DECOMPOSITION TECHNIQUES IN ENERGY RISK MANAGEMENT

OKTAY SÜRÜCÜ

September 2005

DECOMPOSITION TECHNIQUES IN ENERGY RISK MANAGEMENT

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF APPLIED MATHEMATICS
OF
THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

OKTAY SÜRÜCÜ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF FINANCIAL MATHEMATICS

SEPTEMBER 2005

Approval of the Graduate School of Applied Mathematics

Prof. Dr. Ersan AKYILDIZ
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. Hayri KÖREZLİOĞLU
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Assist. Prof. Dr. Esma GAYGISIZ
Supervisor

Examining Committee Members

Assist. Prof. Dr. Esma GAYGISIZ _____

Prof. Dr. Osman SEVAİOĞLU _____

Prof. Dr. Gerhard Wilhelm WEBER _____

Abstract

DECOMPOSITION TECHNIQUES IN ENERGY RISK MANAGEMENT

Oktay Sürücü
M.Sc., Department of Financial Mathematics
Supervisor: Assist. Prof. Dr. Esma Gaygısız

September 2005, 109 pages

The ongoing process of deregulation in energy markets changes the market from a monopoly into a complex one, in which large utilities and independent power producers are no longer suppliers with guaranteed returns but enterprisers which have to compete. This competence has forced utilities to improve their efficiency. In effect, they must still manage the challenges of physical delivery while operating in a complex market characterized by significant volatility, volumetric uncertainty and credit risk. In such an environment, risk management gains more importance than ever.

In order to manage risk, first it must be measured and then this quantified risk must be utilized optimally. Using stochastic programming to construct a model for an energy company in liberalized markets is useful since it provides a generic framework to model the uncertainties and enable decisions that will perform well. However, the resulting stochastic programming problem is a large-scale one and decomposition techniques are needed to solve them.

Keywords: Energy markets, risk management, stochastic programming, decomposition techniques.

Öz

ENERJİ RİSKİ YÖNETİMİNDE DEKOMPOSİZYON TEKNİKLERİ

Oktay Sürücü
Yüksek Lisans, Finansal Matematik Bölümü
Tez Yöneticisi: Yrd. Doç. Dr. Esma Gaygısız

Eylül 2005, 109 sayfa

Enerji marketlerinde hala devam etmekte olan özelleştirme süreci, bu marketlerin monopol yapısını daha kompleks hale dönüştürmektedir. Önceden kazançları garanti olan büyük şirketler ve bağımsız enerji üreticileri bu değişimle birlikte rekabet halinde olan girişimcilere dönüşmektedirler. Bu rekabet, şirketlerin verimliliklerini artırmalarına zorlamaktadır. Fiziksel sevkiyatın yanı sıra; büyük volatilite, hacimsel belirsizlik ve kredi riskiyle tanımlanan bir markette yer almanın zorluklarıyla da mücadele etmektedirler. Böyle bir ortamda risk yönetimi evvelkinden çok daha fazla önem kazanmaktadır.

Riski yönetebilmek için ilk olarak riskin ölçümü yapılmalı ve bundan elde edilen risk miktarı en uygun şekilde kullanılmalıdır. Özelleştirilmiş bir markette yer alan bir enerji firmasının modelini oluştururken stokastik programlama kullanması, bu yöntemin belirsizlikleri modellemede iyi olmasından ve iyi sonuç vermesinden dolayı doğru bir adım olacaktır. Fakat ortaya çıkacak problemin genişliği sonuca ulaşmak için dekomposizyon tekniklerini gerekmektedir.

Anahtar Kelimeler: Enerji piyasaları, risk yönetimi, stokastik programlama, dekomposizyon teknikleri.

To my niece Dide...

Acknowledgments

I would like to take this as an opportunity to express my sincere gratitude to my thesis supervisor, Assist. Prof. Dr. Esma Gaygısız for her patience, constant guidance, valuable advice and encouragement in the completion of this thesis.

I would also like to thank my family and friends for their support, encouragement and concern that they have been giving me all the time.

I should not forget any of the members of the Institute of Applied Mathematics, especially Prof. Dr. Aydın Aytuna and Nejla Erdoğdu, who motivated me throughout this work.

Last but not least, hugs to Ayşegül İşcanoğlu, Başak Şener, Çağrı Ökten, Çekdar Vakıfahmetoğlu, Derya Altıntan, Emre Tiftik, Gül Değirmenci and Yeliz Yolcu Okur for their kindness and much else.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

In the world of business, all firms face a variety of risks, which affect their financial performance and valuation. To stay alive and to improve itself, a firm should consider these risks in depth. At this point, risk management for firms in all business becomes a major subject. This is also supported by the recent surveys, which showed that risk management is ranked by financial executives as one of their most important objectives. The first step to manage risk is to quantify it via risk measures in order to be able to adjust the portfolios of firms in a convenient way. The choice of the risk measure is a crucial factor, since it determines in which way extreme losses are avoided. However, only knowing the amount of risk is not enough. In order to manage risk, quantified risk must be utilized optimally. Firms generally measure their risk relative to their portfolios. In risk management they either maximize expected return of their portfolio subject to constraints on risk of it or minimize risk of the portfolio subject to constraints on expected return of it.

When an energy company in liberalized markets is considered, risk management becomes much more vital, since the company operates in a much more complex market, where there is high volatility in terms of commodity prices and traded volumes. Moreover lacking liquidity, tremendous volatility, non-normal distributions and market incompleteness are further characteristics of energy mar-

kets that differs it from financial markets. Being a major energy market, the electricity market especially distinguishes itself by the transmission constraints and the non-storability of electricity. Because of these characteristics that differ energy markets form financial ones, the risk management ideas developed for financial markets are not directly applicable to them.

Beside this, a portfolio for an energy company may include the generation facilities it owns other than financial instruments e.g. nuclear, hydro, wind, gas and thermal plants. To be able to fulfill its obligations, which may be the contracts that it sold, the company may add some other contracts (contracts of other companies) to its portfolio from an energy market. Hence, the resulting portfolio needs additional attention in term of risk management.

Portfolios of energy firms can be modelled via stochastic programming, since it provides a generic framework to model these uncertainties and enable decisions that will perform well in the general case. In stochastic programming uncertain data is assigned a probability distribution and a mathematical programming model is formed, whose successful solution will yield a decision that will outperform decisions based on deterministic data.

The size of a stochastic programming problem grows proportionally to the number of possible realizations of the uncertain parameters. Since the decisions are taken under uncertainty, it is usually advantageous to consider numerous possible scenarios over a large number of time periods. Thus, the resulting stochastic program is a large-scale one and the solution of it is intractable. The only means available to solve them is via decomposition. Decomposition refers to the strategy of breaking up a large, difficult-to-solve problem into smaller, easier-to-solve problems, such that the solutions to the decomposed problems can be used to

obtain the solution to the larger problem.

This thesis aims to study decomposition techniques to solve large scale stochastic programming models constructed in order to manage risk of a firm in energy markets.

In Chapter 2, detailed information about energy markets is introduced. In particular, causes for deregulation, its effects on these markets and the characteristics of energy prices are provided.

In Chapter 3, the subject of risk management, including risk measures and risk management via optimization, is discussed with the aim of applying it to energy markets.

In Chapter 4, four examples of energy models are presented in order to show the structure of these models and to see how they can be solved.

In Chapter 5, stochastic programming, by which optimization problems under uncertainty are modelled, and decomposition techniques that help to solve large scale stochastic programming problems are introduced.

Finally, the conclusion is given in Chapter 6.

Chapter 2

Energy Markets

Energy is one of the major vital points for business. The reason for this is the extensive usage of it, which gives birth to the construction of markets for energy. Energy markets all around the world are in change due to deregulation, causing these markets to change exceedingly. To have a meaningful understanding, the characteristics of energy markets before and after deregulation are considered in this chapter.

Before deregulation, energy markets were monopoly markets, in which there were no or little freedom for market consumers to chose their electricity suppliers, for example. There were only one or few market players who could directly influence the market, due to this monopoly structure of the markets. Thus, risk management was often reducible to pure price risk management and at least the major utilities were able to transfer the realized economic losses ultimately back onto the consumer base. As a consequence, risk management issues demanded less management attention in companies. The basic structure of an electricity market before deregulation is shown in Figure 2.1.

Deregulation, one of the primal objective of which is to introduce competition in the wholesale market, takes place in energy markets in many parts of the world. The changes in electricity markets were initiated by a realization that generation and distribution functions need not be monopolies, a feeling that public

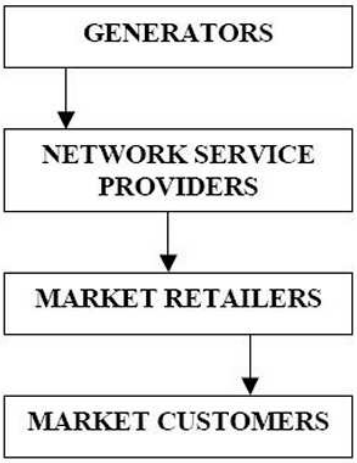

Figure 2.1: *Electricity market participants structure before deregulation.*

service obligations are no longer necessary, the cost reduction potential of competition, increased fuel availability and fuel supply stability, the development of new technologies in power generation and information technology [23].

According to Lai [23], the competitive generation coming with deregulation has four main advantages; cost saving, development of spot market, standardization of market and innovation.

After deregulation, the structure of the markets has changed. There are many market players, and consumers are able to chose their suppliers according to their needs. Figure 2.2, which shows the electricity market participants structure of Australia after deregulation, illustrates this situation. So, large utilities and independent power producers are no longer suppliers with guaranteed returns but enterprisers which have to compete. For example, a new management act became law in Germany on 29 April 1998. Now, German industry as well as private householders have the possibility of choosing their energy supplier for electricity

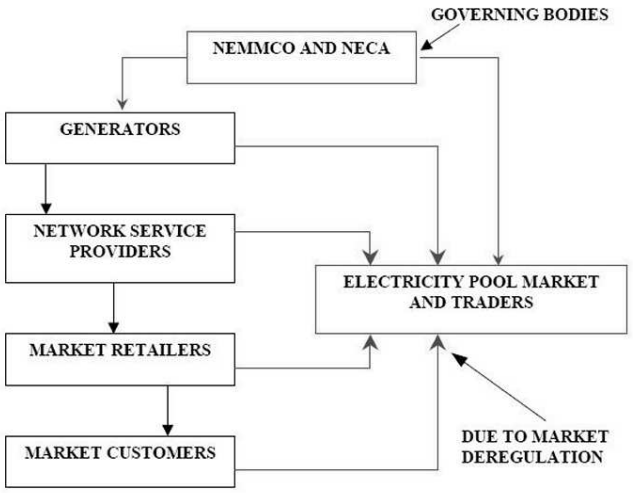

Figure 2.2: *Market participants structure after deregulation (Australia).*

and gas [27]. This competence has forced utilities to improve their efficiency. In effect, they must still manage the challenges of physical delivery while operating in a complex market characterized by significant volatility, volumetric uncertainty and credit risk.

With deregulation, many energy markets have been established all around the world and some major ones are introduced below. When it comes to the situation in Turkey, to stabilize wholesale prices, a transition period program has being applied until the period when energy prices will be totally determined in a competitive environment. Two main scopes of this program are to guarantee the safety of supply by giving opportunity to powerful firms to enter the market and to grant enough time to market participants, so that they can accommodate the competitive environment. At the end of this transition period, it will be possible to establish a energy exchange, where the prices are totally determined by the market, as in developed countries [4].

2.1 Major Energy Markets

International Petroleum Exchange (IPE, London, UK)

A group of energy and futures companies founded the IPE in 1980 in London and the first contract, for Gas Oil futures, was launched the following year. Now, it is Europe's leading energy futures and options exchange. The IPE provides a highly regulated marketplace where industry participants use futures and options to minimize their price exposure in the physical energy market. Over $8 billion daily in underlying value is traded on the IPE [44].

New York Mercantile Exchange (NYMEX, New York,USA)

The New York Mercantile Exchange is the world's largest physical commodity futures exchange and the preeminent trading forum for energy and precious metals. The wide array of trading markets provided by the Exchange include futures and options contracts for crude oil, gasoline, heating oil, natural gas, electricity, gold, silver, copper, aluminum, and platinum; futures contracts for coal, propane, and palladium; and options contracts on the price differentials between crude oil and gasoline, crude oil and heating oil, Brent and West Texas Intermediate crude oil, and various futures contract months (calendar spreads) for light, sweet crude; Brent crude; gasoline; heating oil; and natural gas [40].

Tokyo Commodity Exchange (TOCOM, Tokyo, Japan)

The Tokyo Commodity Exchange was established in 1984 and in 1999, future trading on gas oil was started. Now, it has 10 listed commodities for futures trading including gasoline, kerosene, gas oil and crude oil [41].

Singapore Exchange (SGX, Singapore, Singapore)

Singapore Exchange Limited (SGX) is Asia-Pacific's first integrated securities and derivatives exchange and was established in 1999, following the merger of two established and well-respected financial institutions - the Stock Exchange of Singapore (SES) and the Singapore International Monetary Exchange (SIMEX). Besides security products, derivative products including energy futures are traded in this exchange [42].

Nordic Power Exchange (Nord Pool, Norway)

Norway was the first of the Nordic countries to deregulate its power markets. The Energy Act of 1990, formed the basis for deregulation in the other Nordic countries. Nord Pool was established in 1993, and is owned by the two national grid companies, Statnett SF in Norway (50%) and Affärsverket Svenska Kraftnät in Sweden (50%). It is the world's only multinational exchange for trading electric power [43].

After the constructions of energy markets, the prices are no longer determined by the regulator, but by the market according to demand and supply. There is no correlation between quantity and price. For example, on Nord Pool, the Norwegian power exchange, for one hour more than 100 MWh was treated for a price of over 160 Norwegian crowns. Some hours later, it was possible to achieve prices of less than 50 Norwegian crowns [27]. Thus, in a deregulated market, the monetary expenses are determined by volatility of energy prices on the spot and future exchanges. This forces the energy supplier as well as purchaser to learn the characteristics of energy prices and act in a purposeful manner in the market in order to safeguard favorable power prices. Thus, constructing a portfolio of

power contracts is a vital way to cope with unexpected market movements and so to reduce the magnitudes of the risk. In this way, financial instruments, earlier familiar from financial markets, have also appeared in these markets, such as different types of swap, option, and combination of these instruments. A major characteristic of energy derivative products is that a number of them cannot be found in any other markets. For example, various volumetric options, which have been developed to manage risk associated with meeting the demand in natural gas or power, have no parallel in financial markets.

The main characteristic of energy sources like electricity, oil and gas is that their volatility in terms of commodity prices and traded volumes, which is shown in Figure 2.3. Especially electricity has exhibited enormous price volatility, which is followed by natural gas and petroleum. This high volatility results in risk for both suppliers and consumers. Some of the main reasons why there is such enormous price volatility are:

- the imbalance between supply and demand: demand for energy in the market changes very rapidly, for example the need for electricity is high during summer due to cooling problems;
- political incidents;
- scarcity of energy sources: energy sources like oil and gas come only from some part of the world;
- storage problems: unlike other physical commodity, energy can be stored only in limited quantities;
- weather conditions: this is crucial to produce energy, for example, for a

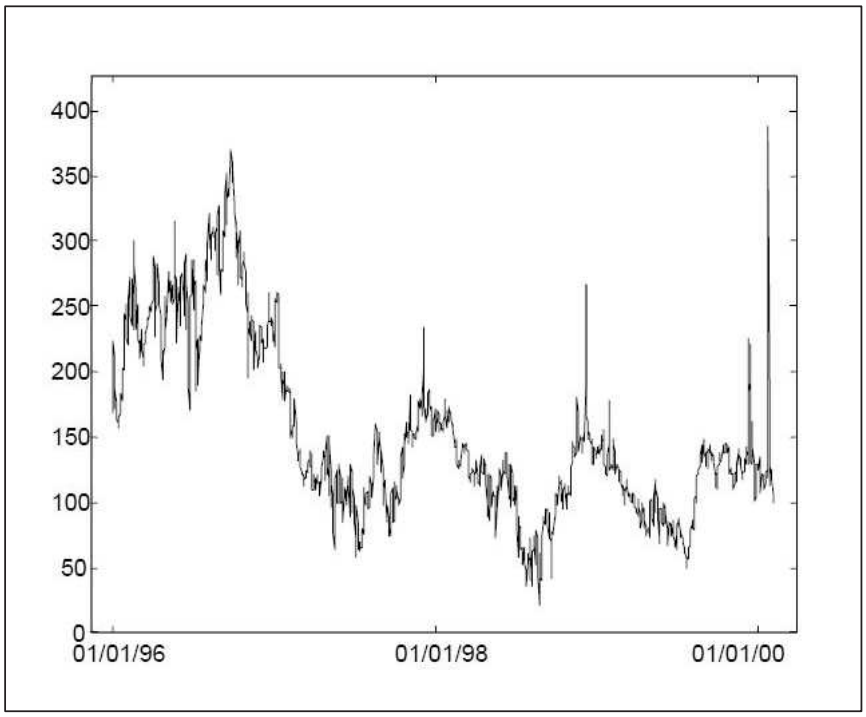

Figure 2.3: *Average daily price per MWh in NOK (Norwegian krone) at Noor Pool*

hydro plant to produce energy rain is the vital point.

Due to this high volatility of energy prices, the energy industry has been considered as boom-bust business. Institutional investors' holdings of energy stocks have shrunk from 30% in the late 1980s to only 6.5% in 2000, which is the result of highly unstable nature of energy sector earnings [21]. This high volatility causes risk management to step into the energy business.

Although the discipline of financial engineering has developed, many techniques for financial markets including commodity markets, most of these techniques should not be simply applied to power markets because of the inherent differences between energy and any other commodity (see Table 2.1).

The number of fundamental price drivers in energy markets are much more

Issue	In Money Markets	In Energy Markets
Maturity of market	Several decades	Relatively new
Fundamental price drivers	Few, simple	Many, complex
Impacts of economic cycles	High	Low
Frequency of events	Low	High
Impact of storage and delivery; the convenience yield	None	Significant
Correlation between short and long-term pricing	High	Low, split personally
Seasonality	None	Key to natural gas and electricity
Regulation	Little	Varies from little to very high
Market activity (liquidity)	High	Low
Market centralization	Centralized	Decentralized
Complexity of derivative contracts	Relatively simple	Relatively complex

Table 2.1: *Differences between money and energy markets* [29].

than in the money markets. For example, unlike in money markets, the supply side in energy markets concerns how to get the actual commodity besides the storage and transfer problems. Each of energy participants deals with a different set of fundamental drivers, which affect the behavior of energy markets, indeed they cause extremely complex price behavior [29].

The impact of economic cycles, which can be measured by mean reversion, is high in money markets, whereas it is low in energy markets. Mean reversion is the process of a market returning its equilibrium level. Unlike energy markets, money markets exhibit relatively weak mean reversion and the rate of mean reversion appears to be related to economic cycles. Hence, it is possible to introduce the state of economy to financial models by means of mean reversion. However, the mean reversion in energy markets appears to be related to the time in which the supply side of the market can react to events, or in which the events go away [29].

In energy markets, there are also two supply drivers, production and storage, and they do not exist in money markets. When considering long-term effects, expectations of market production capacity and cost need to be considered in the long run. However, the technological improvements of drawing gas from the ground, for example, can not be obtained from historical data, but the expectation of it may be expressed by knowledgable traders in determining forward prices [29].

One of the reason why energy markets have much higher volatility than money markets is the storage limitations. Electricity represents the extreme case of this limitation issue. In fact, electricity can not be stored. It is possible for electricity prices to reach levels in multiples of mean price levels. This may happen if all the plants reach their maximum baseload and there is still need for electricity [29].

The correlation between short and long-term pricing is low in energy markets. In fact, energy prices show a split personality, which is due to storage problems. Pilipović, in his book "Energy Risk" [29], states that: "Short-term forward prices reflect the energy currently in storage, while long-term forward prices exhibit the behavior of future potential energy, i.e., energy in the ground."

The usage of derivative contracts for industrial users is to keep plants running. For example, a factory manager protects himself from paying the cost of shutting down and restarting the factory in case there is high prices or lack of available supply, by using derivative contracts. In this context, the net benefit minus the cost for contracts may be defined as the convenience yield, which has a significant role in energy markets.

Seasonality is another major subject in energy markets, especially for natural gas and electricity, whereas it is irrelevant for money markets. For example, heating oil is consumed highly during winter, which leads prices to show great

increase at that season and to return to their mean level during summer. An example of energy asset which peaks during summer is electricity, due to its widely usage for air conditioners at that season.

Energy markets are in the process of deregulation all around the world, and continue to evolve in terms of theoretical sophistication, contract complexity and standardization. Since they are relatively new markets, they suffer from lack of historical price information, as well as relatively small volumes of market activity, which is referred to as an illiquid market [29].

