SNF Report No. 10/09

ESSAYS ON ELECTRICITY MARKETS

by

Linda Rud

INSTITUTE FOR RESEARCH IN ECONOMICS AND BUSINESS ADMINISTRATION
BERGEN, MAY 2009

ISBN 978-82-491-0648-6 Printed version
ISBN 978-82-491-0649-3 Online version
ISSN 0803-4036

FOREWORD

The following report contains my dissertation 'Essays on Electricity Market' which was submitted December 23rd 2008 in partial fulfilment of the requirements for the Philosophiae Doctor (PhD) degree at the Norwegian School of Economics and Business Administration, Department of Finance and Management Science.

My supervisor was professor Kurt Jörnsten, Norwegian School of Economics and Business Administration (NHH). The rest of the committee consisted of professor Einar Hope, NHH, and professor Lars Bergman, Stockholm School of Economics. I also consider professor Mette Bjørndal, NHH as an important, although informal, member of the committee.

The defence took place May 11th 2009. The evaluation committee consisted of professor Richard Green, University of Birmingham, UK (leader of the committee), professor emeritus Thorkell Helgason, Parliament of Iceland, and Dr. Ing. Nina K. Detlefsen, Energinet, Denmark.

May 2009
Linda Rud

Preface

The essays of this dissertation all cover topics of electricity markets, ranging from electricity market design, to more specific problems of participants operating in the market-based electricity market. We will here give a brief outline of the essays of this dissertation.

The first introductory essay, '**Selected Topics on Early Electricity Market Design in Norway**', gives a broad, but also detailed overview of market design issues in establishing the market-based Norwegian electricity market. This paper is anchored in my early research work in the electricity market sector, prior to and in the first years of the Norwegian market reform. At the time the development of a market-based electricity system represented a pioneer reform. In our work, there was thus rather scant directly relevant literature on electricity, calling for the use of general economic theory and experiences from other markets. Basic challenges were also related to identifying the main features and attributes of electricity, and their implications for developing the market system. The essay first reviews the research project that laid important foundations for the Norwegian market reform, addressing basic principles of the competitive electricity market. Further it addresses early market design issues of the main market system building blocks covering day-ahead markets, hedging markets, ancillary services, and the pricing of grid services.

The following four essays are more narrow in their choice of topic, and to a larger extent model and analyze specific problems facing different actors of the market. The central topic of essays 2-4 is congestion. The essays cover methods for handling and pricing congestion, as well as the implication of congestion for price development and investment evaluation. Essay 5 reflects my beginning interest in the two-stage stochastic problems of the electricity market.

Essay 2, '**Capacity Charges: A Price Adjustment Process for Managing Congestion in Electricity Transmission Networks**', takes the viewpoint of the system operator and focuses on a method for handling congestion in a capacitated network. In this paper we suggest a procedure based on capacity charges for managing transmission constraints in electricity networks. The system operator states nodal capacity charges for transmission prior to market clearing. Market clearing brings forth a single market price for electricity. For optimal capacity charges the market equilibrium coincides with that of nodal pricing. Capacity charges are based on technical distribution factors and estimates of the shadow prices of network constraints. Estimates can be based on market information from similar congestion-situations, or they can be adjusted near the optimal values through an iterative process.

Essay 3, ‘**Understanding the Stochastics of Nodal Prices: Price Processes in a Constrained Network**’, takes the view point of market participants operating in the competitive market, and seeks to establish a further understanding of the stochastic nodal prices. Network congestion in competitive electricity markets may be managed by geographically differentiated nodal prices. The stochastics of an unconstrained equilibrium price reflect the underlying fundamentals of demand and supply. The stochastics of nodal prices in addition reflect the consequences of grid congestion. This paper demonstrates how a static three-node model may be combined with a dynamic modelling of fundamental parameters, giving stochastic nodal price processes consistent with the underlying grid. These price processes may be employed in understanding and analysing production, hedging, and investment decisions within a capacitated network system under uncertainty.

Essay 4, ‘**Investment Evaluation in a Constrained Electricity Network with Stochastic Nodal Price Processes**’, further elaborates on the stochastics of nodal prices, and its implications for investment evaluation. The essay studies how the interaction of the competitive market and the capacitated network affects the evaluation of investments in production and grid capacity under uncertainty. In doing so, we pinpoint several potential pitfalls, and discuss principle issues in the use of system prices versus nodal prices; how investments may affect future price processes, the mix of grid and production investments, and the issue of externalities. The issues are illustrated within a three node electricity market model, which is combined with dynamic modelling of underlying fundamental parameters, giving stochastic price processes consistent with the underlying grid.

Essay 5, ‘**A Newsboy Model Perspective on the Power Market: The Case of a Wind Power Producer**’, focuses on the interaction between the day-ahead market and the real-time market, and discusses the optimal bidding and implications of a wind power producer who does not have the ability to predict with certainty his production, nor the ability to adjust production in real-time. The paper discusses how the problem may be interpreted within the classic newsboy model. In a setting of a day-ahead and a real-time market, the results indicate that the optimal sales bid of the wind power producer might diverge from the expected production. This aspect is also found in the context of market optimization, where the uncertainty of the wind power production has direct implications for the optimal level of planned production by other producers.

TABLE OF CONTENTS

ESSAY 1

Selected Topics on Early Electricity Market Design in Norway*

Linda Rud

Abstract

With the electricity market reform following the new Energy Act of June 1990, Norway became one of the pioneers of market-based electricity systems. Market reform called for the development of the basic competitive framework, as well as the design of specific market systems. Being a pioneer, experience on the subject was limited. This paper first reviews the research project that laid important foundations for the Norwegian market reform, addressing basic principles of the competitive electricity market. Further we address early market design issues in the main market system building blocks covering day-ahead markets, hedging markets, ancillary services, and the pricing of grid services.

* This paper gives an overview of market design issues throughout the early years of the electricity market reform in Norway, and is based on research projects I have been engaged in at SAF/SNF in 1988-1998. First and foremost I would like to thank my colleagues and co-authors in these projects throughout the years, that is Einar Hope, Kurt Jørnsten, Kjell Henry Knivsflå, Balbir Singh, and Sigve Tjøtta. In preparing this overview, comments from Petter Bjerksund, Mette Bjørndal, Øystein Gjerde, Einar Hope, Kurt Jørnsten, and Kjell Henry Knivsflå are greatly appreciated. All errors are the sole responsibility of the author.

0 Introduction

With the electricity market reorganization following the new Energy Act of June 1990, Norway became one of the pioneers of market-based competitive electricity systems. Since then, several countries have followed. While the main principles of competition in the electricity trade and regulation of grid activities apply to all markets, different market structures and market designs have been chosen in different countries. This paper follows the establishment of the Norwegian electricity market, with a special focus on early issues in the development of the main organized markets and pricing systems for electricity. As a pioneer in the development of market-based electricity systems, experience on the subject was limited, with a limited base of knowledge to draw from. As such, early market design to a large extent had to draw on general economic theory, and adapt, if possible, experiences from other markets. The new Norwegian electricity market was constructed as an unbundled market, i.e. separating the different functions in different pricing systems or markets. This exposition covers general market design, and the main building blocks of day-ahead markets, hedging markets, ancillary services, and pricing of grid services, and is based on my participation in research projects at SAF/SNF[1].

The research project 'Improving the Efficiency of the Norwegian Electricity Market' led by professor Einar Hope in 1988/89 constituted a central foundation for the Norwegian electricity market reform. Chapter 1 gives an overview of background research leading to the market reform, and of early research issues in establishing the initial framework of a market-based electricity exchange in Norway. Our focus in chapters 2-5 is on early design issues related to the main building blocks of organized exchange of electricity;

- the spot market for electricity;
- organized hedging markets;
- markets and systems for the procurement of ancillary services;
- and systems for allocating constrained grid capacity.

While these 'sub-markets' all have their specialized functions, a common denominator of my work on market design is the acknowledgement of the interaction and interdependency

[1] The research projects were carried out over the period from 1988-1998, at Centre for Applied Research (SAF) and Research in Economics and Business Administration (SNF). The projects have partly been carried out alone, and partly in project teams, which over the different projects have included Einar Hope, Kurt Jørnsten, Kjell Henry Knivsflå, Balbir Singh, and Sigve Tjøtta.

between these sub-markets for the overall efficiency of the electricity market, as illustrated in figure 1. Each chapter is concluded by a brief account of current design of these market functions.

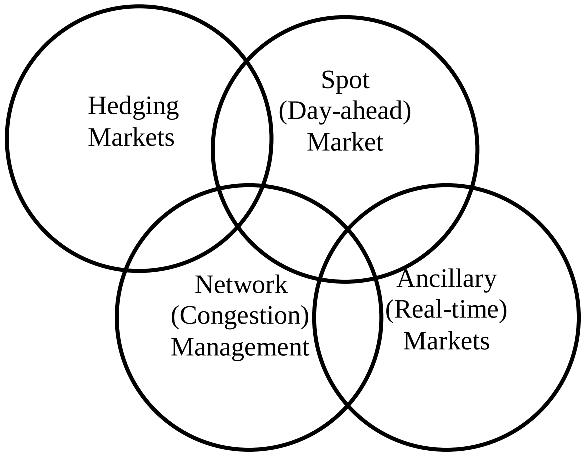


Figure 1 Electricity sub-markets

Chapter 2 focuses on the core markets for short-term allocation of power. As the Norwegian electricity market already had a well-established exchange for short-term allocation of interruptible power between producers, this market gave a flying start for the establishment of a general spot (day-ahead) market. To handle real-time imbalances, a new real-time 'regulation power' market was soon established. The main focus in this chapter is on the efficiency of the chosen market microstructure in these markets, especially with respect to the chosen auction forms and timing of markets, and their efficiency implications for each market, as well as for the interaction between the spot market and the regulation power market.

In a market where uncertainty is a prevalent feature, hedging opportunities are important. Power exchange prior to the reform was largely based on long-term bilateral contracts with a large degree of political and centralized pricing. An early challenge of the reform was to establish more efficient forms of price hedging. Early issues in establishing a futures market for electricity, and the interaction with the spot market are the topics of chapter 3.

With an increasing focus in the mid-nineties on the value of real-time power capacity and ancillary services, attention was drawn to issues of the efficient and sufficient procurement of real-time capacity and ancillary services. The interpretation of security and quality of supply as a public good, and the implications of this for the procurement of ancillary services is the topic of chapter 4. Here the interaction of the spot market and systems or markets for ancillary services is again important.

Electricity is a bundled commodity of energy and transport. The importance of this is reflected in the organization of the overall market structure as commented in chapter 1, but has also important implications for the details of pricing network capacity, including congestion pricing in the spot market. Chapter 5 reviews early research work on congestion pricing and the interaction of the spot (day-ahead) market and network flow.

Chapter 6 concludes this paper by commenting on the efficiency of the market after the reform, and discussing areas for further improvement.

1 The Norwegian Electricity Market Reform

1.1 Background of Reform

At the turn to the 20th century the electricity industry represented a new technology. In developing the vast amount of geographically dispersed hydro-resources in Norway, local communities contributed to electrifying the nation. Gradually electricity became a necessity of modern life, and the sector became a mature industry. Market organization and exchange patterns originating from the pioneer days of electricity gradually adapted to the changes of modern society. In the late 1980's, however, the concern was that the electricity market should be more efficient. Time now seemed ripe to modernize the electricity market.

In the late 1980's the philosophies of how to meet the goal of a more efficient electricity market were quite divergent. On one side the Norwegian Water Resources and Power Board (NVE), a government directorate, had been advocating a plan of 20 regional vertically integrated companies, a plan that was manifested in a new law proposition. On the other side, influenced by the international trend of deregulation and competition within previously publicly driven and regulated sectors[2], emerged the idea of the market as a tool for achieving economic efficiency. In 1988 the Ministry of Finance and the Ministry of Oil and Energy ordered a research project with the objective:

> *'to analyze the possibility for increasing the efficiency of the existing Norwegian electricity system by developing a market-based exchange system, with economically rational participants, and public regulation principles and policies adapted to the special technological and economic conditions of generation and power trade in a hydro-based system'.*[3]

The project was carried out by the research institution SAF[4] (Center for Applied Research) in Bergen, with professor Einar Hope as the leader and main driving force of the project team. Aside from the project premises to keep a focus on the ongoing operation of the market, and a premise of no change as to ownership[5], the project team had free hands to create an idealized market-based electricity system. Project focus was on the principles and economics of market-based power exchange. After an analysis of market inefficiencies of the current

[2] Examples are the sectors of transportation and telecommunications.
[3] Translated into English from quote page 2 in Bjørndalen, Hope, Tandberg and Tennbakk (1989).
[4] The research institute is now named SNF – Research in Economics and Business Administration.
[5] 85% of the Norwegian power sector was publicly owned at the time.

market, an analytical and operational concept for a market-based electricity system was developed with proposals for institutional, organizational, and regulatory measures to be taken to implement the system. The project was documented in the main report Bjørndalen, Hope, Tandberg, and Tennbakk (1989) 'Market-based Electricity Trade in Norway', and the underlying thematic reports of Rud (1989a), Rud and Tjøtta (1989), Strandenes (1989), Lommerud (1989), Vaage (1989), Berg (1989), Rud (1989b), Singh (1989), Amundsen (1989), and Tandberg and Tennbakk (1989).

Subsequently, the former law proposition of 20 vertically integrated energy companies was withdrawn. A new Energy Act (Law of production, transmission, sale and distribution of energy) was approved by the Norwegian Parliament in June 1990, and became operative from January 1 1991. Formulated as it is in general terms, the law's intention of a market-based competitive market is more explicitly given in the underlying proposition, Ot.prp. nr. 43 (1989-90). This proposition to a considerable extent built upon the market based concept and the argumentation and proposals from the SAF-research group. The measures taken by the government to implement the new system also were essentially in conformity with this basic reasoning. With this, the principle of a market-based electricity system became firmly established from an operational and regulatory point of view.

