies, cf. Walkup (1969),

Proposition 2.3.1 *Let* $\operatorname{pos} W := \{t \in \mathbb{R}^{m_2} \mid \exists y \in \mathbb{R}^{n_2}_+ : Wy = t\}$ *and* $D := \{t \in \mathbb{R}^{n_2} \mid \exists \sigma \in \mathbb{R}^{m_2} : \sigma W \leq t\}$. *Then* $\Phi : D \times \operatorname{pos} W \to \mathbb{R}$ *is a real-valued continuous function. In addition, there exist* $\mathbb{R}^{m_2} \times \mathbb{R}^{n_2}$ *matrices* $B_j, j = 1, \ldots, J$ *and full dimensional cones* $\mathcal{K}_j, j = 1, \ldots, J$ *such that*

$$\cup_{j=1}^{J} \mathcal{K}_j = D \times \operatorname{pos} W, \quad \operatorname{int} \mathcal{K}_i \cap \operatorname{int} \mathcal{K}_j = \emptyset, \quad i \neq j,$$

and

$$\Phi(t_1, t_2) = t_1 B_j^{-1} t_2 \quad \forall (t_1, t_2) \in \mathcal{K}_j.$$

For fixed $t_1 \in \mathbb{R}^{n_2}$, *the function* $\Phi(t_1, \cdot)$ *is convex. For fixed* $t_2 \in \mathbb{R}^{m_2}$, *the function* $\Phi(\cdot, t_2)$ *is concave.*

Prerequisites comprise results on the expectation-based model (2.2.4). Lipschitz estimates of the integrand of (2.3.1) can be derived from Proposition 2.3.1 and are found in Römisch (2003).

Proposition 2.3.2 *Assume (A1)–(A2). Then there exist constants* $L_1, L_2, K_1 > 0$ *such that for all* $x, x_1, x_2 \in \mathbb{R}^{n_1}$ *and* μ*-almost all* $\xi, \xi_1, \xi_2 \in \mathbb{R}^N$

(i) $|f(x,\xi_1) - f(x,\xi_2)| \leq L_1 \|x\| \max\{1, \|\xi_1\|, \|\xi_2\|\} \|\xi_1 - \xi_2\|$,

(ii) $|f(x_1,\xi) - f(x_2,\xi)| \leq L_2 \max\{1, \|\xi\|^2\} \|x_1 - x_2\|$,

(iii) $|f(x,\xi)| \leq K_1 \|x\| \max\{1, \|\xi\|^2\}$.

From Wets (1974) and Kall (1976), we get

Proposition 2.3.3 *Assume (A1)–(A3).*

(i) Then $Q_{\mathbb{E}}(\cdot,\mu):\mathbb{R}^{n_1}\to\mathbb{R}$ is a real-valued, convex and Lipschitzian function.

(ii) Suppose further that μ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^N$. Then $Q_{\mathbb{E}}(\cdot,\mu):\mathbb{R}^{n_1}\to\mathbb{R}$ is continuously differentiable.

Remark 2.3.1 *In the case where only the right-hand side is random, results on strong and strict convexity appear in Schultz (1994). Let $q\in\mathbb{R}^{n_2}$ be a known vector, $h:\Omega\to\mathbb{R}^{m_2}$ be a random vector with the corresponding image measure $\mu=\mathbb{P}\circ h^{-1}$. Moreover, let $\bar{Q}_{\mathbb{E}}(\chi,\mu):=\int_{\mathbb{R}^{m_2}}\Phi(q,h-\chi)\mu(dh)$ be a function of the tender variable χ. (i) Assume (A1) and that μ has finite first moment and is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^{m_2}$. Suppose further that there exists a $\sigma\in\mathbb{R}^{m_2}$ such that $\sigma W<q$ component-wise. Then $\bar{Q}_{\mathbb{E}}(\cdot,\mu)$ is strictly convex on any open convex subset $V\subseteq\mathbb{R}^{m_2}$ of the support of μ. (ii) Assume further that there exist a convex open set $V\subseteq\mathbb{R}^{m_2}$, constants $c_1>0, c_2>0$ and a density g of μ such that $g(z)\geq c_1$ for all $z\in\mathbb{R}^{m_2}$ with $dist(z,V):=\inf\{\|z-v\|\mid v\in V\}\leq c_2$. Then $\bar{Q}_{\mathbb{E}}(\cdot,\mu)$ is strongly convex on V.*

The remaining continuity properties are relevant to stability results of the stochastic programs. Such results divide into qualitative and quantitative results. When qualitative stability is brought into focus, joint continuity is of crucial importance. A notion of convergence of probability measures is required, and weak convergence will prove sufficiently general while it still allows for substantial statements. A sequence $\{\mu_n\}\subseteq\mathcal{P}(\mathbb{R}^N)$ is said to converge weakly to $\mu\in\mathcal{P}(\mathbb{R}^N)$, i.e. $\mu_n\xrightarrow{w}\mu$, if for any bounded continuous function $h:\mathbb{R}^N\to\mathbb{R}$, it holds that

$$\int_{\mathbb{R}^N}h(\xi)\mu_n(d\xi)\to\int_{\mathbb{R}^N}h(\xi)\mu(d\xi),$$

cf. Billingsley (1968). Restricting measures to the set

$$\Delta_{r,K}(\mathbb{R}^N):=\Big\{\nu\in\mathcal{P}(\mathbb{R}^N)\Big|\int_{\mathbb{R}^N}\|\xi\|^r\nu(d\xi)\leq K\Big\},$$

we get the following, cf. Kall (1987) and Robinson and Wets (1987),

Proposition 2.3.4 *Assume (A1)–(A2) and let $\mu\in\Delta_{r,K}(\mathbb{R}^N)$ for some $r>2$, $K>0$. Then $Q_{\mathbb{E}}:\mathbb{R}^{n_1}\times\Delta_{r,K}(\mathbb{R}^N)\to\mathbb{R}$ is jointly continuous at (x,μ).*

To arrive at quantitative stability results, we consider a so-called distance with d-structure, given as a uniform distance between expectations of functions from a class of measurable functions. Working with functions that share analytical

properties with the integrand of (2.3.1), we obtain an ideal metric, optimally adjusted to the model, cf. Römisch (2003). Let $\mathcal{F}_2(\mathbb{R}^N) := \{F : \mathbb{R}^N \to \mathbb{R} \mid |F(\xi_1) - F(\xi_2)| \leq c_2(\xi_1, \xi_2)\|\xi_1 - \xi_2\| \ \forall \xi_1, \xi_2 \in \mathbb{R}^N\}$ denote the class of locally Lipschitz continuous functions with constant $c_2(\xi_1, \xi_2) := \max\{1, \|\xi_1\|, \|\xi_2\|\}$ and $\mathcal{P}_2(\mathbb{R}^N) := \{\nu \in \mathcal{P}(\mathbb{R}^N) \mid \int_{\mathbb{R}^N} \|\xi\|^2 \nu(d\xi) < +\infty\}$ the set of measures having finite second moments. For $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^N)$ define the pseudo-metric

$$d_2(\mu, \nu) := \sup_{F \in \mathcal{F}_2(\mathbb{R}^N)} \Big| \int_{\mathbb{R}^N} F(\xi_1)\mu(d\xi_1) - \int_{\mathbb{R}^N} F(\xi_2)\nu(d\xi_2) \Big|.$$

The pseudo-metric is referred to as the Fortet-Mourier metric of second order. The Lipschitz estimates of $Q_{\mathbb{E}}(x, \cdot)$ for all $x \in \mathbb{R}^{n_1}$ can be found in Römisch (2003).

Proposition 2.3.5 *Assume (A1)–(A2) and let $\{x \in X \mid Ax = b\}$ be nonempty and bounded. Then there exists a constant $L > 0$ such that the estimate*

$$sup_{x \in X : Ax = b} |Q_{\mathbb{E}}(x, \mu) - Q_{\mathbb{E}}(x, \nu)| \leq L d_2(\mu, \nu)$$

is valid whenever $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^N)$.

2.4 Structure

The following identities are obtained by straightforward computation, cf. Märkert and Schultz (2005),

$$\begin{aligned} Q_{\mathbb{E}}(x, \mu) + \varrho Q_{\mathcal{D}}(x, \mu) &= \mathbb{E}[f(x, \xi)] + \varrho \mathbb{E}[|f(x, \xi) - \mathbb{E}[f(x, \xi)]|] \\ &= (1 - 2\varrho)\mathbb{E}[f(x, \xi)] + 2\varrho \mathbb{E}[\max\{f(x, \xi), \mathbb{E}[f(x, \xi)]\}], \qquad (2.4.1) \\ Q_{\mathbb{E}}(x, \mu) + \varrho Q_{\mathcal{D}^+}(x, \mu) &= \mathbb{E}[f(x, \xi)] + \varrho \mathbb{E}[\max\{f(x, \xi) - \mathbb{E}[f(x, \xi)], 0\}] \\ &= (1 - \varrho)\mathbb{E}[f(x, \xi)] + \varrho \mathbb{E}[\max\{f(x, \xi), \mathbb{E}[f(x, \xi)]\}], \qquad (2.4.2) \\ Q_{\mathbb{E}}(x, \mu) + \varrho Q_{\mathcal{D}^\eta}(x, \mu) &= \mathbb{E}[f(x, \xi)] + \varrho \mathbb{E}[\max\{f(x, \xi) - \eta, 0\}] \\ &= \mathbb{E}[f(x, \xi)] + \varrho \mathbb{E}[\max\{f(x, \xi), \eta\}] - \varrho\eta. \qquad (2.4.3) \end{aligned}$$

Evidently, the mean-risk objective functions are linear combinations of $Q_{\mathbb{E}}(x, \mu)$ and either

$$Q_{\max}(x, \mu) := \int_{\mathbb{R}^N} \max\{f(x, \xi), Q_{\mathbb{E}}(x, \mu)\}\mu(d\xi) \qquad (2.4.4)$$

or the more simple version

$$Q_{\max,\eta}(x, \mu) := \int_{\mathbb{R}^N} \max\{f(x, \xi), \eta\}\mu(d\xi). \qquad (2.4.5)$$

Remark 2.4.1 *The integrals* (2.4.4) *and* (2.4.5) *make sense for measurable functions only. Let therefore* $x \in \mathbb{R}^{n_1}$. *Since* $\Phi(t_1, t_2)$ *is continuous in* (t_1, t_2) *and* $(q, h - Tx)$ *is linear in* ξ, $f(x, \xi) = cx + \Phi(q, h - Tx)$ *is continuous in* ξ. *By another continuity argument,* $\max\{f(x, \cdot), Q_{\mathbb{E}}(x, \mu)\}$ *and* $\max\{f(x, \cdot), \eta\}$ *are measurable.*

The purpose of this section is to obtain continuity, differentiability and convexity properties of the mean-risk objective functions. Keeping the reformulations (2.4.1)–(2.4.3) in mind and having examined the behavior of $Q_{\max}$, $Q_{\max,\eta}$ and $Q_{\mathbb{E}}$, the desired results will follow immediately. The following investigations concern $Q_{\max}$ only, since the results carry over to $Q_{\max,\eta}$ in a similar fashion and the structural properties of $Q_{\mathbb{E}}$ were already given in Section 2.3

We start by giving Lipschitz estimates of the integrand.

Proposition 2.4.1 *Assume (A1)–(A3). Then for the constants* $L_1, L_2, K_1 > 0$ *of Proposition 2.3.2 and for all* $x, x_1, x_2 \in \mathbb{R}^{n_1}$ *and* μ*-almost all* $\xi, \xi_1, \xi_2 \in \mathbb{R}^N$

(i) $|\max\{f(x, \xi_1), Q_{\mathbb{E}}(x, \mu)\} - \max\{f(x, \xi_2), Q_{\mathbb{E}}(x, \mu)\}| \leq$
$L_1 \|x\| \max\{1, \|\xi_1\|, \|\xi_2\|\}\|\xi_1 - \xi_2\|$,

(ii) $|\max\{f(x_1, \xi), Q_{\mathbb{E}}(x_1, \mu)\} - \max\{f(x_2, \xi), Q_{\mathbb{E}}(x_2, \mu)\}| \leq$
$\max\{L_2 \max\{1, \|\xi\|^2\}, L_{\mathbb{E}}\}\|x_1 - x_2\|$,
where $L_{\mathbb{E}}$ *denotes the Lipschitz constant of* $Q_{\mathbb{E}}(\cdot, \mu)$,

(iii) $|\max\{f(x, \xi), Q_{\mathbb{E}}(x, \mu)\}| \leq K_1 \|x\| \max\{1, \|\xi\|^2\} + |Q_{\mathbb{E}}(x, \mu)|$.

Proof. Apply Proposition 2.3.2 and the following lemma from Donchev (1986): Let $g_i, i = 1, \ldots, I$ be Lipschitzian functions with constants $L_i, i = 1, \ldots, I$ and let g be defined as some continuous selection of $\{g_i\}_{i=1}^I$. Then g is a Lipschitzian function with constant $L = \max_{i=1,\ldots,I}\{L_i\}$. □

Proposition 2.4.2 *Assume (A1)–(A3). Then* $Q_{\max}(\cdot, \mu) : \mathbb{R}^{n_1} \to \mathbb{R}$ *is a real-valued, convex and Lipschitzian function.*

Proof. Throughout the proof we apply proposition 2.4.1. For $x \in \mathbb{R}^{n_1}$,

$$\begin{aligned} |Q_{\max}(x, \mu)| &\leq \mathbb{E}[|\max\{f(x, \xi), Q_{\mathbb{E}}(x, \mu)\}|] \\ &\leq K_1 \|x\|(1 + \mathbb{E}[\|\xi\|^2]) + |Q_{\mathbb{E}}(x, \mu)| < +\infty, \end{aligned}$$

which verifies that $Q_{\max}(\cdot, \mu)$ is real-valued.

Since $\Phi(t_1, \cdot)$ is convex for all $t_1 \in \mathbb{R}^{n_2}$ and $h - Tx$ is linear in x, $f(x, \xi) := cx + \Phi(q, h - Tx)$ is convex in x. Moreover, as both the maximum and the expectation operators preserve convexity, $Q_{max}(\cdot, \mu)$ is convex.

Denote the Lipschitz constant of $Q_{\mathbb{E}}(\cdot,\mu)$ by $L_{\mathbb{E}}$. For $x_1, x_2 \in \mathbb{R}^{n_1}$,

$$\begin{aligned}|Q_{\max}(x_1,\mu) - Q_{\max}(x_2,\mu)| \\ &\leq \mathbb{E}[|\max\{f(x_1,\xi), Q_{\mathbb{E}}(x_1,\mu)\} - \max\{f(x_2,\xi), Q_{\mathbb{E}}(x_2,\mu)\}|] \\ &\leq \mathbb{E}[\max\{L_2 \max\{1, \|\xi\|^2\}, L_{\mathbb{E}}\}]\|x_1 - x_2\|,\end{aligned}$$

where $\mathbb{E}[\max\{L_2 \max\{1, \|\xi\|^2\}, L_{\mathbb{E}}\}] \leq L_2(1 + \mathbb{E}[\|\xi\|^2]) + L_{\mathbb{E}} < +\infty$, and we have established Lipschitz continuity of $Q_{\max}(\cdot,\mu)$. □

Remark 2.4.2 *Together, continuity of $Q_{\mathbb{E}}(\cdot,\mu) + \varrho Q_{\mathcal{R}}(\cdot,\mu)$ and compactness of $\{x \in X \mid Ax = b\}$ ensure that the minimum in* (2.2.8) *is attained.*

Remark 2.4.3 *For $\mathcal{R} = \mathcal{D}$, let $\varrho \in [0, 1/2)$, for $\mathcal{R} = \mathcal{D}^+$, let $\varrho \in [0, 1)$ and for $\mathcal{R} = \mathcal{D}^\eta$ let $\varrho \geq 0$. Consider the reformulations* (2.4.1)–(2.4.3)*. Then $Q_{\mathbb{E}}(\cdot,\mu) + \varrho Q_{\mathcal{R}}(\cdot,\mu)$ is seen to be convex. In the case of only a random right-hand side, both strict and strong convexity is inherited, cf. Remark 2.3.1. To see this, define $Q_{\mathbb{E}}(\chi,\mu)$ as above and define the risk measures $Q_{\mathcal{R}}(\chi,\mu)$ similarly. Provided that the assumptions of the expectation-based case are fulfilled, the mean-risk objective $Q_{\mathbb{E}}(\cdot,\mu) + \varrho Q_{\mathcal{R}}(\cdot,\mu)$ is strictly and strongly convex.*

Example 1 The following example demonstrates that the function $Q_{\mathbb{E}}(\cdot,\mu) + \varrho Q_{\mathcal{D}^+}(\cdot,\mu)$ is in general nonconvex for $\varrho > 1$. Let $\Phi(h - x) = \min\{y \mid y \geq h - x, y \geq 0\} = \max\{h - x, 0\}$ and $f(x,h) = x + \max\{h - x, 0\}$. Assume $\mu(h = 2) = 1/2$, $\mu(h = 3) = 1/2$ and $\varrho = 3$. Now $Q_{\mathbb{E}}(x,\mu) + \varrho Q_{\mathcal{D}^+}(x,\mu) = x - \frac{1}{4}\max\{2 - x, 0\} + \frac{5}{4}\max\{3 - x, 0\}$. Consider the points $x_1 = -1$, $x_2 = 3$ and a convex combination of the two, $x = \frac{1}{2}x_1 + \frac{1}{2}x_2 = 1$. The relation $Q_{\mathbb{E}}(x,\mu) + \varrho Q_{\mathcal{D}^+}(x,\mu) = \frac{13}{4} > \frac{25}{8} = \frac{1}{2}(Q_{\mathbb{E}}(x_1,\mu) + \varrho Q_{\mathcal{D}^+}(x_1,\mu)) + \frac{1}{2}(Q_{\mathbb{E}}(x_2,\mu) + \varrho Q_{\mathcal{D}^+}(x_2,\mu))$ shows that $Q_{\mathbb{E}}(\cdot,\mu) + \varrho Q_{\mathcal{D}^+}(\cdot,\mu)$ fails to be convex. The same data refute convexity of $Q_{\mathbb{E}}(\cdot,\mu) + \varrho Q_{\mathcal{D}}(\cdot,\mu)$ for $\varrho > 1/2$.

Proposition 2.4.3 *Assume (A1)–(A3) and that μ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^N$. Then $Q_{max}(\cdot,\mu) : \mathbb{R}^{n_1} \to \mathbb{R}$ is continuously differentiable.*

Proof. Let $x \in \mathbb{R}^{n_1}$ and denote by $E(x)$ the set of points (q,h,T) in which $\Phi(q, h - Tx)$ is non-differentiable in x. It is shown by Kall (1976) that $\mu(E(x)) = 0$. Hence, $\Phi(q, h - Tx)$ is differentiable in x for μ-almost all $\xi \in \mathbb{R}^N$ and the same holds for $f(x,\xi)$. Moreover, by Proposition 2.3.3, $Q(x,\mu)$ is differentiable. Define the sets

$$E_>(x) := \{\xi \in \mathbb{R}^N \mid f(x,\xi) > Q_{\mathbb{E}}(x,\mu)\},$$

$$E_<(x) := \{\xi \in \mathbb{R}^N \mid f(x,\xi) < Q_{\mathbb{E}}(x,\mu)\},$$
$$E_=(x) := \{\xi \in \mathbb{R}^N \mid f(x,\xi) = Q_{\mathbb{E}}(x,\mu)\}.$$

Then, if $\bigtriangledown_x$ denotes the gradient in x,

$$\bigtriangledown_x \max\{f(x,\xi), Q_{\mathbb{E}}(x,\mu)\} = \begin{cases} \bigtriangledown_x f(x,\xi), & \text{if } \xi \in E_>(x)\backslash E(x) \\ \bigtriangledown_x Q_{\mathbb{E}}(x,\mu), & \text{if } \xi \in E_<(x). \end{cases} \tag{2.4.6}$$

Observe that, according to Proposition 2.3.1,

$$\Phi(q, h - Tx) = \sum_{j=1}^{J} qB_j^{-1}(h - Tx)1_{\mathcal{K}_j(x)}(q,h,T),$$

since $\cup_{j=1}^{J}\mathcal{K}_j(x) = \mathbb{R}^N$ and $\mathcal{K}_i(x) \cap \mathcal{K}_j(x) = \emptyset, i \neq j$, where $\mathcal{K}_j(x) := \{(q,h,T) \mid (q, h - Tx) \in \mathcal{K}_j \backslash \cup_{i=1}^{j-1} \mathcal{K}_i\}$. Hence, the following inclusion applies

$$\begin{aligned} E_=(x) :&= \left\{(q,h,T)\middle| cx + \sum_{j=1}^{J} qB_j^{-1}(h - Tx)1_{\mathcal{K}_j(x)}(q,h,T) = Q_{\mathbb{E}}(x,\mu)\right\} \\ &\subseteq \cup_{j=1}^{J} H_j(x), \end{aligned}$$

where

$$H_j(x) := \{(q,h,T) \mid cx + qB_j^{-1}(h - Tx) = Q_{\mathbb{E}}(x,\mu)\}, j = 1, \ldots, J.$$

For $j = 1, \ldots, J$, we aim to show that $H_j(x)$ has Lebesgue measure zero. Note that

$$H_j(x) = H_j^1(x) \cup H_j^2(x),$$

where

$$H_j^1(x) := \{(q,h,T) \mid q \neq 0,\ \ cx + qB_j^{-1}(h - Tx) = Q_{\mathbb{E}}(x,\mu)\}$$

and

$$H_j^2(x) := \{(0,h,T) \mid cx = Q_{\mathbb{E}}(x,\mu)\}.$$

Letting

$$\begin{aligned} H_j^1(x,q,T) :&= \{h \mid (q,h,T) \in H_j^1(x)\} \\ &= \{h \mid q \neq 0,\ \ qB_j^{-1}h = qB_j^{-1}Tx - cx + Q_{\mathbb{E}}(x,\mu)\}, \end{aligned}$$

it is clear that this is a hyperplane and, thus, has Lebesgue measure zero. Now as $H_j^1(x,q,T)$ is a section of $H_j^1(x)$, $H_j^1(x)$ has Lebesgue measure zero, cf. Theorem 0.25 by Kall (1976). Also, either

$$H_j^2(x) = \emptyset \text{ or } H_j^2(x) = \{0\} \cup \mathbb{R}^{m_2+m_2\times n_1}.$$

In both cases, $H_j^2(x)$ has a dimension lower than $N = n_2 + m_2 + m_2 \times n_1$ and, therefore, has Lebesgue measure zero. Brought together, $H_j(x)$ has Lebesgue measure zero and since μ is absolute continuous with respect to the Lebesgue measure, $\mu(H_j(x)) = 0$. As a consequence, $\max\{f(x,\xi), Q_{\mathbb{E}}(x,\mu)\}$ is differentiable with gradient (2.4.6) for all $\xi \in \mathbb{R}^N \backslash (E(x) \cup E_=(x))$, where $\mu(E(x) \cup E_=(x)) = 0$.

From Proposition 2.4.1,

$$|\max\{f(x_1,\xi), Q_{\mathbb{E}}(x_1,\mu)\} - \max\{f(x_2,\xi), Q_{\mathbb{E}}(x_2,\mu)\}|/\|x_1 - x_2\|$$

is dominated by an integrable constant and Lebesgue dominated convergence implies existence of the gradient

$$\bigtriangledown_x \mathbb{E}[\max\{f(x,\xi), Q_{\mathbb{E}}(x,\mu)\}] = \mathbb{E}[\bigtriangledown_x \max\{f(x,\xi), Q_{\mathbb{E}}(x,\mu)\}].$$

Finally, as $Q_{max}(x,\mu)$ is convex and differentiable, it is continuously differentiable, cf. Theorem 25.5 Rockafellar (1997). □

As in Section 2.3, we restrict measures to the set

$$\Delta_{r,K}(\mathbb{R}^N) := \Big\{\nu \in \mathcal{P}(\mathbb{R}^N)\Big| \int_{\mathbb{R}^N} \|\xi\|^r \nu(d\xi) \le K\Big\}.$$

Proposition 2.4.4 *Assume (A1)–(A2) and let $\mu \in \Delta_{r,K}(\mathbb{R}^N)$ for some $r > 2$, $K > 0$. Then $Q_{\max} : \mathbb{R}^{n_1} \times \Delta_{p,K}(\mathbb{R}^N) \to \mathbb{R}$ is jointly continuous w.r.t. (x,μ).*

Proof. Let $x, x_n \in \mathbb{R}^{n_1}$, $x_n \to x$ and $\mu_n, \mu \in \Delta_{r,K}(\mathbb{R}^N)$, $\mu_n \xrightarrow{w} \mu$. We intent to show

$$Q_{\max}(x_n,\mu_n) \to Q_{\max}(x,\mu), \;\; n \to \infty.$$

Defining $g(\xi) := \max\{f(x,\xi), Q_{\mathbb{E}}(x,\mu)\}$ and $g_n(\xi) := \max\{f(x_n,\xi), Q_{\mathbb{E}}(x_n,\mu_n)\}$, this is equivalent to

$$\int_{\mathbb{R}^N} g_n(\xi)\mu_n(d\xi) \to \int_{\mathbb{R}^N} g(\xi)\mu(d\xi), \quad n \to \infty. \tag{2.4.7}$$

Initially, we aim to show

$$\mu_n \circ g_n^{-1} \xrightarrow{w} \mu \circ g^{-1} \tag{2.4.8}$$

or, equivalently,

$$\int_{\mathbb{R}^N} h(\xi)\mu_n \circ g_n^{-1}(d\xi) \to \int_{\mathbb{R}^N} h(\xi)\mu \circ g^{-1}(d\xi), \quad n \to \infty$$

for any bounded continuous function $h : \mathbb{R}^N \to \mathbb{R}$. Changing variables,

$$\int_{\mathbb{R}^N} h(g_n(\xi))\mu_n(d\xi) \to \int_{\mathbb{R}^N} h(g(\xi))\mu(d\xi), \quad n \to \infty,$$

this follows directly from $\mu_n \xrightarrow{w} \mu$, since $h \circ g_n$ and $h \circ g$ are bounded continuous functions. The proof is completed by applying Theorem 5.4 by Billingsley (1968) according to which (2.4.8) and uniform integrability imply (2.4.7). A sufficient condition for uniform integrability is the moment condition

$$\sup_n \int_{\mathbb{R}^N} |g_n(\xi)|^q \mu_n(d\xi) < +\infty \tag{2.4.9}$$

for some $q > 1$. Now, from Proposition 2.4.1, there exists a constant $K_1 > 0$, such that

$$\begin{aligned} |g_n(\xi)|^q &= |\max\{f(x_n, \xi), Q_{\mathbb{E}}(x_n, \mu_n)\}|^q \\ &\leq (K_1\|x_n\| \max\{1, \|\xi\|^2\} + |Q_{\mathbb{E}}(x_n, \mu_n)|)^q. \end{aligned}$$

The sequence $\{\|x_n\|\}$ has an upper bound, κ_1. Likewise, joint continuity of $Q_{\mathbb{E}}(\cdot, \cdot)$ ensures an upper bound, κ_2, of $\{|Q_{\mathbb{E}}(x_n, \mu_n)|\}$. Consequently, we continue the estimate

$$|g_n(\xi)|^q \leq ((K_1\kappa_1)(1 + \|\xi\|^2) + \kappa_2)^q \leq (4K_1\kappa_1)^q(1 + \|\xi\|^{2q}) + 2^q\kappa_2^q. \tag{2.4.10}$$

Recall that $\mu \in \Delta_{r,K}(\mathbb{R}^N)$ for some $r > 2$. Letting $q = \frac{1}{2}r$, (2.4.10) finally gives us

$$\int_{\mathbb{R}^N} |g_n(\xi)|^q \mu_n(d\xi) \leq (4K_1\kappa_1)^q(1 + K) + 2^q\kappa_2^q < +\infty.$$

Thus, (2.4.9) is verified. □

Proposition 2.4.5 *Assume (A1)–(A2) and let $\{x \in X : Ax = b\}$ be nonempty and bounded. Then there exists a constant $L > 0$ such that the estimate*

$$sup_{x \in X : Ax = b} |Q_{\max}(x, \mu) - Q_{\max}(x, \nu)| \leq L d_2(\mu, \nu)$$

is valid whenever $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^N)$.

Proof. Let $x \in \mathbb{R}^{n_1}$ and $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^N)$. The following evaluations hold

$$
\begin{aligned}
&|Q_{max}(x,\mu) - Q_{max}(x,\nu)| \qquad (2.4.11)\\
&= \Big| \int_{\mathbb{R}^N} \max\{f(x,\xi), Q_{\mathbb{E}}(x,\mu)\}\mu(d\xi) - \int_{\mathbb{R}^N} \max\{f(x,\xi), Q_{\mathbb{E}}(x,\nu)\}\nu(d\xi)\Big|\\
&\leq \Big| \int_{\mathbb{R}^N} \max\{f(x,\xi), Q_{\mathbb{E}}(x,\mu)\}\mu(d\xi) - \int_{\mathbb{R}^N} \max\{f(x,\xi), Q_{\mathbb{E}}(x,\mu)\}\nu(d\xi)\Big|\\
&+ \Big| \int_{\mathbb{R}^N} \max\{f(x,\xi), Q_{\mathbb{E}}(x,\mu)\}\nu(d\xi) - \int_{\mathbb{R}^N} \max\{f(x,\xi), Q_{\mathbb{E}}(x,\nu)\}\nu(d\xi)\Big|.
\end{aligned}
$$

Consider the first term of (2.4.11). Recall the definition of $c_2(\xi_1, \xi_2)$. According to Proposition 2.4.1, for some constant $L_1 > 0$,

$$
\begin{aligned}
&|\max\{f(x,\xi_1), Q_{\mathbb{E}}(x,\mu)\} - \max\{f(x,\xi_2), Q_{\mathbb{E}}(x,\mu)\}|\\
&\qquad \leq L_1\|x\| \max\{1, \|\xi_1\|, \|\xi_2\|\}\|\xi_1 - \xi_2\|\\
&\qquad \leq L_1(\sup_{x\in X: Ax=b}\|x\|)c_2(\xi_1,\xi_2)\|\xi_1 - \xi_2\|
\end{aligned}
$$

for μ-almost all $\xi_1, \xi_2 \in \mathbb{R}^N$ and hence, there exists a constant $L_2 > 0$ such that the function $(1/L_2)\max\{f(x,\xi),\, Q_{\mathbb{E}}(x,\mu)\}$ belongs to the set $\mathcal{F}_2(\mathbb{R}^N)$. Consequently,

$$
\begin{aligned}
&\Big| \int_{\mathbb{R}^N} \max\{f(x,\xi), Q_{\mathbb{E}}(x,\mu)\}\mu(d\xi) - \int_{\mathbb{R}^N} \max\{f(x,\xi), Q_{\mathbb{E}}(x,\mu)\}\nu(d\xi)\Big|\\
&\leq L_2 \sup_{F\in\mathcal{F}_2(\mathbb{R}^N)} \Big| \int_{\mathbb{R}^N} F(\xi)\mu(d\xi) - \int_{\mathbb{R}^N} F(\xi)\nu(d\xi)\Big| = L_2 d_2(\mu,\nu). \qquad (2.4.12)
\end{aligned}
$$

For the second term of (2.4.11), we have

$$
\begin{aligned}
&\Big| \int_{\mathbb{R}^N} \max\{f(x,\xi), Q_{\mathbb{E}}(x,\mu)\}\nu(d\xi) - \int_{\mathbb{R}^N} \max\{f(x,\xi), Q_{\mathbb{E}}(x,\nu)\}\nu(d\xi)\Big|\\
&\qquad \leq \int_{\mathbb{R}^N} \Big| \max\{f(x,\xi), Q_{\mathbb{E}}(x,\mu)\} - \max\{f(x,\xi), Q_{\mathbb{E}}(x,\nu)\}\Big|\nu(d\xi)
\end{aligned}
$$

and applying the inequality $|\max\{a,b\} - \max\{a,c\}| \leq |b-c|$,

$$
\begin{aligned}
&\int_{\mathbb{R}^N} |\max\{f(x,\xi), Q_{\mathbb{E}}(x,\mu)\} - \max\{f(x,\xi), Q_{\mathbb{E}}(x,\nu)\}|\nu(d\xi)\\
&\qquad \leq \int_{\mathbb{R}^N} |Q_{\mathbb{E}}(x,\mu) - Q_{\mathbb{E}}(x,\nu)|\nu(d\xi) = |Q_{\mathbb{E}}(x,\mu) - Q_{\mathbb{E}}(x,\nu)|.
\end{aligned}
$$

From Proposition 2.3.5, we have some constant $L_3 > 0$ such that

$$
|Q_{\mathbb{E}}(x,\mu) - Q_{\mathbb{E}}(x,\nu)| \leq L_3 d_2(\mu,\nu). \qquad (2.4.13)
$$

The final result now follows from (2.4.12) and (2.4.13). □

Remark 2.4.4 *Propositions 2.4.1, 2.4.2, 2.4.3, 2.4.4, 2.4.5 are valid for $Q_{max,\eta}$, also. In fact, with minor modifications, the proofs still hold, replacing $Q_{\mathbb{E}}(x,\mu)$ by the constant η.*

Corollary 2.4.1 *For $\mathcal{R} = \mathcal{D}$, let $\varrho \in [0, 1/2]$; for $\mathcal{R} = \mathcal{D}^+$, let $\varrho \in [0,1]$ and for $\mathcal{R} = \mathcal{D}^\eta$, let $\varrho \geq 0$. Then the statements of propositions 2.4.2, 2.4.3, 2.4.4, 2.4.5 are valid for the mean-risk objective $Q_{\mathbb{E}}(x,\mu) + \varrho Q_{\mathcal{R}}(x,\mu)$.*

2.5 Stability

In many practical applications of stochastic programming, the probability distribution is not fully known. In such cases, the true measure has to be replaced by some suitable estimate, e.g. an empirical measure. Even if the true distribution is indeed fully known, an approximate measure may be required to facilitate computations. This justifies the examination of stability issues in stochastic programming. More specifically, we analyze the situation in which the underlying probability measure is subjected to perturbations. For recent surveys on stability in two-stage stochastic programming, see Römisch (2003) and Schultz (2000).

Considering the problem

$$P(\mu) : \min\{Q_{\mathbb{E}}(x,\mu) + \varrho Q_{\mathcal{R}}(x,\mu) \mid Ax = b, x \in X\},$$

stability results take the form of continuity properties of the optimal value function and the solution set mapping defined as

$$\phi(\mu) := \inf\{Q_{\mathbb{E}}(x,\mu) + \varrho Q_{\mathcal{R}}(x,\mu) \mid Ax = b, x \in X\},$$

$$\varphi(\mu) := \{x \in X \mid Ax = b, Q_{\mathbb{E}}(x,\mu) + \varrho Q_{\mathcal{R}}(x,\mu) = \phi(\mu)\}.$$

The primary result of qualitative stability follows directly from Corrolary 2.4.1 and parametric optimization, cf. Berge (1963), Theorem 1.1, 1.2 and the Maximum Theorem,

Proposition 2.5.1 *Assume (A1)–(A2) and $\mu \in \Delta_{r,K}(\mathbb{R}^N)$ for some $r > 2$, $K > 0$. Suppose further that $\{x \in X \mid Ax = b\}$ is nonempty and bounded. For $\mathcal{R} = \mathcal{D}$, let $\varrho \in [0, 1/2]$; for $\mathcal{R} = \mathcal{D}^+$, let $\varrho \in [0,1]$ and for $\mathcal{R} = \mathcal{D}^\eta$, let $\varrho \geq 0$. Then*

(i) $\phi : \Delta_{r,K}(\mathbb{R}^N) \to \mathbb{R}$ is continuous where $\Delta_{r,K}(\mathbb{R}^N)$ is equipped with weak convergence of probability measures.

(ii) $\varphi : \Delta_{r,K}(\mathbb{R}^N) \to 2^{\mathbb{R}^{n_1}}$ is Berge upper semicontinuous, i.e. for any $\mu \in \Delta_{r,K}(\mathbb{R}^N)$ and any open set $V \subset \mathbb{R}^{n_1}$ with $V \supseteq \varphi(\mu)$, there exists a neighborhood $\mathcal{V}$ of μ in $\Delta_{r,K}(\mathbb{R}^N)$ such that $\varphi(\nu) \subseteq V$ for all ν in $\mathcal{V}$.

Remark 2.5.1 *From Proposition 2.5.1 stability results involving discrete approximations of the true probability measure can be derived in the same way as applies to the expectation-based case. Such approximations include empirical measures or measures that are close to the true one with respect to some (pseudo) metric. The latter is associated with optimal scenario generation and optimal scenario reduction, cf. Römisch (2003), Pflug (2004) and Heitsch and Römisch (2006b). For further reference on scenario generation and reduction, see Chapter 7.*

Example 2 The example shows that the uniform integrability assumption in Proposition 2.4.4 is indispensable for qualitative stability. Let $\Phi(h-x) = \min\{y \mid y \geq h-x, y \geq 0\} = \max\{h-x, 0\}$, $f(x,h) = \max\{h-x, 0\}$ and $\{x \in X \mid Ax = b\} = [0,1]$. Denote by δ_ϵ the measure that places unit mass at $\epsilon \in \mathbb{R}$. Let $\mu = \delta_0$ and $\mu_n = (1-\frac{1}{n})\delta_0 + \frac{1}{n}\delta_n, n \in \mathbb{N}$. Then $\mu_n \xrightarrow{w} \mu$. Consider $\mathcal{R} = \mathcal{D}^+$ and define the optimal value function $\phi(\mu) := \inf\{Q_{\mathbb{E}}(x,\mu) + \varrho Q_{\mathcal{R}}(x,\mu) \mid x \in X, Ax = b\}$. Whereas $\phi(\mu) = 0$, $\phi(\mu_n) = 1 - \frac{1}{n} + \varrho(1 - \frac{2}{n} + \frac{1}{n^2})$. But then $\phi(\mu_n) \nrightarrow \phi(\mu), n \to \infty$. Considering the same data, sufficiency of uniform integrability in the case $\mathcal{R} = \mathcal{D}$ is clear. □

In consequence of the continuity results, quantitative stability is based on the Fortet-Mourier metric of second order.

Proposition 2.5.2 *Assume (A1)–(A2) and let $\{x \in X \mid Ax = b\}$ be nonempty and bounded. Then there exists a constant $L > 0$ such that the estimate*

$$|\phi(\mu) - \phi(\nu)| \leq L d_2(\mu, \nu)$$

is valid whenever $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^N)$.

Proof. Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^N)$. Since $Q_{\mathbb{E}}(x,\mu) + \varrho Q_{\mathcal{R}}(x,\mu)$ and $Q_{\mathbb{E}}(x,\nu) + \varrho Q_{\mathcal{R}}(x,\nu)$ are continuous in x and $\{x \in X \mid Ax = b\}$ is nonempty and compact, $\varphi(\mu) \neq \emptyset$ and $\varphi(\nu) \neq \emptyset$. Therefore, we can choose $x_\mu \in \varphi(\mu)$ and $x_\nu \in \varphi(\nu)$. Together,

$$\begin{aligned}\phi(\mu) &\leq Q_{\mathbb{E}}(x_\nu, \mu) + \varrho Q_{\mathcal{R}}(x_\nu, \mu) \\ &\leq \phi(\nu) + |(Q_{\mathbb{E}}(x_\nu, \mu) + \varrho Q_{\mathcal{R}}(x_\nu, \mu)) - (Q_{\mathbb{E}}(x_\nu, \nu) + \varrho Q_{\mathcal{R}}(x_\nu, \nu))|\end{aligned}$$

and

$$\begin{aligned}\phi(\nu) &\leq Q_{\mathbb{E}}(x_\mu, \nu) + \varrho Q_{\mathcal{R}}(x_\mu, \nu) \\ &\leq \phi(\mu) + |(Q_{\mathbb{E}}(x_\mu, \nu) + \varrho Q_{\mathcal{R}}(x_\mu, \nu)) - (Q_{\mathbb{E}}(x_\mu, \mu) + \varrho Q_{\mathcal{R}}(x_\mu, \mu))|\end{aligned}$$

imply

$$|\phi(\mu) - \phi(\nu)| \leq \sup_{x \in X: Ax = b} |(Q_{\mathbb{E}}(x,\mu) - \varrho Q_{\mathcal{R}}(x,\mu)) - (Q_{\mathbb{E}}(x,\nu) + \varrho Q_{\mathcal{R}}(x,\nu))|.$$

We get the desired result from Corollary 2.4.1. □

Remark 2.5.2 *Under suitable conditions weak convergence of probability measures implies convergence with respect to the Fortet-Mourier metric. For instance, if the set* $\mathcal{F}_2^K(\mathbb{R}^N) := \{\max\{-K, \min\{F(\cdot), K\}\} \mid F \in \mathcal{F}_2(\mathbb{R}^N)\}$ *of truncated functions is a* μ*-uniformity class for large* $K > 0$ *and* $\mathcal{F}_2(\mathbb{R}^N)$ *is uniformly integrable with respect to* $\{\mu_n\}$*, then* $\mu_n, \mu \in \mathcal{P}_2(\mathbb{R}^N)$*,* $\mu_n \xrightarrow{w} \mu$ *implies* $d_2(\mu_n, \mu) \to 0$*. In Römisch (2003), this is utilized to show qualitative stability. Hence, Propositions 2.4.5 and 2.5.2 may be seen as quantifications of Propositions 2.4.4 and 2.5.1.*

2.6 Algorithm

Motivated by the stability results of the previous section, we make the following assumption about the distribution

(A4) The probability measure μ is discrete and has finite support $\{\xi^1, \ldots, \xi^S\}$ with corresponding probabilities $\pi^1, \ldots, \pi^S$.

The assumption (A3) follows automatically. The algorithmic analysis can easily be modified to work without the assumptions (A1) and (A2). In line with the rest of the chapter, however, we assume (A1) and (A2). We refer to a realization of the random data as a scenario and denote it by $\xi^s = (q^s, h^s, T^s), s = 1, \ldots, S$. With a discrete distribution, the problems (2.4.1), (2.4.2) and (2.4.3) have the following deterministic equivalents

$$\min\Big\{cx + (1-2\varrho)\sum_{s=1}^{S}\pi^s\Phi(x,\xi^s) + 2\varrho\sum_{s=1}^{S}\pi^s\max\{\Phi(x,\xi^s), \sum_{s'=1}^{S}\pi^{s'}\Phi(x,\xi^{s'})\}\Big| x \in X, Ax = b\Big\} \tag{2.6.1}$$

in the case of the central deviation,

$$\min\Big\{cx + (1-\varrho)\sum_{s=1}^{S}\pi^s\Phi(x,\xi^s) + \varrho\sum_{s=1}^{S}\pi^s\max\{\Phi(x,\xi^s), \sum_{s'=1}^{S}\pi^{s'}\Phi(x,\xi^{s'})\}\Big| x \in X, Ax = b\Big\} \tag{2.6.2}$$

in the case of the semideviation and

$$\min\Big\{cx+\sum_{s=1}^{S}\pi^s\Phi(x,\xi^s)+ \\ \varrho\sum_{s=1}^{S}\pi^s\max\{\Phi(x,\xi^s),\eta\}\Big|x\in X, Ax=b\Big\}-\varrho\eta \tag{2.6.3}$$

in the case of expected excess of target.