Unlike money markets, energy markets are highly decentralized, thus price depends on location. For example, a megawatt of electricity is priced according to delivery point, whereas the price of a stock in a money market is same everywhere.

Another difference between money and energy markets is the complexity of derivative contracts. Needs of end users makes energy contracts to exhibit a complexity of price averaging and customized characteristics of commodity delivery. Thus the use of derivatives to hedge, which is introduced in the next chapter, is also difficult and complex.

Summary

With the introduction of deregulation, energy markets have turned from monopoly markets to free markets. Electricity, for example, is no longer just a basic commodity to the public but also a trading commodity with the electricity market entering the picture. The main advantage of deregulation energy markets is the increase in the choice of suppliers to the consumers. Market competition will cause the suppliers to provide quality services and competitive pricing. The consumer of electricity, for example, will be expecting lower electricity bills

with the competitive pricing from the retailers and spot market. Prices are no longer determined by the regulator, but by the market and especially electricity shows enormous price volatility, which result in risk for both suppliers and consumers. Furthermore, when an energy company which produces energy through its hydro plants are considered, the risks that it deals with are plenty; weather risk, volume risk, generational risk, etc. Thus, a great deal of attention should be given to risk management in energy markets. However, the strategies from money markets cannot be used in energy markets, due to their characteristic differences explained in this chapter. This forces a risk manager to be much more careful in energy markets. The next chapter concentrates mainly on the subject of risk management, including risk measures and risk management via optimization.

Chapter 3

Risk Management

Uncertainty, which means that what will happen is not known exactly, plays a very important role in the world of finance, as it does everywhere. Since uncertainty is involved, there is a possibility for a trader or for a firm to loose a great deal of money, beside gaining a lot. A rational decision maker would like to take some actions in order to avoid losing, which is done by risk management.

To be able to manage risk, it should be measured first. The choice of the risk measure is a crucial factor, since it determines in which way extreme losses are avoided. In section 3.1 risk measures; variance, semi-variance, Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) are introduced. Risk management is the process of achieving the desired balance of risk and return through particular trading strategies. This balance is achieved via using optimization, where either expected return is maximized given a risk level or risk is minimized given expected return. Section 3.2 deals with risk management via optimization.

3.1 Risk Measures

A risk measure, say ρ, is a mapping between the space X of random variables and a non-negative real number, i.e., $\rho : X \to \mathbb{R}^+ \cup \{0\}$. For example, consider two assets whose returns are X_1 and X_2, and their correspondence risk measures

are $\rho_1 = \rho(X_1)$ and $\rho_2 = \rho(X_2)$. It is possible to compare these assets according to their respective risk value through ρ_1 and ρ_2.

Any acceptable risk measure, ρ, must satisfy the following properties:

- Positive homogeneity: $\rho(\lambda x) = \lambda\rho(x)$ for all random variables x and all positive real numbers λ.
- Subadditivity: $\rho(x + y) \leq \rho(x) + \rho(y)$ for all random variables x and y.

It can be proved that any positively homogeneous functional ρ is convex if and only if it is subadditive ($\rho(\lambda x + (1 - \lambda)y) \leq \rho(\lambda x) + \rho((1 - \lambda)y) = \lambda\rho(x) + (1 - \lambda)\rho(y)$).

Furthermore, if the following properties are satisfied:

- Monotonicity: $x \leq y$ implies $\rho(x) < \rho(y)$ for all random variables x and y;
- Transitional invariance: $\rho(x + \alpha r) = \rho(x) - \alpha$ for all random variables x and real numbers α, and all riskless rates r,

then ρ is a coherent risk measure, which is introduced by Artzner et al. [3].

Through this section, risk measures that can be used in the energy sector will be introduced and discussed whether they are appropriate or not. The first risk measure to be considered is variance.

3.1.1 Variance

The mean-variance portfolio problem (Markowitz [25]) is the standard portfolio optimization approach in the traditional financial markets. In this approach, the portfolio variance is minimized subject to constraint on the expected return.

The variance of a random variable Y is defined as follows

$$V := p_1(y_1 - E)^2 + p_2(y_2 - E)^2 + ... + p_N(y_N - E)^2,$$

where Y is assumed to have finite number of values $y_1, y_2, ..., y_N$ and the probability that $Y = y_i$ is p_i $(i = 1, 2, ..., N)$. Furthermore, the notation E stands for the expected value of Y, which is defined as follows

$$E := p_1 y_1 + p_2 y_2 + ... + p_N y_N.$$

The return on a portfolio consisting of M assets is a weighted sum of M random variables, $R_1, R_2, ..., R_M$. Let μ_i be the expected return of R_i and h_i the weight of the of i-th asset in the portfolio such that $\sum_{i=1}^{M} h_i = 1$. Then, the expected return of the portfolio is

$$E = \sum_{i=1}^{M} h_i \mu_i$$

and the variance is

$$V = \sum_{i=1}^{M} \sum_{j=1}^{M} \sigma_{ij} h_i h_j,$$

where σ_{ij} is the covariance between i-th and j-th assets and defines as

$$\sigma_{ij} := E\{ [R_i - E(R_i)] [R_j - E(R_j)] \}.$$

The advantage of using the variance for describing the risk component of a portfolio, is the simplicity of the computation. However, variance is inappropriate to describe the risk of low probability events and mean-variance decisions are

usually not consistent with the expected utility approach, unless returns are normally distributed. Furthermore, variance is a symmetric measure, i.e., variance includes both up and down deviations from the expected return of the portfolio. Rational investors are interested in minimizing downside risk, since only downside risk is relevant to an investor. In this point, many asymmetric risk definitions are made.

3.1.2 Semivariance

One of the most commonly used downside risk measure is semivariance, which has two types, as below-mean semivariance (SVm), in which the returns that are greater than the expected return are excluded, and below-target semivariance (SVt), in which the returns that are greater than a target return are excluded. In mathematical notations,

$$\begin{aligned} SVm &= \frac{1}{K}\sum_{t=1}^{K}\max[0,(E-R_t)]^2, \\ SVt &= \frac{1}{K}\sum_{t=1}^{K}\max[0,(tar-R_t)]^2, \end{aligned}$$

where K is the number of observations, E is the expected return (mean), tar is the target rate of return and R_t is the asset return during time period t.

The main disadvantage of semivariance over variance is its computational tractability, since the semivariance optimization models using a semicovariance require twice the number of data inputs than the variance model.

Another famous non-symmetric risk measure is Value-at-Risk (VaR), which is widely used in portfolio optimization.

3.1.3 Value-at-Risk

It is defined to be the maximum loss of a portfolio over a given period of time, with a given level of confidence, i.e., VaR_β of a portfolio is the lowest amount α such that, with probability β, the loss will not exceed α in a given period of time.

Let the random variable $l(x, y) \in \mathbb{R}$ be the loss function of a decision variable $x \in \mathbb{R}^n$, which can be seen as a portfolio, and the random vector $y \in \mathbb{R}^m$, which stands for uncertainties that can affect the loss, with distribution $p(y)$. The VaR_β associated with x and any specified probability level $\beta \in (0, 1)$ is given by

$$VaR_\beta(x) := \min\{\alpha \in \mathbb{R} : \int_{l(x,\, y) \leq \alpha} p(y)dy \geq \beta\}.$$

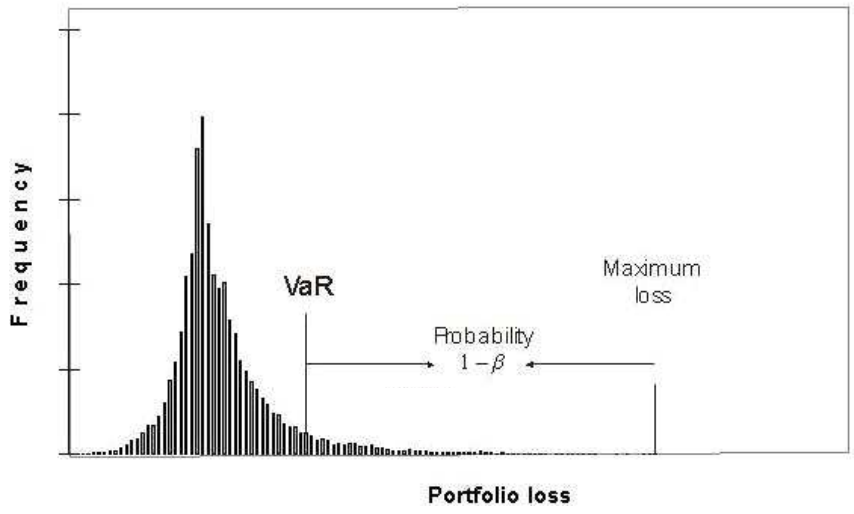

Figure 3.1: *Calculating VaR from portfolio loss distribution*

This definition of VaR is equivalent to the following one:

$$VaR_\beta(X) = q_\beta(-X),$$

where $q_\beta(X)$ is the β-quantile of X and defined as follows

$$q_\beta(X) := \inf\{x \in \mathbb{R} : P[X \leq x] \geq \beta\}.$$

VaR satisfies all the conditions to be a coherent risk measure, but subadditivity, hence it is not a coherent measure of risk. In fact, the notion of coherent risk measures is introduced by Artzner et al. [3] as a response to the deficiencies of VaR. Subadditivity condition is satisfied by VaR for the special case where the joint distribution of return is elliptic. Indeed, VaR is not an acceptable risk measure because of its following properties:

- the losses exceeding VaR are not measured,
- at different confidence levels, conflicting results may be provided,
- it is not sub-additive, which means it is not a coherent risk measure (this lack of property implies that portfolio diversification may lead to an increase of risk, which is totally wrong),
- it is not convex, so it is impossible to use VaR in optimization problem,
- sensitive when applied to discontinuous distributions,
- it has many local extremes leading to unstable VaR ranking [35].

Despite of these problems, VaR is widely used. The reasons for this are:

- it is a compact representation of risk level,
- it applies to any financial instrument and it is expressed in the same unit of measure, namely in lost money,
- it measures downside risk.

In the search for a suitable alternative to value-at-risk, Conditional Value-at-Risk has been characterized as the smallest coherent risk measure to dominate value-at-risk.

3.1.4 Conditional Value-at-Risk

CVaR, developed by Rockafellar and Uryasev [30], is closely related to VaR and defined as the expected value of the losses exceeding VaR, for continuous distributions. For general distributions, including discrete distributions, CVaR is defined as the weighted average of VaR and losses strictly exceeding VaR, i.e., $CVaR_\beta$ is the conditional expectation of losses above VaR_β. In mathematical notations,

$$CVaR_\beta(x) := \frac{1}{1-\beta} \int_{l(x,\, y) \geq VaR_\beta(x)} l(x,y)p(y)dy,$$

where again $l(x,y)$ is the loss associated with the decision vector $x \in \mathbb{R}^n$ and the random vector $y \in \mathbb{R}^m$, and $p(y)$ is the underlying probability distribution of y.

Rockafellar et al. [30] showed that the above definition for CVaR can be characterized by the following convex optimization problem:

$$CVaR_\beta(x) = \min_{(x,\alpha) \in X \times \mathbb{R}} L_\beta(x,\alpha),$$

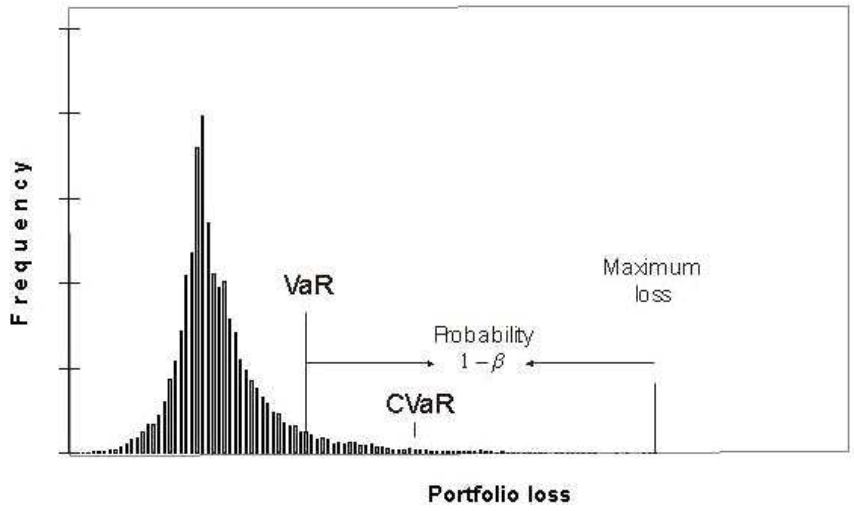

Figure 3.2: *Calculating CVaR from portfolio loss distribution.*

where $L_\beta(x, \alpha)$ is given by

$$L_\beta(x, \alpha) := \alpha + \frac{1}{1-\beta} \int_{y \in \mathbb{R}^m} [l(x, y) - \alpha]^+ p(y) dy,$$

where

$$[\, t \,]^+ := \begin{cases} t, & \text{if } t > 0\,; \\ 0, & \text{otherwise.} \end{cases}$$

The above integration can be approximated by sampling the probability distribution of y according to its density $p(y)$, say $y_1, y_2, ..., y_K$. Then, the corresponding approximation to $L_\beta(x, \alpha)$ is given by

$$\tilde{L}_\beta(x, \alpha) := \alpha + \frac{1}{K(1-\beta)} \sum_{k=1}^{K} [l(x, y_k) - \alpha]^+.$$

Unlike VaR, CVaR is sub-additive and convex. Moreover, it satisfies all the condition to be a coherent risk measure. Due to these powerful properties of it,

CVaR is the most widely used risk measure in energy markets.

Summary

Variance is one of the simplest risk measures used in financial markets. However, it is inconvenience when describing the risk of low probability. Furthermore, its property of being symmetric causes risk managers to look for another measure, since they are interested in minimizing only the downside risk. *Semivariance*, which is a modified version of variance, is a downside risk measure, but it is difficult to compute. Another famous and simple downside risk measure is *Value-at-Risk* (*VaR*). But it is also not a suitable measure to use in energy markets. First of all, according to VaR, portfolio diversification may lead to an increase of risk due to its lack of sub-additivity property. Since it is not sub-additive, it is also not convex, thus not coherent. Moreover, VaR may provide conflicted results at different confidence levels. Another major disadvantage of VaR is that it does not measure the losses exceeding itself, i.e., the losses that might be seen in the tail. A suitable alternative for VaR may be *Conditional Value-at-Risk* (*CVaR*), which is a coherent risk measure. CVaR also quantifies the losses that might be seen in the tail. Unlike VaR, CVaR is convex, so it can be used in optimization problems.

Next section deals with risk management via optimization, in which the quantified risk by a risk measure is either the objective function to be minimized or a constraint to be satisfied.

3.2 Risk Management via Optimization

In the world of finance, all firms construct their own portfolios and adjust it according to their risk perspectives. A Risk managers considers the available assets and decide which ones to buy and to sell. While doing that, he takes the probability of losses into consideration. Different risk measures can be used to quantify losses. The choice of the risk measure depends on the advantages and disadvantages of it in the business that the company does.

In general, portfolio analysis is mainly the process of measuring and achieving the desired risk and return. In classical portfolio theory, there are mainly three subjects to consider: expected return of the portfolio, the risk in that return, and the quantity of each instrument in that portfolio. The latter is the choice variable. Portfolio optimization is simply the search for a vector of quantities of instruments that satisfies a number of constraints and optimizes the objective. When optimizing a portfolio, the objective is either to minimize the risk subject to a constraint on the expected return or to maximize the expected return subject to a constraint on the risk level.

The main feature of portfolio optimization models is the uncertainties they involve. These uncertainties concern the future level of interest rates, yields of stock, exchange rates, inflation, future demand, liabilities, weather conditions, production dispatch, etc..

In this section, different optimization models to manage risk are considered. The first model, best hedge, is an example of hedging for an electricity company. In this example, the concept of hedging and some derivatives are given firstly in order to make the following model understandable. After this example opti-

mization models to manage risk using risk measures variance, Value-at-Risk and Conditional Value-at-Risk are given.

3.2.1 Hedging in Electricity Markets

In a competitive electricity market, some generators selling their power in spot markets face the risk of insufficient spot prices to cover generation costs. On the other hand, consumers face the risk of price variability. Finally, power marketers sell electricity to both wholesale and retail consumers, often at fixed prices. Marketers who buy on the spot market face the risk that the spot market price could substantially exceed fixed prices specified in contracts. Electricity futures and other electric rate derivatives help electricity generators, consumers, and marketers hedge price risk in a competitive electricity market. Other electric rate derivatives include options, price swaps, basis swaps and forward contracts. Futures and options are traded on an exchange where participants are required to post margins to cover potential losses. Other hedging instruments are traded in the "over-the-counter" (OTC) market [33].

Futures contracts are firm commitments to make or accept delivery of a specified quantity and quality of a commodity during a specific month in the future at a price agreed upon at the time the commitment is made. The buyer, known as the long, agrees to take delivery of the underlying commodity. The seller, known as the short, agrees to make delivery. Only a small number of contracts traded each year result in delivery of the underlying commodity. Instead, traders generally offset (a buyer will liquidate by selling the contract, the seller will liquidate by buying back the contract) their futures positions before their contracts mature. The difference between the initial purchase or sale price and the price of

the offsetting transaction represents the realized profit or loss.

There are two types of options, *calls* and *puts*. A call gives the holder, or buyer, the right but not the obligation to buy the underlying commodity at a specific price up to a certain time. A put gives the holder the right, but not the obligation to sell the underlying commodity at a specific price up to a certain time. A call is purchased when the expectation is for rising prices; a put is bought when the expectation is for neutral or falling prices. Furthermore, options on future contracts are also available, where the underlying commodity is a specified future contract.

The basic idea of hedging is to reduce price risk by taking opposite positions in the spot and derivatives markets at the expense of potential reward. For example, a trader, who takes short position in spot market for electricity, should take long position in futures market. When the spot price increases, the profit of the trader will decrease in the spot market, but his profit in the futures market will increase, since future prices exhibit same direction of movement with spot prices. Consequently, his loss from the spot market is decreased due to his gain from the futures market. With a perfect hedge, the loss and the gain are equal, so the total loss is zero. However, perfect hedges are not achievable in electricity markets due to heavily incompleteness characteristic of them. Instead, one may try to find the best possible hedge, which is a hedge that minimizes the risk, under some constraint on the expense of potential reward. Being an appropriate one for electricity markets, CVaR can be used as the measure of risk. Now, the optimization problem is to find the hedge that minimizes the risk, in terms of CVaR under the constraint that the expected return is greater or equal to some threshold R, which is put in order to assure that the expense of reward is not too

high. The mathematical expression of the problem is as follows:

$$\begin{aligned} \min_{x,z,\alpha} \quad & \alpha + \frac{1}{K(1-\beta)} \sum_{k=1}^{K} z_k, \\ \text{subject to} \quad & z_k \geq l(x, y_k) - \alpha, \quad z_k \geq 0, \quad k = 1, ..., K, \\ & -\frac{1}{K} \sum_{k=1}^{K} l(x, y_k) \geq R, \end{aligned} \tag{3.2.1}$$

where $x \in \mathbb{R}^n$, $z \in \mathbb{R}^K$ and $\alpha \in \mathbb{R}$. The contracts to be hedged x^h are held fixed and the other contracts x/x^h are the decision variables to be chosen, such that the best hedge is found.

The best hedge is actually a small scale portfolio optimization, where typically only a part of the whole portfolio is considered, namely, the contracts that are to be hedged. In the following subsections optimization models to manage risk, where the whole portfolio is in consideration, are introduced.

3.2.2 Mean-Variance Portfolio Optimization

Markowitz [25] developed mean-variance model for portfolio optimization in 1952. He minimized the portfolio variance subject to a constraint on the expected return in his model, which is:

$$\begin{aligned} \text{Minimize} \quad & \sum_{i=1}^{M} \sum_{j=1}^{M} \sigma_{ij} m_i m_j, \\ \text{subject to} \quad & \sum_{j=1}^{M} r_j m_j \geq \gamma M_0, \\ & \sum_{j=1}^{M} m_j = M_0, \\ & 0 \leq m_j \leq u_j, \qquad j = 1, 2, \ldots, M, \end{aligned} \tag{3.2.2}$$

where M is the number of assets in the portfolio, σ_{ij} is the covariance between i-th and j-th asset, m_i is the amount of money invested in the i-th asset, r_i is the expected return of the i-th asset, γ is a parameter representing the minimal rate of return required by the investor, M_0 is the initial wealth and u_j is the maximum amount of money which can be invested in the j-th asset. Markowitz [25] used this single stage model to derive an efficient frontier where every portfolio on the frontier maximizes the expected return for a given variance or minimizes the variance for a given expected return. This is usually called the *EV criterion* where E is the expected return and V is the variance of the portfolio.

The generalization of this model to multi-period is studied by Duan Li et al. [24]. They again minimized the portfolio variance subject to a constraint on the expected return and an additional constraint on the reallocation of the wealth among the assets at the beginning of each time period.