The purpose of this chapter is to give an overview the Norwegian market reform. Section 1.2 motivates the main underlying principles of competition in trade and regulation of wires, principles which have proven a common denominator in organizing competitive electricity markets throughout the world. The remainder of this chapter is dedicated to research underlying the specifics of the Norwegian market reform. The former market structure is depicted as a flow diagram in Figure 2. The figure shows the main trading patterns, where traders in the wholesale market consisted of generators, distribution companies, and large end-users. Retail was carried out by local distribution companies. Sections 1.3 – 1.6 review issues of perceived potential sources of market imperfections in this market, and summarize the most important implications for market organization and policy measures. The reform measures were to enforce paramount changes in the underlying institutional and operational structures of the market that induced an inherently competitive and decentralized trade of electricity. Section 1.7 reviews issues of building a new market infrastructure adapted to the requirements of a competitive market. Section 1.8 discusses implications for market participants on various levels of the market. The exposition is partly based on the research documents of the above-mentioned research project, and partly on the proceeding early research projects elaborating on further issues discussed in the infant stage of the market reform.

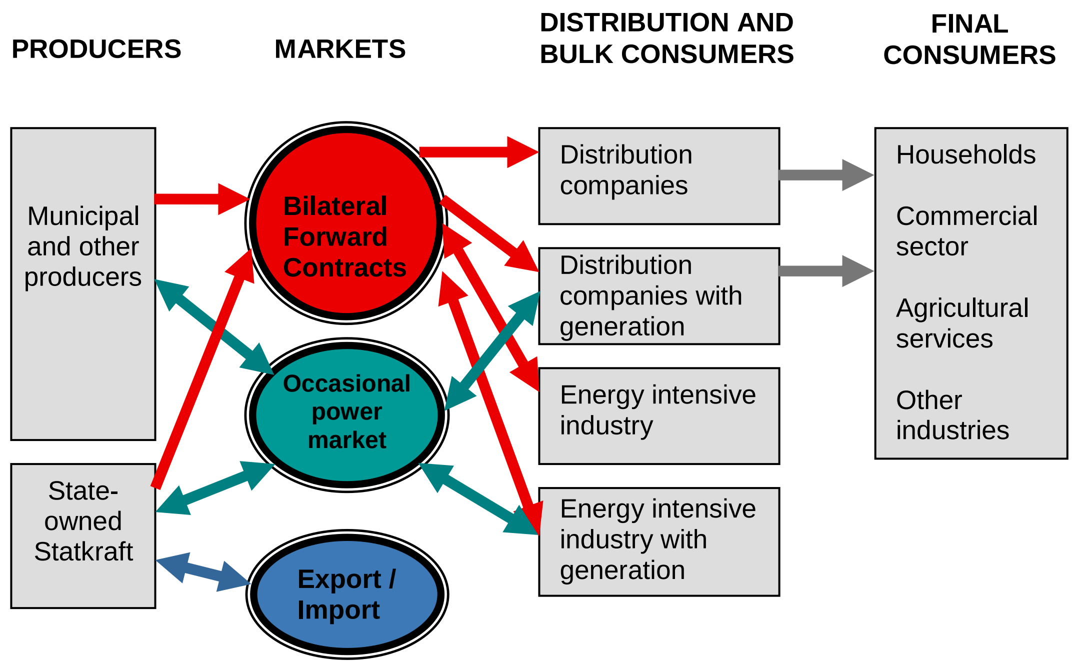


Figure 2 Structure of the Norwegian power sector prior to the reform

1.2 Basic Premises for a Competitive Electricity Market

A competitive market structure is first and foremost a tool for achieving economic efficiency. From basic economic theory, a perfectly competitive market is an efficient market. Such a market is distinguished by several features: a homogeneous product, perfect information, a market structure with a large number of small and non-dominating participants on the supply and demand side, and no barriers to new entry. Markets in this pure form, however, exist only in theory. Potential market imperfections are those of market power and barriers to entry, transaction costs, asymmetric information, public goods, and externalities. The degree and nature of market imperfections affect the efficiency of the market. In restructuring a market to achieve better efficiency, the prevailing market imperfections also have important implications for the potential of competition and necessary measures to achieve this.

For electricity, some sources of market imperfections are inherent to the nature of electricity, and imply general premises for organizing competitive electricity markets. Other market imperfections are market specific and may call for different measures tailored to the underlying structure of the specific market. While the following sections deal with specifics of the Norwegian market of following reform, we will here first discuss general issues of the inherent nature of electricity, which are of importance to the organization of all electricity markets.

Let us first review some basic features of the product. Electricity is supplied and consumed continuously. In measuring the quantity of the product we differentiate between power and energy. Power denotes the energy load at a given instant of time, and is measured in Watt (W). Energy is the total energy load produced or used over a time period, and is measured in Watt hours (Wh). In relation to this, the tradable product electricity is conventionally quoted in terms of *energy* (Wh) produced or consumed within a time period, for example an hour[6,7].

The final delivered product is thus a given quantity of electricity to be supplied in a given time period at a given location. As electricity is essentially a non-storable[8] product, the delivered product is essentially a bundled product of energy and transport. We will see that this interdependence, combined with the fundamental properties of the cost functions of electricity transportation, sets important general premises for organizing markets for electricity. Below we will first discuss basic characteristics and implications for the organization of electricity transportation, before proceeding to the commodity electric energy and the potential for competitive trade in this product. The section is based on Bjørndalen et al. (1989), and Rud and Tjøtta (1989).

1.2.1 The Natural Monopoly Feature of Transportation

Electricity transportation comprises both transmission and distribution. The transmission network basically covers the inter-regionally higher-voltage network, while the distribution network mainly covers the intra-regionally lower-voltage networks[9]. The main distinction is between network infrastructures that service several regions and networks that service only one region. The general cost characteristics are, however, similar for transmission and distribution:

[6] The product normally is quoted in energy. In principle pricing schemes may also include power load related prices, as for example a former Norwegian pricing tariff which charged higher prices for energy when the load was above a certain level. In the following, however, we will follow the convention of quoting electricity in terms of energy only.

[7] Note that 1 kWh = 1000 Wh, 1 MWh = 1000 kWh, 1 GWh = 1000 MWh, and 1 TWh = 1000 GWh.

[8] It is not cost-efficient to store electricity on a large scale, though storage is used small-scale as in batteries.

[9] Note that this voltage-based definition is not accurate, but gives a broad picture of the distinction between transmission and distribution. In considering specific lines, we might find transmission lines of a medium voltage, or distribution lines with higher voltages. Also note that the Norwegian grid in effect is divided into three different levels: the nation-wide grid (from one main area of the country to another, normally 132, 300 or 420 kV), regional grids (from the main grid to the local distribution grids in the area, normally 66 and 132 kV), and the local distribution grids (transport to the final customer).

- *Variable costs* account for a minor part of total transmission and distribution costs. The main variable cost category is that of heat losses. The percentage heat loss is lower in high-voltage lines than in low-voltage lines.
- *Fixed costs* account for the greater part of transmission and distribution costs, and are mainly associated with investments in transmission and distribution lines. Periodical variations over the day and over the year imply that the network has to be dimensioned to meet normal peak demand. It may also be cost-efficient to dimension capacity in relation to future expected increases in demand. Also note that the cost per unit transported is lower when building (and using) higher capacity lines, than when building the equivalent capacity in several lower capacity lines.

In short, the cost functions of transmission and distribution imply that a horizontal integration of such activities is cost-efficient, as the cost functions exhibit economies of scale. The conclusion is that transmission and distribution are natural monopolies for the areas they cover[10]. From a market perspective this means that competition is unsuitable in the supply of transmission and distribution[11]. This means that to achieve cost-efficiency, a single company should provide transportation in the area for which the characteristics of natural monopoly apply. This implies a single nation-wide transmission company, and several regional and/or local distribution companies[12].

The market imperfection of natural monopoly in transmission and distribution has crucial implications for network organization and control. This is important for several reasons: All participants that produce or consume the non-storable product electricity depend on access to the network, and are at the ‘mercy’ of the local grid to which they are connected. A mixture of competitive and natural monopoly activities in a company, furthermore, represents a potential means of exercising market power, either by direct exclusion of competitors, or more subtly and less transparent, by e.g. stating high transport tariffs combined with internal cross-

[10] See Rud and Tjøtta (1989) for a general discussion on this issue.

[11] A sub-additive cost function implies that it is cost-efficient for one production entity to supply the commodity, i.e. that a horizontal integration of the activities is cost-efficient. One company can produce the commodity cheaper than two or more. In general, the implication for a monopoly organization rather than a market organization is, however, not straightforward. First, to induce cost-efficiency in a monopoly, regulation is in general required. Also note that, for example, for mild degrees of sub-additivity, the benefit of potentially lower costs may be outweighed by the efficiency gains of competition. In relation to the technology of electricity transport, however, the conclusion of a natural monopoly and inefficiency of competition, however, seems quite clear.

[12] The actual extension and demarcation of the natural monopoly of transmission networks versus regional networks and versus local networks is, however, a matter of discussion, as discussed in Bjørndal, Hope and Husum (1994).

subsidies of activities. Also note that by holding a monopoly in the supply of transport services, the transmission or distribution company has no inherent incentives for cost-minimization and efficiency. For these reasons public regulation of natural monopolies activities is required.

1.2.2 Competition in Energy Trade

An important premise for a liquid and efficient competitive commodity market is that the traded product is sufficiently homogeneous, so that it is possible to define standardized units or contracts for trading. A fundamental question is whether this applies to electricity. The final delivered product is a quantity of electricity delivered in a given time period, at a given location. By this definition, this final product may in fact be characterized as heterogeneous as it is differentiated by several quality attributes, such as e.g. the probability distribution of interruptions, and variations in voltage. It is now necessary to distinguish between transport and energy. The mentioned quality attributes are first and foremost attributes of the transport part of the bundled good, i.e. of network configuration, management, maintenance, etc. As we saw above, this part of the bundled good cannot be made competitive.

In considering competition in electricity, we must thus turn to the issue of competition in the trading of the pure energy part, i.e. energy supplied within a given time period. Several basic characteristics of the underlying technologies for supplying this part of the good strengthen the notion of electricity as a homogeneous product: There is no direct and identifiable connection between the physical electricity bought and sold on a contract. It is not possible to steer the electricity in the grid to selected buyers, as the flow follows from physical laws and the overall monitoring of system balance. The amount of electricity on the grid at any time may be characterized as a pool of homogeneous energy. This implies that the energy part of the bundled good is highly homogeneous. It is still possible to verify and settle the energy accounts of the individual participants, as the quantity of input to or output from the grid may be measured at each connection point.

The next question is whether the economic characteristics of the production of energy fundamentally support competition. We note that there are high fixed costs related to investments in generation capacity. A large part of variable costs are the fuel costs. For hydro power fuel costs virtually are the alternative cost of water[13]. Cost structures and levels in

[13] The alternative cost of water is the foregone income of later production, and is termed ‘water value’. This is based on the reasoning of Hveding (1968): ‘A kWh in storage should just be withheld, if its value at the moment is exactly the same as the value it can be expected to have at a later point in time’.

general vary for different technologies. Costs and effectiveness also differ as to meeting base load versus meeting peak load demand, and as to the supply of ancillary services. For some technologies there are also step-wise costs connected to ramping and unit commitment requirements. Observing the nature of production cost functions, Rud and Tjøtta (1989) argue on a more general basis that the average cost functions normally rise near full capacity, and that the optimal size of generation plants normally is greatly lower than market size, calling for several generators to serve the market. For hydro generation plants, this especially is the case, as plant size is dimensioned by nature-given hydrological features. For hydropower systems though, it must be noted that there may be externalities within limited hydrological domains, an externality which in principle may be handled by contracts or horizontal integration. To conclude, however, with the requirement of a large number of producers to serve a market, the basic potential for competition did prevail.

1.3 Handling the Natural Monopolies of Transmission and Distribution in a Competitive Norwegian Electricity Market

We now turn to the Norwegian Market. At the time of the reform the transmission network was mostly owned by production companies, and state-owned Statkraft owned nearly 85%. All transmission lines were, however, leased to the Norwegian Power Pool (Samkjøringen av kraftverkene i Norge), which operated the transmission grid and administered the central grid settlement[14]. Regional transmission/distribution and local distribution was carried out by regional and local distribution companies that were publicly (often municipally) owned. The local distribution companies normally integrated the functions of distribution transport and of retail power trading. Their trading function consisted of wholesale power purchases on one side and retail sales on the other. The transport function included e.g. installation, maintenance, investments, metering, technical and economic management of the network, etc. Some distribution companies also had their own production plants as well. An important pre-reform structural feature was moreover the fact that the distribution companies had a manifested institutional monopoly right and obligation to supply their distribution area.

Following the conclusions of natural monopoly in transmission and distribution, the SAF research group recommended several measures. Firstly, competitive and natural monopoly activities were to be separated. On the transmission level, the recommendation was to establish an independent national grid company. Secondly, a regulatory regime had to be

[14] In addition the Norwegian Power Pool operated the occasional power market, which was a short-term market for power. See section 1.5.1.

established to ensure access to the network on non-discriminatory terms, and give incentives for cost-minimization and an economically efficient pricing of network facilities.

The main principles are illustrated in figures 3 and 4. Figure 3 illustrates the underlying grid. In a market-based system, transmission, as an inherently a national natural monopoly, should be organized in a national grid company. As local natural monopolies, the distribution and transmission activities on the regional and local levels, were to be organized in regional and local grid companies. All companies are to be regulated to ensure efficiency and to enforce common carriage. The important implication of these measures is to create a viable environment for competition in trade, where trading patterns are independent of the underlying physics of the grid. This is illustrated in figure 4 which indicates that producers and consumers have the freedom of choosing their trading partners, irrespective of the underlying physics of the network.