Observe that problem (2.6.1) is equivalent to

$$\min\ cx+(1-2\varrho)\sum_{s=1}^{S}\pi^s\theta^s+2\varrho\sum_{s=1}^{S}\pi^s v^s \tag{2.6.4}$$

$$\text{s.t. } \Phi(x,\xi^s)\le\theta^s,\ \sum_{s'=1}^{S}\pi^{s'}\theta^{s'}\le v^s,\ \theta^s\le v^s,\ x\in X,\ Ax=b,$$

$$v^s,\theta^s\in\mathbb{R}, s=1,\dots,S,$$

problem (2.6.2) to

$$\min\ cx+(1-\varrho)\sum_{s=1}^{S}\pi^s\theta^s+\varrho\sum_{s=1}^{S}\pi^s v^s \tag{2.6.5}$$

$$\text{s.t. } \Phi(x,\xi^s)\le\theta^s,\ \sum_{s'=1}^{S}\pi^{s'}\theta^{s'}\le v^s,\ \theta^s\le v^s,\ x\in X,\ Ax=b$$

$$v^s,\theta^s\in\mathbb{R}, s=1,\dots,S$$

and problem (2.6.3) to

$$\min\ cx+\sum_{s=1}^{S}\pi^s\theta^s+\varrho\sum_{s=1}^{S}\pi^s v^s-\varrho\eta$$

$$\text{s.t. } \Phi(x,\xi^s)\le\theta^s,\ \eta\le v^s,\ \theta^s\le v^s,\ x\in X,\ Ax=b$$

$$v^s,\theta^s\in\mathbb{R}, s=1,\dots,S$$

The block structure of the problems (2.6.4) and (2.6.5) does not fit into existing schemes of stochastic programming. In contrast to the traditional two-stage stochastic program, the mean-risk problems contain explicit coupling between the

scenario-dependent variables which prevents most well-known algorithms from working. Nevertheless, by regarding some of the scenario-dependent variables as first-stage variables, we are able to restore a certain degree of separability. It is possible to decompose the problems with respect to stages and work scenario-wise by means of a cutting plane algorithm.

Another cutting plane algorithm is proposed by Ahmed (2004), who derives cuts in a fashion inspired by subgradient methods from convex optimization. Using parametric analysis with respect to the objective coefficients, the efficient frontier can be constructed by varying the trade-off coefficient, ϱ, in the appropriate range. It should be clear from the following that this feature also applies to the present algorithm. The main difference between the two algorithms is the number of cuts added in an iteration and the information within each cuts.

For simplicity, we state the algorithm for central deviation. It works in a similar fashion for semideviation and expected excess of target. The idea is to relax the constraints $\Phi(x, \xi_s) \leq \theta_s, s = 1, \ldots, S$ and iteratively reinforce them by means of optimality cuts. Consider some iteration i of the algorithm and let the corresponding solution be x^i. Then $\Phi(x^i, \xi^s), s = 1, \ldots, S$ is feasible and bounded. For $s = 1, \ldots, S$ let $\sigma^{i,s}$ be a dual optimal solution of $\Phi(x^i, \xi^s)$. If x^i is such that $\theta^{i,s} < \Phi(x^i, \xi^s) = \sigma^{i,s}(h^s - T^s x^i)$ for some $s \in \{1, \ldots, S\}$, then $(x^i, \theta^{i,1}, \ldots, \theta^{i,S})$ is cut off by adding the optimality cut

$$\sigma^{i,s}(h^s - T^s x) \leq \theta^s. \tag{2.6.6}$$

By duality, $\Phi(x, \xi^s) \geq \sigma^{i,s}(h^s - T^s x)$ for all $x \in \mathbb{R}^{n_1}$, showing that the optimality cut is a valid inequality to (2.6.4). The algorithm proceeds as follows

Algorithm 2.6.1

Step 0 (Initialization). Set $i = 0$ and let the current master problem be

$$\begin{aligned} &\min\ cx + (1 - 2\varrho) \sum_{s=1}^{S} \pi^s \theta^s + 2\varrho \sum_{s=1}^{S} \pi^s v^s \\ &s.t.\ \sum_{s=1}^{S} \pi^s \theta^s \leq v^s,\ \ \theta^s \leq v^s,\ \ x \in X,\ \ Ax = b \\ &v^s, \theta^s \in \mathbb{R}, s = 1, \ldots, S. \end{aligned}$$

Step 1 (Solve master problem). Set $i = i + 1$. Solve the current master problem and let $(x^i, v^{i,1}, \ldots, v^{i,S}, \theta^{i,1}, \ldots, \theta^{i,S})$ be an optimal solution (If $\theta^s = -\infty$ or $v^s = -\infty$ for some $s \in \{1, \ldots, S\}$ the variable is ignored in the computation.)

Step 2 (Add optimality cuts). For each $s = 1, \ldots, S$, solve the second-stage problem with $x = x^i$ and let $\sigma^{i,s}$ be a corresponding dual solution. If $\sigma^{i,s}(h^s - T^s x^i) > \theta^{i,s}$ for some $s \in \{1, \ldots, S\}$, add an optimality cut (2.6.6) *to the master problem and return to step 2.*

Step 3 (Termination). Stop. The current solution is optimal.

We obtain a lower bound to the mean-risk problem by means of the relaxation provided by the master problem and eventually also an upper bound by a feasible solution. In that the bounds coincide, the master problem in this fashion produces an optimal solution to the mean-risk problem. Indeed, we have the following,

Proposition 2.6.1 *Assume (A1)-(A2), (A4) and that $\{x \in X \mid Ax = b\}$ is nonempty and bounded. Then Algorithm 2.6.1 terminates with an optimal solution to the mean-risk problem in a finite number of iterations.*

Proof. By remark 2.4.2, an optimal solution to the mean-risk problem (2.6.4) exists. Let x^* be a such solution and let z^* be the optimal value. Since the master problem is a relaxation of (2.6.4), the optimal value z^i of the master problem in iteration i is a lower bound on z^*, i.e.

$$z^i \leq z^*. \tag{2.6.7}$$

Now if in iteration i for some $s \in \{1, \ldots, S\}$, $\sigma^{i,s}(h^s - T^s x^i) > \theta^s$, an optimality cut will be added and the algorithm proceeds. Only a finite number of cuts can be generated since the number of dual extreme points of (2.2.2) is finite. Moreover, no cut will be generated twice as a cut already present would not cut off a current solution. Eventually, for some iteration i', $\Phi(x^{i'}, \xi^s) \leq \theta^{i',s}, s = 1, \ldots, S$ and the algorithm terminates. Upon termination in iteration i', we have

$$z^{i'} = cx^{i'} + (1 - 2\varrho) \sum_{s=1}^{S} \pi^s \theta^{i',s} + 2\varrho \sum_{s=1}^{S} \pi^s v^{i',s} \geq \tag{2.6.8}$$

$$cx^{i'} + (1 - 2\varrho) \sum_{s=1}^{S} \pi^s \theta^{i',s} + 2\varrho \sum_{s=1}^{S} \pi^s \max\Big\{\theta^{i',s}, \sum_{s'=1}^{S} \pi^{s'} \theta^{i',s'}\Big\} \geq$$

$$cx^{i'} + (1 - 2\varrho) \sum_{s=1}^{S} \pi^s \Phi(x^{i'}, \xi^s) + 2\varrho \sum_{s=1}^{S} \pi^s \max\Big\{\Phi(x^{i'}, \xi^s), \sum_{s'=1}^{S} \pi^{s'} \Phi(x^{i'}, \xi^{s'})\Big\} \geq$$

$$cx^* + (1 - 2\varrho) \sum_{s=1}^{S} \pi^s \Phi(x^*, \xi^s) + 2\varrho \sum_{s=1}^{S} \pi^s \max\Big\{\Phi(x^*, \xi^s), \sum_{s'=1}^{S} \pi^{s'} \Phi(x^*, \xi^{s'})\Big\} = z^*$$

since $x^{i'}$ is feasible to (2.6.1). From (2.6.7) and (2.6.8), $z^{i'} = z^*$ and $x^{i'}$ is optimal. □

Note that the above algorithm shows strong resemblance to the ordinary L-shaped method, and, therefore, may suffer from some of the same drawbacks. Of particular importance is the tendency of the solutions to oscillate heavily in early iterations, resulting in slow convergence towards an optimal solution. This drawback is circumvented by regularized decomposition . The basic idea is, in iteration i, to add a regularizing term $\alpha\|x - x^{i-1}\|^2$ with $\alpha > 0$ to the objective in order to penalize divergence from the current solution, x^{i-1}. For further details on the expectation-based case, see Ruszczyńsky (1986) and Birge and Louveaux (1997). To avoid a nonlinear formulation, we implemented a modified version of regularized decomposition and replaced the quadratic term by $0.5\alpha \sum_{j=1}^{n_1} |x_j - x_j^{i-1}|$. Moreover, we used regularization only in the first iterations of the algorithm.

To examine the practicability, Algorithm 2.6.1 was implemented in C++ utilizing procedures from the callable library of CPLEX 9.0, cf. Cplex Optimization Inc. (2006). For testing, we used the linear relaxation of a mixed-integer scheduling problem in chemical production. The problem contains 12 variables and no constraints in the first stage and 204 variables and 182 constraints in the second stage. We ran the algorithm with a varying number of scenarios; 50, 100, 200, 500 and 1000, and with the three different deviation measures, abbreviated CD (central deviation), SD (semideviation) and EE (expected excess). Each of the resulting test problems was run with two versions of the algorithm, one stated exactly as above and one augmented with regularization, referred to as ALG2.6.1 and ALG2.6.1REG, respectively. Results are reported in Tables 2.1, 2.2 and 2.3. For every run we recorded the optimal value, the number of iterations performed, the number of cuts generated and the CPU time spent by the algorithm. All computations were carried out on a Intel Xeon 2.67 GHz processor with 4 GB RAM.

The effect of regularization is moderate in the case of central deviation and semideviation. In general, the number of cuts is reduced, which however does not seem to cohere with fewer iterations and lower computing times. In the case of expected excess of target, the results show small time savings in most runs. Moreover, in all runs the number of iterations and cuts is reduced by the use of regularization.

As a final remark, the mean-risk model can be seen as a scalarization of the bi-criteria optimization problem

$$\min\{(Q_{\mathbb{E}}(x), Q_{\mathcal{R}}(x)) \mid x \in X, Ax = b\}.$$

Optimality in multi-criteria optimization is captured in the concept of efficiency. A point $x^* \in \{x \in X \mid Ax = b\}$ is said to be efficient if there is no other

Table 2.1: Computational results for CD ($\varrho = 1/4$).

		ALG2.6.1			ALG2.6.1REG		
S	Opt. val.	Ite.	Cuts	CPU/s	Ite.	Cuts	CPU/s
50	156.58	33	2103	6.91	35	2050	6.86
100	125.55	32	4202	14.42	30	3600	12.50
200	218.81	28	7534	30.93	28	6341	27.62
500	147.78	28	18677	159.62	33	18942	206.22
1000	326.67	31	41321	594.91	32	39543	635.88

Table 2.2: Computational results for SD ($\varrho = 1$).

		ALG2.6.1			ALG2.6.1REG		
S	Opt. val.	Ite.	Cuts	CPU/s	Ite.	Cuts	CPU/s
50	158.59	32	1956	6.44	33	1847	6.17
100	126.79	31	4044	13.81	30	3398	11.70
200	224.08	28	7105	30.02	26	6958	29.27
500	149.82	28	18445	155.25	30	16905	159.48
1000	335.97	33	38319	618.17	26	35033	609.18

Table 2.3: Computational results for EE ($\varrho = 1, \eta = 150$).

		ALG2.6.1			ALG2.6.1REG		
S	Opt. val.	Ite.	Cuts	CPU/s	Ite.	Cuts	CPU/s
50	161.58	33	2050	6.02	31	1702	5.06
100	124.30	34	4025	12.26	30	3307	9.98
200	277.08	28	7331	21.05	26	6551	18.71
500	147.90	33	20897	143.14	30	19306	96.01
1000	484.70	31	40492	204.94	31	36504	207.11

$x \in \{x \in X \mid Ax = b\}$ that fulfills $Q_{\mathbb{E}}(x) \leq Q_{\mathbb{E}}(x^*)$ and $Q_{\mathcal{R}}(x) \leq Q_{\mathcal{R}}(x^*)$ with at least one inequality strict. The mean-risk model emerges when using a weighting method to generate the efficient frontier. Given $\varrho \in \mathbb{R}_+$, every optimal solution to

$$\min\{Q_{\mathbb{E}}(x) + \varrho Q_{\mathcal{R}}(x) \mid x \in X, Ax = b\} \tag{2.6.9}$$

is efficient. Varying $\varrho \in \mathbb{R}_+$ enables computation of efficient points by solving a family of mean-risk models. Fig. 2.1 shows an example of a mean-risk efficient

frontier, generated by this approach. However, in order to arrive at the entire efficiency set, certain convexity conditions must be met. As $Q_{\mathcal{R}}$ may be non-convex, generally, only a subset of the efficient frontier, the so-called supported part, may be reached when solving (2.6.9) for $\varrho \in \mathbb{R}_+$.

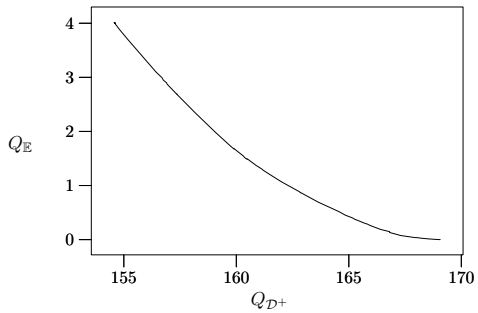


Figure 2.1: Mean-SD efficient frontier for the instance with 50 scenarios.

Example 3 The following example indicates that we cannot be certain that all efficient points are generated by solving (2.6.9) with $\mathcal{R} = \mathcal{D}^+$ for all $\varrho \in \mathbb{R}_+$. Let $\{x \in X \mid Ax = b\} = \mathbb{R}^{n_1}_+$, $\Phi(h - x) = \min\{y \mid y \geq h - x, y \geq 0\} = \max\{h - x, 0\}$ and $f(x, h) = 2x + \max\{h - x, 0\}$. Assume $\mu(h = 2) = 1/2$, $\mu(h = 3) = 1/2$ and $\varrho = 4$. Hence, $Q_{\mathbb{E}}(x) = 2x + \frac{1}{2}(3 - x)^+ + \frac{1}{2}(2 - x)^+$ and $Q_{\mathcal{D}^+}(x) = \frac{1}{2}\max\{\frac{1}{2}(3 - x)^+ - \frac{1}{2}(2 - x)^+, 0\} + \frac{1}{2}\max\{\frac{1}{2}(2 - x)^+ - \frac{1}{2}(3 - x)^+, 0\}$. All points of the interval $[2, 3]$ are efficient points. However, the point 2.5 never appears as a solution to (2.6.9) for any $\varrho \in \mathbb{R}_+$. The same data provides a counterexample in the case $\mathcal{R} = \mathcal{D}$.

Part II

Stochastic programming in power systems

Chapter 3

The development in stochastic recourse models for power production and trading

3.1 Introduction to the power system

This chapter considers the deregulation in the power sector, consisting in the reduction or removal of the government regulations on the industry. To fully understand the subsequent chapters on stochastic programming applications to power systems, we begin this chapter by listing some of the main components of a deregulated power system, including a physical transmission system and a system operator, power producers and consumers, and several markets covering a day-ahead and an intra-day market as well as financial markets.

Transmission system

The *transmission system*, which is also referred to as the grid, is established for the transferring of electric power from place to place, often from a power plant to a smaller station before distributed to the consumers. Electricity is usually transmitted through overhead power transmission lines although occasionally they may be underground. As power transmission lines come with certain capacities, grid congestion is possible. Nowadays, transmission and generation are separated and, most places, transmission is the responsibility of an independent, though state-approved, power system operator. The system operator maintains the grid, provides access to the suppliers, distributors and consumers and collects transmission tariffs. Moreover, the system operator ensures security of supply by balancing the physical delivery of power and the actual consumption.

Producers

Production includes *thermal power*, *hydro-power* and other kinds of renewable power such as wind power. In a thermal power plant, power is produced by units that transform thermal energy, most often by burning fuel, into electrical energy. Thermal power plants are classified according to the type of fuel and generating units. Examples are fossil-fuel powered plants using coal-fired and gas-burning thermal units, nuclear power plants and renewable energy plants. A hydro-power plant consists of a number of water reservoirs, possibly connected in cascade or in a more complex network, and corresponding power stations that contain turbines, also referred to as hydro-power units. As water is released from the reservoirs, it flows to the turbines, in which electricity is generated by converting the potential energy of the water into electrical energy. To some extent, hydro-power generation provides a possibility to "store" energy in the sense that water releases can be put on hold. In many respects, it is beneficial to coordinate thermal power production and hydro-power generation, exploiting the flexibility of hydro-power and stability thermal power. Hydro-thermal production refers to the cooperation of thermal power plants and hydro-power plants, either of the type described above or so-called pumped storage hydro-power plants. In pumped storage plants the units serve to store excess energy by utilizing electricity to pump up water into a reservoir and save it for future hydro-power generation.

In general, producers range from small local stations to large plants covering wide areas. The producers can dispose of power through direct trading or through organized wholesale markets, in which they either act as price-takers or are able to influence market prices.

Consumers

End-users of electricity are industrial or household consumers who buy the commodity directly from power producers, wholesale markets or retailers. Retailers engage in physical trading directly with producers or in wholesale markets.

Markets

Direct physical trading between two parties is done by means of bilateral contracts. However, physical trading is also possible through organized markets that serve different purposes. The *spot market* is a wholesale market, established for the immediate delivery of power from producer to consumer and is, due to the practice of bidding and committing a day in advance, also referred to as the day-ahead market. A market operator collects the supply and demand bids and calculates prices of the following day and the corresponding quantities to be dispatched by both producers and consumers. When organizing a power market with a central-

ized dispatching and pricing mechanism, the power market is also referred to as an electricity pool.

Unlike the spot market, the regulating or the *balancing market* is an intra-day market that handles supply and demand imbalances by means of so-called regulation. Producers and consumers bid on the supply and demand side, respectively, while the balancing is generally handled by the system operator. Less organized markets may comprise markets for investments, reserve markets and markets for other ancillary services.

As a contrast to the markets that imply immediate delivery of the physical commodity, the financial markets for derivatives, such as forwards, futures, contracts of differences and options, serve as instruments for hedging risk. Furthermore, there may be markets for trade in environmental products, the market contracts being electricity certificates.

3.2 From regulated to deregulated markets

We continue this chapter with an overview of stochastic programming models in short-term power production and trading, where special emphasis is placed on the development prompted by the restructuring of the power sector.

Traditional applications of stochastic programming to power systems represent a well-developed research area. The setting is motivated by the wish to control system reliability and is based on centralized operation and regulated markets in which many local producers enjoy monopoly. Since the earlier applications, the markets have been reformed and operating procedures have changed. The area of applications within the new environment of decentralized operation, deregulated markets, and competition is therefore a developing field of stochastic programming.

The present chapter aims at illustrating the impacts of the restructuring on stochastic power optimization problems. We confine ourselves mainly to short-term problems and restrict attention to power production and physical trading. The idea is to explain how traditional models can be adopted to the new environment and newer novel models become highly relevant. To discuss such models from a practical point of view, we include a number of applications, consider computational aspects such as problem sizes and decomposition potential, introduce the most common solution approaches and present some numerical results.

As already indicated, the main field from which the models of this chapter are derived is stochastic programming. Since, however, stochastic programming is closely related to decision analysis, stochastic control theory, Markov decision processes, dynamic programming and optimization of discrete event simulations, the models may stem from any of these fields.

The outline of the remainder of the chapter is as follows. Sections 3.4 and 3.5 are confined to power production and contain the models that are valid before and

after the transition from regulated to deregulated markets, respectively. Section 3.6 investigates the ability of the models to decompose and state the most basic solution approaches. The models for physical power trading, that have mainly arisen with liberalization, are presented in Section 3.7, which also includes the individual solution approaches.

3.3 Stochastic programming electricity models

Muckstadt and Koenig (1977) were among the first to address the unit commitment problem by mathematical programming, and to solve it by Lagrangian decomposition. Reflecting the continuing interest in this problem, Fahd and Sheble (1994) provided an overview of the following solution methods, ranging from enumeration and priority lists over dynamic, linear and mixed-integer programming as well as branch and bound to Lagrangian relaxation.

Given the progress in stochastic programming, the stochastic extension of the deterministic unit commitment problem began to take form with Birge et al. (1994) being the first to formulate the problem in terms of stochastic mixed-integer programming. As part of a larger information system the authors presented a multi-stage model for dynamic unit commitment under uncertainty with respect to electricity load and unit failures.

Early attempts to formulate and solve stochastic versions of short-term power optimization problems further included Bunn and Paschentis (1986), who developed a two-stage recourse model for economic dispatch by allowing under- and over-dispatching in the second stage and Terry et al. (1986), who described the implementation of stochastic dynamic programming for coordinating operation in a large real-world hydro-power system.

The development in stochastic programming electricity models continued throughout the remainder of the century and into the beginning of the next. At present it has become of major interest due to the restructuring of the power sector in the last decades and the following liberalization of the markets that has encouraged efficient operation of the system and competitiveness of the industry.

In general, stochastic power optimization problems are categorized according to their time horizons. *Long-term problems* usually have a time horizon of up to several years, *medium-term problems* of a few months to a few years and *short-term problems* of a day to a week.

Concerning the longer term, most problems involve technology investment and capacity expansion matters, cf. Bienstock and Shapiro (1988) and Gorenstin and Campodonico (1993), and some include environmental planning issues such as pollutant emission control. Formulated as stochastic programs, such problems may contain various sources of uncertainty. The category covers the building of thermal units, the construction of hydro reservoirs and turbines and multi-year

hydro-thermal scheduling, cf. Sherkat et al. (1985). Due to their aggregation level, the long-term problems often allow for the application of the classical solution approaches, Benders' decomposition and dynamic programming.

Problems of the medium-term embrace hydro-thermal coordination problems, cf. Dentcheva and Römisch (1998), and reservoir management problems, cf. Jacobs et al. (1995), the complexity of which calls for more advanced solution methods. In this context, the most significant sources of uncertainty stem from electricity demand and reservoir inflows. As explained by Fosso et al. (1999), the medium-term problems are relevant in their own right, but at the same time provide input data to the short-term models through reservoir storage boundaries, values of stored water etc.

Well-known traditional short-term planning and operation problems include thermal unit commitment and economic dispatch. Short-term hydro-thermal and pure hydro scheduling are often handled as dynamic operation problems on a daily or weekly basis, which may also be the case for hydro unit commitment. Uncertainty in the traditional problems covers electricity demand, unit failures, fuel costs and potentially reservoir inflows. Newer short-term planning problems consist in the day-ahead physical trading and the trading of ancillary services and add market prices uncertainty to the problems.

For an excellent yet more general survey on stochastic programming problems in energy, see Wallace and Fleten (2003), who consider both electricity, oil and gas. The authors present long-term, medium-term and short-term problems and discuss their development with the transition from regulated to deregulated markets. Both physical and financial markets are considered and much of the survey is devoted to risk management. The emphasis is placed on modeling alone and so solution approaches and other computational aspects are only briefly addressed.

3.4 Short-term power production

Due to the previous regulation by legal restrictions, decision-making was usually effected on a centralized level for several producers voluntarily or legally coordinated and obligated to satisfy demand within their areas. Contracts were predominantly bilateral contracts of which the terms were set by administrative decrees, the result being that cost minimization was pursued.

In line with this, the models of this section are traditional short-term power production models that rest on costs minimization subject to demand constraints. The multi-stage stochastic programs are presented in their scenario tree formulations. For further reference on the notation, see Chapter 1.

Thermal unit commitment and hydro-thermal scheduling

As a starting point, we address the *thermal unit commitment* problem from the perspective of a regulated utility. The problem consists in the scheduling of start-ups and shut-downs of the thermal units and the determination of corresponding operation levels such as to minimize the associated costs and meet the electricity demand over some finite time horizon. To formalize this, we let $\mathcal{I}$ index the thermal units. We denote the on/off-decisions of the units $u_i^n, i \in \mathcal{I} \in \{0,1\}, n \in \mathcal{N}$ and let the $p_i^n \in \mathbb{R}_+, i \in \mathcal{I}, n \in \mathcal{N}$ represent the corresponding operating levels. Total costs account for operational fuel costs and start-up costs and are modeled by the functions $FC_i(\cdot), i \in \mathcal{I}$ and $SC_i(\cdot), i \in \mathcal{I}$ in (3.4.1). The functions are occasionally approximated by piecewise linear functions in order for the problem to comply with a linear formulation. According to (3.4.2), generation is subject to lower and upper bounds, $p_i^{min}, p_i^{max}, i \in \mathcal{I}$, resulting from the capacities of the thermal units. To prevent thermal stress and high maintenance costs of the units, minimum up- and down-time constraints apply. These are given by (3.4.3), and (3.4.4), where $\tau_i^{up}, \tau_i^{do}, i \in \mathcal{I}$ denote the minimum up- and down-times. Although not included, must-on/-off constraints can easily be included. A regulated utility is forced to satisfy demand by means of generation only, which leads to the constraints (3.4.5), where the demand is denoted by $d^n, n \in \mathcal{N}$. Finally, the fact that reserves are the responsibility of the utility induces the constraints (3.4.6), where the reserve requirements are denoted by $r^n, n \in \mathcal{N}$. The reserves are so-called spinning reserves that ensure excess system capacity in the case of failure. Uncertainty may relate to both demand and reserve requirements. With this notation, the thermal unit commitment problem formulated as a multi-stage stochastic program takes the form of

$$\min \sum_{n\in\mathcal{N}} \sum_{i\in\mathcal{I}} \pi^n \left(FC_i(p_i^n, u_i^n) + SC_i(u_i^n, u_i^{n-1})\right) \tag{3.4.1}$$

$$\text{s.t. } u_i^n p_i^{min} \leq p_i^n \leq u_i^n p_i^{max}, \;\; i \in \mathcal{I}, n \in \mathcal{N} \tag{3.4.2}$$

$$u_i^{n-\tau} - u_i^{n-(\tau+1)} \leq u_i^n, \;\; \tau = 1, \ldots, \tau_i^{up} - 1, i \in \mathcal{I}, n \in \mathcal{N} \tag{3.4.3}$$

$$u_i^{n-(\tau+1)} - u_i^{n-\tau} \leq 1 - u_i^n, \;\; \tau = 1, \ldots, \tau_i^{do} - 1, i \in \mathcal{I}, n \in \mathcal{N} \tag{3.4.4}$$

$$\sum_{i\in\mathcal{I}} p_i^n \geq d^n, \;\; n \in \mathcal{N} \tag{3.4.5}$$

$$\sum_{i\in\mathcal{I}} (u_i^n p_i^{max} - p_i^n) \geq r^n, \;\; n \in \mathcal{N} \tag{3.4.6}$$

$$p_i^n \geq 0, u_i^n \in \{0,1\}, \;\; i \in \mathcal{I}, n \in \mathcal{N} \tag{3.4.7}$$

The stochastic problem is a direct extension of the deterministic unit commitment problem and can be found in a very similar version by Takriti et al. (1996). The authors only consider demand uncertainty and leave out reserve requirements. Whether reserve constraints should be included in a stochastic model depends on whether failures are explicitly incorporated in the scenarios as decreased capacity or increased demand corresponding to the capacities of disrupted units.

With a fixed configuration of the units, the problem becomes that of distributing the production among them and is known as the *economic dispatch* problem. Both Bunn and Paschentis (1986) and Gröwe et al. (1995) formulate the problem as a two-stage simple recourse problem, where the second stage penalizes deviations between scheduled production and uncertain demand. In solving the problem, both rely on separability of the recourse function and the possibility to evaluate only a one-dimensional integral. However, whereas Bunn and Paschentis (1986) assume normal demand distributions and use discretization of the distribution to solve the economic dispatch problem, Gröwe et al. (1995) approximate the distributions by certain kernel estimators that allow explicit calculation of the integrals.

The thermal unit commitment problem can be extended to the *hydro-thermal scheduling* problem that involves cooperation of thermal power and hydro-power pumped storage plants. Like the thermal unit commitment problem, the problem consists in the scheduling of thermal power production. However, there is a possibility of using electricity to store water in the reservoirs for future hydro-power generation. The objective is therefore to minimize the thermal operational costs as well as the opportunity costs of hydro-power generation that arise since water could be saved for future use. We let $\mathcal{J}$ index the pumped storage plants. The opportunity costs are measured as the future value of stored water, captured by the functions $V_j(\cdot), j \in \mathcal{J}$, as in (3.4.8). Including water values is an attempt to avoid end effects such as the tendency of the multi-stage stochastic programming problem to empty the system in the final stage. We let the notation be the same as before. Moreover, we let the variables $v_j^n, s_j^n, l_j^n \in \mathbb{R}_+, j \in \mathcal{J}, n \in \mathcal{N}$ represent the reservoir discharge levels, pumping levels and storage levels. Apart from the thermal constraints, hydro-thermal scheduling involves bounds on these variables, (3.4.10)-(3.4.12), enforced by $v_j^{min}, v_j^{max}, s_j^{min}, s_j^{max}, l_j^{min}, l_j^{max}, j \in \mathcal{J}$. Reservoir discharges, pumping levels and storage levels are related through the reservoir balance equations, (3.4.13), where $\nu_j^n, j \in \mathcal{J}, n \in \mathcal{N}$ denote reservoir inflows. Finally, demand can be satisfied from thermal generation and hydro-power generation that is not used for pumping. This gives rise to the constraints (3.4.14) in which $\eta_j, \gamma_j, j \in \mathcal{J}$ denote the hydro-power generation and pumping efficiencies, respectively. To simplify the modeling, water discharges and hydro-power generation are often assumed to be proportional. As should be clear from the notation, uncertainty may concern demand and reserve requirements as well as reservoir inflows. Formulated as a stochastic programming problem, the hydro-thermal scheduling

problem of a regulated utility is therefore

$$\min \sum_{n\in\mathcal{N}}\sum_{i\in\mathcal{I}} \pi^n \big(FC_i(p_i^n, u_i^n) + SC_i(u_i^n, u_i^{n-1})\big) - \sum_{n\in\mathcal{N}_T}\sum_{j\in\mathcal{J}} \pi^n V_j(l_j^n) \tag{3.4.8}$$

$$\text{s.t. } u_i^n p_i^{min} \le p_i^n \le u_i^n p_i^{max}, \;\; i \in \mathcal{I}, n \in \mathcal{N} \tag{3.4.9}$$

$$v_j^{min} \le v_j^n \le v_j^{max}, \;\; j \in \mathcal{J}, n \in \mathcal{N} \tag{3.4.10}$$

$$s_j^{min} \le s_j^n \le s_j^{max}, \;\; j \in \mathcal{J}, n \in \mathcal{N} \tag{3.4.11}$$

$$l_j^{min} \le l_j^n \le l_j^{max}, \;\; j \in \mathcal{J}, n \in \mathcal{N} \tag{3.4.12}$$

$$l_j^n - l_j^{n-1} + v_j^n - s_j^n = \nu_j^n, \;\; j \in \mathcal{J}, n \in \mathcal{N} \tag{3.4.13}$$

$$\sum_{i\in\mathcal{I}} p_i^n + \sum_{j\in\mathcal{J}} \eta_j (v_j^n - \gamma_j s_j^n) \ge d^n, \;\; n \in \mathcal{N} \tag{3.4.14}$$

$$\sum_{i\in\mathcal{I}} (u_i^n p_i^{max} - p_i^n) \ge r^n, \;\; n \in \mathcal{N} \tag{3.4.15}$$

$$p_i^n, v_j^n, s_j^n, l_j^n \ge 0, u_i^n \in \{0,1\}, \;\; i \in \mathcal{I}, j \in \mathcal{J}, n \in \mathcal{N}. \tag{3.4.16}$$

Short-term hydro-thermal scheduling models in stochastic programming are found several places in the literature. Dentcheva and Römisch (1998), Gröwe-Kuska et al. (2000), Gröwe-Kuska et al. (2002) and Nowak and Römisch (2000) all suggest models that take the form of the above with a time horizon of a week. Their studies represent different options in modeling. The authors propose to model the operational costs as linear, piecewise linear, piecewise linear convex or convex quadratic, whereas start-up costs can be fixed or down-time dependent. Furthermore, the authors include either minimum down-time constraints or both up- and down-time constraints. None of the studies include a water value function. Instead, to avoid undesired end effects due to a finite time horizon, the final storage levels are constrained. The models are all validated on data from a German utility and solved by the use of stochastic Lagrangian relaxation of the unit coupling constraints, which will be further discussed in the subsequent sections. As a representative example of the model sizes in this section, the model contains 4200 binary variables, 6652 continuous variables and 13441 constraints as a deterministic problem. With 10 scenarios and 756 nodes in the scenario tree formulation of the stochastic problem, the numbers are 18900, 34776 and 60490 and with 100 scenarios and 4200 nodes, the numbers mount to 105000, 193200 and 336100, making decomposition of the stochastic problem highly relevant.

Similar in spirit to the hydro-thermal scheduling models, are the two-stage unit commitment models of Carøe and Schultz (1998) and Gollmer et al. (2000), that are also validated on the German data. Like the above, the models seek to find a unit commitment schedule for thermal units in a hydro-thermal utility. However, since coal-fired units have longer start-up times than gas-burning units, the on-/off-decisions of the coal-fired units are assigned to the first stage, whereas corresponding decisions of the gas-burning units are set aside for the second stage. For other two-stage hydro-thermal planning problems, see also Dentcheva and Römisch (1998) and Nowak et al. (2000), who assign a full schedule in the first stage and a compensation schedule in the second stage.

Hydro scheduling and unit commitment

When *hydro scheduling* is addressed from the point of view of a regulated utility, the problem consists in the spatial distribution of water releases between different hydro reservoirs in order to satisfy electricity demand. Since direct operating costs of hydro-power generation are negligible, the determining of water releases is a matter of striking a balance between the immediate and future value of the remaining reservoir contents. Hence, the overall objective of hydro scheduling is to maximize the value of water stored in the reservoirs (3.4.17) subject to the demand constraints (3.4.23). The challenge is that, when reservoirs are connected by a network, upstream releases contribute to the downstream inflows with some time delay. For simplicity, we assume that the hydro reservoirs are connected in cascade, although more complex networks also exist. We proceed with the same notation as before and let $\mathcal{J}$ index the hydro reservoirs. We model the relation between reservoir discharges and turbine generation in more detail than above and, thus, introduce the variables $w_j^n \in \mathbb{R}_+, j \in \mathcal{J}, n \in \mathcal{N}$ to represent generation explicitly. If ignoring head variation effects on the generation, the relation is usually given by concave functions $G_j(\cdot), j \in \mathcal{J}$ as in (3.4.22), such functions often being approximated by piecewise linear functions to support a linear formulation. The turbine generation is subjected to the bounds (3.4.18) induced by the generation capacities $w_j^{min}, w_j^{max}, j \in \mathcal{J}$. In the reservoir balance equations we include spills $r_j^n, j \in \mathcal{J}, n \in \mathcal{N}$, and like other releases, upstream spills contribute to the downstream inflows. Uncertainty may arise with respect to reservoir inflow and electricity demand, which makes the multi-stage stochastic programming formulation of the hydro scheduling problem take the form

$$\min \; - \sum_{n \in \mathcal{N}_T} \sum_{j \in \mathcal{J}} \pi^n V_j(l_j^n) \tag{3.4.17}$$

$$\text{s.t. } w_j^{min} \leq w_j^n \leq w_j^{max}, \;\; j \in \mathcal{J}, n \in \mathcal{N} \tag{3.4.18}$$

$$v_j^{min} \leq v_j^n \leq v_j^{max}, \;\; j \in \mathcal{J}, n \in \mathcal{N} \tag{3.4.19}$$

$$l_j^{min} \leq l_j^n \leq l_j^{max}, \;\; j \in \mathcal{J}, n \in \mathcal{N} \tag{3.4.20}$$

$$l_j^n - l_j^{n-1} + v_j^n + r_j^n = v_{j-1}^{n-1} + r_{j-1}^{n-1} + \nu_j^n, \;\; j \in \mathcal{J}, n \in \mathcal{N} \tag{3.4.21}$$

$$w_j = G_j(v_j), \;\; j \in \mathcal{J}, n \in \mathcal{N} \tag{3.4.22}$$

$$\sum_{j \in \mathcal{J}} w_j^n \geq d^n, \;\; n \in \mathcal{N} \tag{3.4.23}$$

$$w_j^n, v_j^n, r_j^n, l_j^n \geq 0 \;\; j \in \mathcal{J}, n \in \mathcal{N}. \tag{3.4.24}$$

Although formulated as medium-term hydro scheduling, Jacobs et al. (1995) present a model similar to the above. The model facilitates the development of a stochastic hydro scheduling module of a larger system in an American gas and electricity company. To fully describe the system and the operating restrictions in practice, the model includes an entire network of lakes, reservoirs, water courses, tunnels, junctions and power houses. The company also operates a thermal system, which is why the major costs concern those of avoiding thermal generation. Even for the deterministic model, the number of variables and constraints are in the range of 4650–8686 and 1180–2230, respectively. With as few as 45 scenarios and 64–66 nodes in the scenario tree formulation of the stochastic model, the numbers become 140284–265242 and 35736–68012, which obviously motivates the need for decomposition.

For a general overview of reservoir operation models within stochastic programming and related areas, see Reznicek and Cheng (1991), who include Markov chains, dynamic programming, stochastic programming with recourse and chance constrained stochastic programming.

The *hydro unit commitment* problem introduces individual units and their start-ups and shut-downs into the hydro scheduling problem. It follows that the problem consists in determining on/off-schedules and corresponding generation levels of the units so as to balance current costs and future water values. Most direct operating costs can be ignored such that current costs account for start-up costs alone. Still, hydro start-up costs are much lower than thermal start-up costs. We again consider hydro reservoirs in cascade. We let $\mathcal{J}$ index the hydro reservoirs and stations and let $\mathcal{I}_j, j \in \mathcal{J}$ index the associated turbines. The start-up costs, $SC_i(\cdot), i \in \mathcal{I}$ are added to the objective function (3.4.25). We denote the on/off-decisions of the hydro units $u_i^n \in \{0,1\}, i \in \mathcal{I}_j, j \in \mathcal{J}, n \in \mathcal{N}$ and the generation levels $w_i^n, n \in \mathcal{N}$. Generation must submit to the bounds $w_i^{min}, w_i^{max}, i \in \mathcal{I}$ in (3.4.26). We define the discharges $v_i^n, n \in \mathcal{N}$ on a unit basis and modify the bounds as in (3.4.27) accordingly. The balance equations must be modified in a similar fashion as shown in (3.4.29). The hydro unit commitment problem now

reads

$$\min \sum_{n\in\mathcal{N}}\sum_{j\in\mathcal{J}}\sum_{i\in\mathcal{I}_j}\pi^n SC_i(u_i^n, u_i^{n-1}) - \sum_{n\in\mathcal{N}_T}\sum_{j\in\mathcal{J}}\pi^n V_j(l_j^n) \tag{3.4.25}$$

$$\text{s.t. } u_i^n w_i^{min} \le w_i^n \le u_i^n w_i^{max},\ \ i\in\mathcal{I}_j, j\in\mathcal{J}, n\in\mathcal{N} \tag{3.4.26}$$

$$v_j^{min} \le \sum_{i\in\mathcal{I}_j} v_i^n \le v_j^{max},\ \ j\in\mathcal{J}, n\in\mathcal{N} \tag{3.4.27}$$

$$l_j^{min} \le l_j^n \le l_j^{max},\ \ j\in\mathcal{J}, n\in\mathcal{N} \tag{3.4.28}$$

$$l_j^n - l_j^{n-1} + \sum_{i\in\mathcal{I}_j} v_i^n + r_j^n = \sum_{i\in\mathcal{I}_{j-1}} v_i^{n-1} + r_{j-1}^{n-1} + \nu_j^n,\ \ j\in\mathcal{J}, n\in\mathcal{N} \tag{3.4.29}$$

$$w_i = G_i(v_i),\ \ i\in\mathcal{I}_j, j\in\mathcal{J}, n\in\mathcal{N} \tag{3.4.30}$$

$$\sum_{j\in\mathcal{J}}\sum_{i\in\mathcal{I}_j} w_i \ge d^n,\ \ n\in\mathcal{N} \tag{3.4.31}$$

$$w_i^n, v_i^n, r_j^n, l_j^n \ge 0, u_i^n \in \{0,1\}\ \ i\in\mathcal{I}_j, j\in\mathcal{J}, n\in\mathcal{N}. \tag{3.4.32}$$

To our knowledge, hydro unit commitment has been addressed only few times in the literature. The authors of Hreinsson (1988) and Tufegdzic et al. (1996) present a short-term operation problem of a hydro-power system modeled as a deterministic problem, whereas Philpott et al. (2000) consider the problem subject to uncertain demand and formulates a multi-stage stochastic program. Hreinsson (1988) argues that if the hydro reservoirs are almost unaffected by short-term operations, water scheduling and hydro-power generation can be separated. By addressing only hydro power generation, the problem reduces to something very similar to the thermal unit commitment problem. Tufegdzic et al. (1996) on the other hand describe a full hydro-power generation problem involving unit commitment and water scheduling. The problem arose with the development of an energy management system for the Tasmanian hydro electric commission and the corresponding model is therefore rather detailed, including a hydro network of reservoirs, stations, tunnels and canals and piecewise linear approximations of some constraints and costs. Finally, Philpott et al. (2000) present a hydro unit commitment model similar to the above and illustrates it with a case study from New Zealand. The relations between reservoir discharges and turbine generation are given by concave functions that are combined and approximated by piecewise linear functions.

3.5 Production on market conditions

The lack of organized markets during regulation was argued to make power trading rather inflexible. To achieve economic efficiency, the markets began to reform and a new regulatory administration was introduced for the purpose of promoting competition. Prices were left to be determined by the market conditions through the establishment of a power exchange. Moreover, previous obligations to satisfy demand were replaced by the opportunity of power producers to buy and sell their production at this power exchange. Producers reacted by reorganizing operations to facilitate decentralized decision-making and began maximizing profit, while experiencing an increasing price uncertainty. For a general overview of the changes in the power sector in response to the restructuring, see Hobbs (1995), who includes the growing uncertainty and increased competition.

The models of this section represent the adaption of the traditional power production models to the new environment in terms of profit maximization without demand constraints. As before, the multi-stage stochastic programs are presented using the scenario tree formulation and the corresponding notation is found in Chapter 1.

Thermal unit commitment and hydro-thermal scheduling

We will explain how the traditional models of Section 3.4 can be adapted to the new environment using the thermal unit commitment model for illustration. Changes to the hydro-thermal scheduling model will follow in a similar fashion.

We apply the ideas of Wallace and Fleten (2003). With the possibility of trading in the physical and financial markets, the electricity production can be disposed of through traditional bilateral contracts and through newer market contracts. Furthermore, power producers can purchase electricity from the same markets. We denote by $d^n, n \in \mathcal{N}$ the bilateral contracts and let the variables $y^{+,n}, y^{-,n} \in \mathbb{R}_+, n \in \mathcal{N}$ represent the market contracts for selling and buying, respectively. With no constraints on the market contracts, these provide the flexibility necessary for production, purchases and disposals to match, which leads to the equality constraints

$$\sum_{i \in \mathcal{I}} p_i^n = d^n + y^{+,n} - y^{-,n}, \quad n \in \mathcal{N}. \tag{3.5.1}$$

The revenues of market disposals amount to

$$\sum_{n \in \mathcal{N}} \pi^n \rho^n (y^{+,n} - y^{-,n}), \tag{3.5.2}$$

where $\rho^n, n \in \mathcal{N}$ denote the market prices. By substitution of $y^{+,n}, y^{-,n}, n \in \mathcal{N}$ in (3.5.2), the demand constraints (3.5.1) are eliminated. After substitution, the

constant term of the objective function can be ignored, which corresponds to an assumption of no bilateral trades and pure market disposals. Reserve constraints can also be eliminated from the model, the reason being that reserves market are beginning to appear. As an effect of eliminating demand and reserve constraints, the model decouples with respect to thermal units and, thus, decision-making can be conducted on a single-unit basis. With the introduction of revenues from market disposals, the objective shifts from cost minimization to profit maximization and market price uncertainty comes into play. The thermal single-unit problem is the multi-stage stochastic program

$$\begin{aligned} \max \ & \sum_{n\in\mathcal{N}} \pi^n \big(\rho^n p^n - FC(p^n, u^n) - SC(u^n, u^{n-1})\big) \\ \text{s.t. } & (3.4.2) - (3.4.4) \\ & p^n \geq 0, u^n \in \{0,1\}, \ \ n \in \mathcal{N}. \end{aligned}$$

Takriti et al. (2000) addresses the multi-stage stochastic thermal unit commitment problem and include both buying and selling of electricity to a power pool that serves as a spot market. The spot market is modeled as two additional units; one unit representing buying and one unit representing selling. If the utility buys power, demand decreases, and if it sells power, demand increases as shown above. The bounds on the buying and selling units are automatically imposed, which allows for the substitution. However, the model decoupling is not carried out.