3.2.3 Portfolio Optimization via VaR

There are also many examples of portfolio optimization models that maximizes expected return given a level of risk in literature. Kleindorfer et al. [22] considered multi-period VaR constrained portfolio optimization and its application to the electric power sector. They stated that: "The impact of multi-period risk constraints covering a period on decisions affecting cash flows during shorter periods of time is important for energy management since accounting periods are typically longer than the operational periods over which trading decisions must be taken". For this reason, they used an approach, which they called as equally weighted or uniform disaggregation. In this approach, they take the annual VaR and apply it equally to sub-intervals of the year, which gives the VaR for a day,

for example, as $VaR/365$. Furthermore, to capture the multi-period pooling effect of daily decisions they assumed that the annual cashflows are the sum of 365 daily cashflows and that each of these daily cashflows is normally distributed with mean μ_d and standard deviation σ_d. Then, the typical VaR constraint has the following structure:

$$Pr\{X_t \geq -VaR_d\} = \gamma = Pr\{Y = \sum_{t=1}^{365} X_t \geq -VaR_a\},$$

where X_t denotes the daily cashflow, γ is the risk parameter and Y is the annual cashflow from the portfolio. VaR_d and VaR_a are defined as

$$VaR_d := z(\gamma)\sigma_d - \mu_d, \qquad VaR_a := z(\gamma)\sigma_a - \mu_a$$

with

$$\sigma_d = \frac{\sigma_a}{\sqrt{365}}, \qquad \mu_d = \frac{\mu_a}{365},$$

where $z(\gamma)$ is the z-score of the standardized normal random variable.

They also relaxed the normality assumption of VaR to adapt other distributions with fat-tail characteristics and introduced regularity assumption. This regularity assumption states that for any specified parameter vector (μ, σ, γ), where μ is the mean, σ is the standard deviation of the return function (Π) and γ is the confidence level for VaR, there is a real number $h(\mu, \sigma, \gamma)$ such that $Pr\{\Pi \leq \mu - h(\mu, \sigma, \gamma)\} = 1 - \gamma$. If this regularity assumption is satisfied, then the portfolio‘s VaR is

$$VaR(\Pi) = h(\mu(\Pi), \sigma(\Pi), \gamma) - \mu(\Pi).$$

The portfolio considered here is constructed with the available instruments in the market, such as puts, calls, forward, etc., denoted by Q and the random variable for demand at time t by D_t and the random variable for spot price at time t by P_t. Using these assumptions and notations the VaR-constrained problem for T-period can be stated as follows

$$\begin{aligned} \text{maximize} \quad & E\{\Pi(Q,D,P)\} = E\Big\{\sum_{t=1}^{T}\Pi_t(Q,D_t,P_t)\Big\} = \mu(Q,D,P) \\ \text{subject to} \quad & \sum_{t=1}^{T} E\Big\{\Pi_t(Q,D_t,P_t)\Big\} + VaR_0 \geq h(\mu((Q,D,P)),\sigma((Q,D,P)),\gamma), \end{aligned} \tag{3.2.3}$$

where $\sigma(Q,D,P) = \big(VAR\big[\sum_{t=1}^{N}\Pi_t(Q,D_t,P_t)\big]\big)^{1/2}$ and VaR_0 is the desired annual constrained VaR level at confidence level γ. After constructing the problem as above, Kleindorfer et al. [22] characterize the efficient frontier for such problems.

3.2.4 Portfolio Optimization via CVaR

Doege et al. [13] constructed an optimization model using the risk measure Conditional Value-at-Risk (CVaR). The main structure of this model is as follows:

$$\begin{aligned} \text{maximize} \quad & E[-l(x,Y)] \\ \text{subject to} \quad & CVaR_\beta(x) \leq C, \\ & x \in X. \end{aligned} \tag{3.2.4}$$

The random variable $l(x,Y)$ denotes the loss function of a portfolio x and a given random vector Y representing the future values of stochastic variables.

Furthermore, C is the given upper bound of CVaR and X is the set of all feasible portfolios. Here, CVaR is calculated using the loss function, therefore C can be viewed as an upper bound on the loss function in this context. If J scenarios are generated with realizations $w_1, \ldots, w_J$ of the random variable Y, the above model is equivalent to the following one:

$$\begin{aligned}
\text{maximize} \quad & -\frac{1}{J}\sum_{j=1}^{J} l(x, w_j) \\
\text{subject to} \quad & \alpha + \frac{1}{(1-\beta)J}\sum_{j=1}^{J} z_j \leq C, \\
& z_j \geq l(x, w_j) - \alpha \qquad\qquad j = 1, 2, \ldots, J, \\
& \alpha \in \mathbb{R}, \quad x \in X, \quad z_j \geq 0 \qquad j = 1, 2, \ldots, J.
\end{aligned}$$

The above problems of optimization under uncertainty can be reformulated as *stochastic programming* models.

Summary

Portfolio optimization, which is mainly the process of measuring and achieving the desired risk and return, is the main objective of risk management. While optimizing a portfolio the measure by which risk is quantified should be carefully chosen. The choice depends on the advantages and disadvantages of the measure in the business that the company does. In this chapter, different type of optimization models for different choices of risk measures are presented.

In the next chapter, four examples of energy models, which are constructed to optimize the portfolios of some energy companies, are introduced.

Chapter 4

Examples of Risk Management Models in Energy Markets

In this chapter, four examples of optimization models in energy markets are provided. The first example, introduced by Sen et al. [31], considers a power optimization model that is intended as a decision aid for scheduling and hedging (DASH) in the wholesale power market. The methodology used by Sen et al. [31] is based on a stochastic programming model that selects portfolio positions that perform well on a variety of scenarios generated through statistical modelling and optimization. Sen et al. [31] used nested column generation decomposition method to solve the resulting large scale optimization problem.

The second example, in which the problem of a supplier is considered, is given by Takriti et al. [36]. Since demand is not known in advance, the supplier constructs several demand forecasts and associates a set of possible probability measures with these forecasts. Considering the worst-case probability distribution, he then maximizes the expected profit. The resulting problem is a minimax problem and is solved by branch-and-cut technique.

The third example considers a model for the problem of portfolio optimization of energy contracts under point of view of a generating company and is given by Marzano et al. [26]. In this model, Conditional Value-at-Risk is used as a risk

measure and added to the constraint set of the problem to limit the risk acceptable by the company.

In the last example, which was introduced by Guigues et al. [17], Value-at-Risk approaches are applied on the problem of yearly electric generation management. The same notations as in these four papers are used here for convenience.

4.1 Example I

In the example by Sen et al. [31], decision aid for scheduling and hedging (DASH) model for power portfolio optimization is introduced. Sen et al. [31] state that: "this model provides a tool that helps decision-makers coordinate production decisions with opportunities in the wholesale power market."

The model is constructed in such a way that the financial traders reevaluate their power portfolio, including contracts from electricity and gas markets, at the start of each month. Furthermore, while prices in the electricity market vary on an hourly basis, time is partitioned into two, as *on-peak period*, which cover sixteen hour a day, and *off-peak period* for the rest eight hour. Another assumption made is that the forward contracts are monthly, so that planning for period t refers to some month t in the future. Although, the time scale for spot prices can be hourly, for computational ease, the spot market is treated on a daily basis, allowed to fluctuate according to the on-peak and off-peak periods.

For scenario generation procedures, forward prices for the preceding year are available as hourly quotes which are transformed into on-peak and off-peak average prices. The notation $\pi_{\tau\kappa e}$ is used for forward prices, where π is the price and $\tau \in \{1, 2, ..., 52\}$, $\kappa \in \{1, 2, ..., N\}$ (N denotes the last week in which delivery

will happen) and $e \in \{on, off\}$ denotes the contract week, delivery week and segment, respectively. For example, $\pi_{1,8,on}$ is the price (\$/MWh) on January 7^{th} (i.e., end of first week) for on-peak power delivery starting on March 1^{st}. However returns are used to predict prices, i.e., $r_{\tau,\kappa,e} = \frac{\pi_{\tau+4,\kappa,e} - \pi_{\tau,\kappa,e}}{\pi_{\tau,\kappa,e}}$, where $\tau + 4$ denotes the next month's contract, because of the assumption that there are four weeks in a month. Now, a discrete scenario tree may be formed by grouping returns into subsets for each period and modelling the return process as one that allows probabilistic transitions from one subset to another, over time. For ease of computation, two subsets in each period are considered, which are high and low return states. A sampling based procedure which is guided by the recent observations of the return series is adopted to assign high and low values for the return states. The nominal value which is assigned to each state is the median of the corresponding group of that period. Not to overlook extreme events, a combination of medians and extreme values is used. The precise manner, in which one or the other is chosen depends on a heuristic guided by market conditions prior to running the model. Whenever heuristic produces two nodes that are represented by medians, equal conditional probabilities are used and whenever it produces an extreme value for one path, a conditional probability of $\frac{1}{4}$ is associated with that path.

Another assumption made on gas and electricity returns is that they are perfectly correlated so that a scenario obtained from the electricity forward return tree generates a similar scenario from the gas forward return tree [31]. Although forward prices are arranged on a monthly basis, spot prices must be modelled on a daily basis, as on-peak and off-peak prices. The following formulation is used for spot prices during a delivery month: $r^p_{e,d,\tau,w} = r^f_{e,t,w} + \sigma_{e,t} z_{d,t,w}$, where w is

the node number of the forward scenario tree, $\sigma_{e,t}$ is the standard deviation of spot returns and $r^{f}_{e,t,w}$ is the daily equivalent of the forward return on node w for month t.

With these settings, the model formulation is presented in two parts, which are the financial problem and generation costing problem.

4.1.1 The Financial Problem

In this section, the notation used is introduced, after which constraints for this problem are presented.

Notations

- α : maximum liquidity limit coefficient,
- T : number of periods,
- P : regulated power price,
- J_t : the number of segments in period t,
- ξ : {on-peak, off-peak},
- H_e : hours of one on/off peak segment ($H_e = 16h$ for e =on-peak and $H_e = 8h$ for e =off-peak),
- $p(j)$: peak status (on/off) of segment j,
- YL_{0te} : power forward in long position for delivery period t, peak e held initially,

- YS_{0te} : power forward in short position for delivery period t, peak e held initially,
- YG_{0t} : gas forward for delivery period t held initially,
- PT_t : profit target for period t,
- $PP_{\tau te}$: price of energy forward for delivery period t, peak e at contract period τ,
- $PG_{\tau t}$: price of gas forward for delivery period t, at contract time τ,
- PS_{tj} : price of energy in spot market in period t, segment j,
- D_{tj} : electricity demand in period t, segment j,
- $FP_{\tau te}$: power forward for delivery period t, peak e, signed at contract period τ,
- $FP^{+(-)}_{\tau te}$: power forward in long (short) position for delivery period t, peak e, signed at contract period τ, an upper bound is imposed on this variable,
- $FG_{\tau t}$: gas forward in long position for delivery period t, signed at contract period τ,
- $YP_{\tau te}$: total power forward for delivery period t, peak e held at contract period τ,
- $YP^{+(-)}_{\tau te}$: total power forward in long(short) position for delivery period t, peak e held at contract period τ,
- $YG_{\tau t}$: total gas forward for delivery period t held at contract period τ,

- SP_{tj} : power exchanged with spot market in period t, segment j (positive for purchase and negative for sale),
- ZP_{te} : total power forward cost for delivery period t, peak e,
- ZG_t : total gas forward cost for delivery period t,
- C_{tj} : total generation cost in period t, segment j.

Constraints

$$FP_{\tau te} = FP^{+}_{\tau te} - FP^{-}_{\tau te}, \tag{4.1.1}$$

$$YP_{\tau te} = YP^{+}_{\tau te} - YP^{-}_{\tau te}, \tag{4.1.2}$$

$$YP^{+}_{\tau te} = YP^{+}_{(\tau-1)te} + FP^{+}_{\tau te}, \tag{4.1.3}$$

$$YP^{-}_{\tau te} = YP^{-}_{(\tau-1)te} + FP^{-}_{\tau te}, \tag{4.1.4}$$

$$YG_{\tau t} = YG_{(\tau-1)t} + FG_{\tau t}, \tag{4.1.5}$$

$$\sum_{t\in[\tau,T]} FP^{+}_{\tau te} = \alpha \sum_{t\in[\tau,T]} YP^{+}_{(\tau-1)te}, \tag{4.1.6}$$

$$\sum_{t\in[\tau,T]} FP^{-}_{\tau te} = \alpha \sum_{t\in[\tau,T]} YP^{-}_{(\tau-1)te}, \tag{4.1.7}$$

$$ZP_{te} = \frac{J_t}{2} \sum_{\tau\in[1,t]} PP_{\tau te}\, FP_{\tau te}\, H_e \qquad t \in \{1,2,...,T\}, \tag{4.1.8}$$

$$ZG_t = \sum_{\tau\in[1,t]} PG_{\tau t}\, FG_{\tau t} \qquad t \in \{1,2,...,T\}, \tag{4.1.9}$$

$$\sum_{j\in[1,J_t]} [(D_{tj}\ P - SP_{tj}\ PS_{tj})H_{p(j)} - C_{tj}] - \sum_{e\in\xi} ZP_{te} \geq PT_t \qquad t \in \{1,2,...,T\}, \tag{4.1.10}$$

where $\tau \in \{1,2,...,T\},\ t \in \{\tau, \tau+1,...,T\},\ e \in \xi$ for all the constraints above, if any other specification is not indicated.

Constraint (4.1.3) balances power forward in long position, (4.1.4) in short position and (4.1.5) balances gas forward, all at period τ. Constraint (4.1.6) puts limit on maximum liquidity for long position and (4.1.7) for short position, so they provide a way to control the extent to which a portfolio is allowed to change from one period to the next one. Constraints (4.1.8) and (4.1.9) gives the total power and gas cost for delivery period t, respectively. Finally, constraint (4.1.10) controls the monthly profit target. If this last constraint cannot be satisfied, a penalized objective function is used, which is done by including the target within the objective function via a penalty term.

4.1.2 The Generation Problem

In addition to the previous ones, some new notation is used for the generation problem.

Notations

- I : the set of generators,
- d : index of day,
- $j(d)$: indices of the two segments associated with day d,
- $t(d)$: the period associated with day d,

- ML : maximum acceptable daily loss,
- Gas : the set of gas generators,
- $Coal$: the set of coal generators,
- Nuc : the set of nuclear generators,
- CP_t : coal price for period t,
- NP_t : nuclear fuel price for period t,
- Q_i : maximum generation capacity of generator i,
- q_i : minimum generation capacity of generator i,
- L_i : minimum up time requirement for generator i,
- l_i : minimum down time requirement for generator i,
- $F_i(x)$: consumption function of fuel for generation of x due to generator i ($F_i(x) = a_i + b_i x$, where a_i and b_i are parameters),
- W_{itj} : scheduled outage ($W_{itj} = 0$, if outage is scheduled in period t, segment j for generator i; otherwise $W_{itj} = 1$),
- $\bar{w}_{itj}$: forced outage ($\bar{w}_{itj} = 0$, if outage is forced in period t, segment j for generator i; otherwise $\bar{w}_{itj} = 1$),
- TG_{tj} : total generated power in period t, segment j,
- G_{itj} : power generated by generator i in period t,segment j,

- U_{itj} : operation decisions for generator i in period t, segment j ($U_{itj} = 1$, if generator i is on in period t, segment j; $U_{itj} = 1$, otherwise),
- SG_{tj} : consumption of gas in period t, segment j,
- SC_{tj} : consumption of coal in period t, segment j,
- SN_{tj} : consumption of nuclear fuel in period t, segment j,

Constraints

$$YP_{tte} + SP_{tj} + TG_{tj} = D_{tj}, \tag{4.1.11}$$

$$YG_{tt} = \sum_{j \in [1, J_t]} SG_{tj}, \tag{4.1.12}$$

$$SG_{tj} = \sum_{i \in Gas} F_i(G_{itj}), \tag{4.1.13}$$

$$SC_{tj} = \sum_{i \in Coal} F_i(G_{itj}), \tag{4.1.14}$$

$$SN_{tj} = \sum_{i \in Nuc} F_i(G_{itj}), \tag{4.1.15}$$

$$C_{tj} = \frac{ZG_t}{J_t} + NP_t\ SN_{tj} + CP_t\ SC_{tj}, \tag{4.1.16}$$

$$TG_{tj} = \sum_{i \in I} G_{itj}, \tag{4.1.17}$$

$$q_i\ U_{itj} \leq G_{itj} \leq Q_i\ U_{itj}, \tag{4.1.18}$$

$$U_{itj} - U_{it(j-1)} \leq U_{it\tau} \quad \tau \in \{j+1, ..., \min(j + L_i - 1, J_t)\}, \tag{4.1.19}$$

$$U_{it(j-1)} - U_{itj} \leq 1 - U_{it\tau} \quad \tau \in \{j+1, ..., \min(j + l_i - 1, J_t)\}, \tag{4.1.20}$$

$$U_{itj} \leq W_{itj}, \tag{4.1.21}$$

$$U_{itj} \leq \bar{w}_{itj}, \tag{4.1.22}$$

$$\sum_{j \in j(d)} [(D_{tj}\ P - SP_{tj}\ PS_{tj})H_{p(j)} - C_{tj}] - \frac{\sum_{e \in \xi} ZP_{te}}{J_t/2} + ML \geq 0 \qquad \forall d, t = t(d), \tag{4.1.23}$$

where $t \in \{1, 2, ..., T\},\ j \in \{1, 2, ..., J_t\},\ e = p(j),\ i \in I$ for all the constraints above.

Constraint (4.1.11), where YP_{tte} involves an exchange of electricity (i.e., the net physical amount of electricity exchanged) in month t, stands for the demand-generation-forward-spot relationship, (4.1.12) gives total gas consumption for period t. Constraints (4.1.13), (4.1.14) and (4.1.15) show the gas, coal and nuclear fuel consumption for period t, segment j, respectively. Constraint (4.1.16) gives the generation cost for period t, segment j, (4.1.17) gives the total generated power and (4.1.18) the operating range for each generator. Minimum up- and low-time requirements are given in constraints (4.1.19) and (4.1.20), respectively. (4.1.21) stands for the scheduled and (4.1.22) for the forced outage. Finally, (4.1.23) is the maximum daily loss constraint, which is imposed in order to provide a measure of risk control on the decisions. When the target ML is unattainable, the model may become infeasible. In such instances, the user may include such a measure within a penalized objective function for generation problem. Thus, the two constraints, (4.1.10) and (4.1.23), are risk constraints, which are based on profit targets and failure to meet these targets determines the extent of loss.

The objective function for the generation problem is as follows

$$\max \sum_{t,j} [(D_{tj}\ P - SP_{tj}\ PS_{tj})H_{p(j)} - C_{tj}] - \sum_{t,e} ZP_{te}. \tag{4.1.24}$$

This objective function is scenario dependent, thus it reflects the profit/loss of the spot market activity, as well as the cost of power generation under one scenario. By weighting scenarios by their respective probabilities, one can maximize the overall expected profit, which is the complete objective function.

To solve this stochastic programming model, Sen et al. [31] decomposed this programm into interrelated optimization problems which are motivated by a nested column generation (i.e., Dantzig-Wolfe) type method. The autors made several experimental tests to study the robustness of the approach. The experimental evidence suggests that the stochastic programming approach provides a powerful and robust tool for scheduling and hedging in wholesale electricity markets.

4.2 Example II

The example given by Takriti et al. [36] considers a power producer that is interested in selling its short-term excess capacity in the form of firm contracts as the producer. In a firm contract, the maximum amount of power C, known as capacity, that can be delivered during a single time period is specified. The buyer of the contract has the right but not the obligation to consume an amount of power in the range $[0, C]$ for each time period. The producer charges an initial fee, which is for issuing the contract and reserving the needed capacity. This initial fee is a function of the reserved capacity and is denoted by $\bar{r}$. According to the total consumption of power over the contract duration, the producer gets another charge, which is known as energy charge and denoted by $f(d_t)$, where t is the time period and d_t is the demand at that period. Excess capacity that the

producer is willing to sell is often a result of the conservative planning of electric-power generators in anticipation of existing demand obligations, also known as the native load. The producer's problem consists of selecting bids to maximize its profit given the uncertain nature of the native load as well as that of demand of the received bids. Demand uncertainty is modelled by scenarios, which represent possible future demand patterns. While remaining feasible for each scenario, an optimal production schedule is calculated by minimizing the average cost of operating the system. Due to lack of historical data, to construct the scenario tree and assign the associated probabilities with its nodes, expert opinions, which may differ greatly from one to another, are needed. For example, assume that the problem is to maximize the expected profit over three future scenarios. The associated probabilities for the three scenarios according to the first expert is $p_1 \geq p_2 \geq p_3$, and according to the second one is $p_2 \geq p_1$ and $p_2 \geq p_3$. Then, the set of feasible probabilities can be identified as $p_1 = p_2 = \pi$ and $p_3 = 1 - 2\pi$, where π is a parameter satisfying $1/3 \leq \pi \leq 1/2$. Alternatively, $p \in P$, where $P = \{(p_1, p_2, p_3) : p_1 = p_2 = \pi,\ p_3 = 1 - 2\pi,\ 1/3 \leq \pi \leq 1/2\}$. What left after that is the choice of set of probabilities. As mentioned in subsection 5.3.2, in minimax stochastic programs, the worst set of probabilities are considered, i.e., the probability distribution (on future load scenarios) that causes the expected profit to be at its minimum. To solve the problem, a decomposition-based branch-and-cut strategy, which extends the standard mixed-integer programming branch-and-cut algorithm by using a cutting-plane scheme to approximate the minimax objective function in each iteration, is proposed.