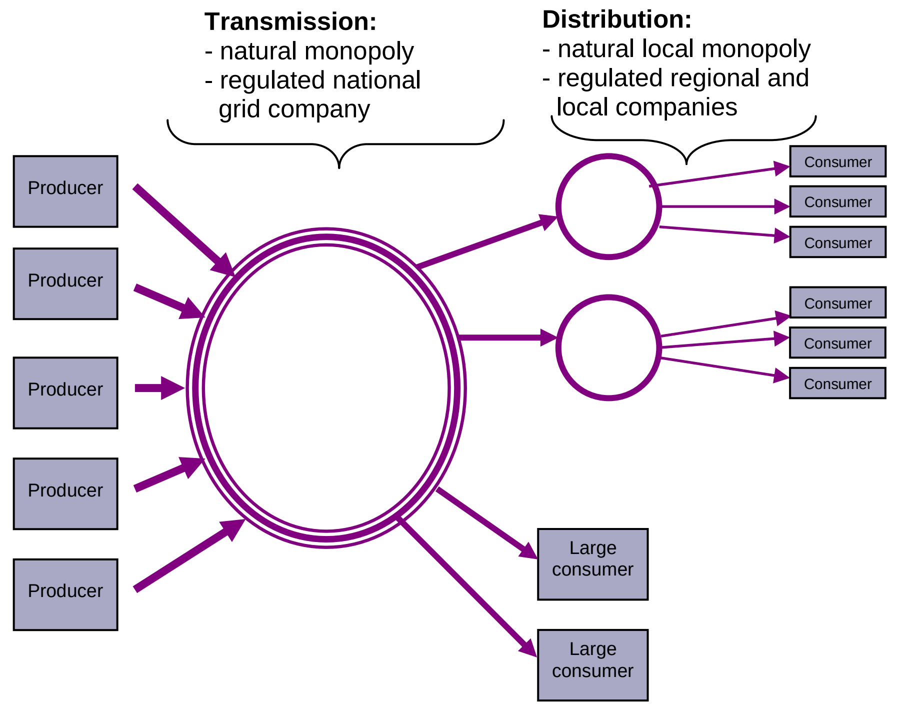


Figure 3 Market structure: Regulated natural monopoly networks

These recommendations were followed up in the Energy Act of June 1990[15]. In the underlying proposition of the new Energy Act, energy transmission and distribution were acknowledged as natural monopolies. The principle of common carriage was laid down for all Norwegian networks. This was obtained through the mechanism in the law of giving out so-called area concessions to grid owners (§4-1). To be granted a concession, the law specifies

[15] An overview and discussion of the main sections of the new Energy Law is given in Rud (1990c).

several types of conditions that may be set, including conditions that in principle grant the basis for regulating the sector. In practice the Norwegian Water Resources and Power Board regulates and administers the concessions. An important condition is the obligation of the concessionaire to connect to the grid and deliver power to the area, and not block deliveries from other parties through the grid. Other conditions may be related to the internal organization of companies, the principles and levels of transport tariffs, and the obligation to submit required information. It should here be noted that already from the start, customers on all levels were given the ability to choose their suppliers freely, with no requirements for small customers as to specific electric meters.

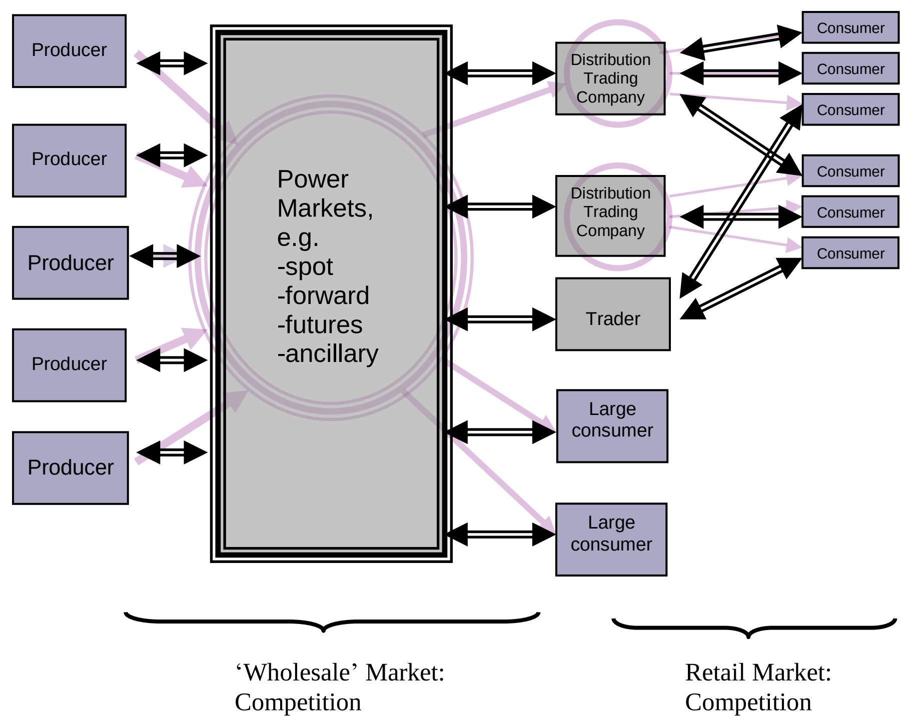


Figure 4: Competitive power markets

For transmission, the former system with central cooperation and arrangements for transmission represented an excellent starting point. The national transmission company was established by splitting the former Statkraft into two companies: Statkraft SF, which became a pure generating company and Statnett SF (The Norwegian Power Grid Company)[16], which became the national grid company. For distribution, the principle of common carriage was laid down, and trading and transport functions were to be separated. Rather than a full

[16] SF means state undertaking or company and indicates a special legal business entity. A SF is wholly state-owned with the state guaranteeing for the company's debt, but otherwise the company can, in principle, act independently.

separation with legally distinct trading and distribution companies, a less rigid measure was implemented, mainly with requirements of an internal divisionalization of the companies, combined with requirements of clearly defined and separated accounts.

1.4 Conditions of Market Power in the Norwegian Market and Implications for Market Structure

Market power on the supply or demand side may in general lower the efficiency of the market. This imperfection may follow from underlying factors determining the participant structure, or from institutionally imposed measures. Though the current market structure displayed a large number of suppliers and traders, several measures were necessary to promote a competitive market structure in Norway.

1.4.1 Market Structure of Energy Generation

For electricity supply, the Norwegian electricity system is almost exclusively based on hydro-power generation. Power production in 1991 was 111 TWh, which was equal to the average production capacity of the system, though the difference in power production between a wet and a dry year could be in the range of 33 TWh. The reservoir capacity was around 80 TWh, and does to some extent provide a buffer against annual and seasonal variations. The maximum generating momentary capacity amounted to about 27.5 GW in 1992.

The number of producers prior to the reform was quite large, with approximately 70 (mostly publicly owned) production/wholesale companies, though where 33 companies owned approximately 96% of the production capacity[17]. Compared to the supply side structure of many other countries, and other industries as well, the Norwegian electricity market was characterized by a large number of participants, making a good starting point for competition.

A issue of concern, however, was the dominating role of the state-owned Statkraft which supplied near 30% of total power production in Norway, owned more than 80% of the transmission grid, organized grid investments, had the exclusive right of foreign trade, and was the largest participant in the occasional power market. The company had also a relatively strong political position and was used by the government as an instrument for implementing parts of social and industrial policies through e.g. subsidized electricity prices. The double role of Statkraft as a transporter and a producer was resolved by the splitting of the company

[17] In addition, several distribution companies had their own production, and there were 94 industry-owned and 21 farmer-owned production facilities generating power, though mostly for own use.

into Statnett SF and Statkraft SF. The administration of foreign exchange followed Statnett SF, which was to be an independent non-profit national grid company (see section 1.7.2).

The overall general conclusion was that market concentration on the supply side was acceptable, even given the relatively strong position of Statkraft SF. However, the provision of adequate competition policies and market surveillance was deemed an important part of the infrastructure of the market-based electricity market, to control and inhibit market power and anti-competitive collaboration. A new regulatory regime was established, with considerably greater emphasis on economic regulation and competition policy issues. The main regulatory body, the Norwegian Water Resources and Power Board, was also to cooperate in matters related to market structure and competition policy with the Norwegian Price Directorate, the (at the time) government directorate for competition policy. To promote contestable competition, entry barriers should be at the minimum (though technically secure) level, thus opening for new participants.

A further issue was the existence of a large number of vertically integrated companies, where power producers supplied their own distribution trading companies. This implied that the internal parties to these transactions were not directly exposed the market, and thus would make internal cross-subsidization possible. As such, the vertically integrated structures were considered a hindrance for the development of a liquid and efficient power market. Through the new energy law it was instituted that vertically integrated companies had to split the organization into a production division and a distribution trading division, with a stated intention that the divisions were to operate as independent entities in the markets.

1.4.2 Market Structure of Wholesale Demand and Retail Market Supply

Market power on the demand side can also be a potential threat to a competitive market. The demand side of the Norwegian wholesale market consisted of large power intensive customers, and of traders that represented retail end-users. The power intensive industry was supplied on special state contracts, and was as such not part of the market reform. The remaining demand was supplied by the traders of the market, mostly distribution companies. In the market the traders thus played a two-sided role, as demand in the wholesale market, and as suppliers in the retail market. In the Norwegian system prior to the reform, the retail trading function was in its entirety carried out by distribution companies. As such the number of traders was large, with as much as about 230 (mostly publicly owned) distribution companies. Sizes, however, ranged from the smallest with 150-200 consumers with an annual consumption of 3-4 GWh, to the largest company serving 280,000 consumers with a 7000

GWh consumption. As such this large a number of trading companies also represented a fine starting point for a competitive market. An issue of concern was, however, the strong vertical ties between production and distribution companies either through ownership or long-term contractual arrangements, and where approximately 75% of total electricity demand from the household and commercial sector was covered by vertically integrated companies.

In the pre-reform system, however, competition was barred in the retail market by the institutionalized monopoly of supply which gave the distribution company the right (and obligation) to supply customers in the regional area of their distribution network. The abolishment of this institutional regional monopoly of sales was therefore an important reform measure. Combined with the new principle of non-discriminatory access to the grid, this provided the basic foundations for future competition in the retail market. Other important measures were the above-mentioned functional disintegration of vertically integrated companies into separate production and distribution trading divisions. Entry barriers for new traders were also lowered. In order to trade power, a trading concession issued by the Norwegian Water Resources and Power Board was needed, while there was free entry for pure brokers. This led to a rapid entry of brokers and traders into the market, where already after a few years a considerable number of new agents had entered the market in the form of independent brokers and traders[18]. It should also be noted that an important expected effect of increased competition in the retail market was an expected improvement in the representation of underlying demand in the market - where formerly the notion of serving an aggregate and rather price insensitive demand, had prevailed.

1.5 Transaction Costs and the Efficiency of Power Trade

The power markets are the core of a competitive market. A major issue of the research underlying the market reform was to analyze the efficiency of existing rules and practices of trading, and to consider alternative market designs for improved efficiency. The market structure prior to the reform was illustrated in figure 2. Below we will describe the former organization, and review the main efficiency problems and suggested measures on a wholesale and retail level. Some comments are also given on the issue of foreign exchange. This section is based on Bjørndalen et al. (1989), Rud (1989b), and Hope, Rud and Singh (1993a).

[18] See Rud (1996) for a discussion of the role of traders and brokers in the electricity market.

1.5.1 Domestic Wholesale Market

The participants in the pre-reform domestic wholesale power market were producers, distribution companies that represented general consumption, and large energy intensive industries. About 90% of the domestic wholesale trading was conducted by long-term bilateral contracts, while the remaining 10% was exchanged on a short-term organized market – the occasional power market'[19].

- *Bilateral forward contracts:* The bilaterally traded forward contracts were, by institutional decree, *firm power contracts* under which the seller had an obligation of full security of delivery. Firm power also had priority in the network. Near 40 % of the firm power contracts were contracts with the state-owned Statkraft with contract prices and terms determined by the Norwegian Parliament, the Storting, and had highly subsidized prices towards the power intensive industry. The contracts were fairly long term, from 5 years for the latest general consumption contracts, to 50 years for the oldest power intensive industry contracts[20]. The contracts could not be resold, and there were stringent restrictions on reselling power[21]. Furthermore, of the remaining 60% of firm power production, a large part was not exchanged in the market, but produced for own use by vertically integrated power companies and large industries. The Statkraft contract terms typically served as a norm for these contract terms, and thereby influenced a very large part of the wholesale market and substantiated the degree of administratively and politically based pricing in the market.

- *The occasional power market:* The exchange on the organized occasional power exchange covered 10 % of domestic wholesale trade. The market was administered and operated by the Norwegian Power Pool (Samkjøringen av kraftverkene i Norge), a power exchange body and national clearing house established in 1971[22]. This market had developed in response to the need for balancing overall energy supply and demand in the system. In principle the spot market was a device for shuffling water in the hydro-power system from producers with excess supply to producers with undersupply in relation to their contracts. By institutional rules these short-term contracts had a lower security of delivery, and had priority in production and network after the firm

[19] In Norwegian: 'Tilfeldigkraftmarkedet'.
[20] The contract period had however gradually been reduced over the last few years in response to changing market conditions and preferences.
[21] An overview of Statkraft firm power contract terms is given in Rud (1989b).
[22] The Norwegian Power Pool also operated the main grid in Norway and administered the Central Grid Settlement.

power contract deliveries. Market participation was restricted to producers with a minimum production capacity of 100 GWh, and Statkraft was in general the largest trader. The exchange operated on market-based equilibrium prices. Once a week the participants quoted supply and demand curves (a vector of price-quantity offers and bids) for delivery in the stated time periods of the following week. On the basis of this information the Norwegian Power Pool derived aggregate supply and demand schedules, and a market-clearing equilibrium price was determined at the intersection of the supply and demand for each defined time period of the coming week.

An important reason for the high share of firm power in the system was that the companies were under a regulatory obligation to cover themselves up with long-term contracts to secure a satisfactory degree of security of supply in the system. This institutionalization of using firm power contracts as the main trading form may partly be interpreted as the way the system had chosen to deal with the inherent uncertainty in the market: On one hand, the firm power contracts offered price hedging through their fixed contract prices. On the other hand the distinction between firm power and occasional power segregated the market into one part with a high degree of supply certainty, and a second part that absorbed the greater part of quantity uncertainty. From this point of view, it may be asserted that this 90% firm power share also reflected the dimensioning of the Norwegian power system with respect to the chosen level of supply with a high security of energy supply. As such this segregation might also be interpreted as a means of handling the perceived low degree of flexibility on the demand side, - though it may be argued that the low flexibility of the demand side in part actually was a side effect of the former organization of the market system. With these rigidities of the supply and demand side, the occasional power market played an important role in providing flexibility in the market.