As already stated, the hydro-thermal scheduling problem is modified in the same way as the thermal unit commitment problem. However, when the demand and reserve constraints are eliminated, the model separates thermal power production from hydro-power generation and decouples further with respect to pumped hydro storage plants.

Hydro scheduling and unit commitment

To illustrate how the traditional models of Section 3.4 can be modified, we consider the hydro scheduling problem. Similar modifications apply to the hydro unit commitment problem.

Since also hydro-power generation can be disposed of by bilateral contracts of a fixed volume and market contracts of a variable volume, the traditional demand constraints are replaced by the equality constraints

$$\sum_{j\in\mathcal{J}} w_j^n = d^n + y^{+,n} - y^{-,n}, \ \ n \in \mathcal{N},$$

and the revenues of market disposals,

$$\sum_{n \in \mathcal{N}} \pi^n \rho^n (y^{+,n} - y^{-,n}),$$

are added to the traditional objective function, which is then based on profit maximization and includes market prices uncertainty. For the same reasons as previously, the demand constraints can be eliminated. If, however, the reservoirs are connected, the balance equations prevent the hydro scheduling problem from decoupling with respect to reservoirs.

The paper of Fosso et al. (1999) gives an overview of the deregulated electricity system in Norway and the corresponding long-term, medium-term and short-term models used for hydro scheduling. In particular, it is explained how water value calculations can be adapted to the system. The long-term model is a stochastic program that is able to handle uncertainties in market prices and reservoir inflows and provides water values for the medium-term models. The authors suggest to calculate marginal water values as derivatives of stochastic dynamic programming value functions with respect to reservoir storage levels. The medium-term models are deterministic scenario problems producing water values for a detailed short-term model that includes nonlinear generation efficiency curves, head variation effects, tunnels branching into pipes etc.

The adaptation of the hydro unit commitment problem to a market-orientated system has not been found in the literature until recently, where a stochastic programming version has appeared by García-González et al. (2006). More interestingly, however, an alternative to handling head variation effects is proposed. A procedure iteratively produces reservoir net heads as a function of storage levels and solves the stochastic programming problem with a relation between reservoir discharge and turbine generation that depends only on a fixed net head.

In Chapter 5, we describe a new hydro unit commitment problem that has become relevant with the introduction of market dispatch. The problem consists in hydro-power production planning that complies with the day-ahead commitments and illustrates the necessity of the demand constraints in the very short term. In contrast to dynamic unit commitment, the aim is to determine a daily production plan that values the current market prices and reservoir inflows against the future. With fixed day-ahead commitments, the production planning becomes a matter of spatial distribution of water among the reservoirs.

3.6 Solution approaches

Thermal unit commitment

The solution approaches to stochastic power production problems are extensions from the deterministic case. Primal approaches mainly rely on LP based branch

and bound, this method being highly supported by the advances in hardware and the development of software implementations. At present, general purpose packages can combine the LP methodology with a variety of options for arranging the branch and bound. The package allows for fast heuristics to obtain initial feasible solutions, specific branching rules and techniques from combinatorial optimization such as valid inequalities. Most importantly, the LP-based branch and bound works with ample enrichment as long as the model is expressed in mixed-integer linear terms. A critical drawback is the full handling of the model which may become prohibitive with the sizes illustrated in the previous sections. This paves the way for decomposition methods.

For an example on the application of LP-based branch and bound to deterministic thermal unit commitment and hydro-thermal scheduling, see Gollmer et al. (1999). The authors argue that the method shows its strengths with complex constraints interconnecting generating units in contrast to dual approaches that have their merits with nonlinearities.

Among the dual approaches, Lagrangian relaxation has proved to be a strong tool because of the algorithmic progress for solving the Lagrangian dual, the usually small duality gap and the advance of fast Lagrangian heuristics to close this gap. For references on the application of Lagrangian relaxation in the deterministic case, see Feltenmark et al. (1997) and Gollmer et al. (1999).

To illustrate Lagrangian relaxation, we develop the Lagrangian dual of the multi-stage stochastic programming version of the thermal unit commitment problem (3.4.1)-(3.4.7) and its scenario tree formulation. The problem is nearly separable with respect to thermal units as only the constraints (3.4.5) and (3.4.6) involve different units. This property may be utilized by *stochastic Lagrangian relaxation* of the unit coupling constraints. Assigning non-negative stochastic Lagrange multipliers that inherit the tree structure from $d^n, r^n, n \in \mathcal{N}$, the Lagrangian function is given by

$$\begin{aligned} L(u,p;\lambda_1,\lambda_2) := & \sum_{n\in\mathcal{N}}\sum_{i\in\mathcal{I}} \pi^n (FC_i(p_i^n,u_i^n) + SC_i(u_i^n,u_i^{n-1})) + \\ & \sum_{n\in\mathcal{N}} \pi^n \lambda_1^n \Big(d^n - \sum_{i\in\mathcal{I}} p_i^n \Big) + \\ & \sum_{n\in\mathcal{N}} \pi^n \lambda_2^n \Big(r^n - \sum_{i\in\mathcal{I}} (u_i^n p_i^{max} - p_i^n) \Big), \end{aligned}$$

where $u = (u_i^n)_{i\in\mathcal{I}n\in\mathcal{N}}$ and $p = (p_i^n)_{i\in\mathcal{I}n\in\mathcal{N}}$ and the corresponding dual function is

$$D(\lambda_1,\lambda_2) := \min_{u,p} \big\{ L(u,p;\lambda_1,\lambda_2) \mid (3.4.2) - (3.4.4) \big\} \tag{3.6.1}$$

The Lagrangian dual now reads

$$\max\big\{D(\lambda_1,\lambda_2)\mid(\lambda_1,\lambda_2)\in\mathbb{R}_+^{2|\mathcal{N}|}\big\}. \tag{3.6.2}$$

Due to integrality constraints, the primal problem is non-convex. The dual problem therefore only provides a lower bound to the primal problem. Still, the Lagrangian relaxation provides a tighter bound than the continuous relaxation.

The problem (3.6.1) decomposes into multi-stage single-unit subproblems. In this fashion the dual function

$$D(\lambda_1,\lambda_2)=\sum_{i\in\mathcal{I}}D_i(\lambda_1,\lambda_2)+\sum_{n\in\mathcal{N}}\pi^n\big(\lambda_1^n d^n+\lambda_2^n r^n\big)$$

is evaluated by solving subproblems of the form

$$D_i(\lambda_1,\lambda_2)=\min_{u_i}\Big\{\sum_{n\in\mathcal{N}}\pi^n\big(\min_{p_i^n}\big\{FC_i(p_i^n,u_i^n)-(\lambda_1^n-\lambda_2^n)p_i^n\mid(3.4.2)\big\}$$
$$+SC_i(u_i^n,u_i^{n-1})-\lambda_2^n u_i^n p_i^{max}\big)\;\Big|\;(3.4.3)-(3.4.4)\Big\},$$

where $u_i=(u_i^n)_{n\in\mathcal{N}}$. The single-unit subproblems are solvable by stochastic dynamic programming. For an outline of the approach, including how to cope with minimum up- and down-times, see for instance Nowak and Römisch (2000).

Since the dual function (3.6.1) is concave and non-differentiable, the Lagrangian dual (3.6.2) was originally solved with subgradient procedures. Currently more refined methods such as cutting plane or bundle methods have been successfully applied. An example is the proximal bundle method used in Dentcheva and Römisch (1998), Nowak and Römisch (2000) and Gröwe-Kuska et al. (2002). Based on function and subgradient information, the method constructs a bundle of linearizations of the dual function. In order to replace the original dual function, the resulting piecewise linear upper approximation is added a proximity term before being minimized.

Mostly the dual solution provided by the cutting plane or bundle method violates the demand and reserve constraints and produces a duality gap. Lagrangian heuristics are therefore used to determine a feasible and hopefully nearly optimal solution of the primal problem. In most cases, the heuristic seeks to find a unit commitment solution that induces economic dispatch. The heuristic suggested in Gröwe-Kuska et al. (2002) starts from an initial set of Lagrange multipliers and perturbs the multipliers such as to obtain primal feasible unit commitment schedules. Most binary variables do not change and can be fixed in order to decrease the size of the problem, whereas the remaining variables are switched off one at a time as long as feasibility persists. Dentcheva and Römisch (1998) provide a bound on the duality gap that depends on the number of demand and reserve

constraints. The duality gaps decreases in the case of a larger power plant or a finer discretization of the probability distribution.

The authors Dentcheva and Römisch (1998), Nowak and Römisch (2000) and Gröwe-Kuska et al. (2002) employ the stochastic Lagrangian relaxation to the stochastic hydro-thermal scheduling problem. This prompts a decomposition into both single-unit thermal and hydro subproblems and opens the possibility of heuristics that exploit the additional flexibility of the hydro-power pumped storage plants. Since the test problem from the German utility involves 25 thermal units and 7 hydro reservoirs, it decomposes into 32 subproblems. For the instance of 100 scenarios and 4200 nodes, the thermal subproblems each have 4200 binary variables, 4200 continuous variables and 8400 constraints and the hydro subproblems 12600 continuous variables and 8500 constraints. The instance can be solved within an optimality gap of approximately 0.1% in about 1000 seconds of computing time. The two-stage problems by Dentcheva and Römisch (1998) and Nowak et al. (2000) are solved on similar lines by Lagrangian relaxation of the unit coupling constraints, the result being a similar decomposition into thermal and hydro subproblems.

Relaxation of unit coupling constraints is also possible by means of the augmented Lagrangian technique as suggested by Carpentier et al. (1996). However, to facilitate decomposition, it is necessary to linearize the Lagrangian. Other dual solution approaches to stochastic power production problems include Lagrangian relaxation of the non-anticipativity constraints, which usually decomposes the problems into a larger number of smaller deterministic subproblems. In this way, Takriti et al. (1996) and Takriti et al. (2000) solve the multi-stage stochastic unit commitment problem by progressive hedging and seem to outperform the above approach. Carøe and Schultz (1998) and Gollmer et al. (2000) solve the two-stage version of the problem by dual decomposition are thereby, in principle, able to close the duality gap by branch and bound. For further references on Lagrangian relaxation of the non-anticipativity constraints, see Chapter 1.

Hydro scheduling

For a long time, stochastic hydro scheduling problems have been solved by *stochastic dynamic programming*, which has also occasionally been used for stochastic hydro unit commitment problems. The application of dynamic programming is supported by the sequential structure of the decision-making process and justified by the fact that the problems have relatively few constraints interconnecting different stages. We attempt to illustrate the dynamic programming recursion on the hydro scheduling problem (3.4.17)-(3.4.24) for reservoirs in cascade. The notation of the preceding sections will be modified to fit the framework. We assume that the uncertain data consists of random demands $d_t, t \in \mathcal{T}$ and random inflows $\nu_{jt}, j \in \mathcal{J}, t \in \mathcal{T}$. For reservoirs in cascade, the storage levels can be redefined

to include delayed inflows from upstream reservoirs. The dynamic programming states are defined by these storage levels, still denoted $l_t = (l_{jt})_{j\in\mathcal{J}}, t \in \mathcal{T}$, and by the uncertain data $\xi_t = (\nu_t, d_t), t \in \mathcal{T}$, where $\nu_t = (\nu_{jt})_{j\in\mathcal{J}}, t \in \mathcal{T}$. The action space consists of reservoir discharges, turbine generation levels and spills $v_t = (v_{jt})_{j\in\mathcal{J}}, w_t = (w_{jt})_{j\in\mathcal{J}}, r_t = (r_{jt})_{j\in\mathcal{J}}, t \in \mathcal{T}$. With these definitions, the expected future value function of stage T compute as

$$F_T(l_{T-1}, \xi_{T-1}) = \mathbb{E}_{\xi_T|\xi_{T-1}}\Big[\max_{v_T, w_T, r_T}\Big\{-\sum_{j\in\mathcal{J}} V_j(l_{jT} - v_{j-1T} - r_{j-1T}) \;\Big|\; (3.4.18) - (3.4.20), l_{jT} - v_{j-1T} - r_{j-1T} + v_{jT} + r_{jT} = l_{jT-1} + \nu_{jT}, j \in \mathcal{J}, (3.4.22) - (3.4.24)\Big\}\Big]$$

and the stage t expected future value as

$$F_t(l_{t-1}, \xi_{t-1}) = \mathbb{E}_{\xi_t|\xi_{t-1}}\Big[\max_{v_t, w_t, r_t}\Big\{F_{t+1}(l_t, \xi_t) \;\Big|\; (3.4.18) - (3.4.20), l_{jt} - v_{j-1t} - r_{j-1t} + v_{jt} + r_{jt} = l_{jt-1} + \nu_{jt}, j \in \mathcal{J}, (3.4.22) - (3.4.24)\Big\}\Big].$$

In dynamic programming, the continuous reservoir storage levels are often discretized in order to facilitate computations. With the discretization of the state space, however, the full dynamic programming approach is known to suffer from the curse of dimensionality and is only able to handle a few reservoirs. In order to restrict the state space, it has been proposed to aggregate reservoirs and power stations or to decompose the dynamic programming problem according to reservoirs. As an alternative, Archibald et al. (2001) suggest that an appropriate definition of the state space for multi-reservoir systems allows for a simplification of the solution method on the basis of information on an optimal policy.

Another way of avoiding the curse of dimensionality is to describe the dynamic programming value function by supporting hyperplanes, which is utilized in *nested Benders' decomposition*. The method has been applied by Jacobs et al. (1995), who suggest a number of algorithmic enhancements that are further explored in Morton (1996). The enhancements include the use of warm start bases, initial cut generation, disaggregated cuts and decision tree traversing strategies. The performance of the enhanced algorithm is tested on a collection of multi-stage stochastic hydro scheduling problems. The algorithm is outperformed by general LP optimizers if applied to single scenario problems but is preferable as the number of scenarios increases. As an example, the previously mentioned problem by Jacobs et al. (1995) decomposes into 45 nodal subproblems of 57–1273 variables and 271–6279 constraints - a significant reduction in size - and is solved in 200–400 seconds of CPU time.

For further inspections of algorithms for multi-stage stochastic hydro scheduling problems, the authors Archibald et al. (1996) compare the revised simplex method, full dynamic programming, dynamic programming decomposition and nested Benders' decomposition. Only the revised simplex method and nested Benders' decomposition provide exact optimal solutions. Furthermore, the authors find the nested Benders' decomposition approach to be the fastest followed by dynamic programming decomposition, the revised simplex method and, finally, full dynamic programming. Although most results favor the nested Benders' decomposition, the approach fails to solve stochastic programs having many stages due to the explosion in the number of nodes.

The major drawbacks of dynamic programming and nested Benders' decompositions are avoided in the *stochastic dual dynamic programming* algorithm proposed by Pereira and Pinto (1991). In a backward run, the algorithm approximates the dynamic programming value function by Benders' cuts induced by trial points that are determined in a forward run by sampling from the set of scenarios. Since state space discretization is unnecessary, the curse of dimensionality is not an issue and because of sampling, the effects of the explosion in the number of nodes are reduced. The stochastic dual dynamic programming, however, is not capable of incorporating market price uncertainty due to lack of convexity in the state space, whereas both dynamic programming and nested Benders' decomposition are applicable to the models for hydro scheduling on market conditions.

The stochastic hydro unit commitment problems have to be solved by dynamic programming, Philpott et al. (2000) being an example, or more general approaches applicable to mixed-integer linear stochastic programming problems. An example of the application of commercial mixed-integer linear programming software is given in Chapter 5, where we consider a multi-stage stochastic hydro unit commitment problem and succeed in solving its scenario tree formulation of 267 scenarios, 11777 nodes, 141846 variables and 261254 constraints to optimality in less than 30 seconds of CPU time.

3.7 Physical trading and bidding

With the liberalization of power markets, new planning problems have arisen and introduced the need for corresponding optimization models. Of great importance are especially the problems of physical trading and bidding in power markets like the spot and the reserve markets. This section presents some of the models that have recently been proposed within the field of stochastic programming and related areas of stochastic optimization. A major difference between the models is the handling of offers or bidding curves submitted to the market. Some models are based on smooth, linear or piece-wise linear bidding curves, while others include a finite number of offers which results in step functions. Another crucial difference

is the extent to which the current market participant has market power. Some models operate on a price-taker assumption, whereas other models assume the market participant is able to influence market prices. If, however, competitors respond to the offers of each other, a game theoretic approach should replace the stochastic programming approach. No general solution approach appears to apply to the newer problems of physical trading and bidding.

The day-ahead market

Most models for *bidding* into the spot or day-ahead market involve one-period sealed auctions in an electricity market that is organized as a pool and for which a uniform clearing price rule applies. Within this framework, Fleten and Pettersen (2005) propose a two-stage stochastic mixed-integer linear program for constructing piece-wise linear bidding curves to be submitted to the Nordic day-ahead market. The problem is addressed from the perspective of a price-taking retailer who supplies to price-sensitive end users under both price and demand uncertainty. Since the problem consists in demand-side bidding, decision-making can be made on an hourly basis. First-stage decisions involve the bids to be submitted, whereas the second stage accounts for the actual dispatch and the settlement of day-ahead and intra-day market costs and risks. The modeling of Chapter 4 resembles this approach in some respects. However, although also assuming a price-taker, the problem is approached from the point of view of a hydro-power producer facing the coordination of day-ahead market trading and hydro-power generation. The complexity of the problem increases in that the time coupling due to reservoir balances and turbine start-up costs forces decision-making to be effected on a daily basis. As before, first-stage decisions relate to day-ahead bidding, whereas the second stage now comprises the dispatch of the system, the settlement of costs and risk and the relatively flexible hydro-power generation to be undertaken in real-time. By formulating the deterministic equivalent of the two-stage stochastic program, the problem of Chapter 5 is solvable by standard mixed-integer linear programming software within small computing times, which also applies to the problem by Fleten and Pettersen (2005). A slightly different two-stage stochastic mixed-integer linear programming model is presented by Nowak et al. (2005). The model describes simultaneous power production and day-ahead power trading of a hydro-thermal producer who is able to influence market prices. The price clearing mechanism is based on the equilibrium between demand and supply and is modeled by the use of logical constraints and binaries. The price clearing serves as the second stage, whereas the first stage consists in bidding and production planning. The producer is subjected to uncertainty with respect to foreign bids, although the question remains as to whether a producer would have market information sufficiently detailed to estimate the necessary distribution. The extensive use of binaries significantly increases the model complexity and in spite of applying de-

composition, the solution time is about 3 hours for a test problem with as few as 10 scenarios.

To illustrate the modeling of bidding curves, define a bid as a price-volume pair (x, p). The problem of selecting both a price p and a volume x is nonlinear. To circumvent this problem, the studies mentioned above all discretize the continuous price range into a finite number of fixed price points $p_1 \leq \cdots \leq p_H$. Feasible bids are then $(x_1, p_1), \ldots, (x_H, p_H)$ of which only the volumes $x_1, \ldots, x_H \in \mathbb{R}_+$ have to be selected. Let the bidding curve be defined by the relation between volume and price, denoted by y and ρ respectively. Now Nowak et al. (2005) suggest the use of hourly block bids which results in the following bidding curve with respect to selling bids

$$y = \begin{cases} x_1 & , \text{if } \; p_1 \leq \rho < p_2 \\ \vdots & \\ \sum_{h' \leq h-1} x_{h'} & , \text{if } \; p_{h-1} \leq \rho < p_h \\ \vdots & \\ \sum_{h' \leq H} x_{h'} & , \text{if } \; \rho = p_H. \end{cases}$$

As should be clear, the bidding curve is a nondecreasing step function. In contrast, Fleten and Pettersen (2005) perform a linear interpolation between the price-volume points and construct the following piece-wise linear nondecreasing bidding curve that is consistent with the rules of the Nordic day-ahead market

$$y = \begin{cases} \frac{\rho - p_1}{p_2 - p_1} x_2 + \frac{p_2 - \rho}{p_2 - p_1} x_1 & , \text{if } \; p_1 \leq \rho < p_2 \\ \vdots & \\ \frac{\rho - p_{h-1}}{p_h - p_{h-1}} x_h + \frac{p_h - \rho}{p_h - p_{h-1}} x_{h-1} & , \text{if } \; p_{h-1} \leq \rho < p_h \\ \vdots & \\ \frac{\rho - p_{H-1}}{p_H - p_{H-1}} x_{H-1} + \frac{p_H - \rho}{p_H - p_{H-1}} x_H & , \text{if } \; p_{H-1} \leq \rho \leq p_H. \end{cases}$$

In Chapter 4, the modeling of the Nordic day-ahead market is further extended to include block bids of a duration longer than an hour.

Although the authors do not apply stochastic programming, Wen and David (2001b) present a related approach to the problem of deriving optimal bidding strategies. In the same spirit as above, the problem concerns a supplier submitting hourly bidding curves to the day-ahead market. Since the supplier has market power, the price clearing mechanism is included in the hourly bidding problem formulated as a nonlinear stochastic optimization problem in which foreign supply curves are uncertain. For the complete bidding strategy, the hourly bidding problem must be coordinated with unit commitment to account for the time coupling

induced by start-up costs. This is done by means of a heuristic. An extension to the same problem is provided by Wen and David (2001a), who consider both large producers and consumers trading in the day-ahead market.

The bidding curves suggested by Wen and David (2001b) and Wen and David (2001a) are linear functions of the form

$$\rho = \alpha + \beta y$$

in the case of selling bids and, similarly, in the case of buying bids. The coefficients $\alpha, \beta \in \mathbb{R}_+$ are optimization variables of the bidding problem. The coefficients of the foreign bidding curves are random variables assumed to have a joint distribution. Again, it is unclear whether it would be possible to estimate the joint distribution of the foreign bidding curves.

For other studies on bidding, Neame et al. (2003) consider a price-taking generator making offers into an electricity spot market under uncertainty in market prices. According to the market rules, offers are restricted to hourly block bids. In that the generator does not affect market prices, the optimal offers reflect the marginal costs of generation, although the market rules only allow an approximation. The solution of the proposed nonlinear programming problem is illustrated in combination with thermal unit commitment and hydro scheduling in a New Zealand river system. Prichard and Zakeri (2003) further address the coordination of hydro scheduling and bidding of a price-taking producer. However, the authors argue that since the costs of hydro-power generation include only opportunity costs of released water, there is no simple way to determine marginal costs and bid accordingly. Instead they suggest to use stochastic dynamic programming for deriving bidding curves on the general form

$$y(\rho), \quad \rho \in [0, \infty).$$

In a different study, Anderson and Philpott (2002) investigate strategies for generators having market power. In formulating the bidding problem of the generator, the bidding curve is modeled as a continuous parametrized curve

$$\sigma = \{(y(k), \rho(k)), 0 \leq k \leq K\},$$

where $y(\cdot), \rho(\cdot)$ are increasing functions. The authors analyze both smooth curves and the step functions that arise in the case of a finite number of bids. The clearing of the spot market is established in a separate network flow model from which spot prices can be obtained as marginal values. The generator is able to influence clearing prices by providing input to this model. Market prices are further affected by random demand and supply of competitors. To encapsulate the effects of uncertainty on the dispatch of the current generator, the so-called market distribution function $\varphi(y, \rho)$ comes into play. This function alone is sufficient to determine the

stochastic behavior of the market and is defined as the probability of not being fully dispatched in the market if offering the volume y at a price of ρ. However, the authors admit the difficulties in estimating such a function. The bidding problem itself is that of selecting the bidding curve that maximizes expected profit in terms of the market distribution function and can be formulated as a nonlinear optimal control problem. Philpott and Schultz (2006) integrate the framework with thermal production and unit commitment in particular. The authors propose two two-stage decision problems for determining optimal offers of either a single or several thermal units. The first stage of the single-unit problem seeks to determine offers of a must-run auction for generating units, while the second stage computes the optimal spot market offers given the results. Considering several generating units, the first stage optimizes the spot market offers and the second stage determines the units to run in order to meet the dispatch. In both cases, the unit commitment part of the problems is solved by dynamic programming, whereas the bidding part is handled by optimality conditions.

A common aim of most bidding problems from the literature is that of selecting bidding curves that maximize expected profit. Profit amounts to

$$P(y, \rho) = y\rho - C(y),$$

where $y\rho$ are revenues from market dispatch and the function $C(y)$ accounts for operational costs and possibly start-up and shut-down costs as well as opportunity costs. Profit may also involve revenues from more long-term physical and financial contracts. Neame et al. (2003) first analyze general cost functions and derive optimality conditions for local optima, and second explore piece-wise linear cost functions and the application of dynamic programming to locate global optima. Anderson and Philpott (2002) consider only smooth operating costs, whereas Philpott and Schultz (2006) extend the situation to allow for a fixed number of jump discontinuities due to start-up and shut-down costs. In both cases, they express the expected profit as an integral with respect to the market distribution function. In particular, if the market distribution function φ is continuous, the expected profit can be calculated as

$$\int_{\sigma} P(y, \rho)\varphi(dy, d\rho).$$

While stochastic programming practitioners usually discretize the distribution of the random components to facilitate computations, Anderson and Philpott (2002) and Philpott and Schultz (2006) seek to solve the problems directly by deriving optimality conditions.

Ancillary services

In relation to the relatively new practice of physical trading and bidding, the deregulation of the markets has called for more ancillary services in the power system, such as spinning reserves, non-spinning reserves, balancing power etc. which in turn has made new markets arise. However, the literature on ancillary service problems is limited.

Among the rather few contributions, the problem of Wen and David (2001b) is extended by Wen and David (2002) to account for the coordinated bidding into the day-ahead energy market and the spinning reserves market. A supplier submits hourly linear supply functions to the two markets separately. Still, from the perspective of a supplier, day-ahead and reserve markets exchanges are inter-dependent due to production capacity limits. A similar problem is addressed by Triki et al. (2005), who consider the simultaneous bidding into the spinning and non-spinning reserve markets, the day-ahead market, an inter-mediate adjustment market and an intra-day balancing market. A price-taking generating unit submits an offer to each of the markets while faced with uncertain prices. The multi-auction problem is therefore formulated as a multi-stage stochastic programming problem in which the different auctions constitute the stages. In most cases the problem is separable with respect to generating units and can be solved with commercial software.

With the restructuring, generation and transmission have been separated and transmission has become the responsibility of an independent system operator, who at the same time ensures security of supply. This involves the task of balancing demand and supply and that of procuring reserves to facilitate the balancing. The paper by Yu et al. (2005) addresses the problem of procuring and pricing reserves in the day-ahead market from the point of view of the power system operator. Hence, social welfare is optimized taking into account the costs of available reserve capacity in the day-ahead market, the costs of the resources actually used for matching demand and supply, the expenses of excess demand as well as the benefits of procuring reserves. Since demand is uncertain, the problem is a stochastic optimization problem. By defining the reliability level as the probability of shortage in reserve capacity, the paper provides a trade-off analysis between the capacity and reliability levels. The application in Chapter 6 includes two models that involve some of the same features in objectives and constraints. However, the stochastic programming models in this chapter seek to determine the optimal purchases of regulating reserves to be used in the real-time balancing of demand and supply. The two models incorporate pay-as-bid pricing and marginal pricing, respectively, since the intra-day balancing market of the particular application is currently undergoing a transition from one pricing scheme to another. The price clearing mechanisms are explicitly included in the linear mixed-integer models by the use of logical constraints and binaries. The two problems are solved with

procedures that utilize the structure of stochastic programming problems but are similar in spirit to those of common practice.

3.8 Risk

Financial risk management is often a high priority for participants in a deregulated electricity market due to the substantial price and volume risk caused by price volatility, demand peaks and supply shortages. To protect themselves from volatility, producers and consumers therefore enter into hedge contract with each other through the financial markets in which contracts have been developed to meet different needs, such contracts including forwards, futures, options and contracts of differences. The risk management deserves further attention in relation to the liberalized power markets that is the case in this chapter.

Chapter 4

Stochastic programming for optimizing bidding strategies of a Nordic hydro-power producer

As discussed in Chapter 3, a highly relevant power optimization problem that has arisen with the deregulation of the power markets in the last decade is the problem of bidding. The current chapter suggests a way of handling the problem using the framework of stochastic programming which appears to be very suitable for the purpose.

Considering a price-taking hydro-power producer participating in the day-ahead power market and for whom market prices are highly uncertain, the chapter provides a model for determining optimal bidding strategies taking this uncertainty into account. In particular, market price scenarios are generated and a stochastic mixed-integer linear programming model that involves both hydro-power production and physical trading aspects is developed. The idea is to explore the effects of including uncertainty explicitly into optimization by comparing the stochastic approach to a deterministic version. The model is illustrated with data from a Norwegian hydro-power producer and the Nordic power market at Nord Pool.

4.1 Introduction

The increased interest in power optimization problems within recent years has been stimulated by the tendency to decentralize and deregulate the power sector. Whereas traditional operating and planning procedures were based on centralized optimization, novel approaches rest on independent optimization of separate power plants. In addition, the liberalization of markets has forced former procedures to conform to more market oriented approaches.

With the pioneering act of 1990 Norway was among the first countries in the world to deregulate and already in 1991 a Norwegian power market was established. From 1996 to 2000 the national Norwegian power market developed into a multi-national Nordic power market that also encompassed the three neighboring countries, Sweden, Finland and Denmark. Today, the Nordic power market has successfully adapted to the new competitive environment and serves as a model for the restructuring of other power markets.

An important component of the Nordic power market is the presence of a power exchange that facilitates physical trading and is effective immediately. The spot market, Elspot, at the Nordic power exchange Nord Pool, takes the form of a pool-based market in which market participants exchange power contracts for physical delivery the following operation day and is referred to as the day-ahead market. In 2004 a total of 167 TWh was exchanged at Elspot, representing 42[1] percent of the overall consumption in the Nordic region.

Inevitably, physical trading is of vital importance in the economic activity of the power sector with the so-called bidding problem being a major challenge in this respect. Bidding involves the submission of sales and purchase bids to the power exchange a day ahead of physical trading. Since day-ahead market clearing prices are determined by the balance between sales and purchase bids, bidding takes place ahead of market clearing and, thus, with only limited information on day-ahead market prices. As a result, bidding is a rather complicated task. In the following, we will refer to the problem of submitting bids to the day-ahead market as the bidding problem.

In the Nordic region, power production comprises thermal power in addition to hydro power and other kinds of renewable power such as wind. Sweden and Finland use nuclear, fossil-fuel and hydro-power production, Denmark makes use of conventional thermal power plants, combined heat and power facilities and accounts for most of the wind power, while almost Norway almost only relies on hydro-power. In general, thermal power production is located in the south, whereas hydro-power production is found in the northern parts of the Nordic region and is transmitted to the heavily populated south. Approving of the comparatively low costs of hydro-power production, the market prefers this energy source over thermal power. The number of hydro-power producers participating in Nord Pool amounts to around 48 and the total hydro-power based generation in the Nordic region is 191 TWh or 49% of the total generation in 2004[2].

A crucial difference between thermal power and hydro-power production is the possibility of "storing" energy by holding water back. If production is disposed of in the power market, this possibility allows the producer to respond to the

[1] Reference: `www.nordel.org` and `www.nordpool.no`

[2] Reference: `www.nordel.org`

development in market prices. When prices are high, water is released and energy is produced and sold immediately, whereas when prices are low, the water is held back and the energy is saved for future disposal at higher prices. Moreover, in contrast to thermal production, hydro-power generation has the ability to quickly start and stop and is therefore typically applied to peak-load demand. The flexibility makes bidding strategies extremely relevant to hydro-power producers.

To address the bidding problem, we will present a model of a price-taking hydro-power producer bidding into the day-ahead market under market price uncertainty. The model is motivated by the needs of small Nordic hydro-power producers although its major parts are applicable in general to price-taking power producers acting in a pool-based day-ahead market.

As already implied, the bidding problem of a hydro-power producer involves both hydro-power generation and day-ahead market exchange. Bidding decisions take place with only limited information on day-ahead market prices, whereas production decisions may be deferred until the information has been fully disclosed. In effect, the decision process is divided into stages and, more specifically, the bidding problem fits the framework of two-stage stochastic programming. The objective is to maximize sales and production profits subject to a number of bidding and operational constraints. The first stage decisions concern day-ahead market exchanges while the second stage includes real-time hydro-power generation. Uncertainty of the day-ahead market prices is represented by a known probability distribution. Formally speaking, the probability distribution is obtained by the modeling of a stochastic process calibrated from historical data.

In practice power planning and operation is often based on deterministic optimization tools, such as the Short-term Hydro Operation Planning (SHOP) model, cf. Fosso et al. (1999), used in Norway. Moreover, the setting up of bidding tables rests on skills and experience of the operating engineers. Until now, uncertainty has not been handled explicitly neither when planning and operating nor when bidding. Current practice is limited to the selection of appropriate critical quantiles of the uncertain data and the following solution of a number of deterministic problems in order for the planner to identify decisions that hedge against adversity at a few probability levels.

The chapter aims at comparing the stochastic approach to a deterministic version. The idea is to explore the effects of including uncertainty explicitly into the optimization model and, in particular, to examine its objective function value and solutions. It has already been shown, cf. Birge and Louveaux (1997), that higher expected profits are obtained by the stochastic approach than by the deterministic version. The difference in profits is however highly relevant. Moreover, we expect the bidding structure of the stochastic approach to differ from that of the deterministic version. Most likely, the deterministic model suggests the use of hourly bids only, whereas the stochastic model combines hourly bids and block bids.

The outline of the chapter is as follows. We explain the composition of the

day-ahead market and model the bidding process in Section 4.2. In Section 4.3 we describe the hydro-power plant and model production. Section 4.4 introduces uncertainty and the resulting stochastic programming model, whereas Section 4.5 is devoted to scenario generation. Section 4.6 illustrates the model with a case study from a Norwegian hydro-power producer and the Nordic power exchange, Nord Pool.

4.2 Day-ahead bidding

Elspot at Nord Pool is a spot market in which contracts for physical delivery the following operation day are exchanged. The power exchange offers an access to the physical market at low transaction costs as well as a possibility of settlement close to real-time operation. The Elspot contracts are power obligations to deliver or receive power of a duration of one hour or longer. Contracts are divided into hourly bids, block bids and flexible hourly bids and all bids consist of a price and a volume. When submitting hourly bids, the procedure is the same for all hours. Sales bids have to be listed in ascending order and purchase bids in descending order according to price. Consistent with the rules, Elspot will make a linear interpolation between the price-volume points to construct the bidding curve. The volume dispatched is determined by the point on the bidding curve that corresponds to the market price and all transactions are settled at market price. Block bids are aggregated bids valid for a number of consecutive hours and associated with only one price and volume. The so-called mean price condition determines whether a block bid is either rejected or accepted as a whole. If the price of a sales bid is less than or equal to the average market price or if the price of a purchase bid is greater than or equal to the average market price of the hours of the block, the bid is accepted. All transactions are settled at the mean price. Flexible hourly bids are basically hourly bids that are accepted in the hour with the highest price, provided this price exceeds a certain threshold. Here, such bids are omitted, since the bids are mainly used by companies able to close down power intensive production. Participants post the price-differentiated bids for all hours of the following operation day before deadline at noon.

It is important to note that the volumes dispatched and the prices at which transactions are settled are unknown until the market has cleared and market clearing prices have been determined. Once this is done, the participants receives a notification of the volumes dispatched. The market price calculations are the same for each individual hour. The bidding curves that are sales or purchase curves, are collocated to an aggregated demand curve and an aggregated supply curve, respectively. The intersection of the demand curve and the supply curve defines a candidate of the unconstrained market price. Through an iterative process, the bidding curves are updated according to certain priority rules that

include accepted block bids and accepted flexible hourly bids until a new unconstrained market price is found. The Nordic grid is however divided into fixed price zones. Sweden, Finland, East and West Denmark are each one zone and Norway is divided into four zones. If the contractual flow between zones does not exceed the allocated grid capacity, the unconstrained market price simply applies to all zones. Otherwise, separate prices are established through counter purchases and corresponding iterations of price calculations in order to relieve grid congestion. We assume a single price applies, which is justified in the case of local physical trading or more generallly by assuming no grid congestion.

The following model should work as a tool of a price-taking hydro-power producer bidding into the day-ahead market and producing in accordance. We assume w.l.g. that the hydro-power producer does not participate in bilateral exchange, but instead disposes of the entire production in the day-ahead and real-time markets. We further assume that the producer only participate in the day-ahead market on the supply side and not on the demand side.

The time horizon of 24 hours is divided into hourly time intervals and is denoted $\mathcal{T} = \{1, \ldots, T\}$ with $T = 24$. From this the set of blocks $\mathcal{A} = \{a_1, \ldots, a_A\}$ is constructed. A block is a number of minimum two consecutive hours and the total number of such blocks within 24 hours is therefore $A = 276$. Examples of blocks are $a_{140} = \{1, \ldots, 7\}$, $a_{165} = \{8, \ldots, 18\}$ and $a_{265} = \{19, \ldots, 24\}$.

Regarding the modeling of the bidding process, the problem of selecting both bid prices and bid volumes is nonlinear. However, we have chosen to work with a linear formulation. The reason is that the problem, even formulated as a mixed-integer linear program, is relatively hard to solve in terms of computing times. Moreover, the linear formulation is amenable to standard mathematical programming software packages and specially designed stochastic programming algorithms. Nonlinearities are avoided by fixing prices in advance such that only volumes have to be selected. Let $\mathcal{H} = \{1, \ldots, H\}$ index the possible bid prices and denote these prices $p_h, h \in \mathcal{H}$, where $p_h \leq p_{h+1}$. The corresponding bid volumes are represented by the variables $x_{ht} \in \mathbb{R}_+, h \in \mathcal{H}, t \in \mathcal{T}$ for hourly bids and $x_{ha} \in \mathbb{R}_+, h \in \mathcal{H}, a \in \mathcal{A}$ for block bids. The variables $y_t \in \mathbb{R}_+, t \in \mathcal{T}$ and $y_a \in \mathbb{R}_+, a \in \mathcal{A}$ are the volumes dispatched, for hourly bids and block bids respectively. The hourly market prices are denoted $\rho_t, t \in \mathcal{T}$ and average market prices for the blocks $\bar{\rho}_a, a \in \mathcal{A}$, where $\bar{\rho}_a = (1/|a|) \sum_{t \in a} \rho_t$.

Disposing of hydro-power production in the day-ahead market, total sales revenues accumulate to

$$\sum_{t \in \mathcal{T}} \rho_t y_t + \sum_{a \in \mathcal{A}} \bar{\rho}_a y_a.$$

Hourly bids are handled in a spirit similar to that of Fleten and Pettersen (2005). The bidding curve is assumed to cover the entire plant, although bidding on a single-reservoir basis causes no further difficulty in modeling. For each hour,

$t \in \mathcal{T}$, the bids $(x_{ht}, p_h), h \in \mathcal{H}$ are interpreted as price-volume points on a bidding curve that determines the relation between volumes bid and volumes dispatched. The curve is constructed by making a linear interpolation between the points, the result being a piecewise linear curve. Thus, in terms of prices, the bidding curve can be expressed as

$$\rho_t = \begin{cases} p_1 + \frac{p_2 - p_1}{x_{2t} - x_{1t}}(y_t - x_{1t}) & \text{, if } x_{1t} \leq y_t < x_{2t} \\ \vdots & \\ p_{h-1} + \frac{p_h - p_{h-1}}{x_{ht} - x_{h-1t}}(y_t - x_{h-1t}) & \text{, if } x_{h-1t} \leq y_t < x_{ht} \\ \vdots & \\ p_{H-1} + \frac{p_H - p_{H-1}}{x_{Ht} - x_{H-1t}}(y_t - x_{H-1t}) & \text{, if } x_{H-1t} \leq y_t \leq x_{Ht} \end{cases}$$

or equivalently, in terms of volumes,

$$y_t = \begin{cases} \frac{\rho_t - p_1}{p_2 - p_1} x_{2t} + \frac{p_2 - \rho_t}{p_2 - p_1} x_{1t} & \text{, if } p_1 \leq \rho_t < p_2 \\ \vdots & \\ \frac{\rho_t - p_{h-1}}{p_h - p_{h-1}} x_{ht} + \frac{p_h - \rho_t}{p_h - p_{h-1}} x_{h-1t} & \text{, if } p_{h-1} \leq \rho_t < p_h \\ \vdots & \\ \frac{\rho_t - p_{H-1}}{p_H - p_{H-1}} x_{Ht} + \frac{p_H - \rho_t}{p_H - p_{H-1}} x_{H-1t} & \text{, if } p_{H-1} \leq \rho_t \leq p_H. \end{cases} \tag{4.2.1}$$

For an illustration, see Fig. 4.1. In the case of sales bids, the bidding curves would naturally be non-decreasing although the market rules does not explicitly dictate this. Monotonicity constraints may therefore in principle be omitted. In practice, decreasing bidding curves are rare and we include the constraints

$$x_{ht} \leq x_{h+1t}, \;\; h \in \mathcal{H} \backslash \{H\}, t \in \mathcal{T}. \tag{4.2.2}$$

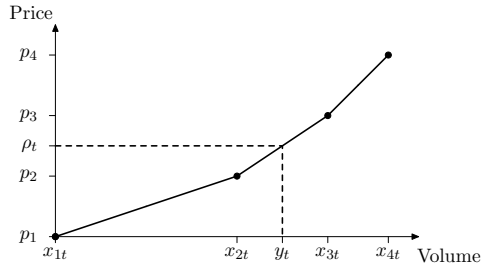


Figure 4.1: Bidding curve of a given time interval $t \in \mathcal{T}$.

For each $b \in \mathcal{A}$, the block bids are $(x_{ha}, p_h), h \in \mathcal{H}$ and the relation between volumes bid and volumes dispatched is

$$y_a = \sum_{j:p_j \leq \bar{\rho}_a} x_{ja}, \;\; a \in \mathcal{A} \tag{4.2.3}$$

i.e. in a given block, the volume dispatched comprises the volumes of accepted bids. For example, consider the block $a_{140} = \{1, \ldots, 7\}$. If two bids are given by $(x_{1,140}, p_1) = (50, 100)$ and $(x_{2,140}, p_2) = (100, 200)$ and the mean price is $\bar{\rho}_{140} = 175$, then only one bid is accepted and the volume dispatched is $y_{140} = 50$.

4.3 Daily hydro-power production

Modeling the hydro-power production side more or less follows the lines of for example Philpott et al. (2000). The section presents a simple but illustrative model of a small Norwegian hydro-power plant. Still, it is relatively straightforward to combine a different modeling of the hydro-power plant with the modeling of the bidding process in Section 4.2. The plant consists of two reservoirs in cascade; a larger upper reservoir and a smaller lower reservoir. There is a time delay between the two reservoirs. The combination of time delay and size differences restricts the flexibility of the system, which contributes to understanding the importance of including uncertainty and will be discussed later. Each reservoir is connected to a power station that contains a single generator. Hydro-power production works as follows. Upstream water reaching the plant flows to the upper reservoir where it is stored until released for generation. When released, the water from the upper reservoir flows to the lower reservoir and is again stored until used for generation. Electricity is generated by converting the potential energy of the water into electrical energy. Water that is not discharged on purpose and used for generation is considered spill. Leaving the plant, the water proceeds downstream. For an illustration, see Fig. 4.2.