4.2.1 The Model

It is assumed that there are J contracts that the producer is dealing with and each of them has a capacity requirement of C_j, $j = 1, 2, ..., J$. As mentioned before $\bar{r}_j$ is the initial fee and $f_j(d_{jt})$ is the payment received by producer, where d_{jt} is the quantity of the commodity delivered during time period t. It is further assumed that $d_{jt} \leq C_j$, since the usage of power that exceeds the capacity results in severe financial penalties. Another assumption is that the number of scenarios is the same for all contracts and is equal to K. The notation d_{jt}^k is used to denote the sampled demand of contract j at time t. The producer expects to receive a payment of r_j, which is the sum of the present revenue of $\bar{r}$ and the expected future revenue of $\sum_{k=1}^{K} p_k \sum_{t=1}^{T} f_j(d_{jt}^k)$. The producer is assumed to have I generating units, and the production of generator i at time period t is denoted by y_{it}^k $i = 1, 2, ..., I$, $t = 1, 2, ..., T$. With these settings, the stochastic programming model is

$$\begin{aligned} \text{maximize}_{x_j, y_{it}^k} \quad & \sum_{j=1}^{J} r_j x_j - \sum_{k=1}^{K} p_k \sum_{t=1}^{T} \sum_{i=1}^{I} g_{it}^k(y_{it}^k) \\ \text{subject to} \quad & \sum_{i=1}^{I} y_{it}^k = d_{0t}^k + \sum_{j=1}^{J} d_{jt}^k x_j, \quad t = 1, ..., T, \; k = 1, ..., K \\ & x_j \text{ binary}, \; y^k \in Y^k, \; k = 1, ..., K, \end{aligned} \tag{4.2.25}$$

where x_j is a binary variable indicating whether contract j is to be accepted $x_j = 1$ or rejected $x_j = 0$. The function $g_{it}^k(y_{it}^k)$ is the cost of producing y_{it}^k units of power and is assumed to be convex. The first constraint of the above problem is to balance supply and demand, where d_{0t}^k represents the native load. Furthermore, $y^k \in Y^k$ denotes the constraints imposed on the production vector,

under scenario k. For example, the production of the operating generator i at time period t cannot exceed or go below its capacity range, i.e., $q_{it}^k \leq y_{it}^k \leq Q_{it}^k$, where $[q_{it}^k, Q_{it}^k]$ is the operating range of unit i.

The model in (4.2.25) is a two-stage stochastic program. The first stage decisions determine the set of contracts to be chosen by assigning x_j the value 0 or 1. The second stage decisions determine an optimal production strategy y_{it}^k in response to the demand dictated by the contracts. The model (4.2.25) assumes a known probability, i.e., $p_k,\ k = 1, 2, ..., K$ are known with complete certainty. As mentioned earlier, a minimax approach is applied to this two-stage stochastic program, which is explained in the following subsection.

4.2.2 Minimax Stochastic Programming

The general form of minimax stochastic programming is

$$\min_x \{c^T x + h(x) \mid x \in X \cap \{0,1\}^J\}, \tag{4.2.26}$$

where $h(x) := \max_p \left\{ \sum_{k=1}^K p_k Q_k(x) \mid (p_1, ..., p_k) \in P \right\}$ and

$$Q_k(x) = \min_y \Big\{ \sum_{t=1}^{T} \sum_{i=1}^{I} g_{it}^k(y_{it}^k) \mid Dy = h_k + T_k x,\ \ y \in Y^k \Big\}.$$

By defining $h(x)$, in which the probabilities that yield maximum expected cost are obtained, the minimax approach is applied to a two-stage stochastic program.

The convexity of g yields $Q_k(x)$ to be convex and finite valued for all $x \in X$. Therefore, the function $h(x)$ is also convex and finite valued over X. The feasible solutions of (4.2.26) are denoted by $x^1, x^2, ..., x^M$; i.e., $X \cap \{0,1\}^J =$

$\{x^1, x^2, ..., x^M\}$. With the help of the property of $h(x)$ being convex, the problem (4.2.26) can be reformulated into the following linear program:

$$\min_{x,\theta}\{c^T x + \theta \mid x \in X \cap \{0,1\}^J,\ (x,\theta) \in S\}, \tag{4.2.27}$$

where $S = \{(x,\theta) \mid h(x^m) + \partial h(x^m)^T(x - x^m),\ m = 1, 2, ..., M\}$ and $\partial h(x^m)$ is a subgradient of h evaluated at (x^m). That is, S is represented using a set of linear constraints, which are binding at all integer solutions x^m. A linear relaxation for (4.2.27) is

$$\min_{x,\theta}\{c^T x + \theta \mid x \in X \cap [0,1]^J,\ (x,\theta) \in \bar{S}\}, \tag{4.2.28}$$

where $\bar{S}$ is defined using a set of valid cuts for S; i.e., $S \subseteq \bar{S}$. A lower bound on the value of (4.2.27) is obtained by solving (4.2.28). Cuts may be added at any point in the branch-and-bound process in order to tighten the approximation $\bar{S}$, as long as they are embedded to all nodes in the tree. Here, binding cuts are added at integer nodes that violate S so that the approximation is exact. As for the root node, it is started without any cuts in the system; i.e., $\bar{S} = \mathbb{R}^{J+1}$, and Benders' decomposition, which is introduced in the next chapter, is used to solve the model. The description of the algorithm is provided in the next subsection.

Branch-and-cut algorithm for solving (4.2.27) [36]

Step 0 Set $\bar{S}$ to $\mathbb{R}^{J+1}$ and $\theta \leftarrow -\infty$. Solve the linear relaxation (4.2.28) as follows

Step a Choose a solution $\tilde{x} \in X \cap [0,1]^J$ and $p \in P$.

Step b For each scenario k, solve the second-stage problem and determine

the optimal objective value $Q_k(\tilde{x})$.

Step c Evaluate $h(\tilde{x}) = \max_p\{\sum_k Q_k(\tilde{x})p_k \mid p \in P\}$. Construct a new cut $\theta \geq h(\tilde{x}) + \partial h(\tilde{x})^T(x - \tilde{x})$ and add it to $\bar{S}$.

Step d If $Q \approx h(\tilde{x})$ or if the maximum number of cuts is reached, go to Step f.

Step e Solve the first-stage problem (4.2.28) and determine a new $\tilde{x}$ and θ. Go to Step b.

Step f This is the end of the root node iteration. Note that $\bar{S}$ contains valid cuts which are added during the solution process. In order to begin the branch-and-bound procedure, define L to be the set of unfathomed problems. Set L so that it has a single element, L^0, which is the linear problem defined by (4.2.28). Set $x^1 \leftarrow \tilde{x}$, n to 1, m^* to -1, and $\bar{z}$ to ∞. As it will become clear, n represents the number of nodes in the tree, m^* points to the optimal node, and $\bar{z}$ serves as an upper bound on the optimal objective value of (4.2.27). Proceed to Step 1.

Step 1 If L is empty, terminate and declare x^{m^*} to be an optimal solution for (4.2.27).

Step 2 Select a problem m from L. Denote the linear program corresponding to this problem by L^m. If L^m is infeasible, set $L \leftarrow L - L^m$, and go to Step 1.

Step 3 Let (x^m, θ^m) be an optimal solution and z^m be the optimum value for problem L^m. If $z^m \geq \bar{z}$, set $L \leftarrow L - L^m$ and go to Step 1.

Step 4 If $x^m \notin \{0,1\}^J$, create two new problems L^n and L^{n+1}, by fixing a non-integer element of x^m to 0 and 1, respectively. Set $L \leftarrow L - L^m$, increment

n by 2, and go to Step 1.

Step 5 Solve the second-stage problems and calculate a new probability measure as in Step b and Step c above. If $c^T x^m + h(x^m) < \bar{z}$, set $\bar{z} \leftarrow c^T x^m + h(x^m)$ and $m^* \leftarrow m$.

Step 6 If $\theta^m = h(x^m)$, set $L \leftarrow L - L^m$ and go to Step 1. Otherwise, update the set $\bar{S}$ by adding a binding cut at (x^m, θ^m); i.e., $\bar{S} \leftarrow \bar{S} \cap \{(x, \theta) \mid \theta \geq h(x^m) + \partial h(x^m)^T (x - x^m)\}$, to all problems $L^m, ..., L^{n-1}$. Set $L \leftarrow L - L^m$. Go to Step 1.

According to Takriti et al. [36], the numerical results obtained with the above algorithm shows that large instances of problems with a finite set of probability distributions can be solved by this approach, with a speed-up ratio of 20-2000 times.

4.3 Example III

An energy company, which looks for a model to optimize its portfolio, is considered in this example, which was introduced by Marzano et al. [26]. In this model, the expected value of the net remuneration is maximized subject to the limit on the Conditional Value-at-Risk acceptable by the company. The model is

as follows:

$$\begin{aligned} \text{maximize} \quad & E_w(f(x,w)) \\ \text{subject to} \quad & C \leq \alpha + \frac{1}{(1-\beta)S} \sum_{s=1}^{S} u_s, \\ & u_s \leq 0, \qquad s = 1, ..., S \\ & u_s \leq f(x,w) - \alpha, \quad s = 1, ..., S \\ & + \text{ some other bounds} \end{aligned} \tag{4.3.29}$$

In this model, CVaR is calculated by using return function, therefore, C is a lower bound for CVaR here. In other words, C is an lower bound for the return function.

The generating company, considered in this example, sells energy through bilateral contract and is obligated to provide energy to the buyer for the price established in the contract. The net remuneration of the company at time t is:

$$\begin{aligned} NR &= p\,x + \pi_t(G_t - x) - C_t\,G_t \\ &= (p - \pi_t)\,x + (\pi_t - C_t)\,G_t, \end{aligned}$$

where $p\,x$ is the incoming cashflow for the energy sold through the contract (p is the price on the contract and x is the amount of contracted energy), $\pi_t(G_t - x)$ is the revenue or expense in the wholesale energy market (π_t is the spot price at time t and G_t is the amount of the generator dispatch) and lastly C_tG_t is the production cost of generating G_t amount of energy (C_t is the unit production cost). The remark that should be made here is that the wholesale and contracts are assumed to be made in the same submarket in this setting, otherwise the spot

price π_t in the submarket where wholesale is made would be different than the spot price in the submarket where contracts are made. For simplicity, this case is ignored in this example.

If T is the time period for which the problem is considered and there are I submarkets, N contracts and S states then the present value of net remuneration per scenario s is

$$PV(NR^s) = \sum_{i=1}^{I} \sum_{n=1}^{N} \sum_{t=1}^{T} \frac{1}{(1+\tau)^{t-1}} \left((\, p_n - \pi_t^{i,s}) \, x_n + (\pi_t^{i,s} - C_t) \, G_t^{i,s} \right).$$

In some time intervals t, x_n will be zero, if that contract's validation period has not started or has finished in that time interval. Rearranging terms of the above expression yields

$$\begin{aligned} PV(NR^s) &= \sum_{n=1}^{N} \theta_n^s \, x_n + \gamma_s, \\ &= \theta^{s^T} \, x + \gamma_s, \end{aligned} \tag{4.3.30}$$

where

$$\theta_n^s = \sum_{i=1}^{I} \sum_{t=1}^{T} \frac{1}{(1+\tau)^{t-1}} \, (\, p_n - \pi_t^{i,s}), \; \gamma_s = \sum_{i=1}^{I} \sum_{t=1}^{T} \frac{1}{(1+\tau)^{t-1}} \, (\pi_t^{i,s} - C_t) \, G_t^{i,s}$$

and $\theta^{s^T} = [\theta_1^s, \theta_2^s, \ldots, \theta_N^s]$, $x = [x_1, x_2, ..., x_N]^T$.

The expected value of the present value of net remuneration is

$$\begin{aligned} E[PV(NR^s)] &= \frac{1}{S}\sum_{s=1}^{S} PV(NR^s) \\ &= \frac{1}{S}\sum_{s=1}^{S}\left[\theta^{s^T} x + \gamma_s\right] \\ &= \frac{1}{S}\sum_{s=1}^{S}\theta^{s^T} x + \frac{1}{S}\sum_{s=1}^{S}\gamma_s \\ &= \Theta^T x + \Gamma, \end{aligned} \tag{4.3.31}$$

where $\Theta := \frac{1}{S}\sum_{s=1}^{S}\theta^s$ and $\Gamma := \frac{1}{S}\sum_{s=1}^{S}\gamma_s$.

Now, the model (4.3.29) is equivalent to the following one

$$\begin{aligned} \text{maximize} \quad & \Theta^T x + \Gamma \\ \text{subject to} \quad & C \leq \alpha + \frac{1}{(1-\beta)S}\sum_{s=1}^{S} u_s, \\ & u_s \leq 0, \qquad \forall s \\ & u_s \leq \theta^{s^T} x + \gamma_s - \alpha, \qquad \forall s \\ & + \text{ some other bounds.} \end{aligned} \tag{4.3.32}$$

Introducing a new variable $\delta_s = \theta^{s^T} x + \gamma_s - \alpha$, the above model can be rewritten

as

$$\begin{aligned}
\text{maximize} \quad & \Theta^T x + \Gamma \\
\text{subject to} \quad & C \leq \alpha + \frac{1}{(1-\beta)S} \sum_{s=1}^{S} u_s, \\
& u_s \leq 0, \qquad \forall s \\
& u_s \leq \delta_s, \qquad \forall s \\
& \delta_s = \theta^{s^T} x + \gamma_s - \alpha, \\
& + \text{ some other bounds.}
\end{aligned} \tag{4.3.33}$$

While modelling the problem for the company, new contracts that are available for the company is ignored. Introducing the notation y for these, say M, new contracts that the company sells during the planning time period into the problem, the present value of the net remuneration (4.3.30) takes the following form

$$\begin{aligned}
PV(NR^s) &= \sum_{n=1}^{N} \theta_n^s \, x_n + \sum_{m=1}^{M} \rho_m^s y_m^s + \gamma_s, \\
&= \theta^{s^T} x + \rho^{s^T} y^s + \gamma_s,
\end{aligned} \tag{4.3.34}$$

where

$$\rho_m^s := \sum_{i=1}^{I} \sum_{t=1}^{T} \frac{1}{(1+\tau)^{t-1}} \, (\, p_m - \pi_t^{i,s})$$

and $\rho^{s^T} = [\rho_1^s, \rho_2^s, ..., \rho_M^s]$, $y = [y_1, y_2, ..., y_M]^T$.

And the new form of the expected value of the present value of net remuner-

ation, which was stated in equation (4.3.31), is

$$\begin{aligned} E[PV(NR^s)] &= \frac{1}{S}\sum_{s=1}^{S} PV(NR^s) \\ &= \frac{1}{S}\sum_{s=1}^{S}\left[\theta^{s^T} x + \rho^{s^T} y^s + \gamma_s\right] \\ &= \frac{1}{S}\sum_{s=1}^{S}\theta^{s^T} x + \frac{1}{S}\sum_{s=1}^{S}\rho^{s^T} y^s + \frac{1}{S}\sum_{s=1}^{S}\gamma_s \\ &= \Theta^T x + \frac{1}{S}\sum_{s=1}^{S}\rho^{s^T} y^s + \Gamma. \end{aligned} \tag{4.3.35}$$

The model (4.3.33) takes the following form:

$$\begin{aligned} \text{maximize} \quad & \Theta^T x + \frac{1}{S}\sum_{s=1}^{S}\rho^{s^T} y^s + \Gamma \\ \text{subject to} \quad & C \le \alpha + \frac{1}{(1-\beta)S}\sum_{s=1}^{S} u_s, \\ & u_s \le 0, \qquad \forall s \\ & u_s \le \delta_s, \qquad \forall s \\ & \delta_s = \theta^{s^T} x + \rho^{s^T} y^s + \gamma_s - \alpha, \\ & + \text{ some other bounds.} \end{aligned} \tag{4.3.36}$$

Marzano et al. [26] used Benders decomposition to solve the above stochastic programming problem and obtained coherent results as expected.

4.4 Example IV

The example which was given by Guigues et al. [17], is an optimization problem for an electrical yearly power management. The company owns electric generation plants, which are nuclear, thermal and hydroelectric power generator plants. In addition to these, demand side management contracts are modelled as a virtual plant called EJP. The objective of the problem is to minimize the production cost over a yearly horizon, to fulfill operating constraints of generation units and the equilibrium between production and demand at each time step. Guigues et al. [17] state that: "the challenge is to ensure robustness of computed optimal production and marginal value face to various uncertainties like customer demand but also water inflows or plants unavailability." In this paper, first a large scale numerical optimization is formulated and solved using Lagrangian relaxation to provide marginal costs on a scenario tree. Then, to compute local feedbacks these marginal costs are used. The robustness of these local feedbacks is shown to be enhanced by a Value-at-Risk (VaR) approach for modelling uncertainties. In the following subsections, the model, the VaR approach and then the application of this approach to the model are introduced.

4.4.1 The Model

The random inputs of the model are the customer's demand, the unavailability of the thermal units and the quantity of natural water inflows. With the help of representation of the random inputs an events as Markov chains, the optimization problem can be formulated as a finite horizon discrete time stochastic control problem on a scenario tree, which represents the behavior of the random inputs

and states. The objective is to minimize the average production cost along this scenario tree. The variable x_n^l stands for the states of the plant production unit l at node n and u_n^l is introduced for the commands applied to this plant. The formulation of the problem is as follows

$$\begin{aligned} \min_{u} \quad & \sum_{l\in L}\sum_{n\in O} \pi_n C_{n,p}^l(x_n^l(p), u_n^l(p)) \\ \text{subject to} \quad & \sum_{l\in L} P_{n,p}^l(x_n^l(p), u_n^l(p)) = D_n(p), \quad \forall n \in O,\ \forall p \in P_n \\ & (x_\cdot^l(\cdot), u_\cdot^l(\cdot)) \in X_l, \quad \forall l \in L. \end{aligned} \tag{4.4.37}$$

The notations used in the above problem are:

- L is the set of plants and O is the set of the nodes,
- π_n is the probability to be at node n,
- P_n is the set of time subdivisions associated to node n,
- $x_n^l(p)$ is the state of plant l at node n and time subdivision (p),
- $u_n^l(p)$ is the control variable of plant l at node n and time subdivision (p),
- $P_{n,p}^l(x_n^l(p), u_n^l(p))$ is the production of plant l in the state $x_n^l(p)$ when command $u_n^l(p)$ is applied to this plant at node n and time subdivision (p),
- $D_n(p)$ is the customer demand at node n and time subdivision (p),
- $C_{n,p}^l(x_n^l(p), u_n^l(p))$ is the production cost when command $u_n^l(p)$ is applied to unit l in the state $x_n^l(p)$,
- X_l is the functional set of constraints on the control and state variables of plant l.

The model above can be reformulated as a linear optimization problem by introducing cost vectors c_i and coupling matrices A_i for $i = 1, 2, 3, 4$ as follows:

$$\begin{aligned} \min \quad & c_1^T u_t + c_2^T u_n + c_3^T x_h + c_4^T x_e \\ \text{subject to} \quad & (u_t, u_n, u_h, u_e) \in T \times N \times H(x_h) \times E(x_e), \\ & A_1 u_t + A_2 u_n + A_3 u_h + A_4 u_e = d \in \mathbb{R}^D, \end{aligned} \tag{4.4.38}$$

where

- u_t is the control variable of classical thermal plants with T as the set of constraints for the thermal plants subsets,
- u_n is the control variable of nuclear thermal plants with N as the set of constraints for the thermal plants subsets,
- x_h is the state variable of hydraulic plants with u_h as the control variable and dynamics described by $u_h \in H(x_h)$,
- x_e is the state variable of EJP contracts with u_e as the control variable and dynamics described by $u_e \in E(x_e)$,
- $d \in \mathbb{R}^D$ is a fixed vector corresponding to the realization of the demand at the different nodes of the scenario tree.