From an efficiency and market point of view, the analyses of these market arrangements of wholesale power trade revealed several severe sources of inefficiency:

- *Political interference:* From a market point of view, the large degree of political interference in trade represented an important source of inefficiency. In a competitive market, prices, quantities, and contract terms are decided through decentralized market transactions, where rational participants weigh marginal benefits up against marginal costs. This makes the transition away from politically dictated contract terms a fundamental step for achieving an efficient and competitive market. Here it was important to promote a widespread understanding of what a well-functioning market can achieve, and equally important what the market cannot achieve: A competitive market is a tool to achieve economic efficiency, and is not suitable for other objectives as e.g. subsidizing geographical areas or groups of people or industries. Moreover, market interference to achieve other goals may impair efficiency. To this end, other supplementary economic and political tools that minimally distort the market are required.

- *High transaction costs:* The bilateral organization of forward trading, and the institutionalized requirement of firm power contracts was held to be the other major source of inefficiency. The inefficiencies may be summarized as high transaction costs[23], information problems and low degrees of transparency, severe rigidness in second-hand trading, the absence of market-based pricing, and in a sense a compulsory level of hedging due to the firm power requirement. Renegotiating contracts in such a system was cumbersome and costly, also making the greater part of the electricity market extremely inflexible. Another aspect was that the system to a large extent reflected a production-oriented pricing, while consumer valuation of electricity and potential flexibility were poorly represented in the pricing process.

To develop a more liquid and thus more efficient market for price hedging, the main proposal was to establish a futures market - with standardized contracts, second-hand trading, and open admittance. Futures markets in principal offer market participants flexibility and liquidity, allowing them to establish positions and close them with ease. Although the futures contract might not be a perfect match to the underlying physical position, the futures contract is an important risk-sharing instrument if the futures price is correlated to the price of the physical position. By collecting and disseminating information via market prices, futures prices also constitute an important source of information. For these reasons the establishment of a futures market in electricity was considered an essential element of the reorganized electricity sector in Norway. Details as to the first stages in establishing a futures market for electricity are treated in chapter 3. Note, however, that unlike for example the initial English market, in Norway, bilaterally negotiated contracts were to live side by side with trade on the organized spot market, which as such was to be a non-mandatory power exchange.

[23] Transaction costs on one hand refer to the direct costs occurred in trading. For trade by bilateral contracts this refers to for example costs of collecting information, negotiating and setting up contracts. From a market point of view there are, however, also implicit costs, as larger transaction costs imply a larger efficiency loss relative to a frictionless market.

The spot market is in many ways the core of an efficient commodity market, and a prerequisite for a well-functioning futures market. Here, the occasional power market, which was founded on market-based principles, represented an early and innovative market arrangement for a sector like electricity, and provided a promising starting point. This market had also given a large group of actors, the producers, a familiarity with principles of market-based exchange. Problems of the current system were mainly related to the restricted admittance to the market, where only power producers that fulfilled their firm power requirements were allowed. The main proposal was thus a further development of the occasional power market into an efficient 'spot' market, with an open admittance, though on necessary non-discriminatory technical and economic terms to traders on all levels. This was also expected to contribute to a better representation of demand in the market.

Though the aspect was not explicitly analyzed, the pre-reform research project also acknowledged the necessity of efficiently dealing with system requirements, including systems for momentary balancing of production and consumption. In implementing the reform, the regulation power market was established. This is the device used for short-term real-time adjustments by the entity that is responsible for the short-run operational co-ordination of production and transportation in the electricity system. As such it is primarily a market for the supply of active production capacity on short notice to secure the operation of the system. To handle any further needs for instantaneous regulation, e.g. because of sudden failures in the system, Statnett was also given the right to intervene directly and call on producers to supply the necessary production capacity, or regulate it down, as the case might be[24].

To sum up, the market set of the wholesale electricity market was to consist of several separate formal and informal markets that supported different market functions:

- an organized spot (day-ahead) market for short-term physical allocation;
- an organized real-time regulation power market for short-notice balancing capacity;
- an organized market for forward contracts, preferably organized as a futures market, to provide instruments for price hedging and risk management;

[24] Some years later further elaborated and more detailed arrangements were to be made for ancillary services. Chapters 2 and 4 discuss some specific issues of research in these areas. Chapter 2 describes and addresses early issues of the market microstructure for the regulation power market. Chapter 4 addresses further aspects of ancillary services as to valuation and principles of dimensioning, and issues related to the coupling of a central surveillance and control role with the decentralized roles in the supply of such capacity.

- and an informal market of bilaterally negotiated contracts serving purposes of tailored needs.

In addition, following the constrained capacity of the network, coordination of power delivery and network operations calls for an interface of congestion pricing and power markets.

Market design and microstructure is important for the resulting liquidity and efficiency of organized markets. Descriptions and research topics on the specifics of market design in these sub-markets are covered in chapters 2 - 5.

1.5.2 Retail Market

Above we have pointed to several measures that all have implications for promoting efficiency and competition in the retail market. The measures of common carriage and the abolishment of the institutional monopoly of distribution companies both promote competition, as consumers now were able to choose their suppliers. An open and more liquid wholesale power market provides market-based and transparent prices and larger flexibility in procuring electricity to sell in the retail market. The drive for efficiency was further expected to be strengthened by the transition to more economically rational participants in the process of adapting to a competitive environment. As such this is in high contrast to the former highly politically influenced distribution companies with political objectives, and politically set retail prices.

1.5.3 Foreign Trade

The former export market was in principle organized as a bilateral market, mainly vis-à-vis Sweden and Denmark. On the Norwegian side Statkraft was given an exclusive right to buy and sell power, however, restricted to terms of occasional power. The capacity of exchange was then in the range of 15 TWh, equivalent to about 15 % of total domestic production capacity. In the bilateral trading transactions Statkraft met the Swedish state-owned Vattenfall as the sole buyer or seller, in other words a virtual bilateral monopoly market. Power exchange with Sweden was priced at the mid-price between the Norwegian occasional power market price, and the relevant merit-order short-run marginal cost scheduled for the Swedish system. Vis-à-vis Denmark the Norwegian import price was set to 110% of the estimated increase in Danish fuel costs, while the export price was set to 75% of reduced Danish fuel costs.

Alternative organization of foreign trade was not an explicit issue of the pre-reform research project, which only discussed the foreign trade bilateral monopoly in a principle manner.

However, Bjørndalen et al. (1989, p. 97) argued that clear efficiency gains probably would be possible through a tighter integrated market based on a common Scandinavian solution. Here we will only briefly outline some main lines of development. The description of the initial post-reform organization of foreign trade is from Hope, Rud and Singh (1993a).

The market reform did bring some changes in foreign trade, opening trade for new players. State control with trade was retained by a provision in the Energy Law that stated that only the state may export or import electricity without a concession to do so. Statnett SF was initially granted the sole concession to organize and administer the foreign power trade[25]. It was emphasized that Statnett's responsibility should be strictly limited to the transportation and coordination functions, and that the company should act neutrally with respect to contracts and agreements for foreign trade. For political reasons, the Norwegian government had been very restrictive as to opening up for long-term contracts for foreign power exchange. In the early 1990's, however, the government did open up for a limited long-term contract power exchange. The foreign power exchange was until January 1996 organized in the following markets:

- *Short-term foreign power market:* This market covered trade for a period up to half a year. The market was administered by Statnett's division for foreign trade, Statnett Utland. Demand and supply from abroad took the form of price-quantity combinations which were fed into the domestic organized spot markets via Statnett's foreign trade division.
- *Long-term concession-free quota market:* The government opened up for a concession-free quota market for power export contracts of up to five years duration within a total quota of 5 TWh annually. There were detailed provisions as to quota limits, the allocation of quotas to market participants, and as to the level of trade with different countries. Contract negotiations on the export contracts took place directly between buyers and sellers. On the Norwegian side the producers organized themselves in larger groups to operate in this market. All trade in quotas took place over the clearing house for trade in quotas administered by Statnett, and all negotiations were to be reported to Statnett.

[25] An exception was made for the Norwegian-Danish power exchange agreements on the equal current cable between Norway and Denmark, where Statkraft SF retained its responsibilities and rights according to those agreements until they were to expire in 1997.

- *Long-term concession-regulated contract market:* To make room for contracts grounded on the benefits of international trade in power the government also opened up for negotiations on other long-term contracts. This was for example related to investment proposals for sub-sea cables which offered new possibilities in reducing Norwegian dependence on existing connections. Another important example would be contracts founded on differences in production properties and technological characteristics in different production systems, e.g. by utilizing the short-run active capacity properties of a hydro-based power production system to export in peak load periods and buy back base load during low-load periods to fill up water reservoirs again.
- *Imports:* There were in principle no limitations with respect to volumes and duration of pure import contracts. However, the transportation and grid consequences of such contracts had to be cleared with Statnett.

This opening of foreign trade based on market negotiated contracts represented a small, but clear step away from the political restrictiveness of foreign trade. Though founded on market based arrangements, this first organization represented a quite detailed and complicated organization, with a relatively high degree of political control reflected in quota terms and the concession system. From a market point of view, important questions also concerned whether it in fact was possible to administer such market arrangements in a completely neutral way.

1.6 Other Issues of Market Imperfections

The above-mentioned potential market imperfections – natural monopoly of transport, transaction costs, and market power issues – in many ways represented the most important factors for determining measures to implement the reform of the Norwegian electricity market. To some degree the issues of asymmetric information, public goods, and externalities also have implications for market organization, but were to some extent connected to issues of refining of the market-based electricity system. We provide a short discussion here pointing to some topics with implications for later developments in market design.

1.6.1 Asymmetric Information in the Electricity Market

An efficient market is characterized by transparent and symmetric information. The high degree of bilateral contracts in the former system contributed to low transparency in the market, though the publicly open Statkraft contracts conveyed the information contained in politically determined contracts terms. The formal occasional power market also signaled the value of short-term power. In this respect market reform measures for establishing and promoting an open and liquid futures market, as well as promoting competition in general, are measures that also increase transparency and information dissemination in the market.

In general, asymmetric information may constitute a major market imperfection. For the hydro power electricity market, the availability of the main resource, water, is important for future price development. Issues of asymmetric information are here especially related to whether any participants have superior market-relevant private information. In this respect concern was especially raised as to information on reservoir levels, a type of information which in principle is private. Requirements of reporting reservoir levels were later introduced, and aggregate reservoir levels became publicized on a regular basis. A further potential issue of asymmetric information is that of bottlenecks in the grid, an issue that was to be handled on a non-discriminatory basis with the establishment of a national grid company.

1.6.2 Public Good Issues

The quality and security of delivery in the grid may in many respects be characterized as a public good. In the pre-reform research on the Norwegian power market, the supply of ancillary services was, however, not recognized as a major issue or problem. It was assumed that arrangements would be made (or continued) to the coordinate the system, and to maintain the quality of delivery and momentary balance of the grid. As such, the regulation power market was soon established after the reform, while detailed arrangements for other ancillary services were to follow some years later[26]. Compared to the development in other countries, the Norwegian development in many ways represented a more step-wise approach towards more elaborated mechanisms and contracts for ancillary services. The development may be interpreted in many ways. Partly, it reflected a confidence that the long traditions and arrangements in handling these issues would be continued, also mirroring the prevailing corporate culture of a strong technological orientation and devotion for the working of the system. Partly, the underlying hydro power technology, combined with a current over-capacity of hydro-power generation capacity, also displayed relatively low ancillary service costs compared with other underlying technologies. The later increased focus on the remuneration of ancillary services in the mid-nineties, might likewise be interpreted to reflect the higher value of ancillary services in a more integrated market.

[26] See chapter 4.

1.6.3 Issues of Externalities in the Electricity Market

In principle several externalities may be identified in the electricity sector. A fundamental issue is that of environmental impact. For hydropower, this is especially related to landscape protection. This issue has traditionally been taken care of through a concessionary system for investments in generation plants and grid investments.

A further issue is the externalities in the grid. A given power flow in transmission may affect congestion in other parts of the net, a feature that has implications for short-term as well as long-term power scheduling. Implications of this source of externality represent research topics that were focused on later in the fine-tuning of the market system, and will be commented upon in chapter 5 which focuses on the pricing of network capacity.

1.7 New Market Infrastructure: Regulatory Regime, Market Institutions, and Organized Market Places

The above measures provide the main framework under which the new competitive market-based electricity system was to arise. To achieve an efficiently working market, further challenges lay in instituting and designing an appropriate new market infrastructure. Important challenges in establishing a new infrastructure were i) to devise and implement a new regulatory regime, ii) to establish and organize central market institutions, and iii) to implement and design the organized markets for electricity. In this section, we will briefly comment upon these issues. Topics related to the design of organized markets will be treated in more detail in chapters 2-5. The exposition of this section is mainly based on Hope, Rud and Singh (1993a) and Rud (1990c).

1.7.1 Regulation and Competition Policy

As a consequence of the transition to a market-based system, an accommodation of regulatory principles and rules was necessary to make the system function efficiently. The general trend of the policy change was to put relatively less emphasis on issues like e.g. system dimensioning and regulations regarding for example contract coverage of producers and distributors. Now more emphasis was put on, on one side market structure issues, competition and market behavior, and on the other hand, the regulation of actors handling natural monopoly functions.

On the institutional level, the traditional regulatory agency, the Norwegian Water Resources and Power Board, continued as the main regulatory board of the power market, though with new regulatory challenges. The new regulatory focus on the competitive climate of the market was also signified by the fact that the Price Directorate (now the Norwegian Competition Authority) was given an active role to play in the formation and enforcement of a competition policy for the electricity industry together with the Norwegian Water Resources and Power Board.