To model the production side, let $\mathcal{J} = \{1, 2\}$ index the reservoirs, let the variables $u_{jt} \in \{0, 1\}, j \in \mathcal{J}, t \in \mathcal{T}$ represent the on/off states of the generators, $w_{jt} \in \mathbb{R}_+, j \in \mathcal{J}, t \in \mathcal{T}$ the generation levels and $v_{jt} \in \mathbb{R}_+, j \in \mathcal{J}, t \in \mathcal{T}$ the corresponding discharges from the reservoirs. Moreover, let the variables $l_{jt} \in \mathbb{R}_+, j \in \mathcal{J}, t \in \mathcal{T}$ be the reservoir storage levels and $r_{jt} \in \mathbb{R}_+, j \in \mathcal{J}, t \in \mathcal{T}$ the spill.

Direct costs of hydro-power production include only start-up costs since operating costs are negligible. Start-up costs amount to

$$\sum_{t \in \mathcal{T}} \sum_{j \in \mathcal{J}} SC_j(u_{jt-1}, u_{jt}),$$

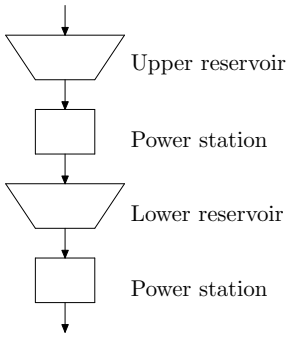


Figure 4.2: Two hydro-power reservoirs in cascade.

with the cost functions being

$$SC_j(u_{jt-1}, u_{jt}) = c_j \max\{u_{jt} - u_{jt-1}, 0\}, \;\; j \in \mathcal{J}, t \in \mathcal{T},$$

where $c_j, j \in \mathcal{J}$ are the unit start-up costs. Note that the cost functions are consistent with a mixed-integer linear formulation. The initial conditions $u_{j0} = u_{j,\text{init}}, j \in \mathcal{J}$ must be added.

Costs also include opportunity costs of releasing water as the water could be stored and saved for future disposal. Such costs are measured as the value of stored water, which is usually available from more long-term models. As estimates of the water values we have used an average of prices of futures contracts and prices on forward contracts.

For reservoirs in cascade the potential for future disposals of stored water depends on the progress of water releases downstream. Since the reservoirs are serially connected, the water value of the upper reservoir accounts for the opportunity of releasing water from both the upper and the lower reservoirs. However, the possibility of releasing water from the upper reservoir depends on the storage level of the lower reservoir.

We begin with the lower reservoir. In computing the value of stored water, we first determine the marginal water value. As a starting point, we have

$$\partial V_2(l_{2t})/\partial l_{2t} = \begin{cases} 0.5FO + 0.5FU & \text{, if } \; l_{2t} = l_2^{min} \\ 0 & \text{, if } \; l_{2t} = l_2^{max}, \end{cases}$$

where FO, FU are the prices on forwards and futures and ∂/∂ denotes the partial derivative. The reasoning is that in case of an empty reservoir the storage contents can be disposed of at any time in the future, whereas in the case of a full reservoir additional storage is spilled. We make a linear interpolation between the end points

so that the marginal water value is given by the linear non-decreasing function

$$\partial V_2(l_{2t})/\partial l_{2t} = (0.5FO + 0.5FU)(l_{2t} - l_2^{max})/(l_2^{min} - l_2^{max})$$

and the water value function can be obtained by integration

$$\int_0^{l_{2t}} (\partial V_2(l'_{2t})/\partial l'_{2t}) dl'_{2t}.$$

To account for the fact that the water value of the upper reservoir depends on the storage level of the lower reservoir, Terry et al. (1986) assume certain operating rules as for instance the upper reservoir must be completely empty before releasing water from the lower or the lower reservoir is always emptied before the upper. We however employ a slightly different approach and determine the water value of the upper reservoir on the assumption that the lower is half full when the water of the upper reservoir reaches the lower. The result is then

$$\begin{aligned}\partial V_1(l_{1t})/\partial l_{1t} = (0.5FO + 0.5FU)\big((l_{1t} - l_1^{max})/(l_1^{min} - l_1^{max}) + \\ (l_{1t} + 0.5l_{2t} - l_2^{max})/(l_2^{min} - l_2^{max})\big), \text{ if } l_{1t} + 0.5l_2 \le l_2^{max}\end{aligned}$$

and

$$\begin{aligned}\partial V_1(l_{1t})/\partial l_{1t} = (0.5FO + 0.5FU)(l_{1t} - l_1^{max})/(l_1^{min} - l_1^{max}), \\ \text{if } l_{1t} + 0.5l_2 \ge l_2^{max}.\end{aligned}$$

To avoid dependency of $\partial V_1(l_{1t})/\partial l_{1t}$ on l_{2t}, we replace it by $l_{2,init}$. Although slightly more complex, the water value function is again obtained by integration.

The water value functions of the upper and lower reservoirs of the case study are displayed in Fig. 4.3. In general, the opportunity costs are given by

$$\sum_{j \in \mathcal{J}} (V_j(l_{j0}) - V_j(l_{jT}))$$

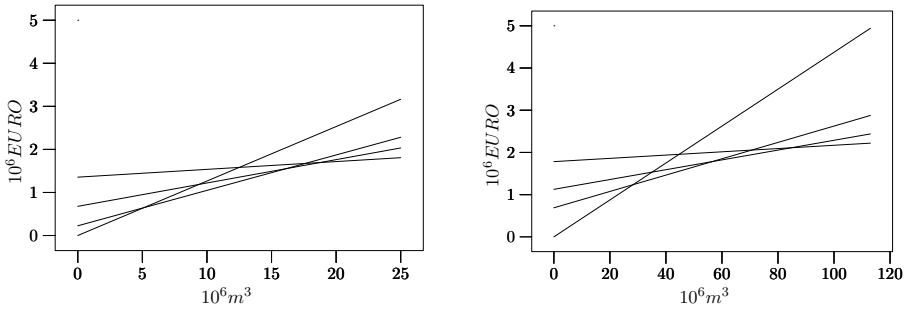
and the concave water value functions are approximated by piecewise linear functions to be consistent with a mixed-integer linear formulation

$$V_j(l_{jt}) = \min_{k \in \mathcal{K}} \{d_{kj}^1 l_{jt} + d_{kj}^2\}, \;\; j \in \mathcal{J}, t \in \mathcal{T},$$

where $d_{kj}^1, d_{kj}^2, k \in \mathcal{K}, j \in \mathcal{J}$ are the coefficients.

The following bounds are imposed on the water discharges. The upper reservoir is either not in operation or operated at maximum capacity, which leads to the constraints

$$w_{1t} = u_{1t} w_1^{max}, \;\; t \in \mathcal{T},$$



(a) Upper reservoir. (b) Lower reservoir.

Figure 4.3: Water value functions for two hydro-power reservoirs in cascade.

where w_1^{max} is the maximum generation level. The lower reservoir, however, can be operated anywhere between its minimum and maximum capacity so that

$$u_{2t} w_2^{min} \leq w_{2t} \leq u_{2t} w_2^{max}, \;\; t \in \mathcal{T},$$

where w_2^{min} and w_2^{max} are the minimum and maximum generation levels. Similar bounds apply to the discharges, i.e.

$$v_j^{min} \leq v_{jt} \leq v_j^{max}, \;\; j \in \mathcal{J}, t \in \mathcal{T}.$$

Here, $v_j^{min}, j \in \mathcal{J}$ and $v_j^{max}, j \in \mathcal{J}$ are the minimum and maximum discharges. Finally, the storage levels have to adhere to the bounds

$$l_j^{min} \leq l_{jt} \leq l_j^{max}, \;\; j \in \mathcal{J}, t \in \mathcal{T},$$

where $l_j^{min}, j \in \mathcal{J}$ and $l_j^{max}, j \in \mathcal{J}$ denote the minimal and maximal storage levels.

As a reservoir has to balance, storage from the previous period and reservoir inflow either appear as storage, discharge or spill. In the case of the upper reservoir the balance equations are

$$l_{1t} - l_{1t-1} + v_{1t} + r_{1t} = \nu_{1t}, \;\; t \in \mathcal{T},$$

in which $\nu_{1t}, t \in \mathcal{T}$ are the inflows from upstream. The initial storage level is $l_{10} = l_{1,\text{init}}$. In the case of the lower reservoir, the balance equations are

$$l_{2t} - l_{2t-1} + v_{2t} + r_{2t} = v_{1t-\tau}, \;\; t \in \mathcal{T},$$

where τ is the time delay between the upper and lower reservoirs. Note that upstream discharges appear as downstream inflows. However, we have assumed

that upstream spill is lost and does not appear downstream. Again, the initial storage level is $l_{20} = l_{2,\text{init}}$.

Generation and discharge are essentially proportional if ignoring that generation efficiency is in fact non-constant and omitting effects of the reservoir storage levels on both generation and discharge. This leads to the constraints

$$w_{jt} = \eta_j v_{jt}, \;\; j \in \mathcal{J}, t \in \mathcal{T},$$

in which $\eta_j, j \in \mathcal{J}$ are the generation efficiency coefficients. For the case study, the assumption of proportional generation and discharge is justified by Fig. 5.3 of Chapter 5.

Imbalances between volumes produced and volumes dispatched in the day-ahead market are settled in an intra-day balancing market. As the day-ahead market should work as a de facto spot market planned imbalances are not allowed. Hence, a hydro-power producer cannot save water and postpone production for disposal in the intra-day balancing market since then the day-ahead market would no longer reflect the physical conditions of the system. The primary focus in the short-term planning of a hydro-power plant is therefore by all means the day-ahead market. Justified by the discussion, we approximate the balancing effects of the intra-day market. We impose a penalty or a reward on imbalances. The penalty is higher than the day-ahead market price and paid if volumes dispatched exceed the volumes produced, i.e. in hours of up-regulation, and the reward is lower than the day-ahead market price and is paid if volumes produced exceed the volumes dispatched, i.e. in hours of down-regulation. By imposing a penalty or a reward, the producer retains flexibility to ramp up or down and to bid this flexible capacity into the balancing market close to real-time. This way of modeling does not prevent the producer from participating in the balancing market, but is alone an attempt to avoid planned imbalances. As penalty and reward, we take the average intra-day market price of hours in which the market has been up-regulated and down-regulated, respectively. Such penalties and rewards would apply in Norway if the producer is always regulated in the same direction as the market. In reality, a producer is very rarely regulated in the opposite direction of the local market. For a different modeling approach, see also Fleten and Pettersen (2005). Let the variables $y_t^{up}, y_t^{do} \in \mathbb{R}_+, t \in \mathcal{T}$ represent the imbalances and let $p_t^{up}, p_t^{do}, t \in \mathcal{T}$ denote the corresponding penalty and reward. Then the total penalty and reward amount to

$$\sum_{t \in \mathcal{T}} (p_t^{up} y_t^{up} - p_t^{do} y_t^{do})$$

and the balance constraints are

$$y_t + \sum_{a \in \mathcal{A}: t \in a} y_a - \sum_{j \in \mathcal{J}} w_{jt} = y_t^{up} - y_t^{do}, \;\; t \in \mathcal{T}.$$

It should be remarked that most problems from the literature on bidding into the day-ahead market simply ignore the trading in a intra-day balancing market and thereby the consequences of a mismatch between production and dispatch.

4.4 Day-ahead bidding under uncertainty

The model of the preceding sections does not take into considerations the uncertainty of the data, which can arise with respect to reservoir inflows and market prices. As the idea is to analyze uncertainty that relates directly to bidding, we ignore the possibility of inflows to be stochastic and confine ourselves to price stochasticity. For a price-taker, the market clearing process is governed by the behavior of other participants and, in that the day-ahead market is organized as a sealed auction, such market conditions are undisclosed. However, market prices are determined by clearing the market, which makes prices unknown at the time of bidding.

Uncertainty can be handled by means of stochastic programming. As decisions are to be made before and after observing uncertainty, respectively, a two-stage recourse model is appropriate. The first stage involves bidding and the second stage concerns production aspects. Since bids are submitted before the market has cleared, prices are unknown at the time of first-stage decision-making. In contrast, second-stage decision-making is put off until the market has cleared and take advantage of the additional information from observing prices. The overall aim is to obtain optimal bidding strategies in terms of expected sales and production profits.

To incorporate uncertainty, we assume market prices $\{\rho_t\}_{t\in\mathcal{T}}$ form a stochastic process on some probability space. We further assume that the multivariate distribution is known and in particular that it is discrete with a finite number of realizations $\mathcal{S} = \{1, \ldots, S\}$ referred to as scenarios. The scenario probabilities are denoted by $\pi^s, s \in \mathcal{S}$ and the corresponding market prices by $\{\rho_t^s\}_{t\in\mathcal{T}s\in\mathcal{S}}$. First-stage decisions $x_{ht}, x_{ha}, h \in \mathcal{H}, t \in \mathcal{T}, a \in \mathcal{A}$ are the volumes bid and should be independent of future market prices. Second-stage decisions $y_t^s, y_a^s, q_t^{up,s}, q_t^{do,s}, v_{jt}^s, w_{jt}^s, l_{jt}^s$, $r_{jt}^s \in \mathbb{R}_+, u_{jt}^s \in \{0,1\}, s \in \mathcal{S}$ are the volumes dispatched and the production decisions. These are allowed to depend on the realization of future market prices, reflected in the scenario superscript. The stochastic program consists in maximizing the expected sales and production profits subject to the bidding and operational constraints. The constraints (4.2.1)-(4.2.3) couple first-stage and second-stage decisions through the relation between volumes bid and volumes dispatched, whereas the constraints (4.4.3)-(4.4.10) apply to second-stage decisions only and model hydro-power production. The two-stage stochastic mixed-integer linear program stated as its deterministic equivalent is then

$$\max \sum_{s\in\mathcal{S}} \pi^s \Big(\sum_{t\in\mathcal{T}} \rho_t^s y_t^s + \sum_{a\in\mathcal{A}} \bar{\rho}_a^s y_a^s - \sum_{t\in\mathcal{T}} (p_t^{up} q_t^{up,s} - p_t^{do} q_t^{do,s}) - \sum_{t\in\mathcal{T}}\sum_{j\in\mathcal{J}} SC_j(u_{jt-1}^s, u_{jt}^s) - \sum_{j\in\mathcal{J}} (V_j(l_{j0}^s) - V_j(l_{jT}^s))\Big) \tag{4.4.1}$$

$$\text{s.t. } (4.2.1) - (4.2.3) \tag{4.4.2}$$

$$w_{1t}^s = u_{1t}^s w_1^{max}, \ \ t \in \mathcal{T}, s \in \mathcal{S} \tag{4.4.3}$$

$$u_{2t}^s w_2^{min} \le w_{2t}^s \le u_{2t}^s w_2^{max}, \ \ t \in \mathcal{T}, s \in \mathcal{S} \tag{4.4.4}$$

$$v_j^{min} \le v_{jt}^s \le v_j^{max}, \ \ j \in \mathcal{J}, t \in \mathcal{T}, s \in \mathcal{S} \tag{4.4.5}$$

$$l_j^{min} \le l_{jt}^s \le l_j^{max}, \ \ j \in \mathcal{J}, t \in \mathcal{T}, s \in \mathcal{S} \tag{4.4.6}$$

$$l_{1t}^s - l_{1t-1}^s + v_{1t}^s + r_{1t}^s = \nu_{1t}^s, \ \ t \in \mathcal{T}, s \in \mathcal{S} \tag{4.4.7}$$

$$l_{2t}^s - l_{2t-1}^s + v_{2t}^s + r_{2t}^s = v_{1t-\tau}^s, \ \ t \in \mathcal{T}, s \in \mathcal{S} \tag{4.4.8}$$

$$w_{jt}^s = \eta_j v_{jt}^s, \ \ j \in \mathcal{J}, t \in \mathcal{T}, s \in \mathcal{S} \tag{4.4.9}$$

$$y_t^s + \sum_{a\in\mathcal{A}: t\in a} y_a^s - \sum_{j\in\mathcal{J}} w_{jt}^s = q_t^{up,s} - q_t^{do,s}, \ \ t \in \mathcal{T}, s \in \mathcal{S} \tag{4.4.10}$$

$$x_{ht}, x_{ha} \in \mathbb{R}_+, h \in \mathcal{H}, t \in \mathcal{T}, a \in \mathcal{A} \tag{4.4.11}$$

$$y_t^s, y_a^s, q_t^{up,s}, q_t^{do,s} \in \mathbb{R}_+, t \in \mathcal{T}, a \in \mathcal{A}, s \in \mathcal{S} \tag{4.4.12}$$

$$v_{jt}^s, w_{jt}^s, l_{jt}^s, r_{jt}^s \in \mathbb{R}_+, u_{jt}^s \in \{0,1\}, j \in \mathcal{J}, t \in \mathcal{T}, s \in \mathcal{S}. \tag{4.4.13}$$

It should be remarked that (4.2.1) and (4.2.3) can be simplified. Consider a fixed $t \in \mathcal{T}$ and $s \in \mathcal{S}$. If the price points $p_h, h \in \mathcal{H}$ are fixed in advance, the realized market price ρ_t^s is located between two adjacent points. The remaining price points are irrelevant for determining the volume to be dispatched. Letting $h(t,s) = \max\{h \in \mathcal{H} : p_h \le \rho_t^s\}$, the point of dispatch (y_t^s, ρ_t^s) is located on the line segment between $(x_{h(t,s)t}, p_{h(t,s)})$ and $(x_{h(t,s)+1t}, p_{h(t,s)+1})$. From this, (4.2.1) is equivalent to

$$y_t^s = \frac{\rho_t^s - p_{h(t,s)}}{p_{h(t,s)+1} - p_{h(t,s)}} x_{h(t,s)+1t} + \frac{p_{h(t,s)+1} - \rho_t^s}{p_{h(t,s)+1} - p_{h(t,s)}} x_{h(t,s)t}, \ \ t \in \mathcal{T}, s \in \mathcal{S}.$$

Similarly, letting $h(a,s) = \max\{h \in \mathcal{H} : p_h \leq \bar{\rho}_a^s\}$, (4.2.3) can be rewritten as

$$y_a^s = x_{h(a,s)a}, \;\; a \in \mathcal{A}, s \in \mathcal{S},$$

the reason being that for a fixed $a \in \mathcal{A}$ and $s \in \mathcal{S}$, the same dispatch is obtained by bidding just below the realized market price.

Remark 4.4.1 *We operate with two different ways of fixing the price points $p_h, h \in \mathcal{H}$ in advance, one way being to fix equidistant price points and the other being to fix price points such that the number of realized market prices $\rho_t^s, t \in \mathcal{T}, s \in \mathcal{S}$ between any two points is always the same. The latter reflects the distribution of market prices and consequently the price graduation is more crude in areas where market prices are less likely. We refer to the price points as distributional. In the last case we examined the effect on the optimal bidding curves of varying the number of price points. In general, the more crude the price graduation, the more crude the bidding curves, although small changes in the number of price points did not alter the curves. Also, we found that equidistant price points may generally induce rather fine bidding curves. However, within the interval of realized of market prices, the bidding curves are in fact cruder than those derived from the distributional price points.*

4.5 Scenario generation

To describe the behavior of day-ahead market prices, the ARMA methodology may be applied. Basically, ARMA processes comprise a specific class of stochastic processes adopted for the analysis of time series. Here, ARMA processes provide a statistical model from which price scenarios can be generated by sampling. A similar approach is applied by Gröwe-Kuska et al. (2000) and Gröwe-Kuska et al. (2002), who generate electricity demand scenarios and by Eichhorn et al. (2005) who consider electricity demand, heat demand and day-ahead market prices. For further details on the ARMA methodology and the application to electricity prices, see Chapter 8.

We briefly sketch the application of the ARMA methodology. An ARMA model can be formulated as

$$\psi(B)\rho_t = \gamma(B)e_t, \;\; t \in \mathbb{Z},$$

where $\psi(B)$ and $\gamma(B)$ are polynomials of the form $\psi(B) = 1 - \sum_k \psi_k B^k$ and $\gamma(B) = 1 - \sum_k \gamma_k B^k$ and B denotes the back-shift operator, i.e. $B^k\rho_t = \rho_{t-k}$. The innovations $\{e_t\}_{t\in\mathbb{Z}}$ are assumed a Gaussian white noise process, i.e. $e_t, t \in \mathbb{Z}$ are independent normally distributed random variables with zero mean and constant variance. The development of the ARMA model follows the steps

1. Identify a statistical model of the historical data.
2. Estimate the parameters of the model.
3. Validate the model.

1. An hourly price profile of a year is given. In the creation of a trial model the data is made stationary by a transformation of the original data. The logarithm is applied to attain a stable variance and the inclusion of the factors $(1-B)$, $(1-B^{24})$ and $(1-B^{168})$ is used to stabilize the mean. The structure of the polynomials is determined by considering seasonalities and investigating the autocorrelations and partial autocorrelations of the transformed data. When refined, the final so-called seasonal integrated ARMA (SARIMA) model is identified as

$$\begin{aligned}(1-B)(1-B^{24})(1-B^{168})\log\rho_t = \\ (1-\gamma_1 B^1-\cdots-\gamma_7 B^7)(1-\gamma_{23}B^{23}-\cdots-\gamma_{25}B^{25}- \\ \gamma_{47}B^{47}-\cdots-\gamma_{49}B^{49})(1-\gamma_{168}B^{168})e_t,\ \ t\in\mathbb{Z}.\end{aligned}$$

2. Having completed the identification, parameter estimates may be computed by the use of maximum likelihood optimization.

3. The model is validated by testing the assumptions of a Gaussian white noise process made on the innovations.

Price scenarios $\{\rho_t^s\}_{t=1}^T, s=1,\dots,S$ can be generated by drawing the starting values $\rho_t, t=-192,\dots,0$ and $e_t, t=-223,\dots,0$ from the historical data and sampling from the independent identically and normally distributed random variables $e_t, t\in\mathbb{Z}$. Monte Carlo sampling has been used to generate a large number of scenarios. An example of some sample scenarios is shown in Fig. 4.4a.

Steps 1-3 are all carried out by the statistical software package SAS, version 8.2, see SAS Institute Inc. (1999). The Monte Carlo sampling has been implemented in C++.

Due to computational limitations, the number of scenarios has been reduced using the scenario reduction approach of Heitsch and Römisch (2006a). The approach rests on the selection and clustering of scenarios that are close with respect to a certain distance. Here, we have used the euclidean distance. For the basic concepts, see Chapter 7. We have implemented the scenario reduction approach in C++.

Remark 4.5.1 *Although in practice the distribution of electricity prices is continuous, it is approximated by a discrete distribution with finite support by sampling a number of scenarios. The quality of the approximation is directly linked to the quality of the scenarios and it is therefore relevant to evaluate the scenario generation method. For practical performance, cf. Chapter 7, the in-sample and the*

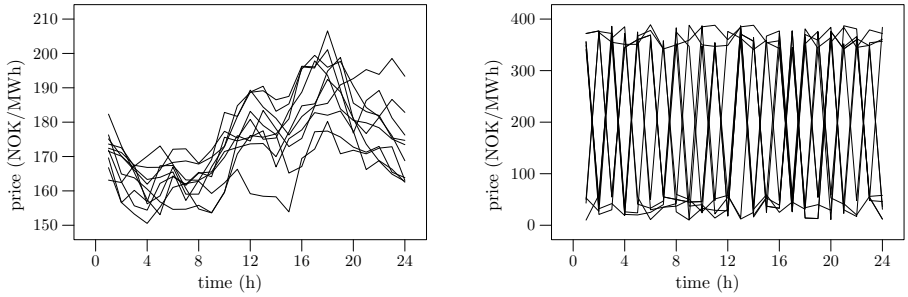
out-of-sample stability was tested. Where the true distribution was needed, so-called back-testing was done using historical data. The values of the mean, the standard deviation and the standard deviation as a percent of the mean are reported in Table 4.1 in case of both in-sample and out-of-sample stability. The

Table 4.1: In-sample stability and out-of sample stability for the problem (4.4.1)-(4.4.13).

	Sce.	Mean	Std. dev.
In-sample	10	33429.62	74.58 (0.22%)
Out-of-sample	10	34716.69	2.02 (0.01%)

stability analysis gives an indication of the number of scenarios to include in order to represent uncertainty. As is clear from the table, the stochastic program was found to fulfill the stability requirements in a satisfying way for as few as 10 scenarios.

For illustration purposes a number of demonstration scenarios have been generated in which the day-ahead market prices are very volatile. An example of some demonstration scenarios is plotted in Fig. 4.4b.



(a) Sample. (b) Demonstration.

Figure 4.4: Day-ahead market price scenarios generated randomly and by sampling from a statistical model.

4.6 Case study

The case study concerns two reservoirs of a small Norwegian hydro-power plant located near Trondheim and run by the company TrønderEnergi. The real dimensions of the reservoirs and the capacities of the power stations are scaled down by

a factor ten for computational reasons. Initial reservoir levels and reservoir inflows are the real data from a typical day of 2005. To generate day-ahead market prices in 2005, Elspot at Nord Pool has provided the real data from 2004 that applied to the Norwegian prize zone NO2, which is the Trondheim area.

The two-stage stochastic programming problem contains 18300 continuous variables and 1440 constraints in the first stage and 590 continuous variables, 48 binary variables and 740 constraints in the second stage. Stated as its deterministic equivalent, (4.4.1)-(4.4.13), the problem is a large-scale mixed-integer linear program solvable by standard software. The problem was solved with the mixed-integer linear programming solver from OPL Studio version 3.7, ILOG Inc. (2003), calling CPLEX 9.0, Cplex Optimization Inc. (2006), on an Intel Xeon 2.67 GHz processor with 4 GB RAM. The time of compiling and solving the problem varied between 1 and 3000 seconds since no special effort was made to make the code efficient. As the problem is a two-stage stochastic mixed-integer linear program special-purpose solution methods such as the dual decomposition algorithm of Carøe and Schultz (1999) might prove useful.

The rest of this section compares the stochastic approach to the bidding problem with a deterministic version. The idea is to explore the effects of including uncertainty explicitly into the optimization model and, in particular, to examine its objective function value and solutions.

To consider the structure of the bids, we solved the bidding problem as a stochastic problem and as a deterministic problem. The former is the two-stage recourse problem and the latter the corresponding expected value problem formed by replacing random prices by expected prices. Moreover, we computed the expected result of using the expected value solution. For more information on the expected value problem and the expected result of using the expected value solution, see also Birge and Louveaux (1997).

Some results are reported in Table 4.2. EVP denotes the expected value problem and RP the recourse problem. H. disp. and B. disp. represent the volume dispatched of hourly bids and block bids, respectively, and Prod. refers to the volumes produced. Finally, Imbal. denotes the imbalances between total volumes dispatched and produced. The table illustrates the difference in the structure of the bids between the expected value problem and the recourse problem. In overall terms, the deterministic model suggests the use of hourly bids, whereas the stochastic model combines hourly bids and block bids.

For details, consider the deterministic bidding problems for one reservoir and for two reservoirs in cascade. In both cases, the structure of the solution is simple and hourly bids are sufficient. With the market price being known in advance, only two hourly bids are relevant. The relevant bids determine the part of the bidding curve that passes through the point given by the optimal production level and the market price. However, the resulting bidding curve may be very sensitive to changes in the market price. In some scenarios the deviations from the expected

market price may enforce heavy balancing as dispatches are far from being met by production. An example is given in Table 4.2 in which the imbalances of a given hour between the total volumes dispatched and the volumes produced are rather large in some scenarios (scenario 3,6 and 10).

For the stochastic bidding problems the structure of the solution is more complex. In the one-reservoir-case, the structure of the solution depends on whether start-up costs are included or not. Without start-up costs hourly bids are sufficient. With start-up costs both hourly bids and block bids are necessary. To explain, note that hourly bids follow prices closely. These are, however, accepted independently of each other and do not consider intertemporalities due to start-ups. Block bids, on the other hand, are valid for a number of consecutive hours and thus tend to support a regular production schedule with less start-ups. In the two-reservoir-case the structure of the solution likewise depends on whether start-up costs are included or not. With the size differences between the reservoirs, the presence of a system bottleneck has further impact on the structure of the solution. A bottleneck occurs when the capacity of the upper reservoir exceeds the capacity of the lower reservoir. Obviously, water is released from both the upper and the lower reservoirs in hours of a high market price. Due to combination of the bottleneck and the time delay between the reservoirs, however, excessive water releases from the upper reservoir may lead to forced releases in the lower reservoir in hours of a low market price. Without a bottleneck hourly bids are sufficient, whereas with a bottleneck both hourly bids and block bids are necessary. An explanation could be the following. Since hourly bids are accepted independently they ignore intertemporalities generated by dependencies between the reservoirs. On the other hand, as block bids are valid for a number of consecutive hours such bids can be used as protection against major price fluctuations over time.

Note that with different formulations of the problems hourly bids and block bids might be relevant both in the deterministic and the stochastic problems. Still, from the present formulation it is clear that uncertainty is one way of justifying the use of block bids and that block bids provide motivation for the inclusion of uncertainty into the bidding problem.

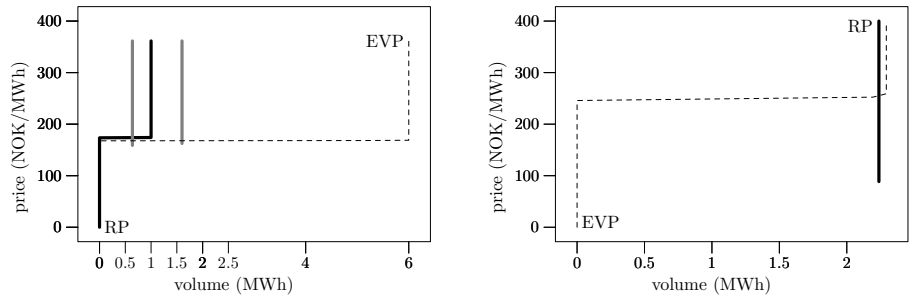
Based on 10 demonstration scenarios and 10 sample scenarios respectively, we have drawn bidding curves in Fig. 4.5. The curves are shown for both the deterministic problem (EVP), cf. the dashed lines, and the stochastic problem (RP), cf. full-drawn lines. The grey lines represent block bids and the black lines represent hourly bids. Although the bidding curves appear as piece-wise constant, the curves are in fact piece-wise linear. Nearly piece-wise constant bidding curves, however, are consistent with the bidding practice of the current application and reflects an almost price insensitive behavior between certain price levels.

The figures illustrate the discussion above. Consider Fig. 4.5b. For an expected market price of 261.90 NOK/MWh the dispatch is the same in the deterministic and the stochastic problems, i.e. 2.24 MWh. Nevertheless, as already stated, the

Table 4.2: Second-stage solutions of a given hour using demonstration scenarios. The columns H. disp. and B. disp. show dispatch generated by hourly bids and block bids, separately. Prod. denotes production. The column Imbal. shows imbalances between total production and dispatch.

	EVP				RP			
Sce.	H. disp.	B. disp.	Prod.	Imbal.	H. disp.	B. disp.	Prod.	Imbal.
1	2.29	0.00	2.24	−0.05	0.00	2.24	2.24	0.00
2	2.29	0.00	2.24	−0.05	0.00	2.24	2.24	0.00
3	0.00	0.00	1.37	1.37	0.00	2.24	2.24	0.00
4	2.29	0.00	2.24	−0.05	0.00	2.24	2.24	0.00
5	2.29	0.00	2.24	−0.05	0.00	2.24	2.24	0.00
6	0.00	0.00	1.37	1.37	0.00	2.24	2.24	0.00
7	2.29	0.00	2.24	−0.05	0.00	2.24	2.24	0.00
8	2.29	0.00	2.24	−0.05	0.00	2.24	2.24	0.00
9	2.29	0.00	2.24	−0.05	0.00	2.24	2.24	0.00
10	0.00	0.00	1.37	1.37	0.00	2.24	2.24	0.00

deterministic bidding curves are very sensitive to changes in the market price. If the price turns out to be 28.00 NOK/MWh (scenario 10) or 389.00 NOK/MWh (scenario 1), the dispatch must be smaller (0.00 MWh) or larger (2.29 MWh), which results in balancing and increased costs. The same situation applies to Fig. 4.5a.



(a) Sample scenarios. (b) Demonstration scenarios.

Figure 4.5: Bidding curves of a given hour.

To examine the effect of including uncertainty on production and sales profits, we solved the expected value problem and the two-stage recourse problem with both the demonstration scenarios and the sample scenarios. For the demonstration scenarios we have fixed equidistant price points and for the sample scenarios

distributional price points. We recorded the optimal value of the recourse problem as well as the expected result of using the expected value solution (EEV). To compare the two we calculated the difference which we refer to as the value of the stochastic solution (VSS). The VSS measures the effect of including stochastic prices explicitly into the bidding problem rather than simply using expected prices. Averages of 10 different runs are reported in Tables 4.3 and 4.4. We focus on the computational results from using the sample scenarios. Although the monetary gains may seem moderate on a per day basis, the gains accumulate with the problem being solved every day of the year. Moreover, recall that the dimension of the power plant has been scaled down by a factor ten. The percentual VSS is 7.93% on average. We conclude that significant profits can be earned by applying stochastic programming.

Table 4.3: Computational results for demonstration scenarios.

S	Opt. val.	EEV	VSS
10	34801.51	33821.17	980.34

Table 4.4: Computational results for sample scenarios.

S	Opt. val.	EEV	VSS
10	33395.45	30992.49	2402.96
30	33420.60	30676.07	2744.53
50	33405.75	30805.37	2600.39
100	33419.60	30566.44	2853.17

It should be remarked that, in terms of the bidding problem, the expected value problem produces rather simple solutions inducing a very optimistic value of the stochastic solution. However, rather than using expected values, current practice often relies upon the selection of appropriate critical quantiles of the uncertain data and the following solution of a number of deterministic problems. This may provide decisions that better hedge against adversity and results in a more moderate value of the stochastic solution. Still, the comparison of the expected value problem with the stochastic programming problem illustrates the effects of including uncertainty explicitly into optimization.

As concluded, solutions are less sensitive to changes in the data in the stochastic approach than in the deterministic version. Still, the expectation-based stochastic program lacks robustness in the sense that solutions tend to be unstable. Almost equal objection function values are obtained by structurally different solutions which indicates a flat objective function. Although structurally different, the solutions are equally good as long as the expectation-based objective criterion is

acknowledged. However, the expectation-based objective criterion can be claimed to ignore both risk attributes and profit distribution issues. In order to obtain a more robust stochastic programming model a risk measure can be included in the objective function. The result is the so-called mean-risk model. The inclusion of the downside risk measure semideviation penalizes deviations from the expected value and has the advantage of being consistent with a mixed-integer linear formulation. With semideviation, the objective function of (4.4.1)-(4.4.13) turns into

$$\sum_{s\in\mathcal{S}} \pi^s z^s - \lambda \sum_{s\in\mathcal{S}} \pi^s \max\Big\{\sum_{\bar{s}\in\mathcal{S}} \pi^{\bar{s}} z^{\bar{s}} - z^s, 0\Big\}$$

with

$$z^s = \sum_{t\in\mathcal{T}} \rho_t^s y_t^s + \sum_{a\in\mathcal{A}} \bar{\rho}_a^s y_a^s - \sum_{t\in\mathcal{T}} (p_t^{up} q_t^{up,s} - p_t^{do} q_t^{do,s}) - \sum_{t\in\mathcal{T}} \sum_{j\in\mathcal{J}} S_j(u_{jt-1}^s, u_{jt}^s) - \sum_{j\in\mathcal{J}} (V_j(l_{j0}^s) - V_j(l_{jT}^s))$$

where λ is a suitable weight. For more information on semideviation and other risk measures that are consistent with a linear formulation, see Chapter 2. The mean-risk problem ensures more stable solutions and a more equal profit distribution among the scenarios. In Table 4.5 we report the expected value and the risk for averages of 10 different runs using sample scenarios. Since the objective function is flat stable solutions can be obtained with only very small profit reductions.

We have solved the mean-risk problem for varying weights $\lambda \in \mathbb{R}_+$. The points defined by the selected weights are efficient points of the bi-criteria mean-risk problem and gives an indication of the efficient frontier, that can be obtained by solving the mean-risk problem for all $\lambda \in \mathbb{R}_+$. Still, even the efficient frontier is insufficient to fully describe the entire efficiency set. For more on the efficient frontier and the efficiency set, see also Chapter 2.

Finally, note that the inclusion of the downside risk measure semideviation does not alter the structure of the solution in a way that contradicts the analysis already made.

Table 4.5: Computational results for semideviation using simulation scenarios.

λ	0.001	1	10	100	1000
Exp. val.	33389.34	33389.34	33192.04	32922.73	32716.62
Risk	107.56	107.56	7.94	1.02	0.00

Chapter 5

Short-term hydro-power production planning by stochastic programming

The present chapter presents a short-term production planning problem not previously addressed in the literature. As in Chapter 4, the problem has become relevant with the restructuring of the power sector and the introduction of the day-ahead market.

Within the framework of multi-stage stochastic mixed-integer linear programming, the aim is to develop a short-term production plan for a price-taking hydro-power plant operating under uncertainty. Current production must however comply with the day-ahead commitments of the previous day which makes short-term production planning a matter of spatial distribution among the reservoirs of the plant. Day-ahead market prices and reservoir inflows are uncertain beyond the current operation day and, hence, water must be allocated among the reservoirs in order to strike a balance between current profits and expected future profits. A case study is based on data from a Norwegian hydro-power producer and the Nordic power market at Nord Pool.

5.1 Introduction

As is also the case in other regions of the world, hydro-power production accounts for a significant share of the total power production in the Nordic countries. As an example, the countries within Nordel produced 191 TWh hydro-power out of a total production of 387 TWh in 2004[1].

[1] Reference: `www.nordel.org`

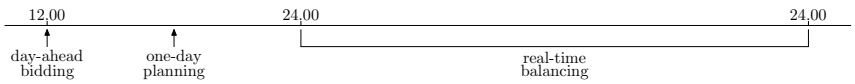


Figure 5.1: Time schedule for short-term power planning and operation.

In the process of planning hydro-power production, problems are usually categorized according to their time horizon; long-term, medium-term and short-term. Short-term hydro-power production planning mainly involves the physical operation of the plant within a time horizon of a day or a week and with a time resolution of an hour or shorter. The most important activities that come into play are

- the day-ahead commitments, i.e. the bidding of the production into a power exchange a day in advance.
- the establishment of a production plan that complies with the day-ahead commitments.
- the intra-day balancing, i.e. the continuous corrections of deviations between the commitments and the actual production.

The focus of the present paper is the establishment of a one-day production plan that complies with the day-ahead commitments of the previous day. From the perspective of this study, the day-ahead bidding has been completed whereas real-time balancing considerations will be postponed until actual production has been observed. For an illustration of the time schedule, see Fig. 5.1.

When the results of the day-ahead auction are known, the classical hydro-power problem, that consists in scheduling of water *through time*, is no longer an issue. The challenge in making a production plan for the following operation day is rather the scheduling of water *through space*, which involves the allocation of production between various parts of the plant to achieve effective and efficient operation. Since, however, short-term planning is strongly coupled to more long-term planning, the value of current decisions must be evaluated against future consequences. Therefore, in determining the spatial distribution of the following operation day, production is usually considered in a longer time span such as seven days.

Day-ahead market prices and reservoir inflows are both subject to data uncertainty caused by non-anticipated market conditions and unpredictable weather situations. We propose the stochastic programming framework to handle this uncertainty. Indeed, as information evolves over time and uncertainty is disclosed in stages, a multi-stage stochastic program is appropriate. In line with the above, the first stage relates to the one-day production plan and the remaining stages to the production of the following six days. It follows that the overall objective

of the stochastic program is to establish a one-day production plan that strikes a balance between current profits and expected future profits subject to a number of operational constraints.

Existing approaches to short-term production planning comprise both simulation and optimization. As simulation is often based on adjusting manual suggestions until a convincing plan is found, this approach is highly user dependent and does not guarantee an optimal plan. On the contrary, optimization represents a relatively impartial way of identifying an optimal plan. Whichever approach is used, most practical applications do not explicitly include data uncertainty.

The outline of the paper is the following. Sections 5.2 and 5.3 present a mixed-integer linear programming problem for the development of a one-day production plan that complies with the day-ahead commitments. Section 5.4 introduces data uncertainty and presents a stochastic programming formulation of the problem, whereas Section 5.5 explains how to generate scenarios that serve as input to the stochastic program. Finally, Section 5.6 illustrates the problem with a case study.

5.2 Short-term hydro-power production

The starting point for modeling is short-term hydro-power production. Modeling is restricted to mixed-integer linear programming and follows along the lines of for example Philpott et al. (2000). For illustration purposes, the case study is kept rather simple and the model concerns only a very small hydro-power plant. Some examples of non-modeled features are constraints that apply to the network of watercourses and junctions, the distinction between baseload and load-following power stations, reserve requirements as well as legal requirements, see Jacobs et al. (1995). Nevertheless, it should be clear that including additional hydrological constraints or modeling a larger hydro-power plant is possible by means of mixed-integer linear programming.

As in Chapter 4, the hydro-power plant of the case study consists of two reservoirs in cascade; a larger upper reservoir and a smaller lower reservoir. However, each reservoir now has its own inflow stream. Furthermore, each reservoir is connected to a power station that contains a number of turbines. As upstream water reaches the plant, it is stored in the reservoirs until released through the turbines. In the turbines, electricity is generated by converting potential energy into electrical energy, before the water proceeds downstream. Water released from the upper reservoir flows to the lower reservoir with some time delay on its way. Water that is not discharged on purpose and used for generation is considered spill. The size differences of the reservoirs restrict the flexibility of the system in that water releases from the upper reservoir may force releases in the lower reservoir or may even lead to spill. For an illustration of the plant, see Fig. 5.2. The model is presented in slightly more general terms than the case study requires.

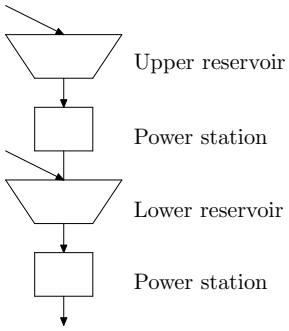


Figure 5.2: Two hydro-power reservoirs in cascade.

The time horizon covers the current operation day, for which a production plan should be made, as well as the following six operation days. Due to the nature of the data, the time horizon is discretized into intervals with the length of an hour and is denoted by $\mathcal{T} = \{1, \ldots, T\}$ with $T = 7 \times 24 = 168$.

To model hydro-power generation in cascade, let $\mathcal{J}$ index the reservoirs and let $\mathcal{I}_j, j \in \mathcal{J}$ index the generators of the connected power stations. For the case study, $\mathcal{J} = \{1, 2\}, \mathcal{I}_1 = \{1\}$ and $\mathcal{I}_2 = \{2\}$. Let the variables $u_{it} \in \{0, 1\}, i \in \mathcal{I}_j, j \in \mathcal{J}, t \in \mathcal{T}$ represent the on/off states of the generators, $w_{it} \in \mathbb{R}_+, i \in \mathcal{I}_j, j \in \mathcal{J}, t \in \mathcal{T}$ the generation levels and $v_{it} \in \mathbb{R}_+, i \in \mathcal{I}_j, j \in \mathcal{J}, t \in \mathcal{T}$ the corresponding discharges from the reservoirs. Also, let the variables $l_{jt} \in \mathbb{R}_+, j \in \mathcal{J}, t \in \mathcal{T}$ be the reservoir storage levels and $r_{jt} \in \mathbb{R}_+, j \in \mathcal{J}, t \in \mathcal{T}$ the spill.

As concerns direct costs of hydro-power generation, operating costs are negligible. However, start-up costs amount to

$$\sum_{t \in \mathcal{T}} \sum_{j \in \mathcal{J}} \sum_{i \in \mathcal{I}_j} SC_i(u_{it-1}, u_{it}),$$

where the cost functions are

$$SC_i(u_{it-1}, u_{it}) = c_i \max\{u_{it} - u_{it-1}, 0\}, \;\; i \in \mathcal{I}_j, j \in \mathcal{J}, t \in \mathcal{T}$$

and the costs per start-up are $c_i, i \in \mathcal{I}_j, j \in \mathcal{J}$. It should be remarked that the formulation can be transformed into a mixed-integer linear formulation. Initial on/off states of the generators are $u_{i0} = u_{i,init}, i \in \mathcal{I}_j, j \in \mathcal{J}$.