When the following settings are made

$$\begin{aligned} u &:= (u_t, u_n, u_h, u_e) \in U = T \times N \times H(x_h) \times E(x_e), \\ f^l(u, x_h, x_e) &:= c_1^T u_t + c_2^T u_n + c_3^T x_h + c_4^T x_e, \\ Au &:= A_1 u_t + A_2 u_n + A_3 u_h + A_4 u_e, \end{aligned}$$

the model (4.4.38) takes the following form

$$\begin{aligned} \min_{u \in U} \quad & f^l(u, x_h, x_e) \\ \text{subject to} \quad & Au = d \text{ given } \in \mathbb{R}^D. \end{aligned} \tag{4.4.39}$$

For a set of independent scenarios, in which different evolutions of the random inputs (demand, inflows for hydro reservoirs and outages of the thermal units) are applied, a generation schedule and its cost are determined. However, the output of the above problem is optimal only on the trajectory of each reserve, so it gives a local optimum. Thus this model needs to be strengthened by introducing a more reliable model of uncertainty on the scenarios. Guigues et al. [17] state that a robust counterpart with the following properties:

1. reduce the volatility of the simulated costs over a continuum set of reasonable scenarios,

2. reduce the number of extreme case optimal strategies (parsimonious use of water reservoir that might not be nearly empty for a long time),

3. reduce the number of very high cost optimal strategies

can be formulated by a Value-at-Risk setting.

4.4.2 The Value-at-Risk Approach

The VaR calculation model considered here is the following:

$$\begin{aligned} \text{maximize} \quad & \gamma \\ \text{subject to} \quad & P(f(r(w), x) \geq \gamma) \geq 1 - \xi, \\ & x \in X, \quad g(x) \geq 0, \end{aligned} \tag{4.4.40}$$

where $f(r(w), x)$ is a concave income functional depending on a random function $r(w)$ and $x \in X \subset \mathbb{R}^n$ is deterministic variable. Furthermore, $\xi \in (0, 1)$ is the confidence level and $g(x)$ is the additional constraint on x.

Guigues et al. [17] show that if $f(r(w), x)$ is taken to be as $c(w)^T x$, then the VaR calculation model (4.4.40) is equivalent to

$$\begin{aligned} \text{maximize} \quad & c(w)^T x \\ hboxs.t. \quad & x \in X, \quad g(x) \geq 0, \end{aligned} \tag{4.4.41}$$

where the uncertainty set chosen for the random vector $c(w)$ is the ellipsoid

$$\{x \in \mathbb{R}^n : (x - E_w[c(w)])^T \Gamma^{-1} (x - E_w[c(w)]) \leq \kappa^2(\xi)\},$$

where $\Gamma_{ij} = cov(c_i(w), c_j(w))$ and $\kappa(\xi)$ is a risk factor depending on the assumptions on the distribution:

$$\kappa(\xi) = \begin{cases} \Phi^{-1}(1 - \xi) > 0 & \text{if } f(r(\cdot), x) \text{ is Gaussian,} \\ \sqrt{\frac{1-\xi}{\xi}} & \text{if } f(r(\cdot), x) \in L^1_{\mathbb{R}} \cap L^2_{\mathbb{R}}. \end{cases}$$

So, VaR approach can be used to calibrate a variance penalty term for a maximization of a random functional.

This approach is applied to the power generation model in two steps, first on the uncertainty on the demand and next on the unavailability of thermoelectric plants. To begin with, the dual problem of (4.4.39) is considered, which is

$$\begin{aligned} \max_{\lambda \in \mathbb{R}^D} \theta(\lambda) &= \max_{\lambda \in \mathbb{R}^D} [\theta_d(\lambda) + \theta_T(\lambda) + \theta_N(\lambda) + \theta_H(\lambda) + \theta_J(\lambda)] \\ &= \max_{\lambda \in \mathbb{R}^D} [\theta_d(\lambda) + \tilde{\theta}(\lambda)], \end{aligned} \tag{4.4.42}$$

where the partial dual functions are:

$$\begin{aligned} \theta_d(\lambda) &= \lambda^T d, \\ \theta_T(\lambda) &= \inf_{u_t \in T} (c_1 - A_1^T \lambda)^T u_t, \\ \theta_N(\lambda) &= \inf_{u_n \in N} (c_2 - A_2^T \lambda)^T u_n, \\ \theta_H(\lambda) &= \inf_{u_h \in H(x_h)} c_3^T x_h - \lambda^T A_3 u_h, \\ \theta_J(\lambda) &= \inf_{u_e \in E(x_e)} c_4^T x_e - \lambda^T A_4 u_e. \end{aligned} \tag{4.4.43}$$

The objective function in the dual problem is grouped into two as $[\theta_d(\lambda) + \tilde{\theta}(\lambda)]$, where θ_d stands for the customer's demand and $\tilde{\theta}$ for the states of the thermal, nuclear and hydro plants.

As mentioned above, first the uncertainty on the demand is handled. To do that, d is supposed to belong to a given uncertainty set Ξ, which is the ellipsoid given by

$$\Xi = \Xi(\bar{d}, \Gamma, \kappa) = \{x \in \mathbb{R}^D \mid (x - \bar{d})^T \Gamma^{-1} (x - \bar{d}) \leq \kappa^2(\xi)\},$$

where $\bar{d} = E_w[d(w)]$, Γ is the covariance matrix given by $\Gamma_{ij} = cov(d_i(w), d_j(w))$ and the risk factor $\kappa(\xi)$ depends on the assumptions made on the distribution of demand. So, problem (4.4.39) is now equivalent to

$$\begin{aligned} \text{minimize}_{u \in U} \quad & f^l(u, x_h, x_e) \\ \text{subject to} \quad & Au = d \in \Xi. \end{aligned} \tag{4.4.44}$$

Solving (4.4.44) by duality amounts to solve $\max_\lambda \theta_R(\lambda)$, where

$$\begin{aligned} \theta_R(\lambda) &= \min_{u \in U,\, d \in \Xi} f^l(u, x_h, x_e) + \lambda^T (d - Au) \\ &= \theta_T(\lambda) + \theta_N(\lambda) + \theta_H(\lambda) + \theta_J(\lambda) + \min_{d \in \Xi} \lambda^T d. \end{aligned}$$

Furthermore, $\min_{d \in \Xi} \lambda^T d = \phi_\Xi(\lambda)$, where ϕ_Ξ is the support function of the uncertainty set Ξ and it is given by:

$$\phi_\Xi(\lambda) := \lambda^T \bar{d} - \kappa(\xi) \sqrt{\lambda^T \Gamma \lambda}.$$

The convention made about the demand, i.e., $d \in \Xi$, allows one to use VaR approach on the problem (4.4.42). So $\theta(\lambda)$ in (4.4.42) can be replaced with $\tilde{\theta}(\lambda) + \gamma^*(\lambda)$, where

$$\begin{aligned} \gamma^*(\lambda) &= \max\ \gamma \\ &= P(\lambda^T d(w) \geq \gamma) \geq 1 - \xi. \end{aligned} \tag{4.4.45}$$

From (4.4.40) and (4.4.41), this VaR approach, which be denoted by VaR_{FA}, reduces to $\max_\lambda \theta_R(\lambda)$.

After the uncertainty of demand, the VaR approach will now be applied on the unavailability of thermoelectric plants. To construct the model, let $\alpha_{j,l}(t)$ denote the probability that thermal group j of thermal unit l works at time step t and $U_{j,l}^t$ the random variable such that $U_{j,l}^t = 1$ if group j works at time step t, otherwise $U_{j,l}^t = 0$. The groups are regularly checked and repaired every m_0 time steps, if necessary. The assumption made here is that between two consecutive checking dates the availability of the units is not changing. The probability $\alpha_{j,l}(t_k)$ depends on the past evolution of the availability of the group j. If at time step t_{k-1}, the group was out of work, then there is a big probability (say $1 - \beta_1^l$ with β_1^l small) that it works at time step t_k and a small probability β_1^l that it is still out of work at time step t_k. Another assumption made here is that the longer the group has been working without failure, the more likely it can break down at the next time step. Thus, there is a decreasing function of m, $\beta_2^l(m)$ such that for any group j of unit l,

$$P(U_{j,l}^{t_k} = 1 \mid \text{group } j \text{ was working from } t_{k-m} \text{ to } t_{k-1}) = \beta_2^l(m).$$

If the state process of a given group is an homogenous Markov chain where the state space is $\{F, W\}$ and F stands for failure state, W for working state, then $\beta_2^l(m) = \beta_2^l$ is fixed and corresponds to the probability for a group of unit l to work on a given period knowing that it was working the period before. The transition matrix for the group of unit l is then given by

$$P_l = \begin{pmatrix} \beta_1^l & 1 - \beta_1^l \\ 1 - \beta_2^l & \beta_2^l \end{pmatrix}.$$

Thus, the probability $\alpha_{j,l}(t_k)$ is given for $k \geq 1$ by

$$\alpha_{j,l}(t_k) = p_F^l(j)P_l^k(1,2) + p_W^l(j)P_l^k(2,2),$$

where $p_W^l(j) = 1 - p_F^l(j)$ and $p_W^l(j)$ is the probability that group j of unit l works at the first time step. Another assumption, that for a given unit l, either all the groups are working or all the groups are out of work at the first time step, is made for the simplicity. Therefore, $\alpha_{j,l}(t_k)$ is j-independent and $\alpha_l(t_k)$ denote the probability that a group of unit l works at time t_k. Furthermore, the scenario tree can be partitioned in subtrees such that the root node and the leaves nodes of a given subtree respectively correspond to time step t_k and t_{k+1} for some $k \in \mathbb{N}$. Thus, the unavailability rates at different nodes of any subtree of this partition are the same for a given unit. Let $O := \cup_{k=1}^m O_k$ be such that O_k are the nodes of the k-th subtree S_k in this partition. Let $T_k := \{(j,p) \mid j \in O_k, p \in P_j\}$. Then the dual thermal subproblem is

$$\begin{aligned} \min \quad & \sum_l \sum_{k=1}^m (c_{1lk} - \lambda_k)^T u_{tlk} \\ \text{subject to} \quad & 0 \leq u_{tlk} \leq \tau_l(k)\tau_T^l(k)P_{max}^l d_k, \end{aligned}$$

where

- P_{max}^l is the maximal available power of thermal unit l,
- $u_{tlk} = (u_{tljp})_{(j,p)\in T_k}$,
- $c_{1lk} = (c_{1ljp})_{(j,p)\in T_k}$,
- $\lambda_k = (\lambda_{j,p})_{(j,p)\in T_k}$,

- $d_k = (d_{j,p})_{(j,p)\in T_k}$,
- $\tau_T^l(k)$ gives the programmed unavailability rates for unit l and the time subdivision of the set T_k,
- $\tau_l(k)$ is the unavailability rate of unit l for the nodes of the set O_k.

The random term in this formulation, $\tau_l(k)$, is in the right hand side of the constraint set as it is the general case for stochastic optimization problems. To transform the random term in to the objective, the setting $\tilde{u}_{tlk} := \frac{u_{tlk}}{\tau_l(k)}$ is made. With this definition the problems takes the following form:

$$\begin{aligned} \min \quad & \sum_l \sum_{k=1}^{m} \tau_l(k)(c_{1lk} - \lambda_k)^T \tilde{u}_{tlk} \\ \text{subject to} \quad & 0 \leq \tilde{u}_{tlk} \leq \tau_T^l(k) P_{max}^l d_k. \end{aligned}$$

The application of the VaR approach on the thermal plant cost/revenue balance with a given confidence level $0 < \xi < 1$ results in the following problem

$$\begin{aligned} \min \quad & \gamma \\ \text{subject to} \quad & P(\sum_l \sum_k \tau_l(k)(c_{1lk} - \lambda_k)^T u_{tlk} \leq \gamma) 1 - \xi, \\ & 0 \leq u_{tlk} \leq \tau_T^l(k) P_{max}^l d_k, \end{aligned} \tag{4.4.46}$$

which is denoted by VaR_{Benef} representing the problem as maximization of the benefits or equivalently minimization of the losses. Now, the unavailability rates, $\tau_l(k)$, needs to be modelled. To do that let $\bar{P}_{max}^l$ be the maximal power of a group in unit l; then, the theoretical maximal power available on thermal unit l is given by $P_{max}^l = n_l \bar{P}_{max}^l$. Thus, the maximal power available of unit l for nodes of the

set O_k is given by

$$\tilde{P}^{l,k}_{max} := \sum_{j=1}^{n_l} U^{t_k}_{j,l} \bar{P}^{l}_{max} = n_l \bar{P}^{l}_{max} \frac{\sum_{j=1}^{n_l} U^{t_k}_{j,l}}{n_l} = P^{l}_{max} \tau_l(k).$$

From the above hypothesis, $n_l \tau_l(k)$ follows the binomial law $B(n_l, \alpha_l(t(k)))$. Then,

$$n_l E[\alpha_l(t(k))] = E[n_l \alpha_l(t(k))] = n_l \alpha_l(t(k)),$$

thus,

$$E[\alpha_l(t(k))] = \alpha_l(t(k)).$$

Furthermore,

$$n_l^2 Var[\alpha_l(t(k))] = Var[n_l \alpha_l(t(k))] = n_l \alpha_l(t(k)) \big(1 - \alpha_l(t(k))\big),$$

thus,

$$Var[\alpha_l(t(k))] = \frac{\alpha_l(t(k)) \big(1 - \alpha_l(t(k))\big)}{n_l}.$$

When defining the random variable $X_{l,k}$ as $\sum_{l,k} \tau_l(k)(c_{1lk} - \lambda_k)^T u_{tlk}$, the expectation and variance of it will be as follows

$$\begin{aligned} E[X_{l,k}] &= \sum_{l,k} \alpha_l(t(k))(c_{1lk} - \lambda_k)^T u_{tlk}, \\ Var[X_{l,k}] &= \sum_{l,k} u^T_{tlk} \, Q_{lk} \, u_{tlk}, \end{aligned}$$

with the matrix

$$Q_{lk} := \frac{\alpha_l(t(k)) \big(1 - \alpha_l(t(k))\big)}{n_l} (c_{1lk} - \lambda_k)(c_{1lk} - \lambda_k)^T.$$

Finally, the VaR approach applied to the problem (4.4.46) will result in the following problem:

$$\begin{aligned} \text{minimize} \quad & \sum_{l,k} \alpha_l(t(k))(c_{1lk} - \lambda_k)^T u_{tlk} + \kappa(\xi)\sqrt{\sum_{l,k} u_{tlk}^T \, Q_{lk} \, u_{tlk}} \\ \text{subject to} \quad & 0 \leq u_{tlk} \leq \tau_T^l(k) P_{max}^l d_k, \end{aligned} \tag{4.4.47}$$

where $\kappa(\xi) = \sqrt{\frac{1-\xi}{\xi}}$.

The summary of what is done until now is as follows: The model for the problem is constructed (4.4.39), whose dual is (4.4.42). To solve it, VaR approach is applied to the dual problem in two steps, firstly, for the demand part and, secondly, for the thermal plant part. The first one resulted in the following problem VaR_{FA}

$$\max_{\lambda} \theta_d^R(\lambda) + \max_{\lambda}(\theta_T + \theta_N + \theta_H + \theta_J)(\lambda)$$

with

$$\theta_d^R(\lambda) := E_w(d(w))^T\lambda - \kappa(\xi_1)\sqrt{\lambda^T \Gamma \lambda},$$

where ξ_1 is the confidence level for VaR_{FA}. The second one resulted in the following problem VaR_{Benef}

$$\max_{\lambda} \theta_d(\lambda) + \max_{\lambda} \theta_T^R(\lambda) + \max_{\lambda}(\theta_N + \theta_H + \theta_J)(\lambda)$$

with

$$\begin{aligned}\theta_T^R(\lambda) \;\; := \;\; & \min \sum_{l,k} \alpha_l(t(k))(c_{1lk} - \lambda_k)^T u_{tlk} + \kappa(\xi_2)\sqrt{\sum_{l,k} u_{tlk}^T \, Q_{lk} \, u_{tlk}} \\ & 0 \leq u_{tlk} \leq \tau_T^l(k) P_{max}^l d_k,\end{aligned}$$

where ξ_2 is the confidence level for VaR_{Benef}. Finally, the combination of these two subproblem is the following mixed problem

$$\max_{\lambda} \theta_{mix}^R(\lambda) = \max_{\lambda} \theta_d^R(\lambda) + \max_{\lambda} \theta_T^R(\lambda) + \max_{\lambda} (\theta_N + \theta_H + \theta_J)(\lambda).$$

Guigues et al. [17] use the dual problem to approximate primal solutions and estimate marginal prices, since numerical simulations have shown that the duality gap is generally quite small. Applying state decomposition method for the dual function and making necessary adaptations for the regularized problems, the optimization problems became solvable. Using three different trees results for all the methods are obtained and discussed by Guigues et al. [17].

Summary

The examples of energy models introduced in this chapter show that the uncertainties involved in the business of energy are plentiful and stochastic programming is a appropriate approach to construct a model for a energy company. This is due to its success despite the complex features of energy markets, such as non-normal returns, the involvement of production assets, non-storability of electricity, complex contracts, etc. In such an environment, using multistage planning programs, which can often be formulated as linear programs with a dynamic matrix structure, is a wise choice. So, the resulting problem is a multistage stochastic linear program. The size of the equivalent deterministic form of multistage stochastic problems can be so large as to appear intractable. In this case, decomposition methods for solving stochastic programs come into play. Stochastic programming and decomposition methods will be introduced in the following chapter.

Chapter 5

Stochastic Programming and Decomposition Techniques

Problems of optimization under uncertainty are characterized by the necessity of making decisions without knowing what their full effects will be. Such problems appear in many areas of application, such as transportation, energy industry, finance, nuclear engineering. For example, an electrical utility has to decide each day how much power to produce without yet knowing the demand, which may depend on weather and season in which a decision is made. Longer range decisions may concern the amount of coal to purchase or the kinds of contracts set up with other utilities.

The uncertainties in a problem have to be represented in such a manner that their effects on present decision making can properly be taken into account. They may often be modelled as random variables to which the theory of probability can be applied.

A framework for modelling optimization problems that involve uncertainty is *stochastic programming*, which takes a probabilistic approach to uncertainty. It was first formally introduced in the 1950's as a branch of mathematical programming by the seminal work of Dantzig [11] and Beale [5]. The goal in stochastic programming is to find some policy that is feasible for all the possible data in-

stances and maximizes or minimizes the expectation of some function of the decision and the random variables.

Stochastic programming models can be classified into two as two-stage and multistage stochastic programming models. In two stage models only the present and the next period decisions are considered. However, most practical decision problems involve a sequence of decisions, rather than two stage, that react to outcomes that evolve over time. These problems can be modelled as multistage stochastic programming problems, which is the extended version of two-stage models to multistage.

The optimization models constructed as stochastic programming problem are large scale models, and decomposition techniques are needed to solve them. Decomposition refers to the strategy of breaking up a large, difficult-to-solve problem into two or more smaller, easier-to-solve problems, such that the solution to the decomposed problems can be used to obtain the solution to the original problem.

Stochastic programming and decomposition methods are the main subject of this chapter.

5.1 Two-Stage Stochastic Programs

The most widely applied and studied stochastic programming models are *two-stage stochastic programs*. In two-stage stochastic program, the decision maker takes some action in the first stage called as *first-stage decisions*, after which a random event occurs affecting the outcome of the first-stage decisions. A recourse decision can then be made in the second-stage that compensates for any bad effects that might have been experienced as a result of the first-stage decisions.

The optimal policy from such a model is a single first-stage policy and a collection of recourse decisions (a decision rule) defining which second-stage action should be taken in response to each random outcome. As an example of a two-stage stochastic program, the farming example in [7] can be considered, where a farmer tries to make an optimal decision on the amounts of various crops to plant. The yields of the crops vary according to the weather. In other words, the farmer must decide on the amounts before knowing the yields. This decision is made in the first stage. After the realization of the random events (here, the yields), a recourse decision can be made in the second stage, which is a sale and purchase decisions of products in this example.

The classical two-stage stochastic programming problem with fixed recourse, can be stated as

$$\begin{aligned} \text{minimize} \quad & z = c^T x + E_\xi[\min q(w)^T y(w)] \\ \text{subject to} \quad & Ax = b, \\ & T(w)x + Wy(w) = h(w), \\ & x \geq 0, \\ & y(w) \geq 0. \end{aligned} \tag{5.1.1}$$

In this model, the decision variables of the first stage are comprised in the vector $x \in \mathbb{R}^{n_1}$. Corresponding to x are the first stage vectors and matrices c, b, and A, of sizes $n_1 \times 1$, $m_1 \times 1$ and $m_1 \times n_1$, respectively. In the second stage, a number of random events $w \in \Omega$ may be realized. For a given realization w, the second stage problem data $q(w)$, $h(w)$ and $T(w)$ become known, where $q(w)$ is $n_2 \times 1$, $h(w)$ is $m_2 \times 1$, and $T(w)$ is $m_2 \times n_1$. The matrix T is usually referred

to as the *technology matrix* and the matrix W of size $m_2 \times m_2$ as the *recourse matrix*, which is assumed to be fixed here in order to be able to characterize the feasibility region in a convenient manner for computation. Otherwise, there may be some difficulties, which are in detail examined in [7].