Without going into details of the new regulatory design, let us will simply illustrate it by referring to some of the early initiatives taken by the regulatory authorities under the new regime. As noted in Hope, Rud and Singh (1993a), these included for example; - work on new methodologies and measures for promoting efficiency and appropriate incentives in the distribution sector; - measures and conditions of area concession regulation to secure common carriage and transparency of tariffs; - measures to divisionalize vertically integrated companies; - accounting measures for distinguishing between natural monopoly activities and competitive activities in distribution companies; - requirements of distribution trading companies to publish their tariffs to secure transparency and market information; - intervention in cases where distribution companies tried to lock in their customers for example by conditions which were deemed anti-competitive from a competition policy point of view; - intervention in cases of alleged collusion among market participants to maintain prices or cooperate with respect to contract conditions; - and special attention was given to Statkraft SF as a potential dominant player in the power markets.

1.7.2 Market Institutions for Power Exchange and Grid Operation

In the proposed electricity trading and transport system there were several central functions which needed to be organized and handled on an aggregate market level. These functions related to the organization of markets and to the organization of transportation. The main central functions may be categorized as:

- *Spot power exchange*: A market place for spot power contracts has to be organized. The main objective of the exchange is to provide an efficient, trustworthy and neutral market place.
- *Futures market exchange:* Likewise a market place for future contracts must be established, with the same basic goals of providing an efficient, trustworthy and neutral market place. As the futures contracts may imply obligations on a longer term, it is also essential to establish market security systems that secure the fulfillment of contracts.

- *Grid management - short-term system coordination:* Aggregate production and consumption has to momentarily be in balance. This central function of short-term system coordination includes the continuous supervision and systems for balancing the network and acquiring necessary capacity for these adjustments. A further issue is that of registering and settling power accounts.
- *Grid management – grid system planning etc:* Grid management also includes longer-term tasks related to e.g. system planning and investments, maintenance programs, and establishing principles for transmission tariffs, etc.

Appropriate market institutions had to be set up for these functions. The organization of market institutions should promote efficient design and operation of each function, and in addition an efficient co-ordination between the functions. A further guideline is that the institutional set-up and specific organization is to be handled in an absolutely neutral and non-discriminatory manner. This also implies that the organizational units, whether power exchanges or system operator, do not participate in ordinary market trading. In this section we will review some comments upon some issues of institutional organization, based on Rud (1990c), Bjørndalen, Hope, Tandberg and Tennbakk (1989), and Hope, Rud, and Singh (1993a).

A general guideline in the overall organization of functions is to choose an organizational structure which promotes efficiency and non-discriminatory procedures for all market participants. It was obvious that a strong integration and coordination is essential in the operation of some of these functions. This implied that necessary measures for operative coordination must be in place, but not necessarily that the functions have to be organized within the same organizational entity. Different characteristics of the functions, for example due to different objectives, whether the function may be made competitive, or due to different regulation requirements, may in principle call for separate organizational entities, with operative arrangements for coordination.

Considering the futures and spot exchange, both exhibit the same organizational goals in promoting an efficient trading place. Further, it was assessed that there were probable cost advantages in administration and information dissemination, as well as lower transaction costs for participants in operating against a single trading exchange organization. As such a coordinated operation, as well as a common organization for exchange in spot and futures was recommended. The conclusion was also supported by the absence of alternative existing commodity or futures exchanges in Norway.

As for the short and long term aspects of grid management, both functions are closely related, but do exhibit important differences. To accomplish short-term system coordination and balancing requirements, contracts and market arrangements with market participants are necessary. The nature of long-term system planning of grid investments, however, reflects the decision making of a central optimizing planner. Still, both functions are characterized as natural monopolies, with common requirements of regulation and control. Close contact and information from daily grid performances are also important input to the planning problem, and the functions are closely related economically, as grid investments affect grid coordination. A common grid and system operator company was thus recommended for these functions.

A further question was the organizational relationship between the power exchange company (spot and futures) and the grid company (short- and long-term grid operation). The close operational relation between spot market clearing and short-term grid management is on one hand obvious. The spot market equilibrium may in general imply an infeasible solution in relation to the capacitated network, calling for adjustments in production or consumption. Adjustments are in principle either handled ex ante in accordance with the chosen congestion management method[27], or real-time using ancillary services. On the other hand, there are clear distinctions. In their nature, the market clearing of the spot market, which is in fact a day-ahead market, and any physically settled organized forward or futures markets, are both concluded prior to delivery. In a sense this also partly applies to the regulation power market, as bids for adjustment are submitted prior to delivery. However, the procurement of short-term ancillary capacity also represents a close borderline between exchange markets and grid management. Still, it may be asserted that there is a clear distinction between the functions of the exchange and the functions of the grid company, and that necessary coordination to a large extent may be organized by the exchange of information on for example possible bottlenecks in the grid, and on delivery obligations of spot, forwards (and possibly futures) contracts. These aspects indicate that a clear delimitation of exchange functions and grid functions is possible, given a well-functioning informational integration.

Our work recommended that the power exchange and the grid management should be organized in separate companies. The recommendation was first and foremost based on a

[27] See chapter 5.

concern as to the integrity of these central functions. Several arguments were voiced: Grid management is a natural monopoly, while the organization of an exchange in principle may be made competitive. This calls for tailored objectives and company management. Furthermore, separate organizations may constitute a reciprocal control. Concern was also that a mixture of these different legal and economic responsibilities might impair the trust in the financial reliability and integrity of the power exchange. Also, as an exchange never should take their own positions in the market, a combination of exchange functions and necessary purchases for balancing the grid within the same company was assessed as highly unsatisfactory.

Thus, for the institutional set-up of these functions, we deemed a legal separation as necessary to retain the financial integrity and independence of the exchange. Two separate market institutions were recommended: - an independent company to administer and supervise spot and futures trade, with a system for effective internal and external controls and safeguards. Here it was noted that the current power exchange company, the Norwegian Power Pool, had valuable experience in operating the existing power markets, making a fine starting point for developing the required infrastructure for the new power exchange house.

The chosen institutional structure was however to be implemented with all these central functions under the national grid company Statnett SF. Statnett SF was established January 1, 1992 as a result of the divesture of the old Statkraft into a production company (Statkraft SF) and a transmission company (Statnett SF). As of July 1, 1992, Statnett SF was assigned the responsibility of administering the foreign exchange of power. Effective from January 1 1993, the Norwegian Power Pool was merged with Statnett. The new company, Statnett Marked Ltd., which now was responsible for operation of the organized power markets, was established as a wholly owned subsidiary company of Statnett SF, and organized as a stock company. Statnett now had responsibilities and functions in three main areas:

> *i) Management of the transmission grid:* Statnett SF owned 80% of the grid, and was given the legal right to lease those parts of the grid that it does not own (mainly regional high-voltage lines integrated with the transmission grid). In this area the functions of Statnett thus included the technical operation of Statnett's own transmission system, the hire of physical transmission lines in systems belonging to and being operated by other companies, the co-ordination of transmission services, and the settlement of transmission services.

ii) Administration of the foreign exchange of power: The subsidiary Statnett Utland now administered foreign trade. See section 1.5.3 for a description of the initial organization of foreign trade after the reform.

iii) Administration and operation of organized markets: The subsidiary Statnett Marked, was responsible for the operation of the organized domestic power markets and foreign exchange. It was organized in four divisions: division for foreign trade, division for domestic trade and market clearing, division for accounting and settlement, and division for contracts.

Though Statnett SF itself, and its different divisions, were under the instruction of a neutral and non-discriminating treatment of participants, as well as of confidential use of all individual participant-based information, there was a strong skepticism amongst market participants that all these functions were assembled under the same organization and its subsidiaries. A special issue of concern was that Statnett Utland which administered the foreign short-term power exchange, was a subsidiary of the same company as Statnett Marked, and at the same time a large trader in the organized markets of Statnett Marked. In this the market participants met a large competitor which was under the same ownership as the power exchange. Concern was also voiced regarding the possible mixture of legal and economic responsibility for the different functions, and the implications for the neutrality and trust in the exchange.

1.7.3 Market Design of Organized Power Markets

A further important part of the market infrastructure was to implement and design the organized market for electricity. Chapters 2-5 takes a closer look into these design issues. Chapter 2 comments on early micro structure issues of the short-term markets, i.e. on one hand the spot market for electricity, 'the daily market', which was developed on an adaptation of the former occasional power market, and on the other hand the regulation power market for short-term adjustments based on proposals from the Norwegian Power Pool. Chapter 3 reviews issues in establishing an organized market for standardized contracts, a market which was to be a forerunner for a futures market for electricity. Chapter 4 follows up on the issue of short-term adjustments and the payment for ancillary services. Chapter 5 looks into the pricing of transportation services and the interaction with the power markets in the case of congestion.

1.8 Implications for Market Participants

With a new competitive environment, a new regulatory regime, and new forms for trading, the market reform represented a dramatic change and placed demanding challenges on trading companies, as well as on transport divisions/companies. The participants now faced crucial challenges in tailoring their objectives, management, and organizational forms to their new role and the new competitive and regulatory environment. The transition away from a former era of politically influenced decision-making, however, also had important implications for the role of politicians, as well as for public owners. Below we will briefly comment on some of the main challenges in adapting to the reform for different groups in the market.

1.8.1 Public Owner and Political Guidelines

An important recognition in the reform process was that the electricity industry reform not only was a question of designing and operationalizing a market-based system; - it was also to a large extent a question of changing the corporate and regulatory culture prevailing in the industry and the political system surrounding it. For central political system levels, such as governmental departments, their main role now was to contribute to measures that enabled the establishment of an efficient market structure, and but not further interrupt with the working of well-functioning markets. For public owners of power companies, the main implication was that their companies should be allowed to act upon normal principles of business, and not pursue any other political agendas.

1.8.2 Market Participants of Electricity Trade

In the new regime all market players that bought or sold electricity now faced a competitive market with major challenges in adapting strategies and management[28]. The market players now consisted of generating companies and end-users, and electricity brokers and traders acting as middlemen between producers and end-users. The latter group of traders had traditionally comprised distribution companies only. In the reorganized market which was

[28] This section is mainly based on Rud (1989a) and Rud (1993) which discuss implications of market-based electricity trade for the organization and management of electricity trading companies. The focus of Rud (1989a) is in particular on the applicability of the accounting information. A problem here was that the former accounting systems were based on the rather special principles of municipal accounting. Rud (1993) comments on new management challenges as to customer orientation, cost efficiency, pricing strategies, risk management, and company organization to meet the new competitive environment.

based on a principle of open entry, new participants soon emerged in the role as independent traders and brokers, thus further sharpening competition[29].

For conventional traders the change was particularly fundamental with the transition from an institutional monopoly to competition. For example, in the former regime, on having closed a poor purchase deal, the distribution company could in principle cover the costs by stating higher retail prices. This was not possible in the new competitive environment. Now competitors might step in and provide better offers, and capture the customers of the distribution trading company. Likewise, customers could on their own initiative search the market for suppliers with better prices. Traders also now had the opportunity of active participation in the organized markets. In general, a fundamental challenge was to deal explicitly with price risk, and balance the risk exposure of purchases against the risk profiles of customer contracts. Moreover, the previous monopoly companies were now forced to adapt to more customer and market-oriented strategies, with more pronounced objectives of cost efficiency, profit maximization, and risk management. Note that this pressure for change and more economic rational participants also was a pressure towards an overall more economically efficient market. A challenge was also for the companies to act more independently of the political system. In the infant stage of the market, it was evidently hard for the local political system to reduce its influence and give the companies the necessary economic and managerial independence.

For producers, trading patterns were also likely to change with the possibility of actively participating in the organized markets of the daily market, the regulation power market, and the weekly market. Though producers were familiar with the former occasional power market, the producers now to a greater extent had to analyze market trends and explicitly monitor their risk position. It may also be noted that basic economic principles implied pricing principles based on a short-run marginal costs. This was in contrast to the many year debate on and use of long-term marginal costs in pricing, a practice that no doubt had led to large inefficiencies in allocation.

[29] On receiving a licence granted by the Norwegian Water Resources and Power Board, traders and brokers could freely operate within the system. The initial license system was merely a matter of registration. Already Fall 1992, some 10 traders were established. Not all new-comers were successful, and several laid down their activities after a short period of time, while new companies were established. See Rud (1996) for a further study of the early development of new trading companies, and their role in promoting competitive trade in the market.

In general the new focus on profit and general business motives in the competitive market had important implications for the adaptation of normal business-oriented management and organizational principles. For example, several distribution companies chose to turn into regular stock holding companies, though without being privatized. Another example of adaptation is related to accounting and informational systems. The existing systems were poorly adapted to the new informational challenges in a competitive market[30]. On one hand this had implications for the practice of individual companies, as it for example, was not common for energy companies to make or use internal management accounts for decision making. On the other hand, it had implications for general accounting practices and regulatory requirements: Municipally owned distribution companies were at the time required to follow municipal principles of accounting[31], which in many ways were founded on a cash principle that to a large degree deviated from normal accounting principles[32]. These principles gave unsuitable information for managing companies in a competitive environment. At the time of the reform, there was already initiated a focus on changing these principals of accounting, and as of 1993 all distribution companies where required to base accounting on the same procedures and rules as for ordinary share holding companies. These are but some examples of adaptations necessary to meet the challenges of the competitive market.

1.8.3 Participants of Electricity Transportation

From a market point of view, the manifestation of a thorough understanding of their role as common independent carriers was crucial both as to Statnett SF and as to the grid-operating distribution companies. Their main objectives should now, simply put, reflect the desire to cost-efficiently operate the grid on a non-profit basis, and offer a non-discriminatory, transparent, and open access to the grid. Their objectives will, however, highly be affected by the imposed regulatory measures, thus the efficiency in this sector is highly dependent on regulatory design. Note also, that though grid operation is a natural monopoly, some functions, such as e.g. metering, in principle can be exposed to competition, so that to meet an objective of cost minimization, make or buy questions for such services may be relevant.

[30] See Rud (1989a) for on issues of economic management in production and distribution companies, and on problems of the currently used municipal accounting principles.