Indirect costs include opportunity costs of releasing water as the water could be stored and saved for future generation and, thus, such costs are measured as the value of stored water. The opportunity costs are

$$\sum_{j \in \mathcal{J}} (V_j(l_{j0}) - V_j(l_{jT})),$$

where

$$V_j(l_{jt}) = \min_{k\in\mathcal{K}}\{d^1_{kj}l_{jt} + d^2_{kj}\},\ \ j \in \mathcal{J}, t \in \mathcal{T}$$

and the coefficients of the concave water value functions are $d^1_{kj}, d^2_{kj}, k \in \mathcal{K}, j \in \mathcal{J}$. The formulation is consistent with a linear formulation. For further remarks on the derivation of the water value functions, see Chapter 4.

The following bounds are imposed on the generation levels

$$w_i^{min}u_{it} \le w_{it} \le w_i^{max}u_{it},\ \ i \in \mathcal{I}_j, j \in \mathcal{J}, t \in \mathcal{T},$$

where $w_i^{min}, i \in \mathcal{I}_j, j \in \mathcal{J}$ and $w_i^{max}, i \in \mathcal{I}_j, j \in \mathcal{J}$ are the minimum and maximum generation levels. The water discharges have to comply with similar bounds, so that

$$v_j^{min} \le \sum_{i\in\mathcal{I}_j} v_{it} \le v_j^{max},\ \ j \in \mathcal{J}, t \in \mathcal{T},$$

where $v_j^{min}, j \in \mathcal{J}$ and $v_j^{max}, j \in \mathcal{J}$ are the minimum and maximum discharges. Finally, the following bounds apply to the storage levels

$$l_j^{min} \le l_{jt} \le l_j^{max},\ \ j \in \mathcal{J}, t \in \mathcal{T},$$

where $l_j^{min}, j \in \mathcal{J}$ and $l_j^{max}, j \in \mathcal{J}$ denote minimal and maximal storage levels.

According to the reservoir balances, inflow and storage from previous periods either appear as discharge, storage or spill in the following period. The upper reservoir balance equations are therefore

$$l_{1t} - l_{1t-1} + \sum_{i\in\mathcal{I}_1} v_{it} + r_{1t} = \nu_{1t},\ \ t \in \mathcal{T},$$

in which $\nu_{1t}, t \in \mathcal{T}$ are the inflows from upstream. The initial storage level is $l_{10} = l_{1,init}$. The lower reservoir balance equations are

$$l_{jt} - l_{jt-1} + \sum_{i\in\mathcal{I}_j} v_{it} + r_{jt} = \sum_{i\in\mathcal{I}_{j-1}} v_{it-\tau} + \nu_{jt},\ \ j \in \mathcal{J}\backslash\{1\}, t \in \mathcal{T},$$

in which $\nu_{jt}, j \in \mathcal{J}\backslash\{1\}, t \in \mathcal{T}$ are the direct inflows from upstream. Note that, for reservoirs in cascade, releases from the upper reservoirs are inflows of the lower reservoirs. It is, however, assumed that upstream spill does not contribute to the downstream inflows. τ is the time delay between the reservoirs. Again, the initial storage level is $l_{j0} = l_{j,init}, j \in \mathcal{J}\backslash\{1\}$.

The generation level is a function of the water discharge from the reservoir and the net water head of the power station. The net water head is the difference

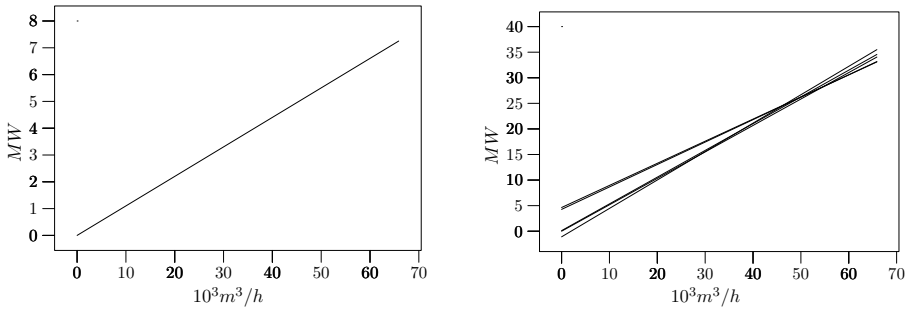
between the headwater elevation and the tailwater elevation and, whereas the former is a function of the reservoir storage level, the latter is a function of the discharge. It is, however, assumed that the net water head only varies with the discharge over the course of the short-term planning horizon. The assumption is justified in the case of relatively small storage level variations compared to the net head, which holds for the case study. By ignoring some head variation effects, the relation between generation and discharge can be approximated by a concave function. Hence,

$$w_{it} = G_i(v_{it}), \;\; i \in \mathcal{I}_j, j \in \mathcal{J}, t \in \mathcal{T},$$

where

$$G_i(v_{it}) = \min_{k \in \mathcal{K}} \{e^1_{ki} v_{it} + e^2_{ki}\}, \;\; i \in \mathcal{I}_j, j \in \mathcal{J}, t \in \mathcal{T}$$

and the coefficients of the concave functions are $e^1_{ki}, e^2_{ki}, k \in \mathcal{K}, i \in \mathcal{I}_j, j \in \mathcal{J}$. For an illustration of the generating functions of the case study, see Fig. 5.3. It is clear that the generation and discharge could in fact be assumed to be proportional in the case study.



(a) Upper reservoir.

(b) Lower reservoir.

Figure 5.3: Generation functions for two different hydro-power turbines.

5.3 Day-ahead market commitments

Norway was among the first countries in the world to undertake the deregulation of the power markets. In the beginning of the nineties a Norwegian power market was established and has since then developed into an overall Nordic power market. An essential component of the power market is the presence of the power exchange that facilitates physical trading activity on a day-per-day basis. The spot market,

Elspot, at the Nordic power exchange, Nord Pool, is a pool-based market in which market participants exchange power contracts for physical delivery the following operation day and is referred to as the day-ahead market. In 2004, a total of 167 TWh or 42 percent of the Nordel power production was traded at Elspot[2].

Elspot contracts are commitments to sell or purchase power of a duration of one hour or longer. To use such contracts, the market participants post price-differentiated bids for every hour of the following operation day before deadline at noon. The hourly market prices are then determined by equilibrium between sales and purchases. Once the market prices have been announced, the market participants receive a notification of the winning bids and the hourly commitments of the following operation day. Real-time operation and physical delivery is done according to the day-ahead commitments to the extent possible.

The value of electricity production should be measured on the basis of day-ahead market prices. We assume w.l.g. that the entire production is sold in the day-ahead market and that bilateral contracts are left out. Day-ahead market sales give rise to the revenues

$$\sum_{t\in\mathcal{T}}\sum_{j\in\mathcal{J}}\sum_{i\in\mathcal{I}_j}\rho_t w_{it},$$

where $\rho_t, t \in \mathcal{T}$ are the day-ahead market prices. By assuming that the producer is a price-taker, market prices can be modeled as exogenous. To justify the price-taker assumption, the producer should be sufficiently small to have only limited market power. In the case of significant market power the concepts of game theory, monopoly or oligopoly becomes important and the complexity of the model may increase.

The production of the first day has to meet the hourly commitments in the day-ahead market. As the day-ahead commitments are fixed a day in advance, these are given as data to the model and may be obtained from more long-term planning models such as the day-ahead bidding model of Chapter 4. This gives rise to the constraints

$$\sum_{j\in\mathcal{J}}\sum_{i\in\mathcal{I}_j} w_{it} = d_t, \;\; t \in \mathcal{T}_1,$$

where $\mathcal{T}_1$ indexes the hours of the first day and $d_t, t \in \mathcal{T}_1$ are the day-ahead commitments of the first day. If a producer relies on portfolio bidding and places compound bids that apply to a number of plants, the modeling of a larger hydropower system comes into play. Still, it should be straightforward to extend the model.

[2] Reference: `www.nordel.org` and `www.nordpool.no`

If production deviates from the day-ahead commitments, the imbalances are settled in an intra-day balancing market. Since, however, the purpose of the model is production planning rather than market exchange, intra-day balancing is not included. Moreover, since planned imbalances are forbidden by the market operators, producers are not allowed to hold back production for the balancing market. Although intra-day balancing is left out of production planning, a producer can always ramp down or, in the case of spare capacity, ramp up while actually producing and thereby participate in the real-time balancing market. For a way of incorporating balancing market considerations, see Chapter 4.

5.4 The stochastic programming model

Both day-ahead market prices and reservoir inflows are rather volatile and hard to predict because of unexpected market conditions and unforeseen weather situations. The model of the previous sections does not reflect the fact that new information about the uncertain data arrives along the planning horizon. Nevertheless, this can be handled by the application of multi-stage stochastic programming.

To facilitate modeling, we assume the uncertain data evolves over time according to a multivariate stochastic process and the probability information is approximated by a scenario tree. Consistent with the notation of multi-stage stochastic programming problems introduced in Chapter 1, the realizations of the uncertain prices and inflows $\{\rho_t, \nu_{1t}, \nu_{2t}\}_{t\in\mathcal{T}}$ are denoted $\{\rho^n, \nu_1^n, \nu_2^n\}_{n\in\mathcal{N}}$.

By assuming that information is revealed only at the beginning of an operation day, the scenario tree consists of seven stages or operation days that each consist of 24 time intervals or hours. The assumption is valid for day-ahead market prices that are announced a day prior to physical delivery. Furthermore, it is justified in the case of daily readings on the reservoir inflow measuring instruments or at least as an effort to limit the size of the scenario tree. An example of a scenario tree is shown in Fig. 5.4.

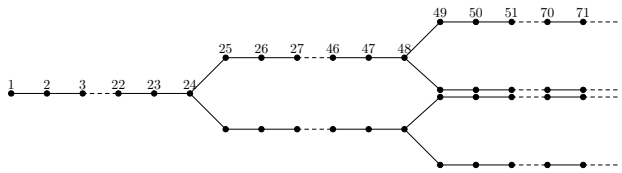


Figure 5.4: An illustration of a scenario tree for day-ahead market prices and reservoir inflows.

The inclusion of seven stages ensures the coupling between short-term and more long-term planning. While the first stage determines the one-day production plan,

the remaining six stages serve to evaluate the impact of the plan on future production. In line with this, the overall objective of the multi-stage stochastic program is to determine the one-day production plan that strikes a balance between current and expected future profits. In its scenario tree formulation, the multi-stage stochastic program is the following

$$\max \sum_{n\in\mathcal{N}}\sum_{j\in\mathcal{J}}\sum_{i\in\mathcal{I}_j}\pi^n\big(\rho^n w_i^n - SC_i(u_i^{n-1},u_i^n)\big)+ \tag{5.4.1}$$

$$\sum_{n\in\mathcal{N}_T}\sum_{j\in\mathcal{J}}\pi^n V_j(l_j^n) - \sum_{j\in\mathcal{J}} V_j(l_{j,init})$$

$$\text{s.t.} \sum_{j\in\mathcal{J}}\sum_{i\in\mathcal{I}_j} w_i^n = d^n,\ \ n\in\mathcal{N}_t, t\in\mathcal{T}_1 \tag{5.4.2}$$

$$u_i^n w_i^{min} \le w_i^n \le u_i^n w_i^{max},\ \ i\in\mathcal{I}_j, j\in\mathcal{J}, n\in\mathcal{N} \tag{5.4.3}$$

$$v_i^{min} \le \sum_{i\in\mathcal{I}_j} v_i^n \le v_i^{max},\ \ i\in\mathcal{I}_j, j\in\mathcal{J}, n\in\mathcal{N} \tag{5.4.4}$$

$$l_j^{min} \le l_j^n \le l_j^{max},\ \ j\in\mathcal{J}, n\in\mathcal{N} \tag{5.4.5}$$

$$l_1^n - l_1^{n-1} + \sum_{i\in\mathcal{I}_1} v_i^n + r_1^n = \nu_1^n,\ \ n\in\mathcal{N} \tag{5.4.6}$$

$$l_j^n - l_j^{n-1} + \sum_{i\in\mathcal{I}_j} v_i^n + r_j^n = \sum_{i\in\mathcal{I}_{j-1}} v_i^{n-\tau} + \nu_j^n,\ \ j\in\mathcal{J}\backslash\{1\}, n\in\mathcal{N} \tag{5.4.7}$$

$$w_i^n = G_i(v_i^n),\ \ i\in\mathcal{I}_j, j\in\mathcal{J}, n\in\mathcal{N} \tag{5.4.8}$$

$$u_i^n\in\{0,1\}, w_i^n, v_i^n, l_j^n, r_j^n \ge 0,\ \ i\in\mathcal{I}_j, j\in\mathcal{J}, n\in\mathcal{N}. \tag{5.4.9}$$

Note that the multi-stage stochastic program may be approximated by a two-stage stochastic program. A natural two-stage approximation would consist of a first stage that determines the one-day production plan and a second stage that comprises the remaining six days.

5.5 Scenario generation

The modeling of day-ahead market prices and reservoir inflows is based on time series analysis. Time series models have the advantage of being derived from limited information and still allow for both forecasting and sampling. As shown

in Chapter 8, the ARMA time series framework provides an appropriate model for hourly prices, whereas the modeling of hourly inflows requires development. Nevertheless, also in this case, we find the ARMA time series models sufficient to demonstrate the usefulness of stochastic programming.

The multivariate stochastic process of hourly day-ahead market prices and water inflows constitutes a time series[3] characterized by seasonal changes and stochastic variations that are mainly due to the market conditions and the weather patterns. In general, a multivariate stochastic process can be modeled as a vector ARMA (VARMA) process

$$\psi(B)\xi_t = \gamma(B)e_t, \;\; t \in \mathbb{Z},$$

where $\psi(B)$ and $\gamma(B)$ are the polynomials $\psi(B) = 1 - \sum_k \psi_k B^k$ and $\gamma(B) = 1 - \sum_k \gamma_k B^k$ with the parameter matrices ψ_k and γ_k and where B denotes the back-shift operator, i.e. $B^k \xi_t = \xi_{t-k}$. ξ_t and e_t are vectors, where $e_t, t \in \mathbb{Z}$ are assumed to be independent normally distributed random vectors with zero mean and constant covariance matrix Σ.

We consider the three-dimensional stochastic process $\{\rho_t, \nu_{1t}, \nu_{2t}\}_{t\in\mathcal{T}}$ of day-ahead market prices and reservoir inflows. An inspection of the cross correlations indicates that the innovations of prices and inflows are uncorrelated and, thus, prices and inflows can be modeled separately. For inflows, the innovations can be assumed only contemporaneously correlated and the inflows can therefore be modeled by a so-called contemporaneous ARMA (CARMA) model. For further references on CARMA models and stream-flows, we refer to Chapter 8. As a result, three univariate ARMA models may be fitted independently and combined to a multivariate model. The development of the three univariate ARMA models follows the steps

1. Identify a statistical model of the data.
2. Estimate the parameters of the model.
3. Validate the model.

1. In the identification step the data is made stationary and the orders of the polynomials are determined. The seasonal ARMA (SARIMA) models are identified as

$$(1 - \psi_1 B)(1 - B)(1 - B^{24})(1 - B^{168})\rho_t = (1 - \gamma_1 B)(1 - \gamma_{24} B^{24})(1 - \gamma_{168} B^{168})e_t, \;\; t \in \mathbb{Z} \quad (5.5.1)$$

[3] Data sources: Nord Pool (prices) and TrønderEnergi (inflows)

for day-ahead market prices and

$$(1-\psi_1^j B)(1-B)\nu_{jt} = (1-\gamma_1^j B-\gamma_2^j B^2)(1-\gamma_{41}^j B^{41})e_{jt},\ \ j=1,2, t\in\mathbb{Z}. \quad (5.5.2)$$

for water inflows.

2. Parameter estimation is based on maximum likelihood optimization.

3. Finally, diagnostic checks are applied to validate the assumptions of independence and normality.

We assume that the innovations of (5.5.1) and (5.5.2), i.e. $\{e_t\}_{t\in\mathbb{Z}}$ and $\{e_{1t}, e_{2t}\}_{t\in\mathbb{Z}}$, are uncorrelated and, since normally distributed, independent. Furthermore, we assume the innovations of (5.5.2), i.e. $\{e_{1t}\}_{t\in\mathbb{Z}}$ and $\{e_{2t}\}_{t\in\mathbb{Z}}$, are only contemporaneously correlated. In order to obtain completely uncorrelated innovations, we let $(e_{1t}, e_{2t})^T = C(\epsilon_{1t}, \epsilon_{2t})^T$, where $CC^T = \Sigma$, C is an upper triangular matrix, Σ is the covariance matrix of the innovations and $\{\epsilon_{1t}, \epsilon_{2t}\}_{t\in\mathbb{Z}}$ are independent normally distributed random vectors with zero mean and constant covariance matrix I_2. A multivariate ARMA model that describes the three-dimensional process of day-ahead market prices and reservoir inflows has then been fitted. A thorough discussion of the application of VARMA and CARMA to the day-ahead market prices and reservoir inflows is found in Chapter 8.

The development of the time series models in steps 1-3, is accomplished by the statistical software package SAS, version 8.2., see SAS Institute Inc. (1999).

There are various approaches for approximating what is actually a continuous probability distribution of the uncertain data by a discrete and finite distribution represented by a scenario tree. For a general overview on the construction of scenario trees, see Chapter 7. With available historical data, is possible to sample from the statistical model and convert the resulting scenario paths into a scenario tree. As a starting point, we therefore generate 1000 scenario paths using Monte Carlo sampling from the multivariate ARMA model. Examples of a few scenario paths are shown in Fig. 5.5. The scenario paths are combined to a scenario tree by applying the conditional clustering algorithm of Heitsch and Römisch (2006b). For multi-stage stochastic linear programs, the method is supported by stability results and consequently, the quality of the constructed scenario tree can be controlled by certain error bounds. Although, the current problem is actually a mixed-integer linear multi-stage stochastic program, we still use the method, the reason being that the conditional clustering is intuitively appealing. Moreover, only few scenario generation methods apply to multi-stage stochastic programs that involves multivariate stochastic processes. The Monte Carlo sampling was implemented in C++, whereas the method of Heitsch and Römisch (2006b) was most kindly made available by the authors.

Remark 5.5.1 *When approximating a continuous probability distribution by a discrete distribution, the quality of the approximation should be evaluated. How-*

ever, the error bounds controlling the quality of the scenario tree construction method of Heitsch and Römisch (2006b) are found to be quite loose. Furthermore, due to the mixed-integer linear formulation, the bounds may not actually be valid. For these reasons, a different evaluation of the practical performance is relevant. We have tested in-sample stability and out-of-sample stability of the stochastic programming problem. For further details on evaluating scenario generation methods and the relation to the particular scenario tree construction method, see Chapter 7. We have generated ten different scenario trees, formulated the stochastic programming problems and solved them. For in-sample stability, we report the optimal values. For out-of-sample stability, we have used back-testing and evaluated the optimal solutions on the historical data. We report the mean, the standard deviation and the standard deviation as a percent of the mean for the optimal values in Table 5.1. The out-of sample values have a smaller standard deviation than the in-sample values since in-sample stability is evaluated on different trees, whereas out-of-sample stability is evaluated on the same tree based on historical data. Note also that the out-of-sample value is much higher than the in-sample value. This would usually be a sign of poor approximate solutions. Here, it could be explained by the fact that, as the historical data includes the weeks of one year, the data does not suitably represent the true probability distribution of a specific week. Unfortunately, due to lack of data, we were unable to consider the specific week in a number of different years. As an overall conclusion, we find the stochastic programming problem to be sufficiently stable with respect to the scenario generation method in use.

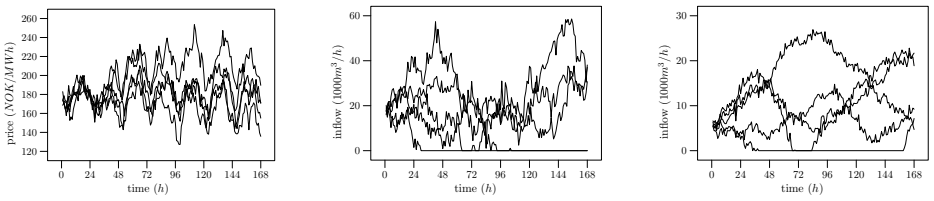
Table 5.1: In-sample stability and out-of sample stability for the problem (5.4.1)-(5.4.9).

	Sce.	Nodes	Mean	Std. dev.
In-sample	264	11681	564162.43	3919.77 (0.69%)
Out-of-sample	273	12096	945564.23	0.00 (0.00%)

In addition to sample scenarios, demonstration scenarios have been generated to illustrate various effects of day-ahead prices and reservoir inflows. The demonstration scenarios allow for prices and inflows to be significantly higher and lower than the real observations of the data.

5.6 Numerical results

The case study is based on data from a small Norwegian hydro-power plant located south of Trondheim and run by the company TrønderEnergi. The reservoir



(a) Day-ahead market prices.

(b) Reservoir inflows, upper reservoir.

(c) Reservoir inflows, lower reservoir.

Figure 5.5: Scenarios generated by sampling from a statistical model.

capacity of all power plants of this company amount to 0.7 TWh, the annual inflows to 1.5 TWh and the generation capacity to 334 MW[4]. Still, the case study concerns only two reservoirs and two power stations with each one turbine which corresponds a reservoir capacity of 24.6 GWh, annual inflows of 69.6 GWh and a generation capacity of 33.9 MW. The initial conditions of the plant are given by the data from a typical week of 2005 and the same applies for the day-ahead commitments. The generation of reservoir inflows for 2005 is based on observations from the year 2004. To generate corresponding day-ahead market prices for 2005, Nord Pool has provided the data from 2004 that applied to the Norwegian prize zone NO2, which includes the Trondheim area. Basically, NO2 consists of nine large hydro-power producers and a number of very small producers. The company TrønderEnergi accounts for 3 percent of the reservoir capacity, 5 percent of the annual inflows and 5 percent of the generation capacity of the area[5] and is therefore considered a price-taker of the area.

The multi-stage stochastic program (5.4.1)-(5.4.9) contains 12 variables, 1 binary and 11 continuous, and 22 constraints per node of the scenario tree, except for a few extra variables and constraints in the first and the last stages. Hence, it is a mixed-integer linear program which size depends on the number of scenarios and nodes of the scenario tree. We have solved the problem with the mixed-integer linear programming solver from OPL Studio version 3.7, ILOG Inc. (2003), calling CPLEX 9.0, Cplex Optimization Inc. (2006), on an Intel Xeon 2.67 GHz processor with 4 GB RAM. Direct application of CPLEX is no longer possible when the number of scenarios and nodes of the scenario tree is further increased. We consider the scenario tree to be sufficiently large to appropriately approximate the probability distribution . Moreover, regarding traditional solution approaches to stochastic hydro scheduling problems, dynamic programming may suffer from the curse of dimensionality and as shown in Archibald et al. (1996) may be un-

[4] Norwegian Competition Authority, 2002

[5] Norwegian Competition Authority, 2002

able to outperform linear programming methods. Also, due to the mixed-integer formulation, the problem is not amenable to nested Benders' decomposition. It may, however, be suitable for some of the general decomposition approaches that apply to multi-stage stochastic mixed-integer linear programming problems. Solution approaches worth investigating comprise those based on Lagrangian relaxation of non-anticipativity constraints, cf. progressive hedging by Løkketangen and Woodruff (1996) or nodal coupling constraints, cf. Dentcheva and Römisch (2004).

While varying the level of clustering in the scenario tree construction method, we have recorded the number of scenarios and nodes of the scenario tree, the total number of variables and constraints of the problem, the optimal objective function value and the computing time spent to solve the problem, cf. Table 5.2. All numbers reported are averages of 10 different runs and all computations are based on sample scenarios. Test runs indicate that the two-stage approximation to the multi-stage stochastic program may provide first-stage solutions of a significant quality in terms of objective function values.

Table 5.2: Computational results for sample scenarios.

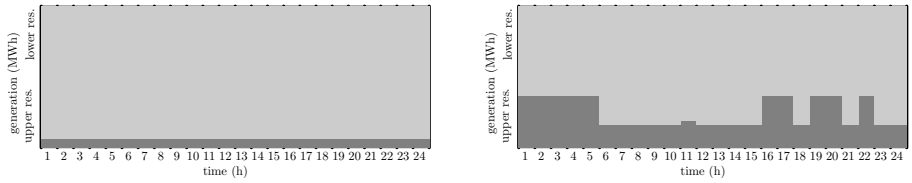
Sce.	Nodes	Var.	Con.	Obj. val./NOK	CPU/s
267	11777	141846	261254	562207.50	28.71
493	43709	525494	965566	567086.91	203.81
782	103144	1239304	2275470	566645.83	764.35

In hydro-power production, planning is coupled in time. The time coupling can be explained mainly by the storage balancing between the reservoirs. The storage balancing and the capacities of the reservoirs determine the spatial distribution of water between the reservoirs of the hydro-power plant. Therefore, the spatial distribution depends on future day-ahead market prices and reservoir inflows. To illustrate this, consider the case where both reservoirs are close to empty. If day-ahead market prices for the following six days are expected to be high, there is a potential for future generation and current production takes place in downstream reservoirs. This ensures future water releases from all reservoirs in cascade and higher future generation levels. If, on the other hand, day-ahead market prices for the following six days are expected to be low, current production is allocated according to water values and start-up costs. Now, consider the case where both reservoirs are almost full. If reservoir inflows for the following six days are expected to be high, current production takes place in downstream reservoirs in order to empty the system and accommodate future generation. If reservoir inflows for the following six days are expected to be low, current production is again allocated according to water values and start-up costs.

The dependence of the spatial distribution on the future day-ahead market prices and reservoir inflow levels is depicted in Figs. 5.6-5.7. The figures repre-

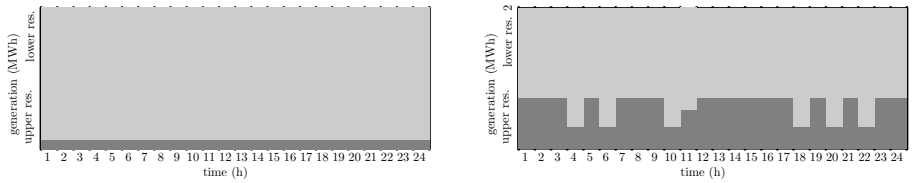
sent the hourly day-ahead commitments, which for the purpose of illustration are chosen to be constant, and the dark and light gray colors show the distribution of generation between the upper and lower reservoirs. The figures are based on demonstration scenarios.

Obviously, mixed effects of future day-ahead market prices and water inflows may appear. If both reservoirs are nearly empty and low day-ahead market prices are expected, current production may take place upstream or downstream depending on future inflow levels as current production is allocated according to water values and start-up costs. In the same fashion, consider the case where both reservoirs are close to full and water inflows are expected to be low. If day-ahead market prices are expected to be low, current production takes place upstream. If day-ahead market prices are expected to be high, there is a potential for future generation and current production takes place both upstream and downstream to ensure higher future generation levels.



(a) High price level. (b) Low price level.

Figure 5.6: Spatial distribution among two reservoirs in cascade. Both reservoirs are close to empty.



(a) High inflow level. (b) Low inflow level.

Figure 5.7: Spatial distribution among two reservoirs in cascade. Both reservoirs are close to full.

Time coupling may also be explained by start-up costs. Furthermore, start-up costs contribute to the complexity of the model by the introduction of binary variables. It is therefore relevant to investigate the importance of including start-up costs. Using sample scenarios, the result of the case study is a production plan with few start-ups. Therefore, to illustrate the effect of start-up costs, we

have generated demonstration scenarios that capture fluctuations in day-ahead prices and reservoir inflows and used these for testing. The results of fluctuating day-ahead market prices and the combination of fluctuating day-ahead market prices and reservoir inflows are displayed in Table 5.3. We have listed the optimal objective function value, the expected number of start-ups and the expected start-up costs. Although start-up costs seem to be limited, the mixed-integer linear formulation still has its relevance. The modeling of a larger hydro-power plant with more turbines or the inclusion of other hydrological constraints might involve an increase in the number of integer variables. Moreover, when increasing the costs per start-up, total start-up costs increase even though the expected number of start-ups decreases.

Table 5.3: Information on start-up costs.

	Obj. val./NOK	Start-ups	Start-up costs/NOK
Prices	1466433.58	0.34	390.53
Prices and inflows	1462328.61	0.22	214.62

By employing the expectation-based objective function criterion, the production planning is conducted in a risk neutral fashion. As most power producers are in fact risk averse, portfolio hedging comes into play. Portfolio hedging is often separated from production planning so that the aim of planning is to maximize the value of the available resources, while hedging alone aims to control the risk of the portfolio. Still, to control risk along with production planning, and in particular to achieve a more uniform profit distribution among scenarios, a risk measure can be appended to the objective function. The result is the so-called mean-risk model. The downside risk measure semideviation penalizes deviations below expected profit and has the advantage of being consistent with a linear formulation. For further references on semideviation, see Chapter 2. We append the risk measure and penalize accumulated deviations to obtain the following mean-risk model

$$\sum_{n\in\mathcal{N}} \pi^n z^n - \lambda \sum_{n\in\mathcal{N}_T} \pi^n \max\Big\{ \sum_{\bar{n}\in\mathcal{N}} \pi^{\bar{n}} z^{\bar{n}} - \sum_{\bar{n}\in path(n)} z^{\bar{n}}, 0 \Big\},$$

where λ is a weight and where

$$z^1 = \sum_{j\in\mathcal{J}} \sum_{i\in\mathcal{I}_j} \rho^1 w_i^1 - \sum_{j\in\mathcal{J}} \sum_{i\in\mathcal{I}_j} S_i(u_{i,init}, u_i^1) - \sum_{j\in\mathcal{J}} V_j(l_{j,init}),$$

$$z^n = \sum_{j\in\mathcal{J}} \sum_{i\in\mathcal{I}_j} \rho^n w_i^n - \sum_{j\in\mathcal{J}} \sum_{i\in\mathcal{I}_j} S_i(u_i^{n-1}, u_i^n), \;\; n \notin \mathcal{N}_T \cap \{1\},$$

$$z^n = \sum_{j\in\mathcal{J}}\sum_{i\in\mathcal{I}_j} \rho^n w_i^n - \sum_{j\in\mathcal{J}}\sum_{i\in\mathcal{I}_j} S_i(u_i^{n-1}, u_i^n) + \sum_{j\in\mathcal{J}} V_j(l_j^n), \;\; n \in \mathcal{N}_T.$$

It should be remarked that it is possible to include a multi-period risk measure, which does not alone focus on accumulated profit but also take into account intermediate time periods. For some examples on multi-period risk measures, see Eichhorn et al. (2005).

We have solved the problem for varying weights and recorded the expected value along with the risk, cf. Table 5.4. By solving for all positive weights, we would obtain the efficient frontier. Still, the efficient frontier does not represent the entire efficiency set and therefore we display only a selected set of weights and their corresponding records. All numbers reported are averages of 10 different runs and all computations are based on simulation scenarios.

Table 5.4: Computational results for semideviation using sample scenarios.

λ	0.001	1	10	100	1000
Exp. val.	563110.56	563110.54	495984.66	467811.71	457764.42
Risk	21906.52	21906.26	1480.32	43.03	0.00

From Table 5.4 it can be seen that when employing the expectation-based objective function criterion in short-term production planning, the downside risk is significant. Also, a reduction in risk requires a considerable reduction in profit. In the long run, however, hydro-power producers have a natural hedge. Whereas prices and inflow are uncorrelated in the short run, they are usually negatively correlated in a longer time span. If inflows decrease, prices tend to increase and compensate for this and vice versa.

Chapter 6

Managing power reserves by two-stage stochastic programming

With the restructuring of the power sector, the reliability of the system has to a great extent become the responsibility of the system operator. This, among other things, includes the balancing of power supply and demand and the management of reserves to facilitate it.

Although the physical transaction of power through bilateral trade or spot market exchange aims at balancing supply and demand, real-time imbalances may still occur due to non-anticipated supply and demand behavior. The real-time balance is the responsibility of the power system operator and is achieved by means of regulation that is purchased in the regulating market. To ensure sufficient regulation in the market, the system operator has the possibility of reserving regulating power in advance. As reserves are purchased prior to actual operation, however, reserve decisions are subject to supply and demand uncertainty. Regulation decisions on the other hand can be deferred until uncertainty has been revealed and the system is operating. In the present paper this is formalized by formulating the problem of managing regulating reserves as a two-stage stochastic program. A case study that concerns the system operator of Western Denmark is discussed in detail.

6.1 Introduction

In a power system that comprises several participants on both the supply and the demand side, it is the task of the system operator to balance production and consumption by means of regulation purchased in the regulating market. A

major part of the balancing is however the management of regulation reserves, as sufficient amounts of regulation are not necessarily available in the market unless reserved in advance. Considering the challenges of uncertain supply and demand, the problem of managing regulating power reserves can be handled by means of stochastic programming. Moreover, previous studies show how similar power optimization problems can be handled successfully by mathematical programming and stochastic programming in particular, see Chapter 3.

The paper is organized as follows. The problem of managing power reserves is presented in Section 6.3. Section 6.4 explains how uncertainty affects reserve management and formulates the problem as a two-stage stochastic program. By assuming a discrete distribution of the random data, the problem is transformed into a large-scale mathematical program that is solved by a specially designed solution procedure in Section 6.5. A specific instance of the problem is addressed using data from the power system operator of Western Denmark and computational results are reported in Section 6.6.

6.2 Power reserves

The project grew out of a collaboration with the former Eltra[1], which is the power system operator of Western Denmark. Due to decentralization of the power generation and deregulation of the power markets, many procedures have either been modified recently or will be within the near future. In particular, Eltra made plans to improve the model on which power reserve management is based, which makes reserves a topic of current interest.

It is necessary to distinguish between different types of reserves

(i) Automatic regulation reserves: Reserves that cover imbalances from the time of appearance until a regulation bid is activated. The reserves are provided by running plants capable of adjusting upwards and consumers capable of adjusting downwards. Activation begins automatically within two to three minutes.

(ii) Manual regulation reserves: Reserves in the form of regulation resources that suppliers are obligated to sell in the regulating market. The reserves are activated manually within 10 minutes.

(iii) Running and available plants: Reserves for ensuring supply in spite of transmission lines or units falling out. Consist of available plants that can be started, running plants that can adjust upwards and consumers that can adjust downwards. Running and available plants are activated either automatically or manually.

[1] now a part of the overall Danish power and gas system operator Energinet.dk, `www.energinet.dk`

(iv) Emergency start plants: Reserves reestablishing the system in case of black-out.

We here confine ourselves to manual regulation reserves. The reason why manual regulating reserves are relevant should be clear from the following. In a typical power system the operator is responsible for balancing supply and demand. Prior to operation, predicted supply and demand are balanced by resources provided by bilateral trades and spot market exchanges. Still, supply and demand often differ from the predictions and imbalances occur when the system is operating. To overcome such real-time imbalances, the system operator compensates suppliers and consumers for adjusting production and demand accordingly. The imbalances are covered by so-called regulation resources, that are traded in a regulating market, established by system operators to serve this purpose. Suppliers either increase or decrease power production above or below the amounts committed through bilateral and spot market contracts and offer the adjustments to the regulating market as so-called up- and down-regulation, respectively. Consumers offer power demand adjustments to the regulating market in a similar fashion. The system operator purchases up-regulation in the case of excess demand and down-regulation in the case of excess supply. In some cases, though, the amounts of resources provided by the regulating market are insufficient to fully cover imbalances. This may happen if

(i) Imbalances are substantial due to extreme supply and demand behavior caused by failure in supplying, unforeseen weather changes leading to unpredicted wind production or non-anticipated heat demand etc.

(ii) Spot market prices are sufficiently high to prevent market participants from saving resources for trading in the less secure regulating market.

(ii) Considerable failures occur during transmission, e.g. important transmission lines fall out.

To ensure that sufficient amounts of resources are available even when facing such critical situations, regulation can be reserved prior to trading in the regulating market. The system operator may secure the right of purchasing regulation with reserver supplier. This right is a type of option in that the system operator pays for the possibility of purchasing regulation in the market without being forced to do so. The reserve supplier is obligated to bid an amount of regulation into the market. As a result, the system operator faces the trade-off between purchasing regulation at the market price only, thereby risking insufficiency of resources, and paying both the market price and an additional fixed price to ensure that regulation is available. We refer to the first and the second type of regulation as direct and reserved regulation, respectively.

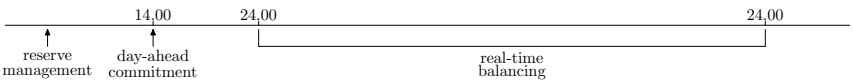


Figure 6.1: Time schedule for the activities involved in power balancing.

The major problem of managing regulating reserves is that of procuring prior to observing supply and demand imbalances. On the one hand, the reserve capacity may turn out to be insufficient to fully cover the imbalances not covered by direct regulation, which means additional regulation resources must be procured elsewhere at considerable costs, otherwise the system simply breaks down. On the other hand, reserves constitute serious costs, which makes excess reserve capacity unwanted. Stochastic programming provides a tool for determining regulation reserve levels that "hedge" against the non-anticipated supply and demand behavior.

The rest of the chapter is motivated by a case study based on the power system of Western Denmark as well as the local and Nordic power markets. In many respects, however, the discussion applies in general to other power systems.

As already explained, different power markets come into play when balancing power supply and demand. To fully understand the interactions of the system operator with the markets, we consider the following time schedule. The system operator procures regulating reserves for a longer time period. In Western Denmark, a formal reserve market has not yet been fully established. Currently, the length of the period is at least one month though a reduction is planned. The remaining actions concern a 24-hour operation day. By noon bids must be submitted to the spot market, which in the Danish case is the Nordic Market at Nord Pool[2]. Having balanced predicted supply and demand, activated bids are announced by 14:00. Finally, from 24:00 to 24:00 actual supply and demand imbalances are continuously corrected by trading in the intra-day regulating market. For the case study, the regulating market is still the local market, although integration into the Nordic market is planned. For an illustration of the time schedule, see Fig. 6.1.

6.3 The problem of managing power reserves

In stating the problem, consider a given planning horizon. Although in practice, the power balance should be maintained at every point in time, the horizon is discretized in order to facilitate computations. Since the bids to the regulating marked have a duration of a number of full hours, the horizon is discretized into hourly time intervals. The finite set of such intervals is denoted by $\mathcal{T} = \{1, \ldots, T\}$.

[2] `www.nordpool.no`

The planning horizon for purchasing regulating reserves may range from one to several months. In Western Denmark some reserve contracts have long durations, whereas other contracts are traded regularly on a monthly basis. The planning horizon of the case study is set to one month, i.e. $T = 24 \times 30 = 720$. However, to increase the flexibility of the system, it is intended to reduce the horizon, as is already the case in Eastern Denmark where reserve contracts are traded on a daily basis.

6.3.1 Procuring reserves

At present, the system operator of Western Denmark procures regulation reserves mostly locally. Although not fully formalized, regulation reserves are traded on a monthly auction or as individual contracts of a longer duration. A regulation reserve bid consists of an offer period, a volume to be available throughout the offer period, a variable price that applies to the portion of the volume that is activated and a fixed price for activation of the bid. We assume for the application that regulation reserves are traded solely on the auction so that the offer period is always one month. According to the above, the system operator pays the variable price for the actual amount of regulation used and the fixed price for the availability of regulation.

Regulation reserves are modeled as follows. Regulation divides into up- or down-regulation, and so do regulation reserves; thus, the superscripts *up* and *do*. The indices $\mathcal{H} = \{1, \ldots, H\}$ are adopted to represent different reserve bids. The variables $x_h^{up}, x_h^{do} \in \{0, 1\}, h \in \mathcal{H}$ indicate whether the reserve bids are activated or not and the prices for activation are denoted $c_h^{up}, c_h^{do}, h \in \mathcal{H}$. Then fixed regulation reserve purchase costs compute as

$$\sum_{h \in \mathcal{H}} (c_h^{up} x_h^{up} + c_h^{do} x_h^{do}).$$

6.3.2 Purchasing regulation

A larger power system may share a common regulating market. The Nordic system operators, the Swedish Svenska Kraftnät, the Norwegian Statnett and the Danish Energinet.dk, have established such a common market, in which Western Denmark was the last to begin integration in January 2006. Still, although full integration is on its way, Western Denmark trades regulation mostly locally. Before July 2006, the regulating market of Western Denmark was a pay-as-bid market. Now, as a step towards integration, the general rule is to use local marginal prices as market prices and finally, Nordic marginal prices should be fully in use by January 2008.

When purchasing regulation and regulation reserves outside Western Denmark, transmission capacity limits may apply. Such limits are due to physical limitations

or political agreements. Since, for the current application, however, regulation is mostly purchased locally, we have omitted transmission capacity limits and in general assumed no grid congestion.

The suppliers to the regulating market are power balance providers submitting bids to the market. Such regulation bids divide into up-regulation and down-regulation bids. Upward regulation makes consumers decrease demand or suppliers increase production (system operators "buy" power) and downward regulation makes suppliers decrease production or consumers increase demand (system operators "sell" power). A regulation bid consists of an offer period, a price and a volume. The offer period may be a number of full hours. During the offer period, the volume is constant whereas the price may vary between hours. We however assume that regulation bids have an offer period of only one hour and thus both the volume and the price are constant. Generally, the up-regulation price is specified as the system spot price (assuming no grid congestion) and a raise, i.e. the up-regulation price is always above the system price. Similarly, the down-regulation price is calculated as the system price (assuming no grid congestion) and a deduction, i.e. the down-regulation price is always below the system price. Since prices are usually given as positive numbers unless the system operator sells up-regulation or buys down-regulation, prices are always assumed to be positive.

Regulation comprises direct purchases in the regulating market and purchases reserved in advance. In the case of reserved purchases, recall that the indices $\mathcal{H} = \{1, \ldots, H\}$ are included to represent different bids. Up- and down-regulation volumes are denoted $\bar{y}_h^{up}, \bar{y}_h^{do}, h \in \mathcal{H}$. We assume no failure of supply and, hence, all reserved purchases will be available in the regulating market. Corresponding prices are denoted $\bar{p}_{ht}^{up}, \bar{p}_{ht}^{do}, h \in \mathcal{H}, t \in \mathcal{T}$. In the case of direct purchases, $\mathcal{K} = \{H+1, \ldots, H+K\}$ are included to index different bids. Up- and down-regulation volumes are denoted $\bar{y}_{ht}^{up}, \bar{y}_{ht}^{do}, h \in \mathcal{K}, t \in \mathcal{T}$ and corresponding prices are $\bar{p}_{ht}^{up}, \bar{p}_{ht}^{do}, h \in \mathcal{K}, t \in \mathcal{T}$. Note that reserved purchases are independent of time, whereas direct purchases are time dependent. Furthermore, for reserved purchases, prices should stay between predefined limits, whereas for direct purchases, prices can vary freely. Finally, a bid is not necessarily activated completely. Actual purchases are represented by the variables $y_{ht}^{up}, y_{ht}^{up} \in \mathbb{R}_+, h \in \mathcal{H} \cup \mathcal{K}, t \in \mathcal{T}$.