The stochastic components of the second-stage data is a vector $\xi^T(w) = (q(w)^T, h(w)^T, T_1(w), ..., T_{m_2}(w))$ with $N = n_2 + m_2 + (m_2 \times n_1)$ components, where $T_i(w)$ is the i-th row of $T(w)$. Let $\Xi \subseteq \mathbb{R}^N$ be the support of ξ, i.e., the smallest closed subset in $\mathbb{R}^N$ such that $P\{\xi \in \Xi\} = 1$. As mentioned, after the realization of w, $q(w)$, $h(w)$ and $T(w)$ become known. Then, the second stage decisions $y(w)$ must be taken. Here, the notation $y(w)$ does not indicate that y is a function of w, but the decisions y are typically not the same under different realizations of w. The second stage decisions are chosen so that the constraints of the above model hold almost surely, i.e., for all $w \in \Omega$, except perhaps for sets with zero probability.

The objective function of (5.1.1) contains a deterministic term $c^T x$ and the expectation of the second stage objective $q(w)^T y(w)$ taken over all realizations of the random event w. For each w, the value $y(w)$ is the solution of a linear program. To stress this, the notion of deterministic equivalent program may be used. For a given realization w, let

$$Q(x, \xi(w)) := \min\{q(w)^T y \mid W\ y = h(w) - T(w)x,\ y \geq 0\}$$

be the optimal value of the second stage problem, which is a well-defined extended real valued function: it takes the value $+\infty$ if the feasible set of the second-stage problem is empty, and the value $-\infty$ if the second-stage problem is unbounded

[32]. Defining the expected value of $Q(x, \xi(w))$ as *recourse function*, which is

$$\mathcal{Q}(x) := E_{\xi} Q(x, \xi(w)),$$

the deterministic equivalent program will be as follows:

$$\begin{aligned} \text{minimize} \quad & z = c^T x + \mathcal{Q}(x) \\ \text{subject to} \quad & Ax = b, \\ & x \geq 0. \end{aligned} \tag{5.1.2}$$

If every solution x that satisfies the first-period constraints, $Ax = b$, is also feasible for the second stage, the stochastic program is said to have *relatively complete recourse*. A special type of relatively complete recourse may often be identified from the structure of W. This form, called *complete recourse*, holds when there exists $y \geq 0$ such that $Wy = t$ for all $t \in \mathbb{R}^{m_2}$. Complete recourse is often added to a model to ensure that no outcome can produce infeasible results. A special type of complete recourse that offers additional computational advantages is *simple recourse*. In a simple recourse problem $W = [I, -I]$, where y and q are divided correspondingly as (y^+, y^-) and (q^+, q^-).

In the case of discrete distributions, there is a finite number of realizations $\xi_k^T = (q_k^T, h_k^T, T_{1k}, ..., T_{m_2 k})$ $(k = 1, 2, ..., K)$, called *scenarios*, with the corresponding probabilities p_k. Then, the recourse function will be: $\mathcal{Q}(x) := \sum_{k=1}^{K} p_k \, Q(x, \xi_k)$, where

$$Q(x, \xi_k) = \min\{q_k^T \; y_k \mid W \; y_k = h_k - T_k x, \; y_k \geq 0\}.$$

Consequently, (5.1.1) becomes

$$\begin{aligned} \text{minimize} \quad & z = c^T\, x + \sum_{k=1}^{K} p_k q_k^T y_k \\ \text{subject to} \quad & Ax = b, \\ & T_k x + W y_k = h_k, \quad k = 1, 2, ..., K, \\ & x \geq 0, \\ & y_k, \geq 0 \quad k = 1, 2, ..., K. \end{aligned} \tag{5.1.3}$$

Summary

Problems under uncertainty concerning only the present and the next period can be modelled via two-stage stochastic programming, which provides the decision maker optimal first-stage decisions and a recourse decision for each random outcome. After taking the first stage decision, a random event occurs affecting the outcome of the first-stage decisions. And according to this outcome the corresponding recourse decision is then taken.

5.2 Multistage Stochastic Program

Most practical decision problems involve a sequence of decisions, rather than two stage, that react to outcomes that evolve over time. Such problems are therefore called *multistage stochastic programming problems*.

The multistage stochastic linear program with fixed recourse has the following

general form:

$$
\begin{aligned}
\text{minimize} \quad & z = (c^1)^T x^1 + E_{\xi^2}\big[\min c^2(w)^T x^2(w^2) + ... + E_{\xi^H}[\min c^H(w)^T x^H(w^H)]...\big] \\
\text{subject to} \quad & W^1 x^1 = h^1, \\
& T^1(w)x^1 + W^2 x^2(w^2) = h^2(w), \\
& \vdots \\
& T^{H-1}(w)x^{H-1}(w^{H-1}) + W^H x^H(w^H) = h^H(w), \\
& x^1 \geq 0;\ x^t(w^t) \geq 0, \qquad t = 2, 3, ..., H,
\end{aligned}
\tag{5.2.4}
$$

where c^1 and h^1 are known vectors, of sizes $n_1 \times 1$ and $m_1 \times 1$, respectively, $\xi^t(w)^T = (c^t(w)^T, h^t(w)^T, T_1^{t-1}(w), ..., T_{m_t}^{t-1}(w))$ is a random N_t-vector and each W^t is a known matrix in $\mathbb{R}^{m_t \times n_t}$. The decisions x depend on the history up to time t, which is indicated by w^t. Here, again it is supposed that Ξ^t is the support of ξ^t. At each stage, realizations of some random variables occur and corresponding decisions are made; i.e., the following sequence of actions ia taken:

decision x^1,

observation ξ^2,

decision x^2,

$\vdots$

observation ξ^H,

decision x^H.

The objective is to design the decision process in such a way that the expected

value of the total cost is minimized while the optimal decisions are allowed to be made at every time period $t = 1, 2, ..., H$.

The deterministic equivalent form of this problem in terms of a dynamic program can be described as follows. At the last stage, H, the values of all problem data are already known and the values of the earlier decision vectors, $x^1, x^2, ..., x^{H-1}$, have been chosen. Therefore, the problem is a simple linear programming problem, which is

$$\begin{aligned} Q^H(x^{H-1}, \xi^H(w)) = \min \quad & c^H(w)^T x^H(w) \\ \text{subject to} \quad & W^H x^H(w) = h^H(w) - T^{H-1}(w)x^{H-1}, \\ & x^H(w) \geq 0. \end{aligned}$$

Letting $\mathcal{Q}^{t+1}(x^t) := E_{\xi^{t+1}}[Q^{t+1}(x^t, \xi^{t+1}(w))]$ for all t, one obtains the recursion for $t = 2, 3, ..., H-1$,

$$\begin{aligned} Q^t(x^{t-1}, \xi^t(w)) = \min \quad & c^t(w)^T x^t(w) + \mathcal{Q}^{t+1}(x^t) \\ \text{subject to} \quad & W^t x^t(w) = h^t(w) - T^{t-1}(w)x^{t-1}, \\ & x^t(w) \geq 0, \end{aligned} \tag{5.2.5}$$

where x^t is used to indicate the state of the system. Other state information in terms of the realizations of the random parameters up to time t should be included if the distribution of ξ^t is not independent of the past outcomes.

The value sought is:

$$\begin{aligned} \text{minimize} \quad & z = {c^1}^T x^1 + \mathcal{Q}(x^1) \\ \text{subject to} \quad & W^1 x^1 = h^1, \\ & x^1 \geq 0. \end{aligned} \tag{5.2.6}$$

Summary

Since most problems under uncertainty involve not only two stages but finitely many of them, the extension of two-stage stochastic programming to multistage should be used to model this kind of problems. In multistage stochastic programming models the same logic in two-stage is used, where the decision maker takes some action in a period, observes the realization of the random variable and takes the next action according to this realization. This process continues until the last period.

In stochastic programming, the expectations depend on the probability distribution of the scenario used, in short they depend on scenarios. So, scenarios are important subjects for stochastic programming and are considered in the next subsection.

5.2.1 Scenarios

When dealing with problems under uncertainty, scenario trees are of great importance, which is also the case in stochastic programming models. A scenario tree is composed of different scenarios, which are sets of representative outcomes of the random events. Figure 5.1 shows an example of a scenario tree for a 3-stage

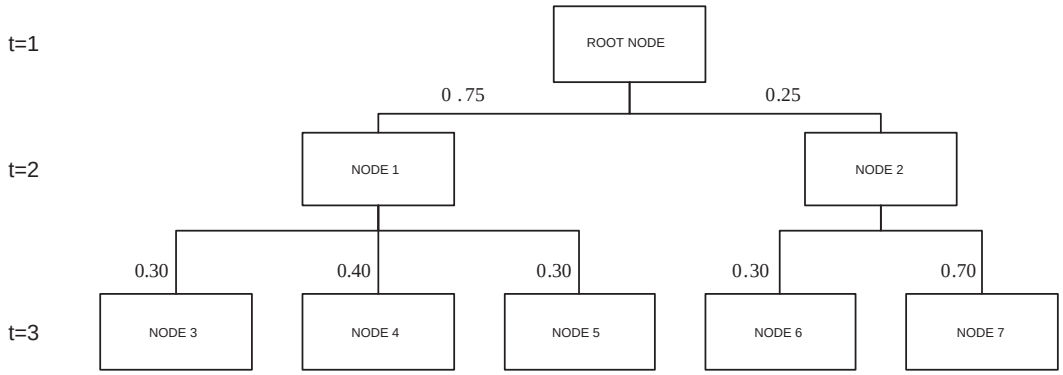

Figure 5.1: *Scenario tree.*

stochastic program. Numbers along the arcs represent conditional probabilities of moving from one node to the next. There are two possible outcomes for a random event (node 1 and node 2) at $t = 2$ with probabilities 0.75 and 0.25, respectively. In other words, c^2, h^2 and T^1 are assumed to have the realization described by node 1 (2), with probability 0.75 (0.25) at $t = 2$. Likewise, there are three possibility for node 1 and two for node 2 at $t = 3$. A scenario consists of a complete path from the root node, which represents the present or a part of the problem that is known, to a single leaf node. Thus, there are five different scenarios in this example. Since the numbers along arcs represent conditional probabilities, the probability of a scenario is the product of conditional probabilities of the corresponding nodes, e.g., the probability of the scenario ending with node 3 is $0.75 \times 0.30 = 0.225$.

Scenario trees give one the flexibility to choose which scenarios need to be considered and their relative importance. Since stochastic programs are dealing with problems under uncertainty, the more possibility of future values the model involves, the more robust solutions are found. Thus, there is a great number of scenarios in stochastic programs, which makes it harder to solve.

5.3 Robust and Minimax Approach

Decision making under uncertainty is a difficult job due to the unknown events of the future, namely uncertainty. As mentioned until now in this chapter, stochastic programming is a useful approach to model the problems under uncertainty. However, even stochastic programming models may give solutions which are in fact not optimal. To strengthen stochastic programming, robust and minimax approaches can be considered. Robust models replaces the expectation term in the stochastic programming models by a term that accounts for both mean and variability. Minimax approach tries to strengthen stochastic programming models by considering the worst case probability distribution that maximizes the cost.

5.3.1 Robust Models

As mentioned previously, two-stage stochastic optimization models minimize the sum of the costs of the first stage and the expected cost of the second stage. A potential limitation of this approach is that it does not account for the variability of the second stage costs and might lead to solutions where the actual second stage costs are unacceptably high. In order to resolve this difficulty, one may try to be more conservative and to reach a compromise between the mean and a risk associated with variability of $Q(x, \xi)$. This may be done by adding the term $\kappa Var[Q(x, \xi)]$ to the objective of the optimization problem, where the coefficient $\kappa \geq 0$ represents a compromise between the expectation and variability of the objective. On the other hand, adding that term destroys the convexity and second stage optimality of two-stage linear program (5.1.1) [32].

For simplicity, the assumption that there is a finite number of scenarios is made here. The problem (5.1.1) takes the following form after adding the term $\kappa Var[Q(x,\xi)]$ to its objective function:

$$\begin{aligned} \text{minimize} \quad & c^T x + \psi(Q(x,\xi_1),...,Q(x,\xi_K)) \\ \text{subject to} \quad & Ax = b, \end{aligned} \tag{5.3.7}$$

where

$$\psi(z) := \sum_{k=1}^{K} p_k z_k + \kappa\Big[\sum_{k=1}^{K} p_k z_k^2 - (\sum_{k=1}^{K} p_k z_k)^2\Big].$$

Now, the objective function above is not convex and the second stage optimality does not hold. To eliminate these undesirable properties, the function $\psi(z)$ may be changed to a *componentwise nondecreasing function*. This is a function $\psi(z) : \mathbb{R}^K \to \mathbb{R}$, which satisfies: if $z \geq \acute{z}$, then $\psi(z) \geq \psi(\acute{z})$ for any $z, \acute{z} \in \mathbb{R}^K$. According to the proposition given by Ruszczyński et al. [32], if the function $\psi(z)$ is componentwise nondecreasing, then the second stage optimality holds for problem (5.3.7) and the optimal value for this problem is the same with the optimal value for the following problem (whenever this problem is feasible):

$$\begin{aligned} \text{minimize} \quad & c^T x + \psi(q_1^T y_1,...,q_k^T p_K) \\ \text{subject to} \quad & Ax = b, \\ & T_k x + W_k y_k = h_k, \\ & x \geq 0,\ y_k \geq 0, \quad k = 1,2,...,K. \end{aligned} \tag{5.3.8}$$

Moreover, if $\psi(z)$ is convex, then the objective function of problem (5.3.7) is also

convex.

For $\psi(z) := \sum_{k=1}^{K} p_k z_k$, problem (5.3.8) coincides with problem (5.1.3). Another possibility is to use a separable function $\psi(z) := \sum_{k=1}^{K} \psi_k(z)$ with one of the following two choices of function $\psi(z)$:

$$\psi(z) := p_k z_k + \kappa p_k (z_k - \alpha)^+, \tag{5.3.9}$$

$$\psi(z) := p_k z_k + \kappa p_k [(z_k - \alpha)^+]^2, \tag{5.3.10}$$

for some $\kappa \geq 0$ and $\alpha \in \mathbb{R}$. For either choice of ψ_k, the corresponding $\psi(z)$ is componentwise nondecreasing and convex. If the parameter α in (5.3.10) is equal to $E[Q(x,\xi)]$ and the distribution of $Q(x,\xi)$ is symmetrical around its mean, then

$$\psi(Q(x,\xi_1), ..., Q(x,\xi_K)) = E[Q(x,\xi)] + \frac{\kappa}{2} Var[Q(x,\xi)].$$

5.3.2 Minimax Approach

Probability distributions of uncertain parameters are never known exactly and can be estimated at best. However, as estimation comes into play, there may be several different distributions, which are all subjective. One way to deal with this problem is to construct a finite set of probability distributions, say $S := \{P_1, P_2, ..., P_l\}$, and to assign probability ρ_i to each P_i, $i = 1, 2, ..., l$. Then, one obtains the unique distribution $P := \sum_{i=1}^{l} \rho_i P_i$, which is an average over possible distributions P_i. However, again a choice of probabilities ρ_i is subjective.

Another way to deal with this problem is to hedge against worst distribution by using *minimax approach* for the main problem. The minimax approach for a

two-stage stochastic problem is

$$\min_x \{c^T x + \max_p \{ \sum_{k=1}^{K} p_k Q_k(x) \mid (p_1, p_2, ..., p_k) \in P\}\},$$

where $Q_k(x)$ is the second stage problem.

As seen from this section, multistage stochastic programs are among the most intractable in numerical computations. Not only does the size of the problem grow as a function of the number of the scenarios (states), but also the problem's structure is difficult to take the advantage of due to numerical instability. That is the reason why decomposition became an important issue in the area of stochastic programming.

5.4 Decomposition

Decomposition refers to the strategy of breaking up a large, difficult-to-solve problem into two or more smaller, easier-to-solve problems, such that the solution to the decomposed problems can be used to obtain the solution to the original problem.

There is an extensive literature on decomposition techniques. Works on this subject has been appearing since the publication of the first papers by Dantzig and Wolfe [10] and Benders [6], whose methods are basically given in this chapter.

Ho and various co-authors (Ho and Manne [18], Ho and Loute [19]) wrote a series of papers about *Dantzig-Wolfe decomposition*, showing that the method terminates with the optimal dual variables and describing ways to recover the optimal primal variables.

Benders decomposition was studied by Abrahamson [1], Wittrock [39] and Scott. If this method is used, the optimal primal decision variables are obtained directly, but the dual variables will have to be recovered later.

The earliest attempt to solve stochastic programs by decomposition was by Van Slyke and Wets [37]. They developed a method, usually referred to in the literature as the *L-Shaped method*, is a form of Benders decomposition. This method was designed for two-stage linear stochastic programs with discrete random variables and finite number of realizations. Birge [8] and Gassmann [16] extended the method to multi-stage, known as *Nested Benders' decomposition*. The algorithms of L-shaped method and Nested Benders' decomposition are given in this chapter.

5.4.1 Benders Decomposition

Benders decomposition was derived in 1962 by J.F. Benders [6] as a technique for solving mixed integer programs. In 1969, Van Slyke and Wets [37] realized that Benders decomposition could be applied to large stochastic programs with a dual angular structure, and they introduced what is called the L-shaped method to obtain exact solutions for these types of problems.

The basic idea of this method is to solve the problems of the following form

$$\begin{aligned} \text{minimize} \quad & Z = c^T x + f^T y \\ \text{subject to} \quad & Ax \qquad = b, \\ & -Bx + Dy = d, \\ & x \geq 0, \;\; y \geq 0, \end{aligned} \tag{5.4.11}$$

which can be partitioned into two problems as follows

$$\begin{aligned} \text{minimize} \quad & Z = cx + z(x) \\ \text{subject to} \quad & Ax = b, \\ & x \geq 0, \end{aligned} \tag{5.4.12}$$

and

$$\begin{aligned} \text{minimize} \quad & z(x) = f^T y \\ \text{subject to} \quad & Dy = d + Bx, \\ & y \geq 0. \end{aligned} \tag{5.4.13}$$

The assumption made here is that the problem (5.4.11) has a finite optimal feasible solution, (x^*, y^*).

The dual of (5.4.13) is

$$\begin{aligned} \text{maximize} \quad & z(x) = \pi^T(d + Bx) \\ \text{subject to} \quad & \pi^T D \leq f^T, \end{aligned} \tag{5.4.14}$$

where π is the dual multiplier corresponding to (5.4.13). Assume that the feasible region of (5.4.14) has p extreme points and q extreme rays and let

$$\begin{aligned} \pi^j,\ j &= 1, 2, ..., p \text{ be the extreme points and} \\ \pi^j,\ j &= p+1, p+2, ..., p+q \text{ be the extreme rays.} \end{aligned}$$

According to the assumption made, the problem (5.4.11) is feasible, therefore the

dual problem (5.4.14) is also feasible with respect to the duality theorem. Thus

$$\pi^{j^T}(d + Bx) \leq 0, \quad j = p + 1, p + 2, ..., p + q.$$

Now, the dual problem can be written as

$$z(x) = \max_{j=1,...,p} \pi^{j^T}(d + Bx)$$

Let

$$z^j(x) := \pi^{j^T}(d + Bx), \ j = 1, 2, ..., p;$$

then

$$z^j(x) \leq z(x) \quad \forall x, \ j = 1, 2, ..., p.$$

Hence, problem (5.4.11) is equivalent to the following full master problem:

$$\begin{aligned} \text{minimize} \quad & Z = c^T x + \theta \\ \text{subject to} \quad & Ax = b, \\ & x \geq 0, \\ & \theta \geq \pi^{j^T}(d + Bx), \ j = 1, 2, ..., p \\ & 0 \geq \pi^{j^T}(d + Bx), \ j = p + 1, p + 2, ..., p + q. \end{aligned} \tag{5.4.15}$$

The original problem (5.4.11) can now be divided into two as master problem and subproblem.

Master problem looks as follows:

$$\begin{aligned} \text{minimize} \quad & c^T x + \theta \\ \text{subject to} \quad & Ax = b, \\ & -G^k x + \alpha\theta \geq g^k,\ k = 1, 2, ..., \\ & x \geq 0, \end{aligned} \tag{5.4.16}$$

where $G^{k+1} = \pi(\hat{x}^k)^T B,\ \ g^{k+1} = \pi(\hat{x}^k)^T d$, k is the number of cuts and

$\alpha = 1,$ if $\pi \in \{\pi^j,\ j = 1, 2, ..., p\}$

$\alpha = 0,$ if $\pi \in \{\pi^j,\ j = p+1, p+2, ..., p+q\}$

Subproblem look as follows:

$$\begin{aligned} z(\hat{x}^k) = \quad & \min f^T y \\ \text{subject to} \quad & Dy = d + B\hat{x}^k, \\ & y \geq 0. \end{aligned} \tag{5.4.17}$$

The lower and upper bounds for z^k are $c^T\hat{x}^k + \hat{\theta}^k$ and $c^T\hat{x}^k + z(\hat{x}^k)$, respectively. Benders decomposition starts with solving the master program with $k = 0$, which is

$$\begin{aligned} \text{minimize} \quad & c^T x + \theta \\ \text{subject to} \quad & Ax = b, \\ & x \geq 0, \end{aligned}$$

and the master problem (5.4.16) at $k > 0$ to obtain $\hat{x}^k,\ \hat{\theta}^k$, and thus z^k_{LB}. Solving

the subproblem, one gets $z(\hat{x}^k)$ and thus z^k_{UB}. Then, if $z^k_{UB} - z^k_{LB} \leq TOL$ the algorithm is terminated and $\hat{x}^k$ is declared to be the optimal solution, where TOL is the given tolerance. Otherwise, a new constraint of type $-G^k x + \alpha\theta \geq g^k$ is added to the master problem at $k = k + 1$, and the new master problem with an additional constraint is solved. This procedure continues until the optimal solution is found.