[31] Municipally owned companies followed the municipal accounting laws as given in the then current 'Kommunal- og arbeidsdepartementets Forskrifter og veiledning for budsjettoppstilling og regnskapsføring i kommunene'.

[32] For example, new loans were registered as revenue, installments as costs, and furthermore, investments were not periodized, so that the full cost was registered at time of investment with no later depreciations.

A further issue was that of optimal size of distribution companies. With some 230 distribution companies serving a population of about 4 million people, and a fairly skewed size distribution, the number of the companies was most likely to be of a suboptimal size[33]. In principle the dimensioning and boundaries of the distribution company should primarily be related to its generic function, i.e. the dimensioning and utilization of the grid.

1.9 Some Concluding Remarks on the Reform Implementation

The reform process led Norway to have one of the, at the time, most advanced market-based electricity systems in the world. As such it may in many ways be characterized as a social experiment on a grand scale with ramifications throughout the economy due to the infrastructural role of the sector. However, though the reform process in principle was quite dramatic, it went surprisingly quick, and without serious drawbacks. The explanations for this success are probably complex and varied, though several circumstances that contributed to this smooth a transition might be mentioned:

- The transition took place in a period with excess supply conditions and excess production capacity. There had been mild winters and above average rainfall through a number of consecutive years, new production capacity was being added as a consequence of earlier investment decisions, and there was low demand from power intensive industries due to low demand for their products. This made the industry conducive to reform, and allowed the market to consolidate before meeting the first trials of dealing with major scarcity.
- The market had the initial benefit of having a decentralized structure in the Norwegian electricity industry, with a large number of market participants, and where the largest actor, though big, still accounted for not more than 30% of the production capacity.
- The occasional power market, established in 1971, being an early and innovative arrangement for a sector like electricity, represented an important starting point and model for the institutional and organizational set-up for the extended market system. Through this market the participants of the electricity market to an extent already were familiar with the nature of market transactions.

[33] Rud and Tjøtta (1989) discuss principles of optimal size based on economies of scale and scope, and conclude that there are likely to be efficiency gains in the horizontal integration of some of the existing distribution companies. An empirical study on the economies of scale and optimal size of Norwegian distribution companies supports the conclusions, see Salvanes and Tjøtta (1990).

- And lastly, the Norwegian market reform process undoubtedly benefited from the general international trends of deregulation and market competition in the electricity industry. It should also be mentioned that the legal and political parts of the reform process were orchestrated with skills and determination by the Norwegian government. By and large, during the first critical years, the government stood firmly up against political pressures to change the speed and direction of the restructuring process. This was important, as without such a firm stand, the process might have collapsed at an early stage.

Now, after 17 years with a market-based system, the Norwegian electricity market reform has largely been termed a successful reform, and paved the way for the world's first cross-country integrated electricity market. We refer to chapter 6 for a further evaluation of the current efficiency of the market, future challenges, and a discussion on areas for further improvement.

1.10 Current Market Organization: Institutional and Regulatory Infrastructure

Since the implementation of the Norwegian electricity market reform, specific market design and regulatory policies have been continually been refined, and are still under the process of adaptation to meet the main future challenges of security of supply in an environmentally sustainable setting. Still, however, the electricity market has in nearly eighteen years worked along the same basic principles as outlined in the initial SAF-project nearly two decades ago. Moreover, our neighboring Nordic counties have followed with similar market reforms, and we now have a fully integrated Nordic market.

As our focus in the remainder of this paper is on the design of market systems, let us here review the main changes and developments in the institutional and regulatory infrastructure of the market[34]. In this respect two main driving forces, - on one hand the Nordic development and subsequent market integration, and - on the other hand changes in financial legislation, to a large extent have enforced major changes in the infrastructure of the market:

- *Nordic market reform:* Other Nordic countries were to follow in the steps of Norway in a market-based restructuring of their electricity markets, with Sweden as the first country after Norway. Following the reform, however, the Norwegian market had, as we saw in section 1.5.3, adopted a relatively restricted framework for international

[34] Sources: The information on the current (per summer 2008) system is mainly found on the websites of Statnett (www.statnett.no), and Nord Pool (www.nordpool.com).

trade, partly mirroring the non-market based organization of their trading partners. To implement a more integrated international trade, Norwegian legislation had to be adapted. The framework for an integrated Nordic market for physical contracts was made possible by Storting Report No. 11 1995/96. Together with the license for cross-border trade awarded by NVE to the power exchange, the foundation for international spot exchange trade was laid on the Norwegian side.

- *Changes in Norwegian legislation:* Several changes in e.g. the Norwegian Securities Act and the Norwegian Exchange Act gave rise to a reorganization of the main institutions of market exchange, and thus also to a more specific regulatory framework for the commodity exchange, the financial exchange and clearing activities.

Following this, major changes in the regulatory and institutional infrastructure were implemented. To put things short, on January 1 1996 Nord Pool (formerly Statnett Marked) was established as a joint Norwegian-Swedish exchange, forming the world's first multinational exchange for trading of electrical power. Finland joined the Nordic power exchange market area in 1998, Western Denmark (Jutland-Funen) in 1999, and to complete the full Nordic market integration, Eastern Denmark (Zealand) joined the Nordic power exchange area in 2000. The integrated market (Norway, Sweden, Denmark, Finland) thus accounts for a yearly production of 397.3 TWh (2007), covering a diverse portfolio of production technologies, with a share of 54% hydro power, 21.8% nuclear power, 21.7% other thermal power, and 2.4% wind power[35].

Following market integration and changes in legislation, the structure of market institutions has been adapted. We find that the internal and external organization of market institutions to a larger extent seems to represent an organization structure with more clearly defined and separate roles, also implying a higher integrity than in the infant market structure. The main market institutions (with a focus on the Norwegian market) now are:

- *Nordic Transmission System Operators:* Each Nordic country has its own transmission system operator, Statnett SF in Norway, Svenska Kraftnät in Sweden, Fingrid in Finland, and Energinet.dk in Denmark. The Norwegian government, represented by the Ministry of Petroleum and Energy, is the owner and ministry responsible for Statnett. The Norwegian Water Resources and Energy Directorate regulates all

[35] Source: Nordel annual statistics, www.nordel.org.

Norwegian grid company's revenues, including Statnett's, and stipulates guidelines for transmission tariffs, and system operator guidelines.

- *Nord Pool ASA*: Nord Pool ASA lists financial electricity contracts and organizes exchange trading in such products. The national grid companies Statnett SF and Svenska Kraftnät each own 50% of Nord Pool ASA. The company is organized as a Norwegian public limited company, authorized by the Norwegian Ministry of Finance as an derivatives exchange under the Exchange Act, and supervised by the Financial Supervisory Authority of Norway (FSAN, 'Kredittilsynet').
- *Nord Pool Spot ASA*: The spot market activities are organized in Nord Pool Spot ASA which runs the day-ahead spot market ELSPOT and the intra-day market ELBAS (the latter through its wholly owned subsidiary Nord Pool Finland Oy). Nord Pool Spot AS is owned with 20% each of Nord Pool ASA, Statnett SF, Svenska Kraftnät, Fingrid and Energinet.dk. Nord Pool Spot holds a license under the Energy Act of Norway, granted by the Ministry of Petroleum and Energy, to operate an organized market place for trades in physically delivered power contracts. Nord Pool Spot is under the supervision of the Norwegian Water Resources and Energy Directorate.
- *Nord Pool Clearing ASA*: As a derivatives clearing house Nord Pool Clearing clears all contracts traded at Nord Pool ASA, and in addition standardized contracts traded in the bilateral financial market which are registered for clearing. Nord Pool Clearing is licensed as a commodity clearing house under the Norwegian Securities Trading Act, authorized by the Ministry of Finance, and is under supervision of FSAN. Nord Pool Clearing ASA which is a public limited company, is a wholly owned subsidiary to Nord Pool ASA.

In addition to the main regulatory body for the power market, The Norwegian Water Resources and Energy Directorate, important regulatory bodies thus now include The Norwegian Competition Authority, and the Financial Supervisory Authority of Norway. All regulatory bodies are in turn subordinated to their respective ministries, i.e. the Norwegian Petroleum and Energy Ministry, and the Ministry of Finance.

2 Spot Markets for Electricity: The Day-Ahead Market and the Regulation Power Market

Two main markets were implemented for short-term physical trade in electricity market, the spot market and the regulation power market. The spot market, at the time called the daily market, is in fact a day-ahead forward market with physical delivery the following day. The regulation power market is a market for the supply of real-time adjustments to balance the power flow. Together these markets form the core of short-term allocation of power. The daily market was an adaptation of the former occasional power market, while the regulation power market was based on proposals from the former Norwegian Power Pool.

During the first years after the market reform, concern was that liquidity was too low in the daily market. Concern was also whether the volume of the regulation power market in comparison should become higher than desired. Our research was at this time concerned with the liquidity and efficiency of these markets. Attention was particularly drawn to the role and efficiency of the chosen micro structure and trading mechanisms, and its implications for the interaction of the day-ahead and the regulation power market. This is the topic of this chapter, which is based on the work of Hope, Rud and Singh (1993b), Knivsflå and Rud (1995a), Knivsflå and Rud (1995b) and Knivsflå and Rud (1997). Section 2.1 gives a closer description of the early design of these markets. Section 2.2 discusses possible sources of inefficiency at the time, while section 2.3 reviews the discussion on possible improvements of market micro structure. Section 2.4 focuses on incentives in trading in the daily market versus the regulation power market, while section 2.5 concludes the chapter with a brief review of the current trading mechanisms.

2.1 Market Descriptions: The Trading Mechanisms

As noted above, Statnett Marked, a subsidiary company of Statnett SF administered the daily market, the regulation power market, and the weekly market. The markets were open for all participants that satisfied certain conditions. As all the markets required physical delivery, the conditions included compliance with technical requirements in addition to the various payable charges that included a fixed entry charge, an annual charge, and a volume charge. Of the total number of participants[36], generators dominated. Trade by distribution companies, bulk consumers, and electricity traders and brokers had however been increasing. On the demand

[36] In 1994 there were about 95 participants.

side, the smaller units traded mostly through electricity traders, as e.g. most households which traded on fixed or variable terms with a distribution company. The exchange-based trading volume for all these markets accounted for approximately 30% of total trade in 1994, while the rest was mainly accounted for by bilateral contracts.

2.1.1 The Spot Market for Electricity

The daily market was the spot market for electricity, where the spot contracts specify delivery the following day. The market was basically an extension and a refinement of the former market for occasional power administered by the Norwegian Power Pool, and was based on the same organization of the power auction.

From the beginning, the basic contract for market clearing and settlement was the so-called 'price section', which was a bundle of several hours. The duration of price sections, i.e. the bundling of hours, was determined by Statnett Marked one week in advance. In 1995 normally 8 – 10 price sections were determined for each weekday Monday – Friday, and less, e.g. 3-4 sections per day in the weekend. From 1996 prices were set by the hour, i.e. the market was cleared for 24 hourly contracts per day, as it is today. Settlement of the contracts is by physical delivery. Any registered deviations in delivery from the hourly contracted volumes are settled over the regulation power market.

The traders submitted their sealed bids and offers to the exchange daily before 12.00[37]. These orders are demand and supply schedules (i.e. limit orders) that specified the price-quantity combinations at which buyers and sellers were willing to trade. Based on these incoming orders, the 'auctioneer', i.e. Statnett Marked, derived the aggregate supply and demand curves for each contract (i.e. for each price section). The market-clearing price and quantity are set so that supply and demand balanced[38]. This common market-clearing price is termed the 'System Price'. The resulting prices and quantities are revealed to the traders, but not the market depth (i.e., orders that are not cleared). The spot contracts are binding for delivery the following day, and specify the amount of energy to be delivered / taken out per hour.

The principles of matching and subsequent price discovery are best illustrated by a stylized numerical example[39]. Table 1 gives the orders for electricity delivered next day between 1200 and 1300 hours. Participants are consumers A, B, and C and producers X, Y, and Z. The spot

[37] The auctions were held daily from May 1992. Before that, the spot auction was held once a week.
[38] In the event of anticipated congestion further procedures were followed. Our focus in this chapter is the auction format aside from the congestion management procedure. The latter is a topic of chapter 5.
[39] The example is from Knivsflå and Rud (1995a).

price the day before is assumed to be NOK 200 per MWh. We here assume that orders are specified over the interval with jumps of NOK 10 per MWh. The trading system transforms incoming orders to continuous functions by connecting the specified combinations of price and quantity with straight lines.

Buy orders volume (MWh)			Buy orders cumulative	Spot price per MWh	Sell orders cumulative	Sell orders volume (MWh)			Potential Trade
A	B	C	(MWh)	(NOK)	(MWh)	**X**	**Y**	**Z**	(MWh)
900	500	0	1400	230	6100	2400	2100	1600	1400
900	800	700	2400	220	5000	2100	1900	1000	2400
900	1000	1000	2900	210	4400	1900	1700	800	2900
900	1100	1200	3200	200	4000	1900	1600	500	3200
900	1400	1400	3700	190	3400	1700	1300	400	3400
900	2200	1900	5000	180	3000	1500	1100	400	3000
1300	2600	2700	6600	170	1700	800	500	400	1700

Table 1 An example of a sealed spot electricity auction where three buyers are matched with three sellers

Buyer A, for instance, wants to buy 900 MWh delivered the next day between 1200 and 1300 hours to secure the running input of electricity in his industry production. If the spot price falls to NOK 170, however, A is willing to boost production and needs 400 MWh more. Seller Z, which could be a power plant, offers to generate 400 MWh at any price clearing the spot market. If the price is NOK 200 per MWh, Z will generate 100 MWh extra; if the price is NOK 210, Z will generate a further 300 MWh; if the price is NOK 220, Z will generate even a further 200 MWh; and finally, if the price is NOK 230, Z is willing to boost his generation by another 600 MWh, in total 1600 MWh.

The equilibrium price is found as the price that maximizes trading volume by matching as many buy orders with as many sell orders as possible. Figure 5 illustrates how the spot price is found in the pre-specified variation interval by requiring that the demand in terms of cumulative buy orders should equal the supply in terms of cumulative sell orders.