Pay-as-bid pricing

Costs of purchasing regulation, whether reserved or direct, consist of up-regulation expenses and down-regulation income

$$\sum_{t \in \mathcal{T}} \sum_{h \in \mathcal{H} \cup \mathcal{K}} (\bar{p}_{ht}^{up} q_{ht}^{up} - \bar{p}_{ht}^{do} y_{ht}^{do}).$$

The following bounds concern reserved purchases

$$y_{ht}^{up} \le \bar{y}_h^{up} x_h^{up}, \;\; y_{ht}^{do} \le \bar{y}_h^{do} x_h^{do}, \;\; h \in \mathcal{H}, t \in \mathcal{T}, \tag{6.3.1}$$

whereas direct purchases must submit to the bounds

$$y_{ht}^{up} \le \bar{y}_{ht}^{up}, \;\; y_{ht}^{do} \le \bar{y}_{ht}^{do}, \;\; h \in \mathcal{K}, t \in \mathcal{T}. \tag{6.3.2}$$

Marginal pricing

For each hour the marginal price is determined as the price of the last bid activated in the direction in which the system is regulated. If the system is up-regulated, the marginal price is the highest price of the activated up-regulation bids. Likewise, if the system is down-regulated, the marginal price is the lowest price of the activated down-regulation bids. The variables $\delta_{ht}^{up}, \delta_{ht}^{do} \in \{0,1\}, h \in \mathcal{H} \cup \mathcal{K}$ indicate whether the regulation bids are activated or not for both reserved and direct regulation. Moreover, the variables $p_t^{up}, p_t^{do} \in \mathbb{R}_+^{n_1}$ represent the marginal prices. In the case of up-regulation, the marginal price is $p_t^{up} = \max\{\bar{p}_{ht}^{up} \delta_{ht}^{up} : h \in \mathcal{H} \cup \mathcal{K}\}$ and in the case of down-regulation, the marginal price is $p_t^{do} = \min\{\bar{p}_{ht}^{do} \delta_{ht}^{do} : h \in \mathcal{H} \cup \mathcal{K}\}$.

Costs of purchasing reserved and direct regulation amount to

$$\sum_{t \in \mathcal{T}} \big(p_t^{up} \sum_{h \in \mathcal{H} \cup \mathcal{K}} y_{ht}^{up} - p_t^{do} \sum_{h \in \mathcal{H} \cup \mathcal{K}} y_{ht}^{do} \big). \tag{6.3.3}$$

Evidently, (6.3.3) is nonlinear. In order to be consistent with a mixed-integer linear formulation, we introduce the variables $\rho_t^{up}, \rho_t^{do} \in \mathbb{R}_+^{n_1}, t \in \mathcal{T}$ and replace (6.3.3) by

$$\sum_{t \in \mathcal{T}} (\rho_t^{up} - \rho_t^{do})$$

and

$$\rho_t^{up} \ge \bar{p}_{ht}^{up} \sum_{h \in \mathcal{H} \cup \mathcal{J}} y_{ht}^{up} - M(1 - \delta_{ht}^{up}), \;\; h \in \mathcal{H} \cup \mathcal{K}, t \in \mathcal{T}, \tag{6.3.4}$$

$$\rho_t^{do} \le \bar{p}_{ht}^{do} \sum_{h \in \mathcal{H} \cup \mathcal{K}} y_{ht}^{do} + M(1 - \delta_{ht}^{do}), \;\; h \in \mathcal{H} \cup \mathcal{K}, t \in \mathcal{T}, \tag{6.3.5}$$

with

$$M = \max \Big\{ \bar{p}_{ht}^{up} \big(\sum_{h \in \mathcal{K}} \bar{y}_{ht}^{up} + \sum_{h \in \mathcal{H}} \bar{y}_t^{up} \big), \bar{p}_{ht}^{do} \big(\sum_{h \in \mathcal{K}} \bar{y}_{ht}^{do} + \sum_{h \in \mathcal{H}} \bar{y}_t^{do} \big) : h \in \mathcal{H} \cup \mathcal{K}, t \in \mathcal{T} \Big\}.$$

For both reserved and direct purchases, the volumes activated cannot exceed the volumes bid

$$y_{ht}^{up} \leq \bar{y}_{ht}^{up}\delta_{ht}^{up}, \;\; y_{ht}^{do} \leq \bar{y}_{ht}^{do}\delta_{ht}^{up}, \;\; h \in \mathcal{H} \cup \mathcal{K}, t \in \mathcal{T}. \tag{6.3.6}$$

Reserved regulation cannot be activated unless the regulation reserve bids are activated

$$\sum_{t\in\mathcal{T}} \delta_{ht}^{up} \leq x_h^{up}, \sum_{t\in\mathcal{T}} \delta_{ht}^{do} \leq x_h^{do}, \;\; h \in \mathcal{H}. \tag{6.3.7}$$

Note that we model the price determination process in terms of mixed-integer linear programming with the introduction of binary variables indicating activation or not. An alternative to avoiding nonlinearities would have been to discretize the continuous price range into a number of price levels and associate with each price level one or more binary variables. The approach is seen in Nowak et al. (2005), who address the bidding problem of a power producer being able to influence market prices. The number of price levels, however, is likely to exceed the number of bids to be considered for activation. For this reason, the above formulation has its advantages.

6.3.3 Balancing

The system operator takes care of the power balancing during actual operation. If actual demand exceeds actual supply, the system operator purchases up-regulation and if supply exceeds demand, down-regulation is purchased. Imbalances between demand and supply are usually fully covered by regulation purchased directly or reserved in advance. In the case of insufficient regulation, however, excess demand and supply may result in irregular in- and out-flows from abroad that are hardly penalized. Such in- and out-flows can be avoided by forcing consumers to decrease demand, by forcing power plants to shut down generating units or by stopping wind turbines, in which cases severe costs are also paid. Let the variables $e_t^{up}, e_t^{do}, t \in \mathcal{T}$ denote excess demand and supply. If $b_t^{up}, b_t^{do}, t \in \mathcal{T}$ denote penalty costs, excess demand and supply give rise to the following costs

$$\sum_{t\in\mathcal{T}} (b_t^{up} e_t^{up} + b_t^{do} e_t^{do}).$$

The power balance constraints are

$$\sum_{h\in\mathcal{H}\cup\mathcal{K}} (y_{ht}^{up} - y_{ht}^{do}) + e_t^{up} - e_t^{do} = \Delta_t, \;\; t \in \mathcal{T}, \tag{6.3.8}$$

where $\Delta_t, t \in \mathcal{T}$ denote the imbalances between demand and supply. If $\Delta_t > 0$, demand exceeds supply and if $\Delta_t < 0$, vice versa. Note that supply includes central, decentral and wind production as well as import and that demand consists of national consumption and export.

6.4 Introducing uncertainty

The problem presented in the preceding sections is deterministic. The real problem, however, involves uncertainties in the data and would therefore be suited for stochastic optimization. Uncertainties relate to the regulating prices and volumes that can be changed until an hour before operation. Moreover, demand and supply uncertainty arises because of failure in supplying, unforeseen weather changes leading to unpredicted wind production or non-anticipated heat demand. Since, however, the system imbalances are caused by demand and supply, we restrict the attention to demand and supply uncertainty.

Uncertainty is handled by means of stochastic programming. Although information evolves over time and a multi-stage stochastic program could be relevant, we approximate the problem by a two-stage stochastic program. We find this approximation sufficient to capture the interplay between reserves and regulation purchases. As reserves must be purchased a month in advance, reserve decisions are first-stage. Decisions have to be made before operation and thus with incomplete knowledge of future supply and demand. In contrast, regulation bids have an activation period of at most ten minutes and can therefore be purchased very close to operation, which makes regulation decisions second-stage. The objective is to minimize reserve costs and expected future regulation and penalty costs.

The uncertain data is represented by a stochastic process on some probability space. To make the problem computationally tractable, we assume a discrete multivariate distribution with finite support. The realizations of uncertainty will be referred to as scenarios, indexed by $\mathcal{S} = \{1, \dots, S\}$ and denoted by $(\Delta_t^s)_{t\in\mathcal{T}\, s\in\mathcal{S}}$. The corresponding probabilities will be denoted by $\pi^s, s \in \mathcal{S}$. First-stage reserve decisions are $x_h^{up}, x_h^{do} \in \{0,1\},\ \ h \in \mathcal{H}$, whereas second-stage regulation decisions are indexed $y_{ht}^{up,s}, y_{ht}^{do,s}, p_{ht}^{up,s}, p_{ht}^{do,s}, e_t^{up}, e_t^{do} \geq 0,\ \ h \in \mathcal{H} \cup \mathcal{K}, t \in \mathcal{T}, s \in \mathcal{S}$.

The two-stage stochastic programming formulation of the regulating reserve management is the following problem, depending on whether pay-as-bid or marginal pricing applies. The extension to market integration should be straightforward.

Pay-as-bid pricing

$$\min \sum_{h\in\mathcal{H}} (c_h^{up} x_h^{up} + c_h^{do} x_h^{do}) + \tag{6.4.1}$$

$$\sum_{s\in\mathcal{S}} \sum_{t\in\mathcal{T}} \sum_{h\in\mathcal{H}\cup\mathcal{K}} \pi^s (\bar{p}_{ht}^{up,s} y_{ht}^{up,s} - \bar{p}_{ht}^{do,s} y_{ht}^{do,s}) +$$

$$\sum_{s\in\mathcal{S}} \sum_{t\in\mathcal{T}} \pi^s (b_t^{up} e_t^{up,s} + b_t^{do} e_t^{do,s})$$

$$\text{s.t. } (6.3.1) - (6.3.2), (6.3.8)$$

$$x_h^{up}, x_h^{do} \in \{0,1\}, \;\; h \in \mathcal{H}$$

$$y_{ht}^{up,s}, y_{ht}^{do,s}, p_{ht}^{up,s}, p_{ht}^{do,s}, e_t^{up,s}, e_t^{do,s} \geq 0, \;\; h \in \mathcal{H} \cup \mathcal{K}, t \in \mathcal{T}, s \in \mathcal{S}.$$

Marginal pricing

$$\begin{aligned} \min \; & \sum_{h \in \mathcal{H}} (c_h^{up} x_h^{up} + c_h^{do} x_h^{do}) + \\ & \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} \pi^s (\rho_t^{up,s} - \rho_t^{do,s}) + \\ & \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}} \pi^s (b_t^{up} e_t^{up,s} + b_t^{do} e_t^{do,s}) \end{aligned} \tag{6.4.2}$$

$$\text{s.t. } (6.3.4) - (6.3.7), (6.3.8)$$

$$x_h^{up}, x_h^{do} \in \{0,1\}, \;\; h \in \mathcal{H}$$

$$y_{ht}^{up,s}, y_{ht}^{do,s}, p_{ht}^{up,s}, p_{ht}^{do,s}, \rho_t^{up,s}, \rho_t^{do,s}, e_t^{up,s}, e_t^{do,s} \geq 0,$$

$$\delta_{ht}^{up,s}, \delta_{ht}^{do,s} \in \{0,1\}, \;\; h \in \mathcal{H} \cup \mathcal{K}, t \in \mathcal{T}, s \in \mathcal{S}.$$

The scenario generation has been kept rather simple. The imbalances between demand and supply constitute a time series and have, thus, been analyzed by means of the field. In order to capture the future behavior of demand and supply imbalances and in particular model the imbalances as a stochastic process, historical data profiles have been used. Because demand and supply show strong correlations over time, the stochastic process is chosen as an autoregressive (AR) process. For proof of concept, the order of the AR process is chosen to be one. The AR process, cf. Box and Jenkins (1976), is the following

$$\Delta_t = \psi \Delta_{t-1} + e_t, \;\; t \in \mathbb{Z}, \tag{6.4.3}$$

where $\{e_t\}_{t \in \mathbb{Z}}$ is a Gaussian white noise process. The time series may be more appropriately described by higher order autoregressive moving average (ARMA) processes, able to describe seasonal variations in the data. For examples on seasonal ARMA models and electricity demand, see Gröwe-Kuska et al. (2000), Gröwe-Kuska et al. (2002) and Eichhorn et al. (2005). Scenarios of future demand and supply imbalances $\{\Delta_t^s\}_{t \in \mathcal{T}, s \in \mathcal{S}}$ are generated by sampling from (6.4.3). To reflect the true probability distribution, a large number of scenarios has been generated using Monte Carlo sampling.

6.5 Solution procedure

The problems (6.4.1) and (6.4.2) may be handled as large-scale mixed-integer linear problems solvable by standard software packages or stochastic mixed-integer linear programs amenable to decomposition approaches such as the integer L-shaped method by Laporte and Louveaux (1993) or the dual decomposition algorithm by Carøe and Schultz (1999). Being able to solve the problems as mixed-integer linear programs is valuable since this approach is very flexible. Adding further linear constraints is uncomplicated and becomes particularly relevant with constraints that introduce time dependencies. However, with the current simplicity of the model, it can be solved by a procedure that utilizes the structure of the problem. The solution procedure is motivated by the current practice of the system operator and applies in the case of both pay-as-bid pricing and marginal pricing.

If formalized, the problems (6.4.1) and (6.4.2) can be stated as

$$\min\Big\{cx + \sum_{s\in\mathcal{S}}\sum_{t\in\mathcal{T}} \pi^s \Phi_t^s(x), x \in \mathbb{B}^{n_1}\Big\},$$

$$\Phi_t^s(x) = \min\{qy | Wy = h_t^s - Tx, y \in \mathbb{R}^{n_2} \times \mathbb{B}^{n_2'}\},$$

where $x \in \mathbb{B}^{n_1}$ represents the first-stage decisions, $y \in \mathbb{R}^{n_2} \times \mathbb{B}^{n_2'}$ represents the second-stage decisions and data vectors and matrices are derived from the problems.

As the number of reserve bids often reduces to less than ten, the set of first-stage decisions is limited and enumeration is possible. The enumeration determines which reserve bids should be activated.

Algorithm 6.5.1 *Enumeration*

Step 0 (Initialization) Let $\bar{z} = \infty$.

Step 1 (Enumeration) Select a first-stage solution, x.

Step 2 (Evaluation) Let

$$\bar{z} = \min\Big\{\bar{z}, cx + \sum_{s\in\mathcal{S}}\sum_{t\in\mathcal{T}} \pi^s \Phi_t^s(x)\Big\},$$

where, for $t \in \mathcal{T}, s \in \mathcal{S}$, $\Phi_t^s(x)$ *is determined by Algorithm 6.5.2. Return to step 1.*

Having fixed the first-stage decisions, the problem decomposes into subproblems according to both scenarios and time intervals. A subproblem is solvable by inspection. After the reserve bids have been activated, the regulating bids are selected one at a time according to the so-called order of merit until imbalances are covered. We state the procedure only for up-regulation as the case of down-regulation will follow in a similar fashion. We assume that the penalty costs of excess supply and demand exceed all bid prices, although the procedure could be easily adjusted to account for other relationships between penalty costs and bid prices.

Algorithm 6.5.2 *Order of merit*

Step 0 (Initialization) If $\Delta_t^s > 0$*, the system must be up-regulated. Let*

$$\mathcal{H}^{up,0} = \{h \in \mathcal{H} : x_h^{up} = 1\}$$

index the activated reserve bids. Available regulation bids are then indexed by $\mathcal{H}^{up,0} \cup \mathcal{K}^0$*. Go to step 1.*

Step 1 (Ranking) Set i=i+1. If $\Delta_t^s > 0$*, activate (fully unless the imbalance is covered by less) the regulation bid* $(\bar{p}_{kt}^{up}, \bar{y}_{kt}^{up,s})$ *defined by*

$$\bar{p}_{kt}^{up} \in \arg\min\{\bar{p}_{ht}^{up} : h \in \mathcal{H}^{up,i-1} \cup \mathcal{K}^{i-1}\}.$$

Set $\mathcal{H}^{up,i} \cup \mathcal{K}^i := \mathcal{H}^{up,i-1} \cup \mathcal{K}^{i-1} \backslash \{k\}$*. If the imbalance is covered, stop. If* $\mathcal{H}^{up,i} \cup \mathcal{K}^i = \emptyset$*, the remaining imbalance is excess demand. Otherwise, return to step 1.*

6.6 Computation results

As already stated, the case study concerns the problem of managing regulating reserves in Western Denmark. The data dates back to June 2006, just before the transition from pay-as-bid pricing to marginal pricing. Hence, we solve the problem with both pay-as-bid pricing (6.4.1) and local marginal pricing (6.4.2). Reserve bids comprise bids of the June auction as well as individual contracts of a potentially longer duration. As the system operator intends to reduce the offer period of reserve bids, we assume that such individual contracts have an offer period of only one month. The reserve bids consist of seven up-regulation bids and one down-regulating bid. The volumes and the fixed prices of the reserve bids are released by Energinet.dk[3]. The variable prices have been randomly generated based on the announced regulating market prices. Regarding regulating bids to the market, ten bids have been constructed. Both volumes and prices have been

[3] www.energinet.dk

randomly generated on the basis of the regulating market prices and the total amounts of up- and down-regulation bid into the market. All data has been provided by Nord Pool[4]. The penalties for excess demand and supply are both set sufficiently high to prevent uncovered imbalances on a regular basis.

With the current data, the problem (6.4.1) contains eight binary variables and no constraints in the first stage and 53280 continuous variables and 13680 constraints in the second stage. The problem (6.4.2) contains the same number of variables and constraints in the first stage and 54720 continuous variables, 12960 binary variables and 39608 constraints in the second stage. The Procedures 6.5.1 and 6.5.2 were implemented in C++ and computations were carried out on an Intel Xeon 2.67 GHz processor with 4 GB RAM.

We have solved the problems (6.4.1) and (6.4.2) with the Procedures 6.5.1 and 6.5.2 and listed the results. For a varying number of scenarios, Table 6.1 displays the average optimal values and CPU times of ten different runs. Obviously, marginal pricing results in a higher optimal value than pay-as-bid pricing. The first column of Table 6.2 shows the total balancing costs divided into reserve costs, regulation costs and penalty costs. Recall that regulation costs consist of up-regulation expenses and down-regulation income and costs may therefore be both positive and negative. The second column of Table 6.2 gives the total imbalances divided into regulation and excess supply and demand along with the reserved regulation that is available but not necessarily activated. Regulation consists of both up- and down-regulation. All numbers are based on 100 scenarios and are averages of ten different runs. It is clear that for both pay-as-bid pricing and marginal pricing, reserves are highly necessary for the optimal covering of imbalances. Finally, Table 6.4 lists the reserve bids and indicate whether activation is effected or not. All ten runs for both pay-as-bid pricing and marginal pricing show the same result and support the use of reserves.

To compare the stochastic programming approach to a deterministic approach, we have solved the expected value problems (EVP), in which stochastic demand and supply imbalances have been replaced by their expected values. Moreover, we have computed the expected results of using the expected value solutions (EEV). For further reference on the expected value problem, see Birge and Louveaux (1997). The average EEVs and CPU times of ten different runs are displayed in Table 6.1. Generally, the EEVs exceed the optimal values of the stochastic programs, which is confirmed. In Table 6.3 the total balancing costs and the total imbalances are divided into reserves, regulation and excess supply and demand. Table 6.4 indicate whether activation of the reserve bids is effected or not. Since imbalances often cancel out on average, no reserve bids are activated in the deterministic case. However, the expected result of using the expected value solutions is a need for a larger amount of direct regulation and if not available, larger excess de-

[4] www.nordpool.no

mand and supply. This is indeed reflected in higher regulation costs, much higher penalty costs and thus higher total costs. The percentual values of the stochastic solutions (VSS), i.e. the percentual saving in costs of using the stochastic solutions rather than the deterministic solutions, are significant. The numbers are in fact in the range of 36-38 percent. Although the VSSs highly depend on the chosen penalty costs, they still demonstrate the relevance of stochastic programming to the managing of regulation reserves.

Table 6.1: Computational results the EVP and the stochastic programming problem. Based on 10 runs.

Sce.	Opt. val. Pay-as-bid pricing	Opt. val. Marginal pricing	CPU/s
100	4.99e+07	5.22e+07	51.37
500	4.94e+07	5.18e+07	257.33
1000	4.94e+07	5.18e+07	512.00
EEV	7.89e+07	8.13e+07	1.04

Table 6.2: Computational results for the stochastic programming problem with 100 scenarios. Based on 10 runs.

		Costs/DKK	Volume/MW
Pay-as-bid pricing	Total	4.99e+07	86060.23
	Reserve	3.46e+06	94320.00
	Regulation	1.32e+06	81545.20
	Excess supply and demand	4.52e+07	4515.03
Marginal pricing	Total	5.22e+07	86060.23
	Reserve	3.46e+06	94320.00
	Regulation	3.68e+06	81545.20
	Excess supply and demand	4.51e+07	4515.03

6.7 Discussion

It could be argued that reserve management affects spot market trading in that purchasing regulating reserves prevents suppliers from disposing of production in the spot market. As the system operator reserves regulation, less production capacity becomes available for the spot market. Still, we have implicitly assumed that the production capacity for the spot market is not seriously affected. The assumption is justified if producers allocate production capacity for the spot market

and the regulating market separately. However, it would be valuable to further investigate the matter by incorporating the present model in a larger model that includes the spot market.

Table 6.3: Computational results for the EVP. Based on 10 runs.

		Costs/DKK	Volume/MW
Pay-as-bid pricing	Total	7.89e+07	86226.71
	Reserve	0e+00	0.00
	Regulation	1.56e+06	78496.00
	Excess supply and demand	7.73e+07	7730.71
Marginal pricing	Total	8.13e+07	86226.71
	Reserve	0e+00	0.00
	Regulation	3.94e+06	78496.00
	Excess supply and demand	7.73e+07	7730.71

Table 6.4: Computational results the EVP and the stochastic programming problem with 100 scenarios. Based on 10 runs.

		Up-reg.			
Price/DKK		288000	625000	384000	330000
Volume/MW		25	16	11	298
Activation	100 sce.	1	1	1	0
	EVP	0	0	0	0
		Up-reg.			Down-reg.
Price/DKK		10298880	714000	966450	152000
Volume/MW		12	21	30	16
Activation	100 sce.	1	1	1	1
	EVP	0	0	0	0

Chapter 7

Scenario generation in stochastic programming electricity models

Being an important part of establishing a stochastic programming model, the remainder of the thesis is dedicated to scenario generation. The idea is to explain the methods used for scenario generation and reduction in Chapters 4 and 5 and put these into perspectives with the existing literature on the subject.

The task of generating the scenarios that serve as input to a stochastic programming model can be handled in many ways, depending on the available information and data and on the required output. In generating scenarios, some of the following requirements could be taken into account. Ideally, the scenarios should

- reflect the actual background and be linked with the purpose of the application.
- approximate the true probability distribution or represent the uncertain data in some other way.
- utilize the information of historical data and at the same time be capable of including exogenous knowledge and extreme situations and capture future effects such as trends and expectations.
- from a computational point of view, result in a problem of a manageable size, that still produces solutions of sufficient quality.

To illustrate different requirements, this chapter gives a selected overview of scenario generation methods from the literature. We present scenario generation methods that are based on experts' opinions or data manipulation alone, methods that aim to represent the probability distribution by matching statistical properties and methods that approximate the probability distribution by sampling from a statistical model. We briefly mention a method suitable for including additional

scenarios. Furthermore, we examine in greater detail methods that reduce the stochastic programming problem to a manageable size, and, finally, we discuss the evaluation of scenario generation methods in terms of solution quality. The main focus is scenario generation methods that have found use in stochastic programming applications to power systems. To complete the overview, however, some major advances in scenario generation will be included. This also serves to justify the choice of the scenario generation methods in Chapter 4 and 5. For a general survey on scenario generation methods, see Dupačová et al. (2000).

Recall from Chapter 1 that the random components of a general stochastic programming problem is described by a random vector ξ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The probability distribution of ξ is given by the image measure $\mu = \mathbb{P} \circ \xi^{-1}$ and is referred to as the probability distribution of the stochastic programming problem. We define ζ as a similar random vector on $(\Omega, \mathcal{F}, \mathbb{P})$ and let the probability distribution of ζ be given by the image measure $\nu = \mathbb{P} \circ \zeta^{-1}$. In the following, we will use ξ and ζ in terms of both two-stage and multi-stage stochastic programming. When specifically addressing only two-stage or multi-stage programs, however, we will adjust notation accordingly. For discrete a approximation with finite support, recall that the approximation of ξ is given by the scenarios $\{\xi^1, \ldots, \xi^S\}$ and corresponding probabilities $\pi^1, \ldots, \pi^S$. When referring to a fixed value such as a single realization or the mean value of ξ, we often denote it $\bar{\xi}$. This chapter aims at approximating the probability distribution of a stochastic programming problem by a discrete distribution with finite support.

7.1 Subjective approaches and data manipulation

Application specific methods for generating scenarios are not necessarily theoretically founded but may rely on subjective judgment, experts' opinion or data manipulation.

An obvious scenario generation method is simple data manipulation. Takriti et al. (1996) present an example that relies alone on historical data but is still distributions-free. The authors generate demand scenarios for a multi-stage stochastic programming formulation of the unit commitment problem subjected to uncertainty of demand and generator failures. The scenarios are based on historical observations under comparable circumstances and are adjusted by incorporating future expectations and demand peaks that corresponds to unavailable generator capacity.

Another method that rests on a rather simple use of historical data is presented by Nowak and Römisch (2000) who develop a multi-stage stochastic programming problem for planning the weekly generation of electric power in a hydro-thermal system under demand uncertainty. Starting from a reference scenario based on real data, a scenario tree is constructed by randomly selecting branching points

and creating identical scenarios that branch from these. Then a Brownian motion is added to each node of the scenario tree.

In contrast to the above, the integration of scenario generation into mathematical programming and, in particular, stochastic programming input formats, has prompted more general scenario generation devices. The stochastic extension of the mathematical programming standard, MPS, the SMPS, can be combined with an external scenario generator. In addition, although relatively simple, the standard itself offers some scenario generation features.

In the initial version of the SMPS, Birge et al. (1987) provide a framework for generating and organizing uncertain data for multi-stage stochastic programming problems. The authors propose an input format that incorporates sampling from a number of univariate distributions by specifying the appropriate parameters. The input format allows for different kinds of dependencies in data, e.g. independence, interstage dependence and dependence on past data.

To overcome some of the shortcomings of the initial version, Gassmann and Schweitzer (2001) have upgraded the SMPS to handle multi-stage stochastic programming problems with chance constraints. Furthermore, the user is now able to choose between a scenario representation or a nodal representation of the data. A multivariate distribution has been added to the built-in distributions of the upgraded version.

To include additional scenarios in a scenario tree, the *contamination method* finds its use. The scenarios may be extreme scenarios or scenarios that reflect exogenous knowledge given by experts. The contaminated probability distribution results from the initial distribution ν and the distribution μ governed by the additional scenarios in the following way

$$(1 - \lambda)\nu + \lambda\mu,$$

where $\lambda \in [0, 1]$. In bounding the optimal value of the multi-stage stochastic programming problem the effect of including additional scenarios can be evaluated. The contamination method does not depend on assumptions on the probability distribution. For further references, see Dupačová et al. (2000).

7.2 Matching statistical properties

Among the scenario generation methods that can be applied in spite of only limited information on the probability distribution, are the methods of matching statistical properties. The overall idea is to generate a discrete approximation such that certain statistical properties of the approximation match statistical properties calculated from data or specified to include future expectations.

The method of Høyland and Wallace (2001) for *matching statistical properties* can handle multivariate distributions in both two-stage and multi-stage stochas-

tic programming. The discrete approximation is found as the solution to a least squares problem that minimizes the distance between the specified properties and the corresponding properties of the approximation. The solution of the least squares problem allows for inconsistencies in the specifications at the expense of a precise matching between the properties. To formulate the least squares problem, let K be the number of properties to match. For $k \in \mathcal{K}$, let w_k be a weight assigned to property k and denote by TAR_k the k'th target property specified. Let the approximate distribution be given by the vectors ξ and π, denote by $TAR_k(\xi, \pi)$ the k'th property of the approximation. Then the problem is

$$\min_{\xi,\pi}\Big\{\sum_{k=1}^{K} w_k(TAR_k(\xi,\pi) - TAR_k)^2 | \pi \text{ is a probability distribution}\Big\}.$$

The properties needed to find a good approximation are highly problem dependent and may be hard to determine. Also, a good match can be difficult to achieve since non-convexity of the problem calls for advanced solution procedures.

In order to match specific statistical properties, Høyland et al. (2003) present a scenario generation method based on *moment matching*. With the purpose of reducing computing times, the least squares problem of Høyland and Wallace (2001) is decomposed by generating univariate distributions separately and combining these to a multivariate distribution. In contrast to the method of Høyland and Wallace (2001), the method of Høyland et al. (2003) relies primarily on transformations.

Starting from an N-dimensional random vector, $\xi = (\xi_1, \ldots, \xi_N)$, the goal is to generate a discrete approximation, whose first four moments and correlations match the specified targets given by the vector TAR and the matrix Σ, respectively. The basics of the matching are the following.

- Construct N discrete random variables.
- Obtain the target moments, TAR, by a cubic transformation.
- Combine the N random variables to an N-dimensional random vector.
- Obtain the target correlations, Σ, by a matrix transformation.

The distributions of N discrete random variables, $\zeta_n, n = 1, \ldots, N$, are generated separately by sampling. Then, to obtain the target moments, the cubic transformation

$$\xi_n = a_n + b_n\zeta_n + c_n\zeta_n^2 + d_n\zeta_n^3$$

is applied before the moments of $\zeta_n, n = 1, \ldots, N$ are matched to the target moments of $\xi_n, n = 1, \ldots, N$ in the same fashion as in Høyland and Wallace (2001). The coefficients a_n, b_n, c_n, d_n are constants. Next, the univariate distributions of

$\xi_n, n = 1, \ldots, N$ are combined to a multivariate distribution of ξ. A specified correlation matrix is achieved by means of the so-called matrix transformation. By applying this transformation, the random vector

$$\zeta = C\xi$$

has correlation matrix Σ, where C is the Cholesky decomposition of Σ. In general, the cubic transformation matches moments but changes correlations, whereas the matrix transformation matches correlations but changes moments. In order for both moments and correlations to almost match the targets, Høyland et al. (2003) propose an algorithm that iterates between random vectors of the correct moments and correct correlations. So far, theoretical convergence of the algorithm is left unproved, but empirical tests are promising.

The moment matching method applies to two-stage stochastic programming. However, although capable of handling correlations, the method is not suitable for preserving time series properties such as correlation over time and the application to multi-stage stochastic programming should be further developed.

The methods of matching statistical properties has been used for stochastic programming applications to power systems in the papers of Fleten et al. (2002) and Fleten and Pettersen (2005), both papers employing moment matching.

7.3 Sampling from statistical models

With sufficient historical data, it may be possible to represent the probability distribution by a statistical model that is suitable for sampling. The modeling of the probability distribution can be done by the use of time series analysis and advanced stochastic processes such as simple and multiple regressions, autoregressive and moving average processes etc. Many examples are found within the application areas of water resources and power systems.

The paper of Jacobs et al. (1995) describes the progress of a Canadian generation scheduling system designed to manage medium-term hydro-power generation in coordination with other energy sources. Special emphasis is placed on the development in modeling the underlying multi-stage stochastic programming problem and the corresponding flow scenarios. It is assumed that different stream-flows are uncorrelated in the medium term due to the Canadian precipitation patterns. Each stream-flow is modeled as a simple *regression* on precipitation. The regression includes future precipitation and thereby provides an opportunity to quickly study the consequences of expected future events. Seasonal flows are disaggregated to monthly flows that are specified in terms of remaining seasonal flows and flows of the previous month.

Due to lack of data, regression may not always be possible and an alternative is *autoregressive and moving average processes*, an example of which is found

Chapter 8. For the use in short-term hydro-power production, both univariate and multivariate distributions of hourly electricity prices and reservoir inflows are modeled. Prices and inflows can be assumed to be uncorrelated in the short term and hence can be modeled separately. For inflows, it is possible to retain the details of the univariate models in the multivariate model by assuming the univariate models are statically related to each other only at the same time. The inflow model that comes to play is a so-called contemporaneous autoregressive moving average process.

In general, with multidimensional distributions, computations quickly become very cumbersome. Another way to overcome this difficulty, is to reduce dimension by *factor analysis* or *principal component analysis*. Both approaches aim to explain the correlation structure of the uncertain data by a small number of independent factors or components. For reference, see for example Infanger (1994).

To arrive at a discrete approximation of the probability distribution, sampling finds its use. The most basic sampling procedure is the *Monte Carlo sampling* in which samples are all assigned the same probabilities. Among many others, the Monte Carlo sampling procedure is applied by Shapiro (2003) who extends the procedure to a conditional sampling procedure that immediately induces the tree structure suitable for multi-stage stochastic programming problems. At the same time, it is shown that conditional sampling provides a valid statistical lower bound for the original stochastic programming problem and that this bound converges if the number of samples tends towards infinity. *Conditional Monte Carlo sampling* is likewise adopted by Shapiro (2006). A more advanced sampling procedure for use in multi-stage stochastic programming is suggested by Pennanen (2005) and employs so-called *integration quadratures*. The procedure applies to a variety of time series models in which the random variables are uniquely determined by uniformly distributed random variables. Among others, such time series models cover autoregressive moving average processes and autoregressive conditionally heteroscedastic processes. Instead of sampling from the original random variables, it is proposed to sample from univariate uniform distributions, combine the univariate sample to a multivariate and, finally, apply the inverse mapping from the uniform to original random variables. The result is a scenario tree that provides a discrete approximation of the distribution. It is shown that as the discretization is made finer, the optimal value of the multi-stage stochastic programming problem converges. The sampling procedure of Pennanen (2005) is further explored in Pennanen and Koivu (2005).

7.4 Tree construction and reduction

The sampling procedures of Section 7.3 all generates scenario paths. Such scenario paths may serve as direct input to two-stage stochastic programming problems

and scenario formulations of multi-stage stochastic programming problems. For tree formulations of multi-stage stochastic programming problems, however, the scenario paths must be converted into scenario trees. As many rely on the same basic principles, we discuss only a few approaches. It should be remarked that for the approaches to work, the scenario paths do not necessarily have to be generated by sampling.

A simple approach to converting paths into trees is to partition the paths according to "high" or "low", "dry", "medium" or "wet" or the like. The root node is assigned all paths and the tree is constructed so that the descendants of a node correspond to a partitioning of the paths assigned to it. The conditional probability of a node is its fraction of paths assigned to the ascending node. For an example of this approach, see Jacobs et al. (1995).

The approach is related to *cluster analysis*. It starts by clustering according to the first components ξ_1 of ξ, continue by conditional clustering according to the second components ξ_2 of ξ and so on. For clustering, a dissimilarity measure such as

$$\sum_{t=1}^{T} w_t \|\xi_t^{s_1} - \xi_t^{s_2}\|$$

is evaluated for each pair of scenario paths $\xi^{s_1} = (\xi_t^{s_1})_{t=1}^T$ and $\xi^{s_2} = (\xi_t^{s_2})_{t=1}^T$, where $w_t \geq 0$ are suitable weights. By clustering according to ξ_1, the result is K_1 clusters $C_1^1, \ldots, C_1^{K_1}$ represented by $\bar{\xi}_1^k, k = \ldots, K_1$. $\bar{\xi}_1^k$ may be the mean value of the scenarios included in cluster k. The probability of $\bar{\xi}_1^k$ is the sum of probabilities of the scenarios included in cluster k. Replacing ξ_1 by $\bar{\xi}_1^k$, the clustering continues for each cluster C_1^k separately. For a survey on cluster analysis from a mathematical programming point of view, see Hansen and Jaumard (1997).

The method we refer to as *optimal scenario reduction* is presented by Heitsch and Römisch (2006b) and is among the approaches founded in probability theory. Starting from a set of scenario paths, forward and backward approaches facilitate the construction and reduction of a scenario tree. Basically, the approaches consist in conditional clustering of scenarios forward or backward in time. However, both approaches are justified by qualitative stability analysis and rely on the following result from Heitsch et al. (2005). We state only the forward approach as the backward is similar in spirit.

For notational convenience, let a general stochastic programming problem be given as

$$\phi(\mu) = \min\{Q(x, \mu) \mid Ax = b, x \in X\},$$

where the dependency of ϕ and Q on μ is emphasized. A stability estimate rests on the L_r-distance

$$\|\zeta\|_r := \big(\sum_{t=1}^{T} \mathbb{E}[\|\zeta_t\|^r]\big)^{1/r} \tag{7.4.1}$$

for some $r \geq 1$, a time dependent variant of this, $\|\cdot\|_{r,t}$, and the filtration distance

$$D(\mu,\nu) := \sup_{\varepsilon\in(0,\beta]} \inf\Big\{ \sum_{t=2}^{T-1} \max\big\{ \|x_t - \mathbb{E}[x_t|\mathcal{F}_t]\|_{r'}, \|x'_t - \mathbb{E}[x'_t\|\mathcal{G}_t]\|_{r'} \big\} : \\ x : Q(\mu,x) \leq \phi(\mu)+\varepsilon, x' : Q(\nu,x') \leq \phi(\nu)+\varepsilon \Big\}, \tag{7.4.2}$$

where r' is problem dependent and where $\mathcal{F}_t$ and $\mathcal{G}_t$ are the σ-algebras generated by ξ^t and ζ^t.

Theorem 7.4.1 *Assume $\{x \in X \mid Ax = b\}$ is bounded, relatively complete recourse locally around ξ, levelboundness locally uniformly at ξ and $\xi \in L_r(\Omega,\mathcal{F},\mathbb{P};\mathbb{R}^{\sum_{t=1}^{T} N_t})$ for some $r \geq 1$. Then there exists positive constants L, α and β such that the estimate*

$$|\phi(\mu) - \phi(\nu)| \leq L(\|\xi - \zeta\|_r + D(\mu,\nu))$$

holds for all $\zeta \in L_r(\Omega,\mathcal{F},\mathbb{P};\mathbb{R}^{\sum_{t=1}^{T} N_t})$ with $\|\xi - \zeta\| \leq \alpha$.

Based on the set of scenario paths ζ, the approach seeks to construct and reduce the scenario tree ξ and at the same time bound the distances (7.4.1) and (7.4.2). In this fashion, the optimal values $\phi(\mu)$ and $\phi(\nu)$ of the multi-stage stochastic programs are kept sufficiently close.

The conditional clustering forward or backward in time make use of certain scenario reduction heuristics referred to as forward selection and backward reduction. As for the clustering, we confine ourselves to the forward approach. Foremost, the scenario reduction calls for a scenario redistribution rule, an optimal reduction problem and a maximal reduction strategy. To state this, let $\mathcal{S}_t^1$ and $\mathcal{S}_t^2$ index the selected and deleted scenarios at stage t. Given $\mathcal{S}_t^1$ and $\mathcal{S}_t^2$, the optimal redistribution is determined by

$$\min\Big\{ l_r(\{\xi^s\}_{s\in\mathcal{S}_{t-1}^1}, \{\xi^s\}_{s\in\mathcal{S}_t^1}) \Big| \sum_{s\in\mathcal{S}_t^1} \pi^{t,s} = 1, \pi^{t,s} \geq 0, s \in \mathcal{S}_t^1 \Big\} = \\ \Big(\sum_{s_2\in\mathcal{S}_t^2} \pi^{t-1,s_2} \min_{s_1\in\mathcal{S}_t^1} \|\xi^{s_1} - \xi^{s_2}\|^r \Big)^{\frac{1}{r}},$$

where l_r is the L_r-minimal metric, also referred to as the Wasserstein distance of order r, cf. Heitsch and Römisch (2006b). The solution provides the redistribution rule of probabilities

$$\pi^{t,s} = \pi^{t-1,s} + \sum_{s_2 \in \mathcal{S}_t^2(s)} \pi^{t-1,s_2}, \;\; s \in \mathcal{S}_t^1,$$

where

$$\mathcal{S}_t^2(s) := \big\{ s_2 \in \mathcal{S}_t^2 | s = s_t(s_2) \big\}, \;\; s_t(s_2) \in \arg\min_{s_1 \in \mathcal{S}_t^1} \|\xi^{s_1} - \xi^{s_2}\|^r.$$

In other words, the probability of a selected scenario is the sum of its former probability and of all probabilities of deleted scenarios that are closest in terms of the L_r-minimal metric. The optimal set $\mathcal{S}_t^2$ for scenario reduction with fixed cardinality $|\mathcal{S}_t^2| = K$ is given by the solution of the optimal reduction problem

$$\min\Big\{ \sum_{s_2 \in \mathcal{S}_t^2} \pi^{t-1,s_2} \min_{s_1 \in \mathcal{S}_t^1} \|\xi^{s_1} - \xi^{s_2}\|^r \Big| \mathcal{S}_t^2 \subset \mathcal{S}_{t-1}^1, |\mathcal{S}_t^2| = K \Big\}.$$

Approximate solutions can be found by the following heuristic, iteratively selecting a scenario at a time.

Algorithm 7.4.1 *(Forward selection)*

Step 0 Let $\mathcal{S}_t^{1,0} := \emptyset$.

Step 1 Set $i = i + 1$ *and let*

$$s^i \in \arg\min_{s_3 \in \mathcal{S}^{2,t-1}} \sum_{s_2 \in \mathcal{S}_t^{2,t-1} \setminus \{s_3\}} \pi^{t-1,s_2} \min_{s_1 \in \mathcal{S}_t^{1,t-1} \cup \{s_3\}} \|\xi^{s_1} - \xi^{s_2}\|^r.$$

$$\mathcal{S}_t^{1,t} := \mathcal{S}_t^{1,t-1} \cup \{s^i\}$$

If $i < K - k + 1$, *return to step 1. Otherwise, go to step 2.*

Step 2 Do optimal redistribution.

The maximal reduction strategy serves to determine the set $\mathcal{S}_t^2$ with maximal cardinality $|\mathcal{S}_t^2|$ such that

$$\sum_{s_2 \in \mathcal{S}_t^2} \pi^{t-1,s} \min_{s_1 \in \mathcal{S}_t^1} \|\xi^{s_1} - \xi^{s_2}\|^r = \|\xi^t - \xi^{t-1}\|_r^r \leq \varepsilon_t.$$

where

$$\xi^{t,s} := \begin{cases} \xi^{t-1,s} & , s \in \mathcal{S}_t^1 \\ \xi^{t-1,s_t(s)} & , s \in \mathcal{S}_t^2. \end{cases}$$

and ε_t is a given tolerance.

In stating the forward scenario generation approach, let $r \geq 1$ and let $\varepsilon_t, t = 2 \ldots, T$ be given. The forward tree construction algorithm is then

Algorithm 7.4.2 *(Forward scenario generation)*

Step 0 *Let $\mathcal{C}_1 = \{\{1, \ldots, S\}\}$ be the 1st stage partitioning of $\{1, \ldots, S\}$ and let $\xi^1 := \zeta$ be the 1st stage scenario tree. Set $t = 0$.*

Step 1 *Set $t = t + 1$. Let $\mathcal{C}_t = \{C_t^1, \ldots, C_t^{K_t}\}$ be the stage-t partitioning of $\{1, \ldots, S\}$. Determine the disjoint index sets $\mathcal{S}_{t+1}^{1,k}$ and $\mathcal{S}_{t+1}^{2,k}$ of stage t+1 with $\mathcal{S}_{t+1}^{1,k} \cup \mathcal{S}_{t+1}^{2,k} = C_t^k$ and the stage t+1 scenario tree ξ^{t+1} with scenarios*

$$\xi_{t_1}^{t+1,s} = \begin{cases} \zeta_{t_1}^{\beta_{t_1}(s)} & , t_1 \leq t+1 \\ \zeta_{t_1}^{s} & , t_1 > t+1, \end{cases}$$

where

$$\beta_t(s) = \begin{cases} s_t^k(s) & , s \in \mathcal{S}_t^{2,k}, k = 1, \ldots, K_{t-1} \\ s & , \text{otherwise}, \end{cases}$$

by applying the maximal reduction strategy to stage t such that $\|\xi^{t+1} - \xi^t\|_{r,t+1} \leq \varepsilon_t$. Set $\mathcal{C}_{t+1} = \{\beta_{t+1}^{-1}(s_1) \mid s_1 \in \mathcal{S}_{t+1}^{1,k}, k = 1, \ldots, K_t\}$. If $t < T$, return to step 1. Otherwise, go to step 2.