Benders decomposition was first applied to stochastic programs with a dual angular structure (see Figure 5.2) by Van Slyke and Wets [37]. The method they developed, L-Shaped method, was designed for two-stage linear stochastic programs with discrete random variables and finite number of realizations.

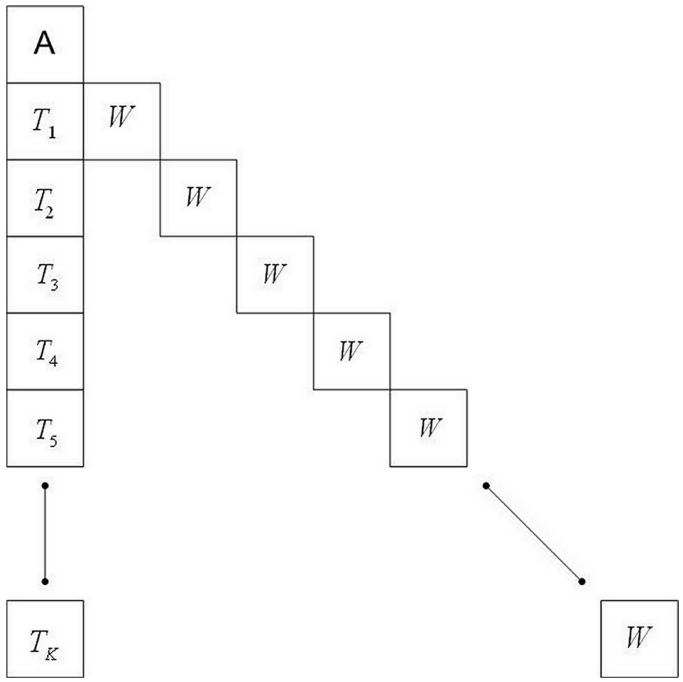

Figure 5.2: *Dual angular structure of a two-stage stochastic program.*

L-Shaped Method

The basic idea of the L-Shaped method is to approximate the nonlinear term in the objective of two-stage stochastic problems (*the recourse function*). The assumption that the random vector ξ has finite support is made. Let $k = 1, 2, ..., K$ index its possible realizations and let p_k be their probabilities. Under this assumption, the deterministic equivalent program is (5.1.3), as shown previously, where one set of second-stage decisions, y_k, is associated to each realization ξ, i.e., to each realization of q_k, h_k and T_k. The structure of the constraints set of this problem is shown in Figure 5.2. This structure has given rise to the name, L-Shaped method for the following algorithm.

L-Shaped Algorithm [7]

Step 0: Set $r = s = \nu = 0$.

Step 1: Set $\nu = \nu + 1$. Solve the following linear program

$$\begin{aligned} \text{minimize} \quad & z = c^T x + \theta && (5.4.18) \\ \text{s.t.} \quad & Ax = b, \\ & D_l x \geq d_l, \qquad l = 1, 2, ..., r, && (5.4.19) \\ & E_l x + \theta \geq e_l, \quad l = 1, 2, ..., s, && (5.4.20) \\ & x \geq 0, \quad \theta \in \mathbb{R}. \end{aligned}$$

If this problem is infeasible, then (5.1.3) is infeasible and the algorithm terminates. Unboundedness can be resolved by the procedure in [37]. Otherwise, let (x^ν, θ^ν) be an optimal solution. If no constraint (5.4.20) is present,

θ^v is set equal to $-\infty$ and is not considered in the computation of x^v.

Step 2: For $k = 1, 2, ..., K$ solve the linear program

$$\begin{aligned} \text{minimize} \quad & w' = e^T\nu^+ + e^T\nu^- && (5.4.21) \\ & \text{s.t. } Wy + I\nu^+ - I\nu^- = h_k - T_k x^v, && (5.4.22) \\ & y \geq 0, \quad \nu^+ \geq 0, \quad \nu^- \geq 0. \end{aligned}$$

where $e^T = (1, 2, ..., 1)$, until, for some k, the optimal value $w' > 0$. If (5.4.21) is unbounded, so is (5.1.3) and the algorithm terminates. Otherwise, let σ^ν be the associated simplex multipliers and define

$$D_{s+1} = (\sigma^\nu)^T T_k \tag{5.4.23}$$

and

$$d_{s+1} = (\sigma^\nu)^T h_k \tag{5.4.24}$$

to generate a constraint (*feasibility cut*) of type (5.4.19). Set $r = r + 1$, add to the constraint set (5.4.19), and return to Step 1. If for all k, $w' = 0$, go to Step 3.

Step 3: For $k = 1, 2, ..., K$ solve the linear program

$$\begin{aligned} \text{minimize} \quad & w = q_k^T y && (5.4.25) \\ & \text{s.t. } Wy = h_k - T_k x^\nu, \\ & y \geq 0. \end{aligned}$$

Let π_k^ν be the simplex multipliers associated with the optimal solution of Problem k of type (5.4.25). Define

$$E_{t+1} = \sum_{k=1}^{K} p_k (\pi_k^\nu)^T T_k, \tag{5.4.26}$$

$$e_{t+1} = \sum_{k=1}^{K} p_k (\pi_k^\nu)^T h_k. \tag{5.4.27}$$

Let $w^\nu := e_{t+1} - E_{t+1}x^\nu$. If $\theta^\nu \geq w^\nu$, stop; x^ν is an optimal solution. Otherwise, set $s = s + 1$, add to the constraint set (5.4.20), and return to Step 1.

First, r, s and ν are set equal to zero in Step 0. In Step 1, ν is increased by one and become 1. Now $r = 0, s = 0, \nu = 1$. The linear program (5.4.18) in Step 1, which has only one constraint, is solved. There are no constraints of the type (5.4.19) and (5.4.20), since $r = 0$ and $s = 0$. If this problem is infeasible, then (5.1.3) is infeasible and the algorithm terminates. Unboundedness can be resolved by the procedure in [37]. Otherwise, the optimal solution for this problem $(x^1, -\infty)$ may be the optimal solution for the main problem (5.1.3). To check this, firstly the feasibility of $(x^1, -\infty)$ is checked. This is done in Step 2 by solving (5.4.21), which tests whether $h_k - T_k x^\nu$ belongs to $posW$ for $k = 1, ..., K$. If not, this means that for some k, $h_k - T_k x^\nu \notin posW$. Then, there must be a hyperplane separating $h_k - T_k x^\nu$ and $posW$. Such a hyperplane, say $\{x|\sigma x = 0\}$, is determined by its normal, σ, which must satisfy $\sigma^T t \leq 0$ for all $t \in posW$ and $\sigma^T(h_k - T_k x^\nu) > 0$. In Step 2, this hyperplane is obtained by taking σ for the value σ^ν of the simplex multipliers of the subproblem (5.4.21) solved in Step 2. If this subproblem is unbounded, so is (5.1.3) and the algorithm terminates. Otherwise, an optimal

solution w' for (5.4.21) is obtained.

By duality, w' being strictly positive is the same as $\sigma^T(h_k - T_k x^\nu) > 0$. Furthermore, $(\sigma^\nu)^T W \leq 0$ is satisfied because σ^ν is an optimal simplex multiplier and, at the optimum, the reduced cost associated with y must be non-negative. Hence, σ^ν has the desired property.

If the solution of the linear program (5.4.21) is zero ($w' = 0$) for all states, it means that the point $(x^1, -\infty)$ is in the feasible region of the main problem (5.1.3). Otherwise (if $w' > 0$ for at least one state), the point $(x^1, -\infty)$ is not in the feasible region, thus not optimal. In this case, this solution is cut off from the feasible region by feasibility cut. Feasibility cut is introduced by constructing D_{s+1} and d_{s+1}. This cut is added to the constraint set (5.4.19) and r is increased by 1, i.e., r becomes 1. After that, one must return to Step 1 and solve the problem (5.4.18) with the new added constraint. This process is done until a feasible solution is found (i.e., $w = 0$ for all states).

Then what must be done is to check whether this feasible solution is optimal. Solving (5.1.2) is equivalent to solve

$$\begin{aligned} \text{minimize} \quad & c^T x + \theta \\ \text{subject to} \quad & \mathcal{Q}(x) \leq \theta, \\ & x \in K_1 \cap K_2, \end{aligned} \tag{5.4.28}$$

where $K_1 := \{x | Ax = b, x \geq 0\}$ and $K_2 := \{x | Q(x, \xi) < +\infty\}$. To check the optimality, the feasible solution x^ν is substituted to its place in the constraint set of the linear program (5.4.25) in Step 3 and solved repeatedly for each states

$k = 1, 2, ..., K$, yielding optimal simplex multipliers π_k^ν. From duality, for each k,

$$Q(x^\nu, \xi_k) = (\pi_k^\nu)^T (h_k - T_k x^\nu).$$

Moreover, by convexity of $Q(x, \xi_k)$, it follows from the subgradient inequality that

$$Q(x, \xi_k) \geq (\pi_k^\nu)^T h_k - (\pi_k^\nu) T_k x.$$

Taking the expectation of these two relations yields

$$\mathcal{Q}(x^\nu) = E(\pi^\nu)^T (h - T x^\nu) = \sum_{k=1}^{K} p_k (\pi_k^\nu)^T (h_k - T_k x^\nu)$$

and

$$\mathcal{Q}(x) \geq E(\pi^\nu)^T (h - Tx) = \sum_{k=1}^{K} p_k (\pi_k^\nu)^T h_k - (\sum_{k=1}^{K} p_k (\pi_k^\nu)^T T_k) x.$$

By $\theta \geq \mathcal{Q}(x)$, it follows that a pair (x, θ) is feasible for (5.4.28) only if $\theta \geq E(\pi^\nu)^T (h - Tx)$, which corresponds to (5.4.20) where E_l and e_l are defined in (5.4.26) and (5.4.27).

If (x^ν, θ^ν) is optimal for (5.4.28), then $\mathcal{Q}(x^\nu) = \theta^\nu$. This happens when $\theta^\nu = E(\pi^\nu)^T (h - T x^\nu)$, which justifies the termination criterion in Step 3. So, it is the optimal solution if $\theta \geq \mathcal{Q}(x)$. Otherwise, if $\theta^\nu < \mathcal{Q}(x)$, a new constraint of type (5.4.20) is generated and added to the constraint set of the problem (5.4.18). Now, s is set to $s + 1$. After that, one must return to Step 1 and continue this process until $\theta^\nu \geq \mathcal{Q}(x)$, which gives the optimal solution of the main problem (5.1.3).

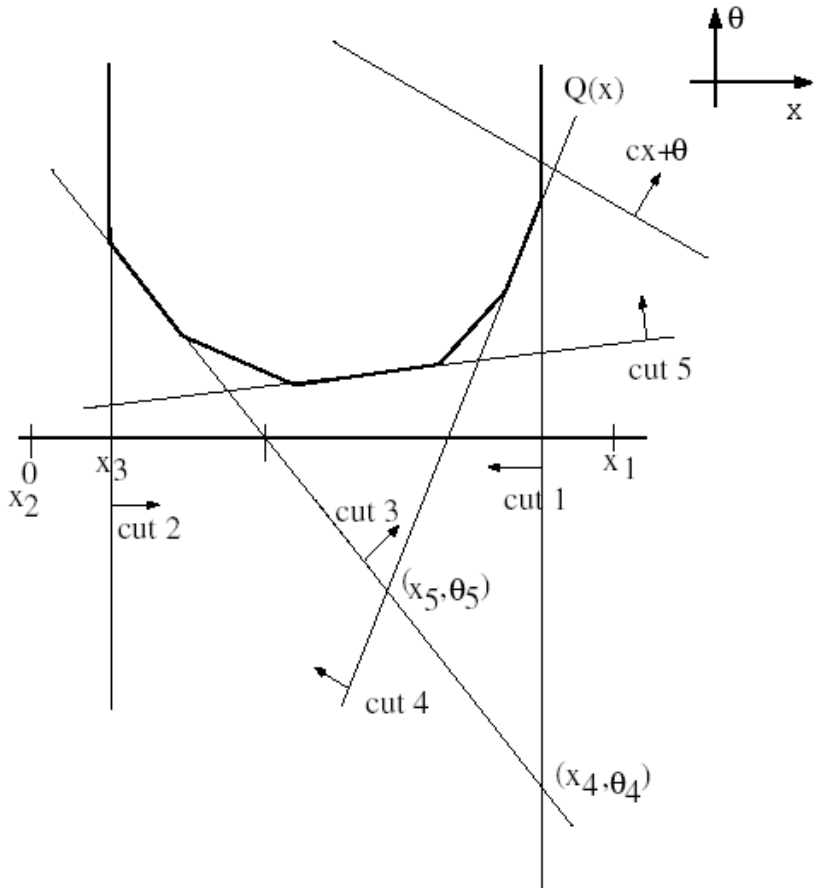

Figure 5.3: *Example of the process of the L-Shaped decomposition algorithm* [20].

The example in Figure 5.3 can be useful in understanding the L-Shaped decomposition algorithm. The five first solutions and cuts are shown. The initial $\hat{x}^1$ was chosen arbitrarily. Cuts 1 and 2 are feasibility cuts and the rest optimality cuts. $\hat{\theta}_1 = \hat{\theta}_2 = \hat{\theta}_3 = -\infty$.

In Step 3 of the L-Shaped method, a single cut is generated (optimality cut). However, it is possible to generate several cuts (one cut for each realization in the second stage), instead of one. The multicut version of the L-Shaped method is given below.

Multicut Version [7]

Step 0: Set $r = \nu = 0$ and $s_k = 0$ for all $k = 1, 2, ..., K$.

Step 1: Set $\nu = \nu + 1$. Solve the following linear program

$$\begin{aligned} \min \quad & z = c^T x + \sum_{k=1}^{K} \theta_k && (5.4.29) \\ \text{s.t.} \quad & Ax = b, \\ & D_l x \geq d_l, \qquad l = 1, 2, ..., r, && (5.4.30) \\ & E_{l(k)} x + \theta_k \geq e_{l(k)}, \quad l(k) = 1, 2, ..., s(k),\ k = 1, 2, ..., K, && (5.4.31) \\ & x \geq 0. \end{aligned}$$

Let $(x^\nu, \theta_1^\nu, ..., \theta_K^\nu)$ be an optimal solution of the above problem. If no constraint (5.4.31) is present for some k, θ_k^ν is set equal to $-\infty$ and is not considered in the computation of x^ν.

Step 2: As before.

Step 3: For $k = 1, 2, ..., K$ solve the linear program (5.4.25). Let π_k^ν be the simplex multipliers associated with the optimal solution of Problem k. If

$$\theta_k^\nu < p_k (\pi_k^\nu)^T (h_k - T_k x^\nu), \qquad (5.4.32)$$

define

$$\begin{aligned} E_{s(k)+1} &:= p_k(\pi_k^\nu)^T T_k, \\ e_{s(k)+1} &:= p_k(\pi_k^\nu)^T h_k, \end{aligned}$$

and set $s(k) = s(k) + 1$. If (5.4.32) does not hold for any $k = 1, 2, ..., K$, stop; x^ν is an optimal solution. Otherwise, return to Step 1.

Numerical experiments show that with this multicut version, the number of iterations is reduced. Furthermore, the multi cut approach is expected to be more effective when the number of realizations K is not significantly larger than the number of first-stage constraints m_1 [7].

The L-Shaped method can be also extended to the multistage method presented in the next section.

Nested Benders' Decomposition Method

For multistage problems, an appropriate horizon length, H, is difficult to determine. The horizon should be distant enough for present decisions to reflect the future accurately, but short enough to allow for efficient computation. It has been shown that, under appropriate conditions, finite horizon solutions can be found that are close to an infinite horizon optimum [8].

The general multistage problem (5.2.4) can be modelled as a dynamic program with stages $1, 2, ..., H$ and states $y^t = T^t x^t$ for $t = 1, ..., H - 1$. The following

problem is an equivalent form of (5.2.4):

$$\begin{aligned} \varsigma(y^{t-1,\xi^t}) = \min \quad & (c^t)^T x^t + z^{t+1}(y^t) \\ \text{subject to} \quad & W^t x^t = \xi^t + y^{t-1}, \\ & T^t x^t = y^t, \\ & x^t \geq 0, \end{aligned} \tag{5.4.33}$$

where $z^t(y^{t-1}) := E_{\xi^t}[\varsigma(y^{t-1,\xi^t})]$. When $t = H$, then $T^H = 0$ and $z^{H+1}(0) = 0$.

The extension of problem (5.4.18) to the multistage is the following

$$\begin{aligned} \text{minimize} \quad & (c^t)^T x^t + \theta^t & (5.4.34) \\ \text{s.t.} \quad & W^t x^t = \xi^t + T^{t-1} x^{t-1}, & \\ & D_l^t x^t \geq d_l^t, \qquad l = 1, ..., r, & (5.4.35) \\ & E_l^t x^t + \theta^t \geq e_l^t, \quad l = 1, ..., s, & (5.4.36) \\ & x^t \geq 0. & \end{aligned}$$

Figure 5.4 is an example of a three period problem with two realizations of the random vector in each period. There are 4 scenarios in period three. Each period-three scenario has an *ancestor* scenario in period two and each period-two scenario has two *descendant* scenarios in period three.

In general, the random vector for scenario j in period t is represented by ξ_j^t for $j = 1, 2, ..., K^t$. The relaxation of (5.4.34) for period t and scenario j after r_j^t

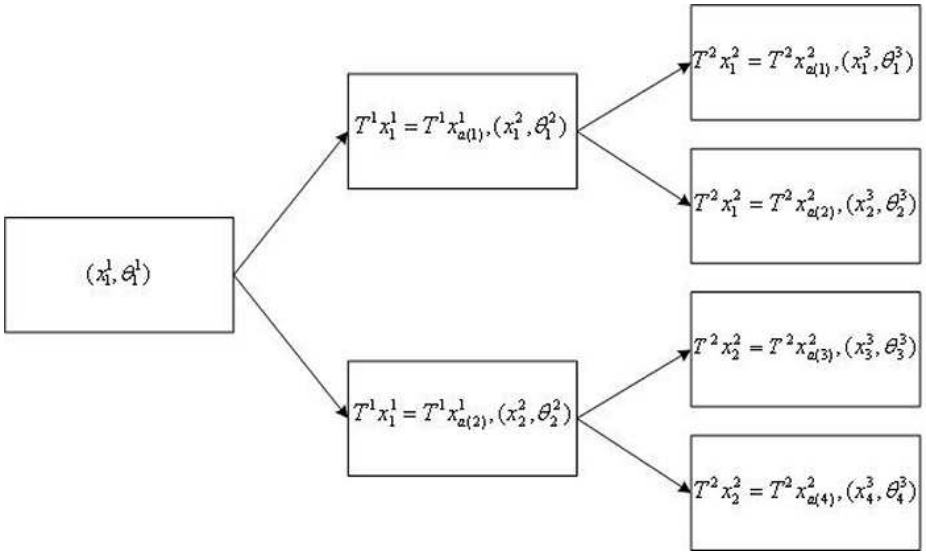

Figure 5.4: *Stages and decisions for a three-period problem* [8].

constraints (5.4.35) and s_j^t constraints (5.4.36) being added is

$$\begin{aligned}
\text{minimize} \quad & (c^t)^T x_j^t + \theta_j^t && (5.4.37)\\
\text{s.t.} \quad & W^t x_j^t = \xi_j^t + T^{t-1} x_{a(j)}^{t-1}, && (5.4.38)\\
& D_{l,j}^t x_j^t \geq d_{l,j}^t, \qquad l = 1, 2, ..., r_j^t, && (5.4.39)\\
& E_{l,j}^t x_j^t + \theta_j^t \geq e_{l,j}^t, \quad l = 1, 2, ..., s_j^t, && (5.4.40)\\
& x_j^t \geq 0,
\end{aligned}$$

where $a(j)$ is the ancestor scenario of j in period $t-1$, and $x_{a(j)}^{t-1}$ is the current solution of the $a(j)$ scenario problem in period $t-1$.

Nested Benders' Decomposition Algorithm [8]

Step 0: Solve (5.4.37) for $t = 1$ where $\theta^1 = 0$, $r^1 = s^1 = 0$, and (5.4.38) is replaced by

$$W^1 x^1 = h^1.$$

Set $\theta_j^t = 0$ and $r_j^t = s_j^t = 0$ in problem (5.4.37) for all t and scenarios j at t.