The cumulative buy orders represent the demand function for electricity, and the cumulative sell orders represent the supply function. Both demand and supply orders have been made continuous by connecting the price-quantity numbers in Table 1 with straight lines. The highest possible volume, which is the criterion for an equilibrium, is obtained in the interval [190, 200], and it follows that the equilibrium volume is $MWh^* = \frac{(b_h - b_l)s_l - (s_h - s_l)b_l}{(b_h - b_l) - (s_h - s_l)} = 3563.64$, and the equilibrium spot price is $P^* = p_1 + \frac{p_h - p_l}{s_h - s_l}(MWh^* - s_1) = 192.73$, where b_h and s_h are the cumulative buy orders and cumulative sell orders at the price $p_h = 200$, and b_l and s_l are the cumulative buy orders and cumulative sell orders at the price $p_l = 190$[40]. By setting the spot price equal to NOK 192.73 per MWh, it is possible for the auctioneers to match 3563.64 MWh in demand with 3563.64 MWh in supply. Accordingly, the spot price for electricity delivered the following day between 1200 and 1300 hours falls from NOK 200 at the auction yesterday to NOK 192.73 today.

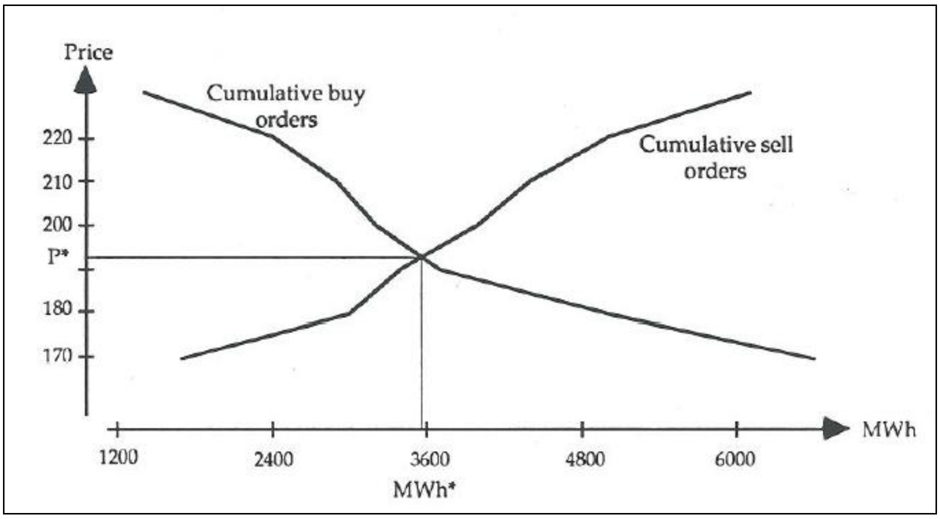


Figure 5 The equilibrium spot price for electricity given the order book specified by Table 1

Executed orders for each participant are the purchases of 900A, 1318.18B, 1345.45C, and the sales of 1754.55X, 1381.82Y, and 427.27Z. These are determined according to priority [e.g. for B we have $1318.18 = 1100 + (1400\text{-}1100)\frac{(200\text{-}192.73)}{(200\text{-}190)}$]. All buy orders above P*, and all sell orders below P*, are executed. As the auctioneers take the orders as straight lines, some buy orders placed at the price NOK 190 and some sell orders at the price NOK 200 are also executed. After the auction procedure is complete, the electricity exchange's information system displays the outcome to the public.

Together with obligations on forward contracts the spot market contracts specify the planned power flow for the following day. However, due to capacity constraints, the power flow may not be feasible. In the event of anticipated bottlenecks, a capacity charge is stipulated in the spot market, and the spot market equilibrium is adjusted to render a feasible planned power flow: In submitting their orders, the agents have specified the location at which the energy is to be delivered. If the equilibrium power flow is not feasible, the market is geographically segmented according to the anticipated bottleneck. The capacity charge is set by adjusting the market equilibrium of the daily market: Spot prices are raised in deficit areas, and lowered in

[40] We have inserted for $p_h = 200$, $p_l = 190$, $s_h = 4000$, $s_l = 3400$, $b_h = 3200$, and $b_l = 3700$.

surplus areas. The difference is the implicit *capacity charge*, which is set to balance the energy flow between the deficit and the surplus area. In this chapter, we will concentrate on general aspects of the microstructure of the market, and assume a market without bottlenecks, while some issues of congestion management are covered in chapter 5.

2.1.2 The Regulation Power Market

After the spot market is cleared, the delivery obligations of buyers and sellers are fixed for the following day. This follows from the spot contracts and forward contracts. All delivery obligations are reported to the grid control, i.e., to Statnett. The electricity demand, however, depends heavily on unforeseen factors such as temperature. If, for instance, the temperature falls unexpectedly, the consumers might increase the consumption of electricity for heating relative to the expected demand stipulated when the energy exchange closed the day before. Similarly, the generation of electricity may also depend on unforeseen factors such as technical defaults. Thus, in the course of delivery, actual input to the grid may deviate from contractual input, and actual output from the grid may deviate from contractual output. As contracts state the obligation to take out or generate a specified amount of *energy*, natural load variations also necessitate adjustments.

At each moment of time, however, the power flow has to balance. To allow such unexpected corrections relative to the contractual commitments, an overall momentary balancing of the energy flow is carried out by the grid coordinator Statnett. This real-time adjustment of the power flow involves increasing or reducing the load on a short notice. To do so, the grid coordinator has to contract for the necessary capacity for adjustment, and arrangements must be made in which producers and consumers on short-term instructions from the central grid control are obliged to increase or decrease their supply and demand.

The mechanism through which Statnett procures capacity to adjust the power flow is called the regulation power market. The mechanism allocates and prices capacity for short-term adjustment. *Active* participants that supply short-term adjustment capacity were at the time primarily producers who could adjust their load at 15 minutes' notice and satisfy the necessary specified technical conditions. We however noted that it might be efficient that real-time flexible consumers supplied active capacity, as in principle load adjustment to meet an unexpected increase (reduction) in consumption may be met by reducing (increasing) consumption by other consumers, as well as by increasing (reducing) generation.

Adjustment orders for the following day are submitted daily prior to 1930 hours the day before delivery, that is, after the spot market has closed, but prior to delivery the next day. The market is operated on an hourly basis so that separate orders are quoted for each hour of the following day. Statnett compiles a priority list based on submitted orders. Clearing virtually takes place continuously during the delivery day, in which the system operator consecutively chooses the required adjustment capacity based on a merit order selection. The regulation price, i.e. the compensation for rescaling, is determined as the price of the marginal order used when making the net adjustment. Subsequently, all deviations between actual load and contractual load are recorded and settled by the regulation power price. As such the remaining electricity market participants may be characterized as *passive* participants of the regulation power market.

We will illustrate the mechanism by a simplified example. Table 2 gives an example of orders submitted to the separate auction detecting the regulation power market price. In order to illustrate the settlement procedure in terms of payment and power exchange, we have assumed that the ex ante total trade in the market has been settled by the spot market alone.

Bids for adjustment (MWh)			Cumulative bids (MWh)	Price of adjustment per MWh (NOK)
X	Y	C		
400	200	400	1000	230
200	200	400	800	220
0	200	200	400	210
0	0	0	0	200
0	0	0	0	190
-200	-200	0	-400	180
-400	-400	-600	-1400	170

Table 2 The auction of short-term changes in production /consumption after the electricity exchange has closed.

If the price exceeds NOK 210 per MWh, producer Y is signaling to the regulation power market that he is willing to expand production on a 15 minutes' notice by 200 MWh. In the same way, consumer C is willing to reduce his consumption of electricity compared to the contractual consumption by 200 MWh. Consumer C could be a distribution company that has sold contracts on terms of 'interruptible' power. If the price is NOK 170, producer X and Y are willing to reduce their production of electricity by 400 MWh each, whereas consumer C is willing to expand consumption by 600 MWh.

The price in the regulation power market $P^{\bullet}$ depends on the adjustment $\mathrm{MWh}^{\bullet}$ which is the difference between the actual load $\mathrm{MWh}_{\mathrm{A}}$ and the contractual load MWh^{*}. It seems reasonable to expect that if $MWh_A > MWh^*$, then $P^\bullet > P^*$, reflecting higher costs of producing on short notice. Conversely, if $MWh_A < MWh^*$, then $P^\bullet < P^*$, as agents who adjust the load only, are expected to be compensated for lost profits. In the following, we will illustrate the principles of adjustment when MWh_A differs from MWh^*. The case where $MWh_A > MWh^*$ is considered separately from the case where $MWh_A < MWh^*$, but the same principles for market clearing, of course, apply.

Realized Load is Higher than Contractual Load

Due to mainly an unexpected fall in temperature, let us assume that the total load adjustment needed is $MWh^\bullet = MWh_A - MWh^* = 4058.43 - 3563.63 = 494.80 MWh$. To balance the deviation, the central grid control contacts the agents with the best bids for short-term adjustments in load, and instructs them to increase production or reduce consumption accordingly (i.e. according to Table 2). For a more instructive presentation, let us now in our example assume that the bids are made continuous[41]. Then the price is $P^\bullet = p_1 + \frac{p_h - p_l}{s_h - s_l}(MWh^* - s_1) = 212.37$[42]. This means that if $MWh_A > MWh^*$, then $P^* < P^\bullet$. The price is higher in the regulation power market than in the spot market, probably because the marginal cost of producing electricity on a short-term notice is higher than the marginal cost of producing electricity according to spot commitments. Table 3 summarizes the adjustments and payments of all participants.

Consumers / producers	Spot contract (MWh)	Deviation (MWh)	Adjustment (MWh)	Power flow (MWh)	Contract payments (NOK)	Deviation payment (NOK)	Adjustment (NOK)	Total payments (NOK)
A	-900.00	-385.71	0.00	-1285.71	-173 457.00	-81 913.23	0.00	-255 370.23
B	-1318.18	32.47	0.00	-1285.71	-254 052.83	6 895.65	0.00	-247 157.18
C	-1345.45	0.00	247.40	-1098.05	-259 308.58	0.00	52 540.34	-206 768.24
X	1754.54	0.00	47.40	1801.84	338 152.49	0.00	10 066.34	348 218.83
Y	1381.82	0.00	200.00	1581.82	266 318.17	0.00	42 474.00	308 792.17
Z	427.27	-141.56	0.00	285.71	82 347.75	-30 063.10	0.00	52 284.65
	0.00	-494.80	494.80	0.00	0.00	-105 080.68	105 080.68	0.00

Table 3 Realized load from 1200 to 1300 hours is higher than contractual load (no transaction costs)

Consumer A consumes 385.71 MWh more than he should according to his spot trade. Consumer B consumes 32.47 MWh less. Producer Z, having suffered a breakdown, delivers 141.56 MWh less than his obligation. Consumer A could be a distribution company whose customers have increased consumption because of the unexpected fall in temperature. The

[41] The orders to the Norwegian regulation power market are, however, not made continuous functions.
[42] We have inserted for $p_h = 220$, $p_l = 210$, $s_h = 800$, and $s_l = 400$.

active participants of Producers X and Y are required to generate 47.4 and 200 MWh extra respectively, while Consumer C is instructed to reduce consumption by 247.4 MWh. All deviations are settled by $P^{\bullet} = 212.37$.

The *average* price paid differs among consumers: A pays in average pays NOK 198.62, B pays NOK 192.23, and C pays NOK 188.30. Consumers B and C in average pay a price that is lower than the spot price P^*, since they actually have contributed to balancing the market by reducing the total load of the grid. On the other hand, Consumer A pays more than P^*, since he, by consuming more than the contracted volume, is 'punished' by having to pay a higher price for the extra production by generators and the reduced consumption of other consumers. In the same way, the average price actually received for the electricity produced differs: Producer X receives NOK 193.25, Y receives NOK 195.21, and Z receives NOK 183.00. Producer Z receives an average price lower than the spot price P^*, as Z, by delivering a lower production than contracted, increases the net deviation of the market.

Realized Load is Lower than Contractual Load

Suppose that a reduction in the load of 478.25 MWh is required to balance the system. It can be managed by reducing production, increasing consumption, or a combination of both. According to Table 2, the price in the regulation market is: $P^{\bullet} = p_1 + \frac{p_h - p_l}{s_h - s_l}(MWh^* - s_1) = 179.22$[43]. If $MWh_A < MWh^*$, then the price in the regulation market is lower than the price in the spot market, i.e. $\mathrm{P}^* > \mathrm{P}^{\bullet}$, as participants unable to fulfil their obligations have to compensate the other side of the market for lost business opportunities. Based on the bids of Table 2, Table 4 summarizes the adjustments and payments of all participants.

Consumer A consumes 289.21 MWh less than contracted, B consumes 43.6 MWh more than contracted, whereas Producer Z produces 232.64 MWh more than contracted. Producers X and Y are required to reduce production by 215.65 MWh each, while Consumer C is instructed to increase consumption by 46.95 MWh.

[43] We have inserted for $p_h = 180$, $p_l = 170$, $s_h = -400$, and $s_l = -1400$.