Step 2 *Let $\mathcal{C}_T = \{C_T^1, \ldots, C_T^{K_T}\}$ be the stage-T partitioning of $\{1, \ldots, S\}$. Construct a scenario tree ξ having K_T scenarios such that $\xi_t^k := \zeta_t^{\beta^t(s)}$ for $s \in C_T^k, k =$*
$1, \ldots, K_T, t = 1, \ldots, T$.

It is shown that if ξ is constructed according to Algorithm 7.4.2, then we have the bound

$$\|\xi - \zeta\|_r \leq \sum_{t=2}^{T} \|\xi^t - \xi^{t-1}\|_{r,t}^r \leq \sum_{t=2}^{T} \varepsilon_t.$$

As the L_r-distance can be bounded by including conditions into Algorithm 7.4.2, so can the filtration distance and we are thereby able to control the precision of process.

The optimal scenario reduction has been introduced to the GAMS Distribution 20.6 and is contained in the library SCENRED.GAMS/SCENRED, cf. GAMS Software GmbH (2002). To test the approach, Gröwe-Kuska et al. (2003) consider a hydro-thermal generation system of a German utility. In one case study, scenario

paths are given by historical data of electrical loads and prices and in another by electrical loads sampled from a statistical model. Both case studies show encouraging computational result.

The forward scenario generation method using the forward selection approach is applied in Chapter 5. The scenario paths are samples from a stochastic process describing the future development in hourly electricity prices and reservoir inflows. The resulting scenario tree serves as input to a multi-stage stochastic programming problem for the short-term production planning of a Nordic hydro-power plant.

As already stated, the optimal scenario reduction integrates the scenario tree reduction within the construction. However, the method can be used for scenario reduction alone and then applies to both two-stage and multi-stage stochastic programming problems. This was already suggested by Dupačová et al. (2003) and Heitsch and Römisch (2006a), who proposed a variant of Algorithm 7.4.1 that rests on forward selection of single scenarios. An application to two-stage stochastic programming can be found in Chapter 4. The scenarios are samples from a statistical model of hourly electricity prices, while the corresponding two-stage stochastic programming problem determines the commitment of a Nordic hydro-power producer bidding into the electricity market. For another example on forward selection of single scenarios in two-stage stochastic programming and power management, see Gröwe-Kuska et al. (2000).

The approach by Heitsch and Römisch (2006b) is closely related to the *optimal discretization* approach by Pflug (2004). Both approaches aim at generating a scenario tree that is optimal in the sense that it minimizes the distance between the optimal values corresponding to the true and the approximate distributions, respectively. Moreover, the distances used by Heitsch and Römisch (2006b) and Pflug (2004) are both deriven from the same so-called Wasserstein distance.

Since the sampling procedures of Section 7.3 as well as the scenario tree construction and reduction methods are all used to generate stochastic programming input before solving the problem, the methods are classified as external. As a contrast to the external sampling procedures, the next section discusses internal sampling procedures.

7.5 Internal sampling

When using the scenario generation methods that approximate the probability distribution by external sampling, it is implicitly assumed that the distribution is known and the problem is solved with an adaption of a deterministic solution procedure. In contrast, scenario generation methods based on internal sampling are integrated into a more general framework that iteratively refines an approximation, possibly by incorporating sampling into the solution procedure. The components of the iterative framework are

- Scenario generation by sampling.
- Formulation of a stochastic programming problem that takes the scenarios as input.
- Optimization, potentially interfaced with sampling.

The *stochastic quasi-gradient method* by Ermoliev and Gaivoronski (1992) represents one way of incorporating sampling into a solution method. The algorithm iteratively determines new solutions to a stochastic programming problem by moving in the direction of the gradient, which is estimated by sampling a few points or potentially only one point in each iteration.

Partly based on this, Higle and Sen (1991) suggested another algorithmic interface between optimization and sampling, referred to as *stochastic decomposition*. The cutting plane algorithm applies to two-stage stochastic linear programs with complete recourse and combines the principles of the deterministic decomposition procedures such as Benders' decomposition method with those of stochastic approximation procedures such as quasi-gradient methods. Inspired by Benders' decomposition, the algorithm successively generates cuts that are statistical estimates of supports of the objective function. Like quasi-gradient methods, each iteration of the algorithm requires the solution of a subproblem at only one sample point. Convergence is established in the sense that certain subsequences of estimated supports accumulate at the support of the original objective function. Conditional stochastic decomposition by Higle et al. (1994) extends stochastic decomposition to a multi-cut version, in which the cuts support the conditional expected objective function terms.

A very similar way of interfacing Benders' decomposition with sampling is presented by Infanger (2005). Instead of using a single sample point in each iteration of the algorithm, a limited number of sample points is used in an importance sampling based heuristic. The method is applied to an expansion planning problem of a network of electric utilities with uncertainty in the availability of transmission lines and generators.

Like stochastic decomposition is interfaced with Benders' decomposition for two-stage stochastic programming problems, *sequential importance sampling* can be combined with nested Benders' decomposition for multi-stage stochastic programming problems.

In its basic form, the sequential importance sampling algorithm serves to construct scenario trees. The initial tree structure is defined through a scenario tree nodal matrix. As the algorithm progresses one stage at a time, scenario paths are generated by a conditional resampling that is consistent with the scenario tree nodal matrix. The resamplings are usually conducted by conditional Monte Carlo sampling from a statistical model as described in Section 7.3. Given the resulting scenario tree, the stochastic programming problem is formulated and solved.

Then, nodal values of an importance sampling criterion are evaluated at the current stage. A new tree structure is defined through an update of the scenario tree nodal matrix such that the new scenario tree will be enriched with the relevant scenario paths at this stage according to the importance sampling criterion.

Dempster (2004) employs expected value of perfect information EVPI as importance sample criterion. The stage t EVPI is defined as

$$\psi_t(x^{t-1}, \xi^t) := \bar{\Phi}_t(x^{t-1}, \xi^t) - \Phi_t(x^{t-1}, \xi^t),$$

where

$$\Phi_t(x^{t-1}, \xi^t) = \min_{x_t} \mathbb{E}\big[c_t x_t + \Phi_{t+1}(x^t, \xi^{t+1}) | \mathcal{F}_t\big], \tag{7.5.1}$$

$$\bar{\Phi}_t(x^{t-1}, \xi^t) = \mathbb{E}\big[\min_{x_t} c_t x_t + \Phi_{t+1}(x^t, \xi^{t+1}) | \mathcal{F}_t\big] \tag{7.5.2}$$

are the optimal values of the stage t stochastic programming problem and the expected value of the associated wait-and-see problem, cf. Birge et al. (1987). The nodal values of EVPI are evaluated when solving the discrete approximations (7.5.1) and (7.5.2). Using resampling, each descendant of a node with a small or a large EVPI is replaced by one new or a number of new descendants, respectively. Since the EVPI measures the value of future information, a small EVPI indicates that future scenario information cannot be effectively utilized and that a deterministic future will suffice. In the process of importance sampling, both solutions and objective function values are shown to converge to those of the original problem when the number of descendants tends towards infinity. For a small sample size, the error in objective function values can be estimated, provided an upper bound on the original problem is at hand.

A parallel version of the EVPI-based sequential importance sampling algorithm is developed by Dempster and Thompson (1999). The parallel version combines parallel nested Benders' decomposition and a parallel approach to evaluating nodal EVPI.

From test runs, the EVPI-based sequential importance sampling is shown to require quite long computing times. Some success has been obtained with the parallel version, although the authors admit that further progress is still possible.

Whereas an external scenario generation method can be used with an explicit scenario reduction method to limit the number of scenarios before solving a problem, the internal scenario generation method of importance sampling can be viewed as an implicit attempt to reduce the potentially very large dimensionality of stochastic programming problems by only including scenarios that are important in the process of solving.

7.6 Evaluating scenario generation methods

For computational purposes, the true probability distribution is mostly approximated by a finite number of scenarios and the quality of the approximation is therefore determined by the quality of the scenarios. To assess the suitability of a scenario generation method, Kaut and Wallace (2003) suggest some quality requirements that should be fulfilled and show how to test the requirements.

The definition of a "good" approximation is to be understood in various ways. In a statistical sense, the discrete distribution should converge to the true distribution as the number of realizations tends towards infinity. Due to computational tractability, however, convergence may be hard to achieve. From an optimization point of view, the approximation should be evaluated by the quality of the solutions and the objective function values. Continuity of the solution set mapping and the optimal value function address such stability issues in a theoretical fashion. Still, to facilitate computations, this often involves empirical measures that converge as the number of observations tends towards infinity. From a practical viewpoint, a "good" approximation is given by a limited number of representative scenarios.

We consider a given stochastic programming problem, its true probability distributions denoted by μ and an approximation denoted by ν. Using the true or approximate distributions, we refer to the corresponding problems as the true or approximate problems. To evaluate the quality of the solutions, define the following error in the true objective function value at the optimal solutions of the approximate and true problems

$$e(\mu,\nu) := Q(\arg\min_x Q(x,\nu),\mu) - Q(\arg\min_x Q(x,\mu),\mu). \tag{7.6.1}$$

The optimal solutions are not compared directly. The reason for comparing the solutions by their objective function values is that the objective function is often flat and different solutions have very similar values. The quality requirements are based on the idea of a small error, (7.6.1).

We can compare the practical stability requirements to theoretical continuity results from the literature. According to Pflug (2004), we obtain the following continuity estimate with respect to the Fortet-Mourier metric of order r. If the functions $Q(x,\cdot)$ are Lipschitz continuous with a constant of order r uniformly in x, we get the estimate

$$e(\mu,\nu) \leq L \sup_x |Q(x,\mu) - Q(x,\nu)| \leq L_r d_r(\mu,\nu)$$

for some constants L, L_r and where d_r denotes the Fortet-Mourier metric or order r. The continuity estimates provides an upper bound on (7.6.1).

In view of this, the optimal scenario reduction of Section 7.4 seeks to minimize to an upper bound on the error (7.6.1) in that

$$e(\mu,\nu) \leq L \sup_x |Q(x,\mu) - Q(x,\nu)| \leq L_1 l_1(\mu,\nu) \leq L_1 \|\xi - \zeta\|_1$$

for some constants L, L_1. The inequalities are valid due to the equivalence of the Wasserstein l_1 and Fortet-Mourier d_1 metrics of first order. The optimal scenario reduction can therefore be regarded as an attempt to minimize the error (7.6.1). The upper bound, nevertheless, is often quite loose for problems of manageable sizes. Furthermore, in minimizing the upper bound, the link between the optimization problem and the scenario generation method is lost and the practical evaluation of the scenario generation method maintains its relevance.

Likewise, the convergence of scenario generation methods such as the conditional Monte Carlo sampling and the sampling via integration quadratures in Section 7.3 does not preclude the need for testing the practical performance of a number of representative scenarios.

The suggested quality requirements of a scenario generation method are *in-sample stability* and *out-of-sample stability*. In-sample and out-of-sample stability require that the discrete approximations $\nu_k, k = 1, \ldots, K$ of a given distribution μ produce optimal solutions with the same objective function values of the approximate and true problems, respectively. Therefore, denote by $x_k^*, k = 1, \ldots, K$ the optimal solutions of the discrete approximations. Then in-sample stability can be stated as

$$Q(x_{k_1}^*, \nu_{k_1}) \approx Q(x_{k_2}^*, \nu_{k_2}), \quad k_1, k_2 = 1, \ldots, K$$

and out-of-sample stability as

$$Q(x_{k_1}^*, \mu) \approx Q(x_{k_2}^*, \mu), \quad k_1, k_2 = 1, \ldots, K.$$

To test in-sample stability, the approximate problems have to be solved. To test out-of-sample stability, the true objective function should be evaluated at a given solution, which requires full knowledge of the true distribution. If the true distribution is known, simulation is a possibility. Otherwise, another scenario generation method can provide a reference distribution representing the true distribution or if historical data is available, back-testing may be appropriate. It should be remarked that in-sample stability does not imply out-of-sample stability or vice versa and testing both types is recommended. In addition to being stable, a scenario generation method should produce *unbiased solutions* in the sense that the solution of the approximate problem should be a solution to the true problem. To test this, denote by x_ν^* and x_μ^* the optimal solutions of the approximate and the true problem. Then the requirement of unbiased solutions reads

$$Q(x_\nu^*, \mu) \approx Q(x_\mu^*, \mu).$$

Computing the optimal value of the true problem can be done by the use of a reference distribution from an unbiased scenario generation method or by utilizing all available historical data.

The way to improve performance of an in-stable or biased scenario generation method depend on the type of scenario generation method employed. For sampling methods, the number of scenarios should be increased or the sampling method itself should be improved. For methods of matching statistical properties, an option is to increase the number of statistical properties to match or the number of scenarios, although the latter is not guaranteed to work.

Examples of testing practical performance of the scenario generation methods are presented in Chapters 4 and 5. In Chapter 4 electricity price scenarios have been generated by sampling from a statistical model. Tests for in-sample and out-of-sample stability of the solutions are found to be satisfiable for a sample of limited size. Likewise, in Chapter 5, scenario paths that consist of electricity prices and reservoir inflows have been generated by sampling and converted into a scenario tree by optimal scenario reduction. Again, tests for stability do well even for limited sample sizes.

Chapter 8

Uncertainty modeling for the short-term management of hydro-power systems

As the selected overview of Chapter 7 indicates, there are several methods for generating scenarios in stochastic programming. With extensive historical data records, one possibility is to represent the probability distribution of the uncertain data by a statistical model suitable for sampling. This method is especially useful for handling uncertain data that develops over time by means of time series analysis. The sample paths can be used directly as scenarios or scenario paths in two-stage and multi-stage stochastic programming or converted into the scenario trees for use in multi-stage stochastic programming. Moreover, the method is found as a part of both external and internal scenario generation methods, as was explained in Chapter 7.

This chapter aims to determine time series models that are relevant to the short-term management of hydro-power systems and further illustrates the abilities of the models to capture the development in the uncertain data over time. Results from the Nordic power exchange Nord Pool and a Norwegian power plant are presented.

8.1 Introduction

In the short-term management of a hydro-power plant, the uncertainty of the future surroundings is a major challenge. In many respects, uncertainty of the inflows to the reservoirs of the plant is essential. Moreover, uncertainty with respect to the electricity demand was of vital importance in the traditional setting. Demand uncertainty, however, has become less prominent with the deregulation

of electricity markets, as a plant is no longer obligated to satisfy demand, but can resort to market exchange. The exchange through short-term markets has called for profit maximization, and market price uncertainty has become highly relevant.

Reservoir inflow uncertainty mainly stems from non-anticipated precipitation and melt of snow. Market price uncertainty is driven by demand and supply. Demand uncertainty is mostly caused by temperature unpredictability and unpredicted customer behavior, whereas supply uncertainty may be due to unexpected failures. Nevertheless, a common feature of inflows and prices is that current observations show strong dependencies on past observations, and therefore the stochastic processes of inflows and prices can be handled by means of time series analysis. The time series analysis serves to gain insight into the empirical time series, to model the underlying stochastic processes and the development of data over time in particular, and to understand future data values.

For modeling uncertainty in electricity prices and water stream-flows, there are several frameworks within the ares of engineering, economics and statistics. Engineering approaches include bottom-up models for power systems and neural networks for the modeling of hydrological processes. Statistical methods most notably embrace time series analysis, possibly combined with other statistical tools. Finally, econometric approaches to analyze time series are mostly developed to analyze top-down models such as aggregated models for electricity prices.

One of the most common time series models is the ARMA model and its variants; for instance the integrated ARMA model, the fractional integrated ARMA model, the seasonal ARMA model, the vector ARMA model, the transfer model and the contemporaneous model, the ARCH and the GARCH models, some of which will be discussed in this chapter.

If stochastic processes are appropriately represented by time series models, the models can be used to forecast and simulate future data values. The short-term management of hydro-power systems may be based on forecasts and simulations directly or the procedures may be used as input to other instruments such as optimization tools. Often, optimization tools are based on mathematical programming and take the form of large-scale deterministic or stochastic programs. For examples of time series models in optimization, see Eichhorn et al. (2005), Gröwe-Kuska et al. (2000), Gröwe-Kuska et al. (2001), Gröwe-Kuska et al. (2002) and Chapters 4 and 5.

The chapter is organized as follows. In Section 8.2 an ARMA model and some extensions are introduced. The ARMA time series analysis is applied to day-ahead electricity prices and reservoir inflows by first analyzing the univariate time series in Subsections 8.2.1 and 8.2.2, and secondly the multivariate time series in Section 8.3 and Subsection 8.3.1.

8.2 Univariate ARMA modeling

An empirical time series consists of observations made at equidistant time points, so that an observation is dependent on the past development of the process and therefore on past observations. *Autoregressive moving average* (ARMA) processes that were introduced by Box and Jenkins (1976) constitute a class of stochastic processes for the purpose of analyzing such time series. An ARMA process consists of two parts; an *autoregressive* (AR) part and a *moving average* (MA) part, that, loosely speaking, describe the deterministic and stochastic components of the process, as it evolves over time.

To formalize the concept of ARMA processes, note that a weakly stationary process is defined as a process of random variables with constant mean value and autocovariances that are invariant under translation in time. Moreover, a white noise process is defined as a process of uncorrelated random variables with mean values zero and constant variances. We will refer to these random variables as the innovations. The, an ARMA process can be defined as follows, cf. Andersen and Blæsild (2003a) and Andersen and Blæsild (2003b).

Definition 8.2.1 *Let $\{e_t\}_{t\in\mathbb{Z}}$ be a white noise process with variance σ^2 and let $\{\xi_t\}_{t\in\mathbb{Z}}$ be a weakly stationary stochastic process that solves*

$$\psi(B)\xi_t = \gamma(B)e_t, \;\; t \in \mathbb{Z},$$

where $\psi(z) = 1 - \sum_{k=1}^{p} \psi_k z^k$ and $\gamma(z) = 1 - \sum_{k=1}^{q} \gamma_k z^k$ are polynomials with $p, q \in \mathbb{N}_0$, $\psi(z) \neq 0$ and $\gamma(z) \neq 0$ for $|z| \leq 1$, $\psi(z)$ and $\gamma(z)$ have no common roots, and B is the back-shift operator, i.e. $B^k\xi_t = \xi_{t-k}$. Then $\{\xi_t\}_{t\in\mathbb{Z}}$ is called an ARMA process of order (p, q) with parameters $\sigma^2, \psi_1, \ldots, \psi_p, \gamma_1, \ldots, \gamma_q$.

An extension of an ARMA process that is not a weakly stationary process is an *integrated* ARMA process (ARIMA) defined as below.

Definition 8.2.2 *Let $\{\xi_t\}_{t\in\mathbb{Z}}$ be a stochastic process, $d \in \mathbb{N}_0$, $\Delta = 1 - B$ and $\{\Delta^d\xi_t\}_{t\in\mathbb{Z}}$ an ARMA process of order (p, q). Then $\{\xi_t\}_{t\in\mathbb{Z}}$ is called an ARIMA process of order (p, d, q).*

A *seasonal* ARMA process (SARMA) is obtained as a special case of an ARMA process that exhibits periodic behavior.

Definition 8.2.3 *Let $\{e_t\}_{t\in\mathbb{Z}}$ be a white noise process with variance σ^2 and let $\{\xi_t\}_{t\in\mathbb{Z}}$ be a weakly stationary stochastic process that solves*

$$\psi(B)\Psi(B^s)\xi_t = \gamma(B)\Gamma(B^s)e_t, \;\; t \in \mathbb{Z},$$

where $\psi(z) = 1 - \sum_{k=1}^{p} \psi_k z^k, \gamma(z) = 1 - \sum_{k=1}^{q} \gamma_k z^k, \Psi(z) = 1 - \sum_{k=1}^{P} \Psi_k z^k$ and $\Gamma(z) = 1 - \sum_{k=1}^{Q} \Gamma_k z^k$ are polynomials with $p, q, P, Q, s \in \mathbb{N}_0$, $\psi(z) \neq 0, \Psi(z) \neq$

$0, \gamma(z) \neq 0$ *and* $\Gamma(z) \neq 0$ *for* $|z| \leq 1$, $\psi(z)$ *and* $\gamma(z)$ *have no common roots,* $\Psi(z)$ *and* $\Gamma(z)$ *have no common roots, and* B *is the back-shift operator, i.e.* $B^k \xi_t = \xi_{t-k}$. *Then* $\{\xi_t\}_{t\in\mathbb{Z}}$ *is called a SARMA process of order* (p, q, P, Q, s) *with parameters* $\sigma^2, \psi_1, \ldots, \psi_p, \Psi_1, \ldots, \Psi_P, \gamma_1, \ldots, \gamma_q, \Gamma_1, \ldots, \Gamma_Q$.

Regarding lack of stationarity, a similar extension of a SARMA process to an *integrated* SARMA process (SARIMA) applies.

Definition 8.2.4 *Let* $\{\xi_t\}_{t\in\mathbb{Z}}$ *be a stochastic process,* $d, D \in \mathbb{N}_0$, $\Delta = 1 - B, \Delta_s = 1 - B^s$ *and* $\{\Delta^d \Delta_s^D \xi_t\}_{t\in\mathbb{Z}}$ *a SARMA process of order* (p, q, P, Q, s). *Then* $\{\xi_t\}_{t\in\mathbb{Z}}$ *is called a SARIMA process of order* (p, d, q, P, D, Q, s).

It is straightforward to generalize the SARMA and SARIMA processes to a larger class of seasonal processes by taking in more than one season. In the following we assume that the involved white noise processes are Gaussian.

Having observed the process $\{\xi_t\}_{t\in\mathbb{Z}}$ until time t, $\bar{\xi}_{t'}, t' \leq t$, it is interesting to predict or forecast future values of the process, $\xi_{t+l}, l = 1, 2, 3, \ldots$. The predicted value of ξ_{t+l} from time t is denoted $\bar{\xi}_t(l)$, where l is referred to as the lead time. In a stationary process such as an ARMA or a SARMA process, the forecasting relies on a mean square error approach that consists in minimizing the squared difference between ξ_{t+l} and $\bar{\xi}_t(l)$ in the conditional distribution given $\bar{\xi}_{t'}, t' \leq t$, i.e.

$$\mathbb{E}[(\xi_{t+l} - \bar{\xi}_t(l))^2 | \{\bar{\xi}_{t'}\}_{t'\leq t}].$$

The minimum is attained for the so-called minimum mean square predictor

$$\bar{\xi}_t(l) = \mathbb{E}[\xi_{t+l} | \{\bar{\xi}_{t'}\}_{t'\leq t}].$$

If $\{\xi_t\}_{t\in\mathbb{Z}}$ is an $ARMA(p, q)$ process, and $\{\zeta_t\}_{t\in\mathbb{Z}}$ is an $ARIMA(p, d, q)$ process, the forecast is determined by

$$\bar{\xi}_t(l) = \Delta^d \bar{\zeta}_t(l)$$

and likewise, if $\{\xi_t\}_{t\in\mathbb{Z}}$ is a $SARMA(p, q, P, Q, s)$ process and $\{\zeta_t\}_{t\in\mathbb{Z}}$ is a $SARIMA(p, d, q, P, D, Q, s)$ process, then

$$\bar{\xi}_t(l) = \Delta^d \Delta_s^D \bar{\zeta}_t(l).$$

In an ARMA process $\{\xi_t\}_{t\in\mathbb{Z}}$ observed until time t, $\bar{\xi}_{t'}, t' \leq t$, the simulating of future values $\xi_{t+l}, l = 1, 2, 3, \ldots$ rests on drawing samples from the conditional distributions given $\bar{\xi}_{\bar{t}}, \bar{t} \leq t$. In particular, a simulated value of ξ_{t+l} from time t is given by a sample from the distribution of the random variable denoted by $\xi_t(l)$ and determined by

$$\xi_t(l) \sim \sum_{k=1}^{p} \psi_k B^k \xi_{t+l} + e_{t+l} + \sum_{k=1}^{p} \gamma_k B^k e_{t+l} \Big| \{\bar{\xi}_{t'}\}_{t'\leq t},$$

where the innovations are independent random variables since they constitute a Gaussian white noise process. Simulating future values of an ARIMA, a SARMA or a SARIMA process is similar in spirit to the way of forecasting.

A standard application of the ARMA time series analysis comprises the general steps of model identification, parameter estimation, model control along with in-sample and out-of-sample validation of the model.

Model identification: Identifying the class of models starts with detecting possible seasonalities. If periodic behavior is expected, the period is typically determined by the way the data has been collected and is therefore closely related to the design of the experiment. In this respect, either the ARMA or SARMA model comes into play.

Under the assumptions made, both ARMA and SARMA processes are completely determined by their mean value and autocovariances. Further model identification may therefore be based on the study of the empirical autocorrelation function and the partial autocorrelation function, where the latter is a function of the former.

The next step is to consider stationarity issues. A weakly stationary process such as an ARMA or a SARMA process is characterized by a constant mean value and an autocovariance function invariant under translation in time. Moreover, the empirical autocorrelation function and partial autocorrelation function decay exponentially to zero, whereas a slow decay to zero is a sign of non-stationarity. If the original empirical time series does not appear to be stationary, the property may apply to the differences. The original series then constitutes an ARIMA or a SARIMA model.

When the class of models has been identified, the order of the corresponding ARMA or SARMA process must be determined. For an ARMA process of order (p, q), the following is valid and can be used as a guideline. If $p \neq 0$, the autocorrelation function is a sum of exponentially decreasing terms and exponentially damped sine waves, whereas if $p = 0$ and the MA process is of order q, the lagged-k autocorrelations are zero for $k > q$. In the same fashion, if $q \neq 0$, the partial autocorrelation function is nearly a sum of exponentially decreasing terms and exponentially damped sine waves, whereas if $q = 0$ and the AR process is of order p, the lagged-k partial autocorrelations are zero for $k > p$. Similar results hold for a SARMA process of order (p, q, P, Q, s) where $p + sP$ replaces p and $q + sQ$ replaces q. Furthermore, for a SARMA process, the autocorrelations corresponding to the lags $1, 2, 3, \ldots$ behave like an ARMA process of order (p, q) and the autocorrelations corresponding to the lags $1s, 2s, 3s, \ldots$ behave like an ARMA process of order (P, Q).

Parameter estimation: With an appropriate model at hand, the parameters have to be estimated. By accepting that the model is stationary, the most common estimating methods include maximum likelihood estimation, in which the log likelihood function is maximized with respect to the parameters. Other

common methods are the unconditional least square estimation, in which the sum of squares is minimized, and the conditional least square estimation, in which the sum of squares is also minimized, but under the assumption that unknown past values equal zero.

Model control: It remains to be shown that the model is appropriate for describing the empirical time series and the fit of the model has to be validated. The validation is based on the residuals, i.e. the differences between actual observations and predicted values from the fitted model. For the residuals to follow a white noise process, they must have zero mean value, constant variance and be uncorrelated. The inspection of their autocorrelation and partial autocorrelation functions is constructive in the sense that it indicates how an initial model should be altered. Furthermore, the fit of the model can be assessed using the so-called Ljung-Box statistics that offer a test of the residuals being uncorrelated.

In-sample and out-of-sample testing: The final model can be used to forecast or simulate future values of the data. In particular, the ability of the model to forecast can be evaluated by comparing a forecast with real data of the identification period, referred to as in-sample testing, or with real data of a validation period, referred to as out-of-sample testing. If compared, out-of-sample and in-sample forecast errors should be similar. The simulation abilities of the model can be tested by comparing distributional characteristics of a large sample to those of the real data in the identification or validation period.

Concerning the fit of ARMA time series models to day-ahead market prices and reservoir inflows, we rely on the framework presented above and explain each step in detail. All computations have been carried out by the statistical software package SAS, version 8.2, cf. SAS Institute Inc. (1999).

8.2.1 Day-ahead market prices

The recent tendency to restructure and deregulate electricity markets has stimulated most power producers to shift the objective from costs minimization and demand satisfaction to profit maximization alone, which has made the behavior of electricity prices of foremost importance. In the short term, power producers may participate in spot markets such as the day-ahead market for disposing of physical production.To participate in the day-ahead market, bids must be submitted a day in advance, as already indicated by the name. The market cannot clear and the clearing prices of the following day cannot be announced until afterwards. The clearing prices, also referred to as day-ahead market prices, are therefore uncertain at the time of bidding. Even if disposing of physical production by bilateral contracts, the market price uncertainty strongly influences the situation, since in most cases, the price of a bilateral contract reflects the expected development in day-ahead market prices. The forecasting or simulation of the day-ahead market prices of tomorrow is essential to market participants in general. The develop-

ment of bidding strategies may simply be based on reliable price forecasts and simulations, or the procedures may be used in combination with other instruments. Moreover, for power producers in particular, the prediction of day-ahead market prices plays a major role in the short-term planning of the power plant, since physical production is disposed of mostly when prices are profitable.

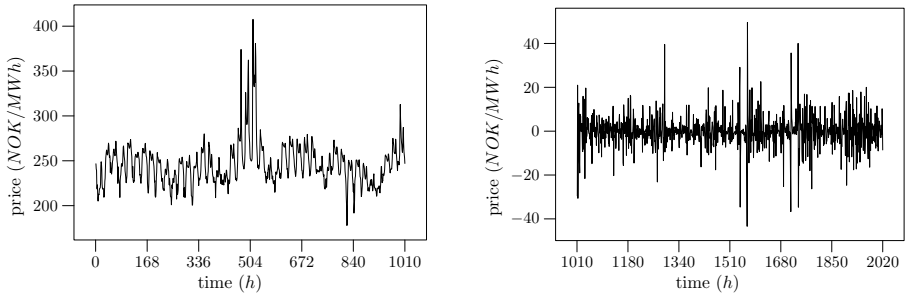
For applications of the ARMA framework to predict day-ahead electricity prices, we refer to the following contributions within the literature. The authors of Contreras et al. (2003) present SARIMA models for day-ahead electricity prices of the Spanish and Californian markets, whereas Nogales et al. (2002) fit transfer and regression models in which electricity load has been included as an explanatory variable and Garcia et al. (2005) focus on GARCH models of the same markets. Another application of the GARCH framework is presented by Escribano et al. (2002), who also consider time-dependent jumps in electricity prices and considers the markets of Argentina, Australia, New Zealand, U.S., Spain as well as the Nordic market. In a similar spirit, the authors of Haldrup and Nielsen (2004) develop a model from within the class of long memory models. The ARFIMA model for the Nordic day-ahead electricity market includes regime switching.

The ARMA time series analysis of day-ahead market prices is performed on observations from the spot market Elspot at the Nordic power exchange Nord Pool, and more precisely from the so-called price zone NO2, which is the Trondheim area in Norway. The empirical time series contains hourly observations from the year of 2004, of which the first 40 weeks are used for model identification and the following 10 weeks for model validation.

Model identification: To identify an appropriate model, the first step is to detect seasonalities. As day-ahead market prices are partly driven by electricity demand which exhibits a daily and a weekly pattern, both daily and weekly periodic behavior is to be expected. This is supported by the fact that the day-ahead market clears every day of the week except in weekends. The periodicities are visible from Fig. 8.1a. The class of SARMA models provides a base for identifying a model.

A non-constant mean indicates non-stationarity of the time series data. This is further justified by empirical autocorrelations that decay very slowly to zero. The factors $(1 - B)$, $(1 - B^{24})$ and $(1 - B^{168})$ are included in order to stabilize the mean, $(1 - B^{24})$ and $(1 - B^{168})$ to remove seasonality, and the process of differences can then be accepted as being stationary, cf. Fig. 8.1b. Experiments was made with a logarithmic tranformation to stabilize the variance. However, the best results were obtained without this transformation. The original process can therefore be described by a SARIMA model.

Inspecting the empirical autocorrelation and partial autocorrelation functions, the order of the process of differences can be determined. The functions are shown in Figs. 8.2. The autocorrelations corresponding to the lags $1, 2, 3, \ldots$ give an indication of an $ARMA(2, 2)$ process. The autocorrelations of the lags $24, 48, 72, \ldots$



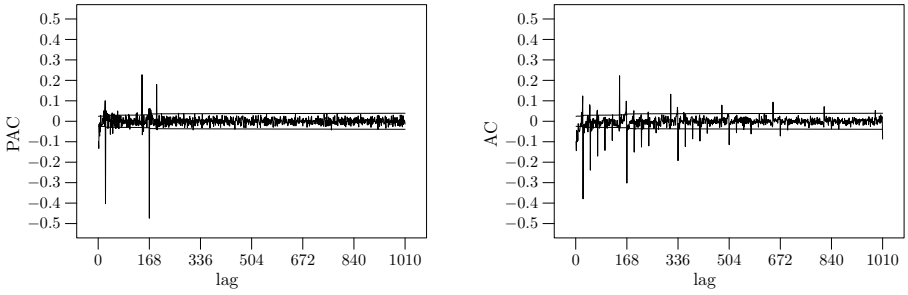
(a) Real. (b) Differences.

Figure 8.1: Hourly day-ahead market prices.

show evidence of an $MA(2)$ process, as the lagged-$24k$ autocorrelations are zero for $k > 2$ and the partial autocorrelation function is exponentially decreasing. Finally, the autocorrelations of the lags $168, 336, 504, \ldots$ are indications of an $MA(1)$ process, as the lagged-$168k$ autocorrelations are zero for $k > 1$ and the partial autocorrelation function is exponentially decreasing. Although not clearly visible in Fig. 8.2a, the $ARMA(2,2)$ process causes the autocorrelation function to peak in the neighborhood of the lag $24, 48, 72, \ldots$ and the $MA(2)$ process causes it to peak at the lags $144, 192, 312, 360, 480, 528$. The initial proposal of a model is

$$(1 - \psi_1 B - \psi_2 B^2)(1 - B)(1 - B^{24})(1 - B^{168})\xi_t = \\ (1 - \gamma_1 B - \gamma_2 B^2)(1 - \gamma_{24} B^{24} - \gamma_{48} B^{48})(1 - \gamma_{168} B^{168}) e_t, \;\; t \in \mathbb{Z}$$

Inspecting the autocorrelation function and the partial autocorrelation function



(a) Autocorrelation. (b) Partial autocorrelation.

Figure 8.2: Autocorrelation functions for day-ahead market prices.

of the residuals, the model can be further refined to

$$(1-\psi_1 B-\psi_2 B^2)(1-B)(1-B^{24})(1-B^{168})\xi_t = \\ (1-\gamma_1 B-\gamma_2 B^2)(1-\gamma_{23}B^{23}-\gamma_{24}B^{24}-\gamma_{25}B^{25} \\ -\gamma_{47}B^{47}-\gamma_{48}B^{48}-\gamma_{49}B^{49})(1-\gamma_{168}B^{168})e_t,\ \ t\in\mathbb{Z}.$$

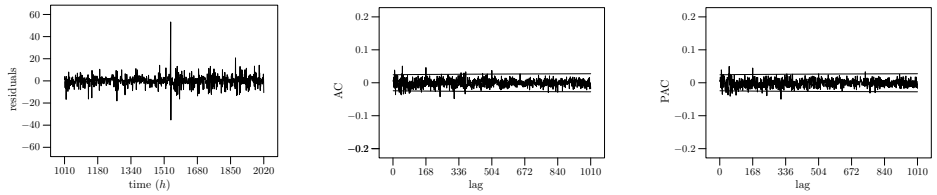
The model is an extension of those of the preceding sections. It is, nevertheless, sufficiently general to include the main characteristics of day-ahead market prices.

Parameter estimation: Parameter estimates are obtained by the use of maximum likelihood estimation. Estimates based on data of the model identification period can be found in Table 8.1.

Table 8.1: Maximum likelihood estimates for day-ahead market prices

Parameter	ψ_1	ψ_2	γ_1	γ_2	γ_{24}	γ_{25}
Estimate	0.3120	0.3758	0.3942	0.5382	−0.0826	0.6888
Parameter	γ_{26}	γ_{47}	γ_{48}	γ_{49}	γ_{168}	σ
Estimate	−0.0748	0.0553	0.1834	0.0388	0.9810	5.0487

Model control: To validate the model, the assumption of a white noise process on the residuals must be confirmed. A plot of the residuals is given in Fig. 8.3a. It should be clear that the mean value can be assumed to be zero and that the variance appears to be constant. Furthermore, the autocorrelation and partial autocorrelation functions of the residuals, cf. Figs. 8.3b and 8.3c, are both close to zero as is the case for a white noise process. The Ljung-Box statistics back up the fit of the model.



(a) Residuals. (b) Autocorrelation. (c) Partial autocorrelation.

Figure 8.3: Residuals of day-ahead market prices.

Out-of-sample testing: Before using forecasts and simulations as tools for planning purposes, both procedures are suitable for further validation of the model. Out-of-sample tests are performed and, hence, the model is tested on the data of the validation period by forecasting and simulating into this period.

Table 8.2: Weekly forecast errors for day-ahead market prices.

Week	41	42	43	44	45
MPE	0.73	−0.34	0.09	−0.11	−0.16
MAPE	3.00	1.72	2.51	2.95	2.19
MSE	72.32	36.22	58.82	436.67	47.15
Week	46	47	48	49	50
MPE	0.48	0.25	0.51	−0.52	−0.25
MAPE	1.87	1.80	2.08	2.25	2.66
MSE	31.81	44.39	37.51	45.88	61.95

Table 8.3: Weekly descriptive statistics for day-ahead market prices.

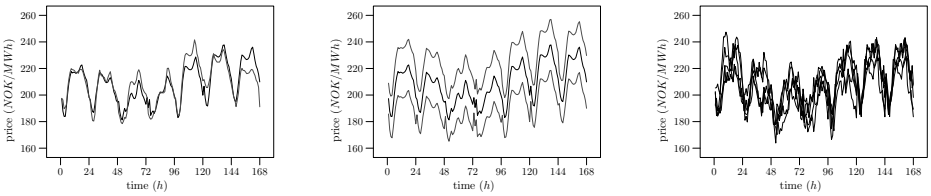
	Week	41	42	43	44	45
Sim.	Mean value	228.46	245.55	242.18	246.32	239.12
	Std. dev.	16.26	13.84	11.87	11.35	11.91
Real	Mean value	229.68	244.55	240.96	246.57	238.59
	Std. dev.	11.46	6.97	4.75	18.67	12.29
	Week	46	47	48	49	50
Sim.	Mean value	227.95	241.05	236.40	226.20	215.93
	Std. dev.	14.74	13.22	13.87	13.45	14.68
Real	Mean value	229.10	241.83	238.95	225.23	215.65
	Std. dev.	14.67	10.37	11.50	14.49	13.57

The forecast errors of the validation period, i.e. the weeks 41-50 or the hours 6721-8400, are reported in Table 8.2. The mean percentual error (MPE), mean absolute percentual error (MAPE) as well as the mean square error (MSE) are displayed on a weekly basis. It should be remarked that the estimation of the validation period has been conducted on a 24-hour basis using an adaptive approach, cf. Contreras et al. (2003) and Nogales et al. (2002). The estimation of the first 24 hours, i.e. the hours 6721-6744, is based on data of the hours 1-6720. Moving the time window 24 hours, the estimation of the next 24 hours, i.e. the hours 6745-6768, is based on data of the hours 24-6744 etc. The forecast errors are seen to be rather small and therefore the model is suitable for forecasting.

Descriptive statistics for a simulation of 1000 samples are shown in Table 8.3 along with the same information on the real observations of the validation period. Concerning the preservation of the descriptive statistics, the mean value is well preserved, whereas the standard deviation is generally over-estimated.

Forecast and simulation: Starting from the end of the validation period, forecasts and simulations can be generated further into the future. As an example

of a short-term use of the procedures, hourly prices are forecasted and simulated a week ahead, i.e. into week 51. A plot of the forecasts and the real observations can be found in Fig. 8.4a. Moreover, the forecast and its confidence intervals are plotted in Fig. 8.4b. Examples of a few simulated sample paths are displayed in Fig. 8.4c.



(a) Forecast and real data.

(b) Confidence interval of the forecast.

(c) Simulation.

Figure 8.4: Forecasted and simulated day-ahead market prices.

8.2.2 Reservoir inflows

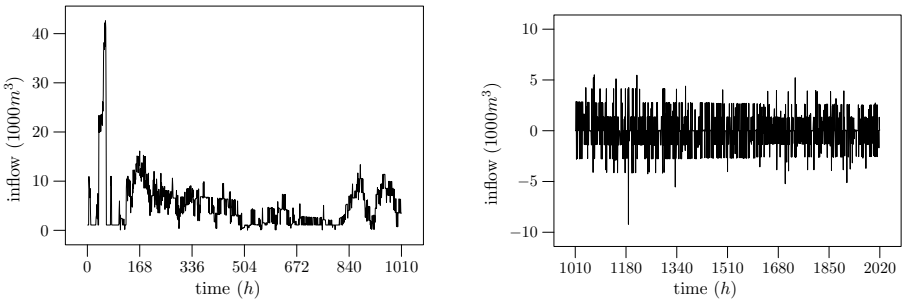
As water inflows to hydro reservoirs are often highly uncertain, the forecasting or simulation of future values is crucial in the short-term planning of hydro-power plants. In the management of reservoirs, the scheduling of water releases depend among other things on the current reservoir levels. Water releases in turn affect unit commitment and production level decisions, which makes the knowledge of future reservoir inflows very valuable. Early applications of the ARMA framework to forecast and simulate stream-flows can be found in Camacho et al. (1985a), Camacho et al. (1985b), Camacho et al. (1987), Salas et al. (1980) and Salas et al. (1985). Newer applications comprise Kuo and Sun (1996), who establish a section model for ten-days average stream-flows in which each section is a separate ARMA model. Concerning more long-term models, Mohan and Vedula (1995) fit a SARIMA model for forecasting monthly inflows into a reservoir system, whereas Jardim et al. (2001) simulate monthly stream-flows from a periodic ARMA model to alleviate mid-term planning in a hydro-power plant.

For the ARMA time series analysis of water inflows to reservoirs, the data consists of hourly observations from the same hydro-power plant located in the area of Trondheim in Norway and run by the company TrønderEnergi. The data dates back to 2004, which is again divided into a model identification period of 40 weeks and a model validation period of 10 weeks.

Typically, even smaller hydro-power plants consist of more than one reservoir. Therefore, it is highly relevant to consider multiple reservoir inflow series. Inflows to different reservoirs may stem from the same stream or from different streams.

Here, we consider two reservoir inflow series from two different streams and initially handle the series individually. We show a model for one of the inflow series should be fitted, as fitting the other may be done in a similar fashion. The first inflow series corresponds to a reservoir named Samsjøen, the second inflows series to Håen.

Model identification: Consider the inflow to Samsjøen. The data does not immediately disclose any obvious short-term seasonalities of the reservoir inflows. Hence, the starting point of model identification is stationarity. Highly non-constant mean value and variance reveal non-stationarity of the time series data, which is further verified by slowly decreasing empirical autocorrelations. By experimenting with logarithmic transformation and differences, the inclusion of factor $(1 - B)$ was found most suitable in obtaining stationarity. In particular, the aucorrelation functions decreased more quickly without the logarithmic transformation. Hence, an appropriate model should be found within the class of ARIMA models. The original empirical time series and the series of differences are displayed in Fig. 8.5.

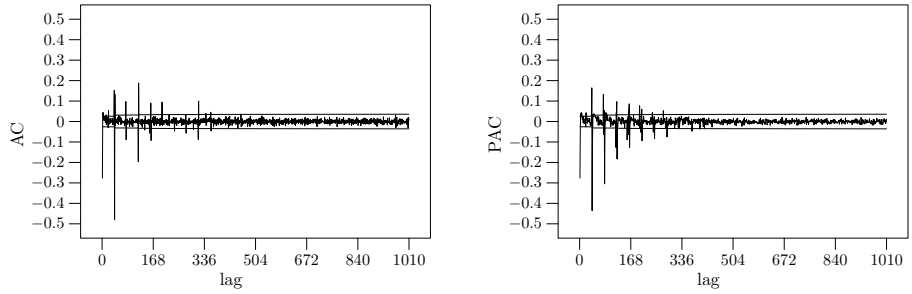


(a) Real. (b) Differences.

Figure 8.5: Hourly reservoir inflows.