Step 1: If the period 1 problem is infeasible, then the problem is infeasible. Otherwise, let $\bar{x}^1$ be the current optimal solution of (5.4.37) for $t = 1$. Use $\bar{x}^1$ as an input in (5.4.38) for $t = 2$. Solve (5.4.37) for $t = 2$ and all ξ_j^2 $(j = 1, 2, ..., K^2)$.

If any period two problem is infeasible, then add a feasibility constraint (5.4.39) for $t = 1$, resolve for $t = 1$, and return to Step 1.

Otherwise, let $t = 2$ and go to Step 2.

Step 2: a) Let the current period t optimal solution be $\bar{x}_j^t$ for $j = 1, 2, ..., K^t$. Solve (5.4.37) for $t + 1$ and all $j = 1, 2, ..., K^{t+1}$ using the appropriate ancestor solution $\bar{x}_j^t$ in (5.4.38).

b) If any period $t + 1$ problem is infeasible, add a feasibility constraint to the corresponding ancestor period t problem and resolve that problem.

If the period t problem is infeasible, let $t = t - 1$.

If $t = 1$, go to Step 1.

Otherwise, return to Step 2a.

Otherwise, return to Step 2a.

Otherwise, all period $t + 1$ problems are feasible.

If $t \leq T - 2$, let $t = t + 1$ and return to 2a.

Otherwise $(t = T - 1)$, remove any remaining $\theta_j^\tau = 0$ restrictions for all periods τ and scenarios j at τ and, for each of these, let the current value of θ_j^τ be

$\bar{\theta}_j^\tau = -\infty$. Go to Step 3.

Step 3:a) Find $E_{l,j}^t$ and $e_{l,j}^t$ for a new constraint (5.4.40) at each scenario t problem (5.4.37) using the current period $t+1$ solutions. The vector $E_{l,j}^t$ is calculated as $-\sum_{d(j)} \pi_{d(j)}^{t+1} T^t$ and $e_{l,j}^t = \sum_{d(j)} \pi_{d(j)}^{t+1} \xi_{d(j)}^{t+1}$ where $d(j)$ scenarios are the descendants of j and $\pi_{d(j)}^{t+1}$ is an optimal dual vector in the $d(j)$ descendant problem.

b) If some j satisfies

$$\bar{\theta}_j^t < e_{l,j}^t - E_{l,j}^t \bar{x}_j^t, \tag{5.4.41}$$

then add the new constraint (5.4.40) to each period t problem (5.4.37) for which (5.4.41) holds. Solve each period t problem (5.4.37). Use the resulting solutions $(\bar{x}_j^t, \bar{\theta}_j^t)$ to form (5.4.38) for the corresponding descendant period $t+1$ problems (5.4.37) and resolve each period $t+1$ problem (5.4.37).

If $t < T-1$, let $t = t+1$ and go to Step 2a.

Otherwise, return to Step 3a.

Otherwise, $\bar{\theta}_j^t = e_{l,j}^t - E_{l,j}^t \bar{x}_j^t$ for all scenario j at t.

If $t > 1$, let $t = t-1$ and return to Step 3a.

Otherwise, the current solutions $\bar{x}_j^\tau$, $\tau = 1, 2, ..., H$ form an optimal solution of (5.2.4).

Summary

Benders decomposition is a widely used major technique to solve mixed integer programs. Van Slyke and Wets [37] applied this technique to stochastic programming problems with a dual angular structure and because of the angular structure of the constraint set it is named as L-shaped method. The extension of L-shaped method to multistage is done by Birge [8] and Gassmann [16] and

named as the Nested Benders decomposition.

5.4.2 Dantzig-Wolfe Decomposition

Another important decomposition technique is *Dantzig-Wolfe decomposition* developed by Dantzig and Wolfe [10]. Dantzig-Wolfe decomposition is related to Benders decomposition in that it is equivalent to performing Benders decomposition on the dual of some linear program. As Benders decomposition is an iterative procedure in which a new row is added to the master program after every iteration, Dantzig-Wolfe decomposition is an iterative procedure in which a new column is added to the master program after every iteration. Dantzig-Wolfe decomposition can be applied to problems with block angular structure.

The basic idea of this method is to solve the problems of the following form

$$\begin{aligned} \text{minimize} \quad & Z = c_1^T x_1 + c_2^T x_2 \\ \text{subject to} \quad & A_1 x_1 + A_2 x_2 = b, \\ & B_1 x_1 \qquad\quad = b_1, \\ & \qquad\quad B_2 x_2 = b_2, \\ & x_1, \geq 0 \qquad x_2 \geq 0, \end{aligned} \tag{5.4.42}$$

where A_1, A_2, B_1 and B_2 are matrices with sizes $m \times n_1, m \times n_2, m_1 \times n_1$ and $m_2 \times n_2$, respectively. Furthermore, c_1, c_2, x_1 and x_2 are vectors of sizes $n_1 \times 1, n_2 \times 1, n_1 \times 1$ and $n_2 \times 1$, respectively.

For the sake of simplicity, it is assumed that

$$\begin{aligned} & S_1 := \{x_1 | x_1 \geq 0, \quad B_1 x_1 = b_1\} \\ \text{and} \qquad & S_2 := \{x_2 | x_2 \geq 0, \quad B_2 x_2 = b_2\} \end{aligned}$$

are bounded. Let $W_1 := \{x_{11}, x_{12}, ..., x_{1K_1}\}$ and $W_2 := \{x_{21}, x_{22}, ..., x_{2K_2}\}$ be the sets of all extreme points of the convex polyhedron S_1 and S_2, respectively. So, any point $x \in S_j$ can be written as $x = \sum_{k=1}^{K_j} \lambda_{jk} x_{jk}$, where $\sum_k \lambda_{jk} = 1$ and $\lambda_{jk} \geq 0$ $(k = 1, 2, ..., K;\ j = 1, 2)$. Furthermore, let

$$\begin{aligned} P_{jk} &= A_j x_{jk} \\ f_{jk} &= c_j^T x_{jk} \end{aligned}$$

for $k = 1, 2, ..., K_j$ and $j = 1, 2$. Then, the following full master program is equivalent to (5.4.42)

$$\begin{aligned} \text{minimize} \quad & \sum_{k=1}^{K_1} f_{1k}\lambda_{1k} + \sum_{k=1}^{K_2} f_{2k}\lambda_{2k} \\ \text{subject to} \quad & \sum_{k=1}^{K_1} P_{1k}\lambda_{1k} + \sum_{k=1}^{K_2} P_{2k}\lambda_{2k} = b, \\ & \sum_{k=1}^{K_1} \lambda_{1k} = 1, \\ & \sum_{k=1}^{K_2} \lambda_{2k} = 1, \\ & \lambda_{1k} \geq 0, \qquad k = 1, 2, ..., K_1, \\ & \lambda_{2k} \geq 0, \qquad k = 1, 2, ..., K_2. \end{aligned} \tag{5.4.43}$$

This master program is completely equivalent to the original. It has only $m+2$ rows, compared to the $m+m_1+m_2$ rows of the original problem. It also has as many columns as the sum of the numbers of extreme points of polyhedrons S_1 and S_2, i.e., K_1+K_2 columns. To solve this full master program, *simplex method* can be used. Simplex method basically looks to the extreme points of the feasible set one by one to check whether it is the optimal solution. However, in our case, the number of extreme points (K_1+K_2) may be very large. Instead of checking all these points, a technique called *column generation* is used. In this technique, rather than tabulating all columns, columns are created to enter the basis as they are needed.

Since the master program has $m+2$ equation constraints, a feasible basis will consist of $m+2$ columns, that is, these columns are linearly independent, and the unique solution of the constraint equations obtained by setting to zero those variables associated with all other columns is nonnegative. If the simplex method is used in performing the calculations, there will also be the $m+2$ vector of prices $(\pi;\hat{\pi})$, the m-vector π associated with the first m constraints and the 2-vector $\hat{\pi}=(\hat{\pi}_1,\hat{\pi}_2)$ with the remaining two. The inner product of the price vector with any column of the basis must be equal to the cost associated with that column; in the case of master program this relation can be written as

$$\begin{aligned} \pi P_{1k}+\hat{\pi}_1 &= f_{1k}, \\ \pi P_{2k}+\hat{\pi}_2 &= f_{2k}. \end{aligned}$$

One step of the simplex method iteration for solving the master program would be performed as follows: find a column of the constraint matrix whose reduced

cost is negative, that is, for which

$$f_{jk} - \pi P_{jk} - \hat{\pi}_j < 0, \tag{5.4.44}$$

where j=1,2. Add this column to the current basis, and delete one column from the basis in such a way that the new basis is still feasible. If no column satisfying the above inequality can be found, then the current solution solves the master problem. Otherwise, the simplex method gives the appropriate rules for the removal of a column from the basis and for the calculation of the new prices associated with the new basis, with which the next iteration step can begin.

Dantzig-Wolfe Decomposition Algorithm

Step 1: Assume that an initial basic feasible solution for the master program is available, with basis matrix B and simplex multipliers (price vectors) $(\pi, \hat{\pi})$. Using the simplex multipliers, solve the following subproblems

$$\begin{aligned} \text{minimize} \quad & z_j = (c_j - \pi A_j)x_j \\ \text{subject to} \quad & B_j x_j = b_j, \\ & x_j \geq 0, \qquad j = 1, 2, \end{aligned}$$

obtaining solutions $x_j(\pi) = (\hat{x}_1, \hat{x}_2)$ and optimal objective values $(\hat{z}_1, \hat{z}_2)$.
Step 2: Compute

$$\theta_j = \hat{z}_j - \hat{\pi}_j,$$

for $j = 1, 2$. If for all j, $\theta_j \geq 0$, then the solution is optimal for the master program. Thus, the algorithm is terminated.
Otherwise ($\theta_j < 0$), form the new column as

$$\begin{pmatrix} A_j x_j(\pi) \\ 1 \end{pmatrix}.$$

Add this column to the basis and form a new basis and new prices using the rules of simplex method and return to Step 1.

Dantzig-Wolfe Decomposition Applied to Stochastic Programming

Dantzig-Wolfe decomposition can be applied to problems with block angular structure (see Figure 5.5). Once observed that the desired structures for Dantzig-Wolfe decomposition and Benders decomposition are duals of each other, Dantzig-Wolfe decomposition approach can be derived from the L-shaped method by taking duals [7].

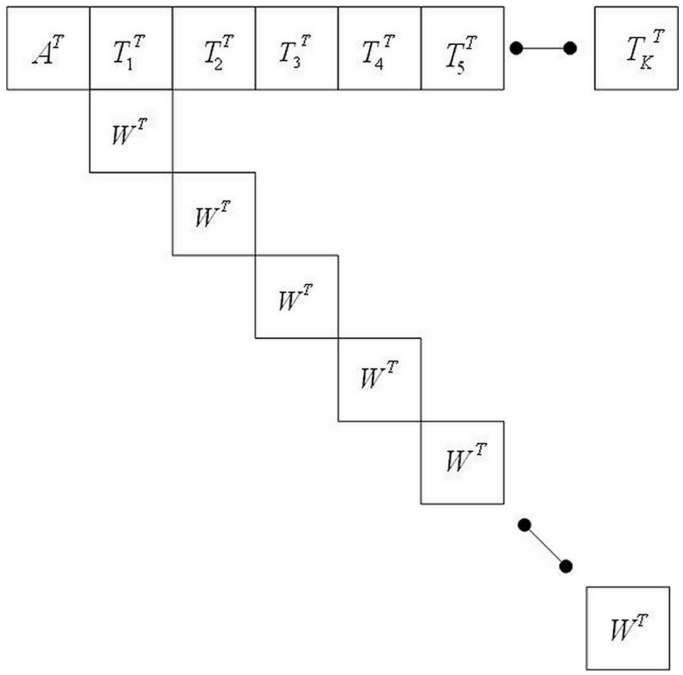

Figure 5.5: *Block angular structure of a two-stage stochastic program.*

Dantzig-Wolfe Algorithm for Stochastic Programming Models [7]

Step 0: Set $r = s = \nu = 0$.

Step 1: Set $\nu = \nu + 1$ and solve the linear program in (5.4.45), which is the dual

of (5.4.18).

$$\text{maximize} \quad \zeta = \rho^T b + \sum_{l=1}^{r} \sigma_l d_l + \sum_{l=1}^{s} \pi_l e_l \tag{5.4.45}$$

$$\text{s.t.} \quad \rho^T A + \sum_{l=1}^{r} \sigma_l D_l + \sum_{l=1}^{s} \pi_l E_l \leq c^T, \tag{5.4.46}$$

$$\sum_{l=1}^{s} \pi_l = 1, \tag{5.4.47}$$

$$\sigma_l \geq 0, \;\; l = 1, 2, ..., r, \quad \pi_l \geq 0, \;\; l = 1, 2, ..., s.$$

Let the solution be $(\rho^\nu, \sigma^\nu, \pi^\nu)$ with a dual solution, (x^ν, θ^ν).
Step 2: For $k = 1, 2, ..., K$ solve (5.4.48).

$$\begin{aligned} \text{maximize} \quad & w = \pi^T (h_k - T_k x^\nu) \\ & \text{s.t. } \pi^T W \leq q^T. \end{aligned} \tag{5.4.48}$$

If any infeasible problem (5.4.48) is found, stop and evaluate the formulation. If an unbounded solution with extreme ray σ^ν is found for any k, then form new columns D_{r+1}, d_{r+1} and set $r = r + 1$, then return to Step 1.

If all problems (5.4.48) are solvable, then form new columns E_{s+1} and e_{s+1} as in (5.4.26) and (5.4.27). If $e_{s+1} - E_{s+1} x^\nu - \theta^\nu \leq 0$, then stop; $(\rho^\nu, \sigma^\nu, \pi^\nu)$ and (x^ν, θ^ν) are optimal in the original problem (5.4.18). If $e_{s+1} - E_{s+1} x^\nu - \theta^\nu > 0$, set $s = s + 1$, and return to Step 1.
See Dantzig and Thapa [12] for a deeper treatment of Dantzig-Wolfe decomposition.

Summary

Dantzig-Wolfe decomposition, which is a useful technique for solving stochastic programming models, is closely related to Benders decomposition in that it is equivalent to performing Benders decomposition on the dual of some linear program.

Chapter 6

Conclusion

In this work, risk management for an energy company in deregulated markets is considered. To form an adequate background the characteristics of energy markets, changes in them and risk management are considered. Different types of models constructed for energy companies in order to manage risk are studied, which showed that stochastic programming is an appropriate approach for these models. Finally, decomposition techniques to solve resulting large-scale stochastic programming models are introduced.

Future work for this thesis should further consider the models constructed so far, and construct a stochastic programming model for an energy company. The model should be solved using decomposition techniques and the results should be interpreted.

References

[1] Abrahamson, P. G., *A nested decompsition approach for solving staircase linear programs* , Technical Report SOL 83-4, System Optimazation Laboratory, Stanford University, 1983.

[2] Acerbi, C., and Tasche, D., *On the coherence of expected shortfall*, Working paper, 2002.

[3] Artzner, P., Delbaen F., Eber, J. M., and Heath, D., *Coherent measures of risk*, Mathematical Finance 9, 1993, pp. 203-228.

[4] Bağci, H., *Enerji borsaları ve türev araçlarının bu piyasalarda kullanımı*, Türkiye Enerji Forumu, Vadeli İşlemler ve Opsiyon Borsası A.Ş., 2004.

[5] Beale, E. M. L., *On minimizing a convex function subject to linear inequalities*, J. Royal Statistical Society, 1955.

[6] Benders, J. F., *Partitioning procedures for solving mixed-variables programming problems*, Numerische Mathematik 4, 1962, pp. 238-252.

[7] Birge, J. and Louveaux, F., *Introduction to Stochastic Programming*, Springer, 1997.

[8] Birge, J. R., *Decompostion and partitioning methods for multistage stochastic linear programs* , Operations Research 33, 1985, pp. 989-1007.

[9] Church C., *The Time Decay in Calculating Value-at-Risk* ,Master Thesis, 2003.

[10] Dantzig, G. B. and Wolfe, P., *Decompostion principle for linear programs* , Operations Research 8, 1960, pp. 101-111.

[11] Dantzig, G. B., *Linear programming under uncertainty*, Management Science 1, 1955, pp. 197-206.

[12] Dantzig, G. B. and Thapa, M. N., *Linear Programming I: Introduction*, Springer- Verlag, 1997.

[13] Doege, J., Lüthi, H.-J., and Schiltknecht, P., *Risk management of power portfolios an valuation of flexibility*, preprint, ETH Zurich, 2003.

[14] Eichhorn, A., Römisch, W., and Wegner, I., *Polyhedral risk measures in electricity portfolio optimization*, PAMM Proceedings in Applied Mathematics and Mechanics, 2004.

[15] Eichorn, A. and Römisch, W., *Polyhedral risk measures in stochastic programming*, Preprint 20004-05, Department of Mathematics, Humboldt-University Berlin, 2004.

[16] Gassmann, H. I., *MSLiP: A computer code for the multistage stocahstic linear programming problem* , Mathematical Programming 47, pp. 407-423.

[17] Guigues, V., and Ndiaye P. M., *A value-at-risk approach for robust management of electricity power generation*,

[18] Ho, J. K. and Manne, A. S., *Nested decomposition for dynamic models* , Mathematical Programming 6, 1974, pp. 121-140.

[19] Ho, J. K., and Loute, E., *An advanced implementation of the dantzig-wolfe decomposition algorithm for linear programming*, Mathematical Programming 20, 1981, pp. 303-326.

[20] Kall, P., and Wallace, S. W., *Stochastic Programming*, Springer, 1997.????

[21] Khindanova, I., and Atakhanova, Z. *Stable modelling in energy risk management* , Mathematical Methods of Operations Research , 2002, pp. 225-245.

[22] Kleindorfer, P. R., and Li, L., *Multi-period var-constrained protfolio optimization with applications to the electric power sector*, Working paper, Wharton Risk Management and Decision Processes Center, University of Pennsylvania, Philadelphia, PA.

[23] Lai, L. L., *Power system restructuring and deregulation*, John Wiley, 2001.

[24] Li, D., and Ng, W. L., *Optimal dynamic portfolio selection: Multi-period mean-variance formulation*, Mathematical Finance 10, 2000, pp. 387406.

[25] Markowitz, H. M., *Portfolio selection* , Journal of Finance 7, 1952, pp. 77-91.

[26] Marzano, L. G. B., Melo, A. C. G., and Souza, R. C., *An approach for portfolio optimization of energy contracts in the brazilian electric sector* , Power Tech Conference Proceedings, 2003 IEEE Bologna.

[27] Mohr, M., and Unger, H., *Economic reassessment of energy technoogies with risk-management techniques* , Applied Energy 64, 1999, pp. 165-173.

[28] Ojanen, O., Makkonen, S., and Salo, A., *A multi-criteria framework for the selection of risk analysis methods at energy utilities*,International Journal of Risk Assesment and Management 5, 2005.

[29] Pilipović, D., *Energy Risk: Valuing and Managing Energy Derivatives*, New York: McGraw-Hill, 1998.

[30] Rockafellar, R. T., and Uryasev, S., *Optimization of conditional value-at-risk* , Journal of Risk 2, 2000, pp. 21-41.

[31] Sen, S., Yu, L., and Genc, T., *A stochastic programming approach to power portfolio optimization*, Stochastic Programming E-Print Series, http://www.speps.info, 2003.

[32] Ruszczyński, A., and Shapiro, A., *Stochastic Programming, Handbooks in Operations Research and Management Science 10*, NorthHolland, Amsterdam, 2003.

[33] Stoft, S., Belden T., Goldman C., and Pickle S. *Primer on electricity futures and other derivatives* , LBNL-41098, January 1998.

[34] Studer, G., *Value-at-risk and maximum loss optimization* , RiskLab: Technical Report, Zurich, December, 1995.

[35] Szegö, G., *Measures of risk*, Journal of Banking and Finance 26, 2002, pp. 1253-1272.

[36] Takriti, S., and Ahmed, S., *Managing short-term electricity contracts under uncertainty: a minimax approach*, Technical Report, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, 2002.

[37] Van Slyke, R., and Wets, R. J.-B., *L-shaped linear programs with application to optimal control and stochastic oprimization*, SIAM Journal of Applied Mathematics 17, 1969, pp. 638-663.

[38] Harper, D., http://www.investopedia.com/articles/04/092904.asp, September 29, 2004.

[39] Wittrock, R. J., *Advances in a nested decomposition algorithm for solving staircase linear programs* , Technical Report SOL 83-2, System Optimazation Laboratory, Stanford University, 1983.

[40] New York Mercantile Exchange, http://www.nymex.com.

[41] Tokyo Commodity Exchange, http://www.tocom.or.jp.

[42] Singapore Exchange , http://www.sgx.com.

[43] Nordic Power Exchange , http://www.nordpool.no.

[44] The International Petroleum Exchange of London Ltd., http://www.theipe.com.