To determine the order of the model, the empirical autocorrelation and partial autocorrelation have been drawn in Fig. 8.6. The autocorrelations of the lags $1, 2, 3, \ldots$ suggest an $ARMA(1, 2)$ model, whereas the autocorrelations of the lags $41, 82, 123, \ldots$ and $120, 240, 360, \ldots$ both point to a $MA(1)$ process. Refining the model further, an appropriate end result is

$$(1 - \psi_1 B)(1 - B)\xi_t = (1 - \gamma_1 B - \gamma_2 B^2) \\ (1 - \gamma_{41} B^{41})(1 - \gamma_{119} B^{119} - \gamma_{120} B^{120} - \gamma_{21} B^{121})\epsilon_t, \;\; t \in \mathbb{Z}.$$



(a) Autocorrelation. (b) Partial autocorrelation.

Figure 8.6: Autocorrelation functions for reservoir inflows.

Parameter estimation: Parameter estimates are obtained using maximum likelihood estimation. Estimates based on data of the model identification period are found in Table 8.4.

Table 8.4: Maximum likelihood estimates for reservoir inflows.

Parameter	ψ_1	γ_1	γ_2	γ_{41}
Estimate	0.9879	1.2698	−0.3095	0.8385
Parameter	γ_{119}	γ_{120}	γ_{121}	σ
Estimate	0.1729	−0.1742	−0.0691	3049.4570

Model control: The model is validated by testing the assumption of a white noise process on the residuals and is confirmed by the behavior of the autocorrelation and partial autocorrelation functions as well as by the Ljung-Box statistics, none of which is displayed here.

Out-of-sample testing: Further validation of the model is based on forecasts and simulations. Again, we do out-of-sample testing. Forecast errors are shown in Table 8.5. We find forecasts to be useful for high inflow weeks, i.e. weeks 41 and 45-50. However, for low inflow weeks, i.e. weeks 42-44, the forecasts are rather poor. The descriptive statistics of the simulations, cf. Table 8.6, show that the mean is more or less preserved, whereas the standard deviation is highly overestimated.

Forecast and simulation: With the validated model at hand, short-term forecasts and simulations of hourly inflows can be made. Plots of the forecast and the real data as well as the forecast and its corresponding confidence interval a week ahead are shown in Fig. 8.7a and 8.7b. Examples of a few simulated sample paths are shown in Fig. 8.7c.

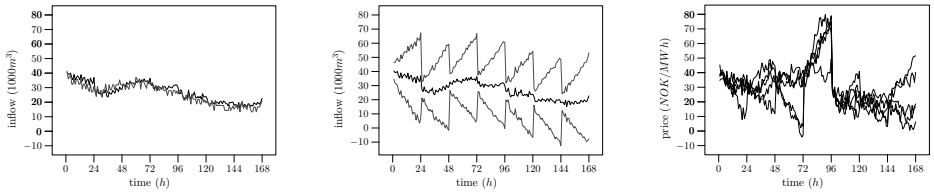
Model identification and parameter estimation: Consider now the inflow

Table 8.5: Weekly forecast errors for reservoir inflows.

Week	41	42	43	44	45
MPE	-7.74	-52.36	15.88	6.78	9.22
MAPE	16.10	116.10	115.78	57.10	15.15
MSE	12.78×10^6	13.89×10^6	12.95×10^6	10.73×10^6	62.16×10^6
Week	46	47	48	49	50
MPE	1.97	0.31	-5.22	3.32	5.46
MAPE	11.76	10.05	11.07	10.55	9.68
MSE	46.75×10^6	9.73×10^6	6.36×10^6	7.17×10^6	22.97×10^6

Table 8.6: Weekly descriptive statistics for reservoir inflows

	Week	41	42	43	44	45
Sim.	Mean value	19053.45	9346.40	9077.62	6458.55	50638.26
	Std. dev.	10224.38	10116.10	10102.11	10348.29	10348.29
Real	Mean value	17896.71	6323.36	8995.71	6960.00	44862.86
	Std. dev.	4818.84	3597.40	4677.30	3386.63	17653.88
	Week	46	47	48	49	50
Sim.	Mean value	50638.26	24185.43	19574.73	17772.61	38566.49
	Std. dev.	12882.12	10196.85	10178.67	9570.64	10407.45
Real	Mean value	45686.57	24573.21	19284.64	21884.79	40509.64
	Std. dev.	8933.57	6164.27	6311.45	8864.11	11268.43



(a) Forecast and real data.

(b) Confidence interval of the forecast.

(c) Simulation.

Figure 8.7: Forecasted and simulated reservoir inflows.

to Håen. The model fitted for the corresponding inflow series is

$$(1-\psi_1 B)(1-B)\xi_t = (1-\gamma_1 B-\gamma_2 B^2)(1-\gamma_{41}B^{41})e_t,\ \ t\in\mathbb{Z},$$

with the parameter estimates of Table 8.7.

Table 8.7: Maximum likelihood estimates for reservoir inflows

Parameter	ψ_1	γ_1	γ_2	γ_{41}	σ^2
Estimate	0.9759	1.4352	−0.5554	0.8344	740.3185

8.3 Multivariate ARMA modeling

Univariate ARMA processes can be generalized to multivariate ARMA processes, see Tiao and Box (1981) and Jenkins and Alavi (1981). The generalization of ARMA processes to higher dimensions is referred to as *vector* ARMA (VARMA) processes.

Definition 8.3.1 *Let $\{e_t\}_{t\in\mathbb{Z}}$ be a $N\times 1$ white noise process with $N\times N$ covariance matrix Σ and let $\{\xi_t\}_{t\in\mathbb{Z}}$ be a $N\times 1$ weakly stationary stochastic process that solves*

$$\psi(B)\xi_t = \gamma(B)e_t, \;\; t\in\mathbb{Z},$$

where $\psi(z) = I_N - \sum_{k=1}^{p}\psi_k z^k$ and $\gamma(z) = I_N - \sum_{k=1}^{q}\gamma_k z^k$ are polynomials with $p,q\in\mathbb{N}_0$, $\psi(z)\neq 0$ and $\gamma(z)\neq 0$ for $\|z\|\leq 1$ and B is the backshift operator, i.e. $B^k\xi_t = \xi_{t-k}$. Then $\{\xi_t\}_{t\in\mathbb{Z}}$ is called a VARMA process of order (p,q) with $N\times N$ parameter matrices $\Sigma,\psi_1,\ldots,\psi_p,\gamma_1,\ldots,\gamma_q$.

If the parameter matrices are diagonal and the innovations are uncorrolated, the multivariate VARMA model collapses to *independent* univariate ARMA models. As a hybrid between the multivariate and the univariate models, the *contemporaneous* ARMA (CARMA) model arises if the parameter matrices are diagonal, but the innovations are allowed to be contemporaneously correlated, see Salas et al. (1980) and Hipel and McLeod (1993).

Definition 8.3.2 *Let $\{e_t\}_{t\in\mathbb{Z}}$ be a $N\times 1$ white noise process with $N\times N$ covariance matrix Σ and let $\{\xi_t\}_{t\in\mathbb{Z}}$ be a $N\times 1$ stochastic process that solves*

$$\psi(B)\xi_t = \gamma(B)e_t, \;\; t\in\mathbb{Z},$$

where $\psi(z) = I_N - \sum_{k=1}^{p}\psi_k z^k$ and $\gamma(z) = I_N - \sum_{k=1}^{q}\gamma_k z^k$ are polynomials with $p,q\in\mathbb{N}_0$, and $N\times N$ parameter matrices $\psi_1,\ldots,\psi_p,\gamma_1,\ldots,\gamma_q$ that have zeros off the diagonal and B is the backshift operator, i.e. $B^k\xi_t = \xi_{t-k}$. If the univariate processes $\{\xi_{n,t}\}_{t\in\mathbb{Z}}$ that solve

$$\psi_{nn}(B)\xi_{n,t} = \gamma_{nn}(B)e_{n,t}, \;\; t\in\mathbb{Z}$$

are ARMA processes of order (p,q) and the corresponding white noise processes $\{e_{n,t}\}_{t\in\mathbb{Z}}$ are allowed to be contemporaneously correlated, then the multivariate process $\{\xi_t\}_{t\in\mathbb{Z}}$ is called a CARMA process of order (p,q) with $N\times N$ parameter matrices $\Sigma,\psi_1,\ldots,\psi_p,\gamma_1,\ldots,\gamma_q$.

Motivated by the fact that the VARMA model is difficult to estimate, especially for higher order models in which the number of parameters is large, the more simple CARMA model finds its relevance. The main advantage of the CARMA model is its decoupling of the multivariate model into univariate models that facilitates the application of specially designed parameter estimation procedures based on the univariate estimates. In spite of its simplicity, the CARMA model incorporates correlation among the univariate models by allowing the innovations to be instantaneously correlated.

Salas et al. (1980) propose an alternative formulation of the CARMA model as a VARMA model. Let $\{\xi_t\}_{t\in\mathbb{Z}}$ be a CARMA process of order (p, q) with parameter matrices $\Sigma, \psi_1, \ldots, \psi_p, \gamma_1, \ldots, \gamma_q$ and an upper triangular matrix C that satisfies $CC^T = \Sigma$. Let $\{\epsilon_t\}_{t\in\mathbb{Z}}$ be a $N \times 1$ white noise process with covariance matrix I_N. Then $\{\xi_t\}_{t\in\mathbb{Z}}$ is a $N \times 1$ weakly stationary stochastic process that solves

$$\psi(B)\xi_t = \gamma(B)C\epsilon_t, \;\; t \in \mathbb{Z}.$$

Hence, $\{\xi_t\}_{t\in\mathbb{Z}}$ is a VARMA process of order (p, q) with parameter matrices $I_N, \psi_1, \ldots, \psi_p, \gamma_1 C, \ldots, \gamma_q C$.

Multivariate ARMA time series modeling includes the same general steps as for univariate models, which is model identification, parameter estimation, model control and validation as well as forecasting and simulation.

Model identification: To fix the structure of a multivariate model, the dependencies between the univariate series must be identified. Whether the series exhibit contemporaneous, unidirectional, bidirectional relationships, mixtures of these or no dependencies at all, dependencies are disclosed by the cross-correlations of the series. For contemporaneous models, the cross-correlations are zero except at lag zero, for unidirectional relationships, the cross-correlations are zero at negative lags but significant at the remaining ones, whereas for bidirectional relationships the cross-correlations are significant at all lags. When the structure of the model has been fixed, the order can be determined. For contemporaneous ARMA models, univariate procedures can be used. For vector ARMA models in general, the cross-correlation matrices and the partial correlation matrices complement the autocorrelation and partial autocorrelation functions in the order determination.

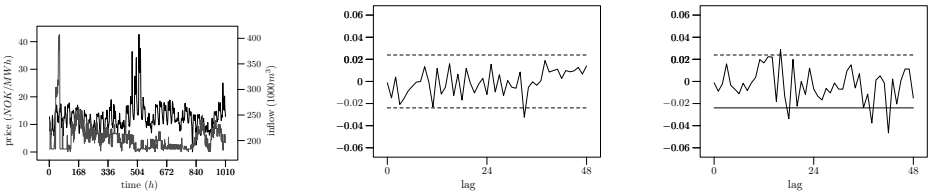
Parameter estimation and model control: Parameter estimation, model control, forecasting and simulation in the multivariate case is an extension that in the univariate case.

8.3.1 Simultaneity of prices and inflows

If hourly day-ahead market prices and reservoir inflows are considered simultaneously, the time series is multidimensional. We illustrate the fitting of a multivariate model with the day-ahead market price series and the two reservoir inflow series of the preceding sections.

Examples from the literature on VARMA modeling in stream-flows include Camacho et al. (1985a) and Salas et al. (1980) and on CARMA modeling especially Camacho et al. (1985b) and Salas et al. (1985).

Model identification: The VARMA model could be adopted to fit the multivariate model. The three-dimensional VARMA model, however, simplifies as day-ahead market prices and reservoir inflows may be assumed uncorrelated in the short term. This is supported by the plot of day-ahead market prices versus reservoir inflows in Fig. 8.8a. It should be clearer from Figs. 8.8b and 8.8c, which display cross-correlations between the residuals of prices and inflows from one of the streams. Corresponding cross-correlations between the residuals of prices and inflows from the other stream are similar. The figures reveal uncorrelated residuals. Consequently, day-ahead market prices and reservoir inflows can be handled separately. For the day-ahead market prices, we refer to the preceding sections.



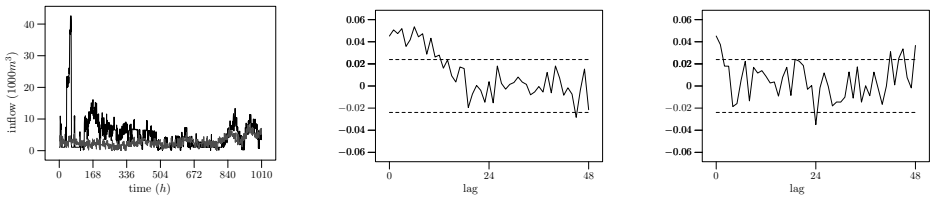
(a) Prices and inflows. (b) Serial cross correlation. (c) Serial cross correlation.

Figure 8.8: Correlation between day-ahead market price and reservoir inflow series.

We proceed with the reservoir inflows. Inflow series from different streams are correlated and cannot be treated separately. The correlation between the two inflow series should be visible from Fig. 8.9a. Therefore, the simultaneous modeling of multiple inflow series, in principle, relies upon the VARMA model. Nevertheless, unidirectional dependencies often occur in the case of connected streams, whereas, if the streams are not connected, contemporaneous dependencies may apply. Within reasonable limits, the cross-correlations between the residuals of the two inflow series depicted in Figs. 8.9b and 8.9c support the use of a CARMA model.

Parameter estimation: We use the univariate estimates, although a multivariate estimation procedure, applying only to CARMA models, may reduce the variance of the estimates. Some parameter estimates were given in Tables 8.4 and 8.7. Moreover, Table 8.8 shows the maximum likelihood estimates for the covariance matrix of the innovations Σ and the corresponding upper triangular matrix C.

Model control: The univariate models have been validated in the preceding sections.



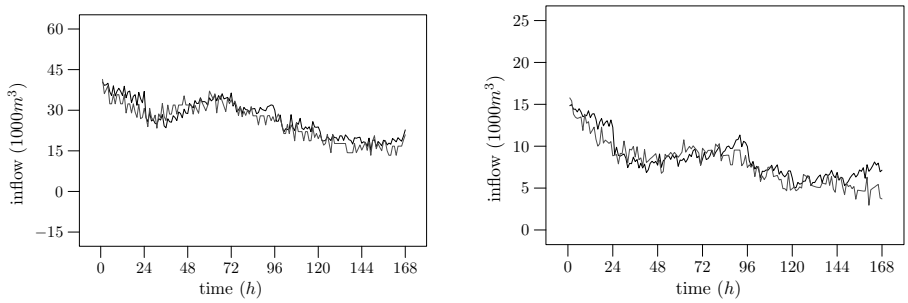
(a) Inflows. (b) Serial cross correlation. (c) Serial cross correlation.

Figure 8.9: Correlation between reservoir inflows series.

Table 8.8: Maximum likelihood estimates for reservoir inflow.

Parameter	Σ_{11}	Σ_{12}	Σ_{21}	Σ_{22}
Estimate	9270446	104030	104030	547131
Parameter	C_{11}	C_{12}	C_{21}	C_{22}
Estimate	3044.74	0.00	34.17	738.89

Forecast and simulation: Forecasts and simulations can be based on the alternative formulation of the multivariate model. The forecasts are shown in Fig. 8.10 and the simulations in Fig. 8.11. The correlation between the two univariate inflow series should be visible from both the forecasts and the simulations.

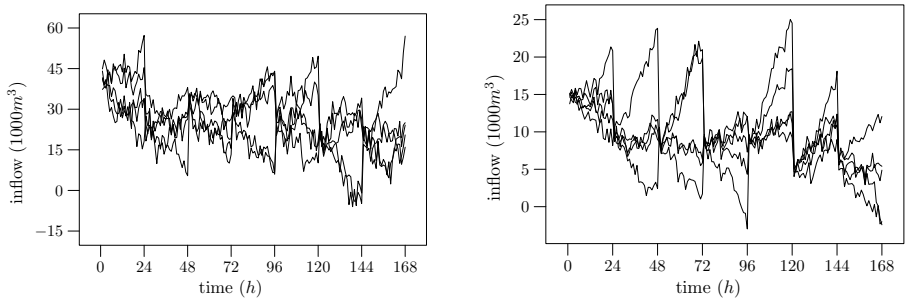


(a) Upper reservoir. (b) Lower reservoir.

Figure 8.10: Forecasted and real reservoir inflows.

8.4 Conclusions

The results of the previous sections illustrate the ability of ARMA time series models to forecast and simulate hourly day-ahead electricity prices and reservoir



(a) Upper reservoir. (b) Lower reservoir.

Figure 8.11: Simulated reservoir inflows.

inflows. As already indicated, forecasts and simulations are suitable as input for optimization problems such as deterministic and stochastic planning problems that future values of data as input.

An example of using simulated day-ahead electricity prices in a stochastic planning problem is found in Chapter 4. The problem consists in day-ahead bidding of a hydro-power plant under price uncertainty and simulated hourly day-ahead market prices serve as scenarios in a corresponding stochastic program. In a similar fashion, Chapter 5 uses simulated hourly day-ahead market prices and reservoir inflows as scenarios in a stochastic program that determines the spatial distribution of water among reservoirs of a hydro-power plant.

In practice, power producers rely upon so-called fundamental models that involve detailed physical relationships. Although the forecast performance may in general be superior to time series models, the fundamental models are not capable of simulating and the time series models still have their relevance.

The literature has previously reported on the success of forecasting or simulating hourly day-ahead market prices by ARMA models and the current results show the same potential. However, ARMA models have been used mostly to describe monthly and annual stream-flows, whereas the use of ARMA models to forecast or simulate hourly stream-flows has occurred only rarely in the literature. By a direct application of the ARMA framwork it may not be possible to capture intermittency and the current results show a moderate performance with respect to the hourly stream-flows.

Although the ability to forecast or simulate day-ahead market prices and reservoir inflows can be improved, we consider ARMA time series models sufficient for illustrating the usefulness of stochastic programming, as in Chapters 4 and 5.

8.4.1 Further improvement

To improve the time series models for reservoir inflows, we suggest to combine a long-term model, that is able to capture memory, persistence and periodicity properties, with a short-term model suitable for describing intermittency. A possible outline is the following.

Since the ARMA framework has been successfully applied for modeling more long-term stream-flows, performance may improve with a time resolution of a day or a week. This point of view is supported by the ability of the ARMA processes to model the general trend in the data, as seen in Figs. 8.7a and 8.10. The idea is therefore to describe the aggregated daily or weekly inflows by means of ARMA processes in the same way as the hourly inflows were described. With a daily time resolution, the "long-term" model is based on the means

$$\bar{\xi}_{\lceil t/24 \rceil}, t \in \mathbb{Z}$$

and similarly for a weekly time resolution.

The daily or weekly data then has to be disaggregated into hourly data. Several disaggregation techniques can be found in the literature. Still, many techniques are not suitable for hourly disaggregation due to the intermittency of stream-flows, whereas others are unable to preserve the time series properties at the disaggregated level. However, the availability of both "long-term" (daily or weekly) and short-term (hourly) data gives rise to models of different time resolutions, and the aggregation-disaggregation reduces to a matter of combining the models. The long-term and the short-term models can be combined by considering

$$\zeta_t = \xi_t / \bar{\xi}_{\lceil t/24 \rceil} \text{ or } \zeta_t = \xi_t - \bar{\xi}_{\lceil t/24 \rceil}, t \in \mathbb{Z}$$

in the short-term model, cf. Eichhorn et al. (2005) and Gröwe-Kuska et al. (2001).

The short-term model allows for a partitioning of the data set into smaller sets. Stream-flows can be categorized according to the seasons or according to dry or wet periods of time. The categories may be determined by clustering methods, each cluster being used to construct a standard profile or a stochastic submodel based for instance on regression on time.

The time series models for day-ahead market prices may likewise be improved by combining a long-term model with a short-term model. Fig. 8.4a indicates that the ARMA processes are able to model the development in prices except for the general trend in the data. It may be therefore be relevant to establish a more long-term model based on weekly or daily means and a short-term model in which prices are partitioned into smaller sets. Prices may contain the categories spring, summer, fall and winter and for each of them workday, Saturday and Sunday.

.

Bibliography

Ahmed, S. (2004). Mean-risk objectives in stochastic programming. . `http://hera.rz.hu-berlin.de/speps/contents04.html`.

Andersen, A. H. and P. Blæsild (2003a). *Time Series Analysis*, Volume 1. Århus, Denmark: University of Aarhus.

Andersen, A. H. and P. Blæsild (2003b). *Time Series Analysis*, Volume 2. Århus, Denmark: University of Aarhus.

Anderson, E. J. and A. B. Philpott (2002). Optimal offer construction in electricity markets. *Mathematics of Operations Research 27*(1), 82–100.

Archibald, T. W., C. S. Buchanan, K. I. M. McKinnon, and L. C. Thomas (1996). Nested Benders decomposition and dynamic programming for reservoir optimization. *Journal of the Operations Research Society 50*, 468–479.

Archibald, T. W., C. S. Buchanan, L. C. Thomas, and K. I. M. McKinnon (2001). Controlling multi-reservoir systems. *European Journal of Operational Research 129*, 619–626.

Benders, J. F. (1962). Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik 4*, 238–252.

Berge, C. (1963). *Topological spaces*. Edinburgh: Oliver & Boyd.

Bienstock, D. and J. F. Shapiro (1988). Optimizing resource acquisition decisions by stochastic programming. *Management Science 34*(2), 215–229.

Billingsley, P. (1968). *Convergence of Probability Measures*. New York: Wiley.

Birge, J. R. (1985). Decomposition and partitioning methods for multistage stochastic linear programs. *Operations Research 33*(5), 989–1007.

Birge, J. R., M. A. H. Dempster, H. I. Gassmann, E. A. Gunn, A. J. King, and S. W. Wallace (1987). A standard input format for multiperiod stochastic linear programs. *COAL Newsletter 17*, 1–19.

Birge, J. R. and F. Louveaux (1997). *Introduction to Stochastic Programming*. New York: Springer.

Birge, J. R. and F. V. Louveaux (1988). A multicut algorithm for two-stage stochastic linear programs. *European Journal of Operational Research 34*, 384–392.

Birge, J. R., S. Takriti, and E. Long (1994). Intelligent unified control of unit commitment and generation allocation. Technical report, Department of Industrial and Operations Engineering, The University of Michigan.

Box, G. E. P. and G. M. Jenkins (1976). *Time Series Analysis: Forecasting and Control*. San Fransisco: Holden Day.

Bunn, D. W. and S. N. Paschentis (1986). Development of a stochastic model for the economic dispatch of electric power. *European Journal of Operational Research 27*, 179–191.

Camacho, F., A. I. McLeod, and K. W. Hipel (1985a). Contemporaneous autoregressive - moving average (CARMA) modeling in hydrology. *Water Resources Bulletin 21*(4), 709–720.

Camacho, F., A. I. McLeod, and K. W. Hipel (1985b). Developments in multivariate arma modeling in hydrology. In *International Symposium on Multivariate Analysis of Hydrologic Process*, Fort Collins, Colorado.

Camacho, F., A. I. McLeod, and K. W. Hipel (1987). Multivariate contemporaneous arma model with hydrological applications. *Hydrology and Hydraulics 1*(2), 141–154.

Carøe, C. C. and R. Schultz (1998). A two-stage stochastic program for unit commitment under uncertainty in a hydro-thermal power system. citeseer.ist.psu.edu/caroe98twostage.html.

Carøe, C. C. and R. Schultz (1999). Dual decomposition in stochastic integer programming. *Operations Research Letters 24*, 37–45.

Carøe, C. C. and J. Tind (1998). L-shaped decomposition of two-stage stochastic programs with integer recourse. *Mathematical Programming Series A 83*(3), 451–464.

Carpentier, P., G. Cohen, J. C. Culioli, and A. Renaud (1996). Stochastic optimization of unit commitment: A new decomposition framework. *IEEE Transactions on Power Systems 11*(2), 1067–1073.

Contreras, J., R. Espinola, and F. J. Nogales (2003). Arima models to predict next-day electricity prices. *IEEE Transactions on Power Systems 18*(3), 1014–1020.

Cplex Optimization Inc. (2006). ILOG CPLEX 10.0 User's Manual.

Dantzig, G. B. and P. Wolfe (1960). Decomposition principles for linear programs. *Operations Research 8*(1), 101–111.

Danzig, G. B. (1955). Linear programming under uncertainty. *Management Science 1*(3/4), 197–206.

Dempster, M. A. H. (2004). Sequential importance sampling algorithms for dynamic stochastic programming. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) 312*, 94–129, 312–313.

Dempster, M. A. H. and R. T. Thompson (1999). EVPI-based importance sampling solution procedures for multistage stochastic linear programmes on parallel MIMD architectures. *Annals of Operations Research 90*, 161–184.

Dentcheva, D. and W. Römisch (1998). Optimal power generation under uncertainty via stochastic programming. In *Stochastic Programming Methods and Technical Applications*, Volume 458 of *Lecture Notes in Economics and Mathematical Systems*, pp. 22–56. Springer-Verlag.

Dentcheva, D. and W. Römisch (2004). Duality gaps in nonconvex stochastic optimization. *Mathematical Programming 101*, 25–53.

Donchev, A. L. (1986). On the regularity of the kuhn-tucker curve. *SIAM Journal of Control and Optimization 24*, 169–176.

Dupačová, J., G. Consigli, and S. W. Wallace (2000). Scenarios for multistage stochastic programs. *Annals of Operations Research 100*(1-4), 25–53.

Dupačová, J., N. Gröwe-Kuska, and W. Römisch (2003). Scenario reduction in stochastic programming: An approach using probability metrics. *Mathematical Programming Series A 95*, 493–511.

Eichhorn, A. and W. Römisch (2005). Polyhedral risk measures in stochastic programming. *SIAM Journal of Optimization 16*(1), 69–95.

Eichhorn, A., W. Römisch, and I. Wegner (2005). Mean-risk optimization of electricity portfolios using polyhedral risk measures. In *PowerTech Conference Proceedings*, St. Petersburg. IEEE.

Ermoliev, Y. M. and A. A. Gaivoronski (1992). Stochastic quasigradient methods for optimization of discrete event systems. *Annals of Operations Research 39*(1), 1–39.

Escribano, Á., J. I. Peña, and P. Villaplana (2002). Modeling electricity prices: International evidence. Economics Ser. 80, Working Papers, Universidad Carlos III, Departamento de Economía.

Fahd, G. N. and G. B. Sheble (1994). Unit commitment literature synopsis. *IEEE Transactions on Power Systems 9*(1), 128–135.

Feltenmark, S., K. C. Kiwiel, and P.-O. Lindberg (1997). Solving unit commitment problems in power production planning. In *Operations Research Proceedings* (U. Zimmermann ed.)., pp. 236–241. Berlin: Springer-Verlag.

Fleten, S.-E. and T. K. Kristoffersen (2006). Short-term hydropower production planning by stochastic programming. In press, Computers and Operations Research.

Fleten, S.-E. and T. K. Kristoffersen (2007). Stochastic programming for optimizing bidding strategies of a Nordic hydropower producer. *European Journal of Operational Research 181*, 916–928.

Fleten, S.-E. and E. Pettersen (2005). Constructing bidding curves for a price-taking retailer in the Norwegian electricity market. *IEEE Transactions on Power Systems 20*(2), 701–708.

Fleten, S.-E., S. W. Wallace, and W. T. Ziemba (2002). Hedging electricity portfolios via stochastic programming. In C. Greengard and A. Ruszczyński (Eds.), *Decision Making Under Uncertainty: Energy and Power*, Volume 128 of *IMA Volumes on Mathematics and Its Applications*, pp. 71–93. New York: Springer-Verlag.

Fosso, O. B., A. Gjelsvik, A. Haugstad, B. Mo, and I. Wagensteen (1999). Generation scheduling in a deregulated system. The Norwegian case. *IEEE Transactions on Power Systems 14*(1), 75–81.

GAMS Software GmbH (2002). GAMS/SCENRED Dokumentation.

Garcia, R. C., J. Contreras, M. van Akkeren, and J. B. C. Garcia (2005). A garch forecasting model to predict day-ahead electricity prices. *IEEE Transactions on Power Systems 20*(2), 867–874.

García-González, J., E. Parrilla, and A. Mateo (2006). Risk-averse profit-based optimal scheduling of a hydro-chain in the day-ahead electricity market. In press, European Journal of Operational Research.

Gassmann, H. I. and E. Schweitzer (2001). A comprehensive input format for stochastic linear programs. *Annals of Operations Research 104*, 89–125.

Gollmer, R., A. Möller, M. P. Nowak, W. Römisch, and R. Schultz (1999). Primal and dual methods for unit commitment in a hydro-thermal power system. In *Proceedings of the 13th Power Systems Computation Conference*, Trondheim, pp. 724–730.

Gollmer, R., M. P. Nowak, W. Römisch, and R. Schultz (2000). Unit commitment in power generation - a basic model and some extensions. *Annals of Operations Research 96*(1-4), 167–189.

Gorenstin, B. G. and N. M. Campodonico (1993). Power system expansion planning under uncertainty. *IEEE Transactions on Power Systems 8*(1), 129–155.

Gröwe, N., W. Römisch, and R. Schultz (1995). A simple recourse model for power dispatch under uncertain demand. *Annals of Operations Research 59*, 135–164.

Gröwe-Kuska, N., H. Heitsch, and W. Römisch (2003). Scenario reduction and scenario tree construction for power management problems. In *PowerTech Conference Proceedings*, Volume 3, Bologna, pp. 7–14. IEEE.

Gröwe-Kuska, N., K. C. Kiwiel, M. P. Nowak, W. Römisch, and I. Wegner (2002). Power management in a hydro-thermal system under uncertainty by Lagrangian relaxation. In A. R. C. Greengard (Ed.), *Decision Making under Uncertainty: Energy and Power*, Volume 128 of *IMA Volumes in Mathematics and its Applications*, pp. 39–70. New York: Springer-Verlag.

Gröwe-Kuska, N., M. P. Nowak, and I. Wegner (2001). Modeling of uncertainty for the real-time management of power systems. In M. Grötschel, S. Krumke, and J. Rambau (Eds.), *Online Optimization of Large-Scale Systems*, pp. 621–645. Berlin: Springer.

Gröwe-Kuska, N., W. Römisch, and M. P. Nowak (2000). Power management under uncertainty by Lagrangian relaxation. In *Proceedings of the 6th International Conference Probabilistic Methods Applied to Power Systems PMAPS 2000*, Volume 2, INESC Porto.

Haldrup, N. and M. O. Nielsen (2004). A regime switching long memory model for electricity prices. In press, Journal of Econometrics.

Hansen, P. and B. Jaumard (1997). Cluster analysis and mathematical programming. *Mathematical Programming 79*(1-3), 191–215.

Heitsch, H. and W. Römisch (2006a). A note on scenario reduction for two-stage stochastic programs. Preprint 335, DFG Research Center Matheon "Mathematics for key technologies", Humboldt University Berlin.

Heitsch, H. and W. Römisch (2006b). Scenario tree modelling for multistage stochastic programs. `http://edoc.hu-berlin.de/docviews/abstract.php?lang=ger&id=27364`.

Heitsch, H., W. Römisch, and C. Strugarek (2005). Stability of multistage stochastic programs. Preprint 255, DFG Research Center MATHEON "Mathematics for key technologies", Humboldt University Berlin.

Higle, J. L., W. W. Lowe, and R. Odio (1994). Conditional stochastic decomposition: An algorithmic interface for optimization and simulation. *Operations Research 42*(2), 311–322.

Higle, J. L. and S. Sen (1991). Stochastic decomposition: An algorithm for two-state linear programs with recourse. *Mathematics of Operations Research 16*(3), 650–669.

Hipel, K. W. and A. McLeod (1993). *Time-series Modelling of Water Resources and Environmental Systems*. Elsevier.

Hobbs, B. F. (1995). Optimization methods for electric utility resource planning. *European Journal of Operational Research 83*, 1–20.

Høyland, K., M. Kaut, and S. W. Wallace (2003). A heuristic for moment-matching scenario generation. *Computational Optimization and Applications 24*(2-3), 169–185. Stochastic programming.

Høyland, K. and S. W. Wallace (2001). Generating scenario trees for multistage decision problems. *Management Science 47*(2), 295–307.

Hreinsson, E. B. (1988). Optimal short term operation of a purely hydroelectric system. *IEEE Transactions on Power Systems 3*(3), 1072–1077.

ILOG Inc. (2003). OPL Studio 3.7 User's Manual.

Infanger, G. (1994). *Planning under uncertainty. Solving large scale stochastic linear programs*, Volume 108. Andover, Hempshire: The Scientific Press.

Infanger, G. (2005). Monte Carlo (importance) sampling within a benders decomposition algorithm for stochastic linear programs. *Annals of Operations Research 39*(1), 69–95.

Jacobs, J., G. Freeman, J. Grygier, D. Morton, G. Schultz, K. Staschus, and J. Stedinger (1995). SOCRATES: A system for scheduling hydroelectric generation under uncertainty. *Annals of Operations Research 59*, 99–133. Models for planning under uncertainty.

Jardim, D. L. D. D., M. E. P. Maceira, and D. M. Falcao (2001). Stochastic streamflow model for hydroelectric systems using clustering techniques. In *IEEE Porto Power Tech*, Porto, Portugal.

Jenkins, G. M. and A. S. Alavi (1981). Some aspects of modeling and forecasting multivariate time series. *Journal of Time series Analysis 2*, 1–47.

Kall, P. (1976). *Stochastic Linear Programming*. Berlin: Springer.

Kall, P. (1987). On approximations and stability in stochastic programming. In J. Guddat, T. H. Jongen, B. Kummer, and F. Nožička (Eds.), *Parametric Optimization and Related Topics*, pp. 387–407. Berlin: Akademie-Verlag.

Kall, P. and S. W. Wallace (1994). *Stochastic Programming*. Chichester: Wiley.

Kaut, M. and S. W. Wallace (2003). Evaluation of scenario-generation methods for stochastic programming. In press, Pacific Journal of Optimization.

Kristoffersen, T. K. (2005). Deviation measures in linear two-stage stochastic programming. *Mathematical Methods of Operations Research 62*(2), 255–274.

Kristoffersen, T. K. (2007). The development in stochastic programming models for power production and trading. Working paper.

Kristoffersen, T. K. and C. Schaumburg-Müller (2007). Power reserve management by two-stage stochastic programming. Submitted.

Kuo, J.-T. and Y.-H. Sun (1996). An ARMA-type section model for average ten-day streamflow synthesis. *Water Resources Management 10*, 333–354.

Laporte, G. and F. V. Louveaux (1993). The integer L-shaped method for stochastic integer programs with complete recourse. *Operations Research Letters 13*, 133—142.

Løkketangen, A. and D. L. Woodruff (1996). Progressive hedging and tabu search applied to mixed integer (0,1) multistage stochastic programming. *Journal of Heuristics 2*, 111–128.

Louveaux, F. V. and R. Schultz (2003). Stochastic integer programming. *Handbooks in OR & MS 10*, 213–266.

Märkert, A. and R. Schultz (2005). Deviation measures in stochastic integer programming. *Operations Research Letters 33*, 441–449.

Mohan, S. and S. Vedula (1995). Multiplicative seasonal arima model for longterm forecasting of inflows. *Water Reources Management 9*, 115–126.

Morton, D. P. (1996). An enhanced decomposition algorithm for multistage stochastic hydroelectric scheduling. *Annals of Operations Research 64*, 211–235.

Muckstadt, J. A. and S. A. Koenig (1977). An application of Lagrangian relaxation to scheduling in power generation systems. *Operations Research 25*(3), 387–403.

Mulvey, J. M. and A. Ruszczyński (1995). A new scenario decomposition method for large-scale stochastic optimization. *Operations Research 43*(3), 477–490.

Neame, P. J., A. B. Philpott, and G. Pritchard (2003). Offer stack optimization in electricity pool markets. *Operations Research 51*(3), 397–408.

Nogales, F. J., J. Contreras, and A. J. Conjeno (2002). Forecasting next-day electricity prices by time series models. *IEEE Transactions on power systems 17*(2), 342–348.

Norkin, V. I., Y. M. Ermoliev, and A. Ruszczyński (1998). On optimal allocation of indivisibles under uncertainty. *Operations Research 46*, 381–395.

Norkin, V. I., G. C. Pflug, and A. Ruszczyński (1998). A branch and bound method for stochastic global optimization. *Mathematical Programming Series A 83*(3), 425–450.

Nowak, M. P., R. Nürnberg, W. Römisch, R. Schultz, and M. Westphalen (2000). Stochastic programming for power production and trading under uncertainty. Submitted.

Nowak, M. P. and W. Römisch (2000). Stochastic Lagrangian relaxation applied to power scheduling in a hydro-thermal system under uncertainty. *Annals of Operations Research 100*(1-4), 251–272.

Nowak, M. P., R. Schultz, and M. Westphalen (2005). A stochastic integer programming model for incorporating day-ahead trading of electricity into hydro-thermal unit commitment. *Optimization and Engineering 6*, 163–176.

Ogryczak, W. and A. Ruszczyński (1999). From stochastic dominance to mean-risk models: Semideviations as risk measures. *European Journal of Operational Research 116*, 33–50.

Ogryczak, W. and A. Ruszczyński (2001). On consistency of stochastic dominance and mean-semideviation models. *Mathematical Programming 89*, 217–232.

Pennanen, T. (2005). Epi-convergent discretizations of multistage stochastic programs. *Mathematics of Operations Research 30*(1), 245–256.

Pennanen, T. and M. Koivu (2005). Epi-convergent discretizations of stochastic programs via integration quadratures. *Numerical Mathematics 100*(1), 141–163.

Pereira, M. V. F. and L. M. V. G. Pinto (1991). Multi-stage stochastic optimization applied to energy planning. *Mathematical Programming 52*, 359–375.

Pflug, G. C. (2004). Optimal tree generation for multiperiod financial optimization by optimal discretization. *Mathematical Programming 89*(2), 251–271.

Philpott, A. and R. Schultz (2006). Unit commitment in electricity pool markets. *Mathematical Programming Series B*(108), 313–337.

Philpott, A. B., M. Craddock, and H. Waterer (2000). Hydro-electric unit commitment subject to uncertain demand. *European Journal of Operational Research 125*, 410–424.

Prekopa, A. (1995). *Stochastic Programming*. Dordrecht: Kluwer.

Prichard, G. and G. Zakeri (2003). Market offering strategies for hydroelectric generators. *Operations Research 51*(3), 602–612.

Reznicek, K. and T. C. E. Cheng (1991). Stochastic modeling in reservoir operations. *European journal of Operational Research 50*, 235–248.

Robinson, S. M. and R. J.-B. Wets (1987). Stability in two-stage stochastic programming. *SIAM Journal of Control and Optimization 26*(6), 1409–1416.

Rockafellar, R. T. (1997). *Convex Analysis*. Princeton.

Rockafellar, R. T. and R. J.-B. Wets (1991). Scenarios and policy aggregation under uncertainty. *Mathematics of Operations Research 16*(1), 119–147.

Römisch, W. (2003). Stability of stochastic programming problems. . In A. Ruszczyńsky and A. Shapiro (Eds.), *Handbooks in Operations Research and Management Science*, Volume 10 of *Stochastic Programming*, pp. 483–554. Amsterdam: Elsevier.

Römisch, W. and R. Schultz (2001). Multi-stage stochastic integer programs - an introduction. `http://edoc.hu-berlin.de/docviews/abstract.php?id=26658`.

Ruszczyńsky, A. (1986). A regularized decomposition method for minimizing a sum of polyhedral functions. *Mathematical Programming 35*, 309–333.

Salas, J. D., J. W. Delleur, V. Yevjevich, and W. L. Lane (1980). *Applied modeling of hydrologic time series*. Littleton, Colorado: Water Resources Publications.

Salas, J. D., G. D. Tabois, and P. Bartolini (1985). Approaches to multivariate modeling of water resources time series. *Water Resources Bulletin 21*, 683–708.

SAS Institute Inc. (1999). SAS Onliedoc.

Schultz, R. (1994). Strong convexity in stochastic programs with complete recourse. *Journal of Computational and Applied Mathematics 56*, 3–22.

Schultz, R. (2000). Some aspects of stability in stochastic programming. *Annals of Operations Research 100*, 55–84.

Schultz, R. (2003). Stochastic programming with integer variables. *Mathematical Programming 97*(1–2), 285–310.

Shapiro, A. (2003). Statistical inference of multistage stochastic programming problems. *Mathematical Methods of Operations Research 58*(1), 57–68.

Shapiro, A. (2006). On complexity of multistage stochastic programs. *Operations Research Letters 34*(1), 1–8.

Sherkat, V. R., R. Campo, K. Moslehi, and E. O. Lo (1985). Stochastic long-term hydrothermal optimization for a multireservoir system. *IEEE Transactions on Power Apparatus and Systems 104*(8), 2040–2049.

Slyke, R. M. V. and R. Wets (1969). L-shaped linear programs with applications to optimal control and stochastic programming. *SIAM Journal of Applied Mathematics 17*(4), 638–663.

Takriti, S., J. R. Birge, and E. Long (1996). A stochastic model for the unit commitment problem. *IEEE Transactions on Power Systems 11*(3), 1497–1508.

Takriti, S., B. Krasenbrink, and L. S.-Y. Wu (2000). Incorporating fuel constraints and electricity spot prices into the stochastic unit commitment problem. *Operations Research 48*(2), 268–280.

Terry, L. A., M. V. F. Pereira, T. A. A. Neto, L. E. C. A. Pinto, and P. R. H. Sales (1986). Coordinating the energy generation of the Brazilian national hydrothermal electrical generating system. *INTERFACES 16*(1), 16–38.

Tiao, G. C. and G. E. P. Box (1981). Modeling multiple time series with applications. *Journal of American Statistical Association 75*, 802–816.

Triki, C., P. Beraldi, and G. Gross (2005). Optimal capacity allocation in multi-auction electricity markets under uncertainty. *Computers and Operations Research 32*, 201–217.

Tufegdzic, N., R. J. Frowd, and W. O. Stadlin (1996). A coordinated approach for real-time short term hydro scheduling. *IEEE Transactions on Power Systems 11*(4), 1698–1272.

Walkup, D. W. (1969). Stochastic programs with fixed recourse: The equivalent deterministic program. *SIAM review 16*(3), 309–338.

Wallace, S. W. and S.-E. Fleten (2003). Stochastic programming models in energy. In A. Ruszczyński and A. Shapiro (Eds.), *Stochastic programming*, Volume 10 of *Handbooks in OR & MS*, pp. 637–677. North-Holland: Elsevier.

Wen, F. and A. K. David (2002). Coordination of bidding strategies in day-ahead energy and spinning reserve markets. *Electrical Power and Energy Systems 24*, 251–161.

Wen, F. S. and A. K. David (2001a). Optimal bidding strategies for competitive generators and large consumers. *Electrical Power and Energy Systems 23*, 37–43.

Wen, F. S. and A. K. David (2001b). Strategic bidding for electricity supply in a day-ahead energy market. *Electric Power Systems Research 59*, 197–206.

Wets, R. J.-B. (1974). Stochastic programs with fixed recourse: The equivalent deterministic program. *SIAM review 16*, 309–338.

Wets, R. J.-B. (1983). Stochastic programming: Solution techniques and approximation schemes. In A. Bachem, M. Groetschel, and B. Korte (Eds.), *Mathematical Programming*, Volume 1982 of *State-of-the-Art*, pp. 506–603. Springer.

Yu, C. W., X. S. Zhao, F. S. Wen, C. Y. Chung, T. S. Chung, and M. X. Huang (2005). Pricing and procurement of operating reserves in competitive pool-based electricity markets. *Electric Power Systems Research 73*, 37–43.

Subject index

Notation index

Symbols