Consumers/ producers	Spot contract (MWh)	Deviation (MWh)	Adjustment (MWh)	Power flow (MWh)	Contract payments (NOK)	Deviation payment (NOK)	Adjustment (NOK)	Total payments (NOK)
A	-900.00	289.21	0.00	-610.79	-173 457.00	51 831.49	0.00	-121 625.51
B	-1318.18	-43.60	0.00	-1361.78	-254 052.83	-7 813.88	0.00	-261 866.71
C	-1345.45	0.00	-46.95	-1392.40	-259 308.58	0.00	-8 414.26	-267 722.84
X	1745.54	0.00	-215.65	1538.89	338 152.49	0.00	-38 648.25	299 504.24
Y	1381.82	0.00	-215.65	1166.17	266 318.17	0.00	-38 648.25	227 669.91
Z	427.27	232.64	0.00	659.91	82 347.75	41 693.16	0.00	124 040.90
Sum	0.00	478.25	-478.25	0.00	0.00	85 710.77	-85 710.77	0.00

Table 4 Realized load from 1200 to 1300 hours is lower than contractual load (no transaction costs)

The average price paid differs significantly among consumers: Consumer A pays NOK 199.13, B pays NOK 192.30, and C pays NOK 192.27 per MWh. Consumer B pays an average price that is lower than the spot price P^*, as he, by consuming more than contracted, helps the central planning unit to avoid reducing production. B is compensated according to the regulation price. Consumer A pays a price that is higher than P^* because he, by consuming less than expected, has to pay for the necessary reduction in the production of electricity. Similarly, Producer Z, having increased production and thus having contributed to increasing the net deviation of the market, receives an average price of NOK 187.97 per MWh that is lower than the spot price P^*. Producers X and Y cover the deviations and receive respectively an average price of NOK 194.62 per MWh and NOK 195.23 per MWh, both higher than the spot price P^*.

2.1.3 A Characterization of the Trading Systems

In a broad sense the trading mechanism designates the means by which prices are detected and resources allocated in an organized market. Its crucial function is to transform the latent supply and demand of buyers and sellers into realized transactions. The key to this transformation is price discovery, i.e. the process of finding market clearing prices. No trading mechanisms are in principle alike, they differ in the types of orders permitted, the times at which trading can occur, the quantity and quality of information made available at the time of order submission, and the reliance on market makers to provide liquidity. The underlying objective in electricity market design is to establish trading systems and mechanisms that contribute to maximize the efficiency of the overall electricity market. The objective of our work at the time was to provide a preliminary discussion of whether alternative design of the trading mechanisms could contribute to increased efficiency, with a focus on possible principle effects of alternative mechanism. To structure the discussion, we will here briefly characterize the trading mechanisms within some main categories.

One set of basic distinctions are whether the auction is order-driven or quote-driven, and whether the auction is periodic or continuous. These auction categories may briefly be defined as follows[44]:

- *Order-driven mechanisms:* In an order-driven system, it is the incoming orders that drive the auction and thus the pricing. The system operates either continuously or periodically. In the first type, known as a *continuous auction*, traders submit orders for immediate execution through bilateral, or in a liquid market, multilateral, matching. The system is continuous since orders are executed upon arrival. The second type is known as a *periodic auction*. Submitted orders are here stored for execution and cleared at the same time at a single market clearing price.
- *Quote-driven trading mechanisms:* In a quote-driven system, participants can obtain firm price quotations from market-makers before submission. Thus, it is the quotes that drive incoming orders. On order submission, the buyer/seller need not wait for order execution, but instead trades immediately with a market-maker. Being a continuous trading system, the mechanism is characterized by a sequence of bilateral transactions, at possibly different prices.

Auctions of both the daily market and the regulation power market are driven by the incoming submitted supply and demand bids, and may thus be characterized as order-driven mechanisms. The daily market is clearly periodic, since orders are stored and the market is cleared at a single point of time, at a single market clearing price. Bids of the regulation power market are also stored for execution, and cleared during the delivery hour in question at a single market price. As such it too is a periodic order driven mechanism.

A further basic distinction is the extent to which the trading mechanism reveals information about orders submitted to the exchange:

- *Sealed auction:* The trading mechanism is classified as sealed if traders are prevented from observing incoming orders. In a sealed auction only the resulting price and volume are displayed.
- *Open auction:* The trading mechanism is classified as open if traders can observe the orders when they are placed in the order book by the auctioneers.

[44] Exchanges may in principle also employ a combination of these mechanisms, for example with a periodic, order-driven auction at the start of the trading day, followed by a continuous order- or quote-driven mechanism for the rest of the trading day.

Periodic auctions tend to be sealed, whereas continuous auctions tend to be more open. For both the daily market and the regulation power market no information on the incoming orders is publicized prior to market clearing. Both may thus be characterized as sealed auctions.

Finally, it should be pointed out that the market set-up of a separate day-ahead market and a separate market for real-time adjustment capacity represents an unbundled electricity market system where energy and reserve capacity are unbundled into separately priced commodities.

2.2 Potential Sources of Inefficiency

Together the daily market and the regulation power market form the markets for short-term delivery of power. The markets, however, fill different functions. The daily market is the core spot market for planned short-term allocation of power. The regulation power market is a residual market with the role of pricing and procuring real-time adjustment capacity, and of handling, pricing and settling deviations between actual deliveries and contracted deliveries[45]. An efficient and liquid spot market for electricity is vital to the efficiency of the electricity market. In principle the optimal equilibrium spot price equals short-term marginal cost, which in turn equals the short-term marginal benefit of consumption. The spot price thus reflects the marginal value and relative scarcity of power, and is important in the short-term allocation of power as a reference price for all use and production of electricity. Note also that a well-functioning spot market is a prerequisite for a well-functioning futures market for power.

At the time, the trading volume on the regulation power market seemed to be relatively high. Concern was whether the high trading volume in the regulation power market came at the expense of the liquidity of the spot market, and whether the overlapping of the two markets would undermine the liquidity and efficiency of the spot market in particular, and the efficiency and functioning of the electricity market in general. A basic question in the quest for improving efficiency was to explain and identify the causes of such a market development in the short-term markets. It is highly probable that the development followed from a multifaceted range of market issues, for example the explicit design of the organized markets, specific features of regulation policy or the general market structure, as well as frictions in the transition from the former system. It is also probable that the different issues were highly interrelated, so that such a development was a product of all these features. For example, in the transition away from the former bilateral contract dominated market system, the market

[45] Physical delivery may in the Norwegian electricity system be contracted by the daily market, by bilateral contracts, and at the time also by the contracts of the weekly market which were settled by physical delivery.

design and realized liquidity of the organized forward or futures market may have had a major impact on the liquidity of the spot market; A poorly functioning futures or forward market may cause a greater use of bilateral contracts, which in turn may affect the liquidity in the spot market. This dependence is however reciprocal, as an efficient futures market also depends upon the existence of a liquid and efficient spot market. The liquidity of both markets may further rely on the chosen market design, as well as issues of competition in the general market structure.

While there are many factors which in sum might have accounted for the poor liquidity of the spot market, the focus of our work in Knivsflå and Rud (1995a), Knivsflå and Rud (1995b), and partly Hope, Rud and Singh (1993b), was on the chosen microstructure of the short-term markets. The question asked was whether alternative market design and trading mechanisms could contribute to a more efficient market in general, and in particular a more efficient interaction of the two short-term delivery markets. Before turning to the discussion of alternative trading mechanisms, let us however review some of the perceived problems. Hope, Rud and Singh (1993b) evaluated the current functioning of the markets, and suggested two sources of inefficiency which were thought to be attributed to features of the chosen trading mechanism:

- *The interface between the daily market and the regulation power market:* Figure 6 illustrates the timing between the daily market and the regulation power market. Orders to the daily market had to be placed before 1200 hours the day before delivery. Shortly after, the daily market was cleared. Delivery on the spot contracts occurs 12-36 hours later. After spot market clearing, any necessary adjustments in consumption and production are in principle assigned to the regulation power market.

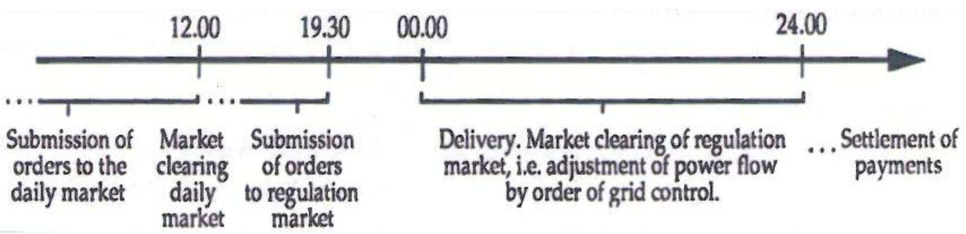


Figure 6 The sequence of events in the daily market and the regulation power market

One concern was here whether the time lag itself between the clearing of the daily market and delivery could contribute to the enhanced volume of the regulation power market. In the 12-36 hours from spot market clearing to delivery, the market receives new information, e.g. on temperature and rainfall. The equilibrium contract obligations of the daily market, however, stand firm. The participants have no possibility to undertake any desired contract adjustments to adjust for this new information in production plans or consumption plans. Any deviations, unplanned, as well as any planned discrepancies due to new information, will thus all be handled and registered as deviations in the regulation power market. Compared to a situation with better pre-delivery adjustment possibilities, this set-up was assessed as a factor that possibly contributed to amplifying the level of deviations between the actual load and contractual load, and to enhancing the trading volume of the regulation power market.

Note also that in allocating their trade among the short-term markets, the participants compare prices, costs of trading, as well as possibilities of strategic bidding in the daily market versus the regulation power market. A concern was here whether there were other features of the current trading design that gave the participants incentives to canalize more trade to the regulation power market, i.e. by willfully causing deviations and this way offering capacity in the regulation power market rather in the spot market. The basic question was whether there were issues, such as features of the trading mechanisms, the timing of markets, or differences in market admittance, which promoted such incentives.

- *Market information from the current trading mechanism:* The spot price conveys information on the market value of the commodity. In an efficient power market the spot price reflects the value of power in alternative use, as well as the cost of supplying additional power on the margin. For hydro-power plants the main short-term marginal cost is not a cash-based cost, but rather the ‘water value’, which is the alternative cost of water. For the individual power plant, this value partly follows from conditions of the plant itself, and partly from the expected future market conditions. The spot price conveys information on the market’s aggregate assessment of the current value, and thus implicitly the expected future value of water. As such, the spot price is an important input to the production decisions of the individual producers. In this the spot price also plays an important role in coordinating the production decisions of different producers. We note that trading mechanism of the daily market reveals the resulting equilibrium price and quantity, but offers no information during the process of market clearing. At the time of trading, i.e. when submitting bids, the available market data is mainly given by the previous spot prices. Different designs of the trading mechanism may, however, convey

> different levels of information during the process of trading. The question now asked was whether changes in trading mechanisms, or policies of information disclosure might improve the informational content of prices and also the efficiency of resource allocation.

On this basis, our question was whether alternative trading mechanisms could contribute to improved efficiency in the short-term power markets.

2.3 Alternative Trading Mechanisms

The trading mechanisms of the short-term markets were an adaptation of the mechanism used in the former market for occasional power. In the initial start-up phase the use of a well-known mechanism might have had beneficial effects on the liquidity of the spot market. However, in the early 1990's the liquidity was rather low in the daily market, while the trading volume of the regulation power market was possibly too high. The question now raised was whether alternative trading mechanisms could improve the efficiency of the spot market, and offer better price discovery, lower explicit and implicit transaction costs[46], and confine the markets to their specialized functions of respectively short-term planned loads versus real-time adjustments.

Several adjustments were considered. Our focus was mainly on the pricing mechanism, the number and frequency of auctions, and the dissemination of information through the trading system. We considered that several of the alternative designs for trading mechanisms potentially could have contributed to enhance market efficiency, provided a starting point of a more mature and liquid market. For example, by increasing the auctions of the daily market from one to two, the participants could earn more flexibility in constructing their portfolio of planned short-term contracts, and tentatively minimize the deviation volume of the regulation power market. However, we perceived that in a market with an initially low liquidity, the danger was that such a market segmentation on the contrary would lower liquidity in both markets. Based on our analysis, which was qualitative and preliminary in nature, we therefore did not recommend any of the considered adjustments for the daily market or regulation

[46] Note that the efficiency of a market place is closely related to the liquidity and transaction costs of the market. Trading costs on one hand include trading charges, which in turn reflect the efficiency and liquidity of the exchange. On the other hand trading costs include the costs of the participants in e.g. bidding, seeking information, etc. The hypothetical frictionless market with no transaction costs is in theory the most efficient market, and also the market of maximum liquidity. To see this, note that the existence of transaction costs will shift supply and demand curves, and result in lower trading volumes, and hence lower liquidity. Market restructuring that implies lower transaction costs may thus contribute to enhance liquidity as well as efficiency.

power market to be carried out at the time due to the current poor liquidity. The basic encountered problem was that though several arguments indicated that the considered alternative mechanisms could be efficient in a more liquid market, in all likelihood the mechanisms would behave poorly given a starting point of a relatively thin market. The fundamental question was how to achieve the transition to the more liquid spot market. This, however, was left as an open question. Below we refer some of the main arguments of our preliminary study of the mechanisms.

Periodic versus Continuous Trading Mechanisms

Both trading mechanisms were periodic and order-driven. An alternative considered was whether a continuous mechanism could contribute to a more efficient market in terms of better price discovery and lower transactions costs. In a continuous mechanism market clearing takes place during a defined time period, which in principle can be extended close up to the time of delivery. Within this period buyers and sellers submit their orders, in the form of a desired buy or sell quantity at a given price. In an order-driven continuous auction, contracts are cleared as soon as a corresponding order is submitted. In a quote-driven continuous mechanism, market makers further provide extra liquidity by quoting binding buying and selling offers on which the buyer or seller immediately can close the contract. In both these forms of continuous mechanisms the contracts would thus be consecutively entered into. The contracts can thus be characterized as sequence of bilateral transactions at possibly different prices.

By utilizing a continuous mechanism market participants receive market information on trading volumes and the prices of the consecutively concluded contracts[47]. With an extended trading period starting early and reaching close up to the time of delivery, the mechanism also offers the opportunity to adjust and tailor short-term contract positions to any new information received close up to delivery. In this way the continuous auction might contribute to less deviations, and thus a lower regulation power market volume. A further effect is related to the differences in ‘waiting costs’: The continuous auction provides immediate trading, whereas the traders in a periodic auction must wait until pre-specified times for market clearing and execution of orders. That is, in a periodic auction, the trader faces the risk of changes in the value of a commodity between the time he submits his bid, and the time the

[47] The issue of sealed versus open auctions is discussed below. If the continuous auction in addition is implemented as an open auction, the information disseminated to the market would also include information from the order book.

trade is completed. If there is a rapid inflow of new information, waiting may be costly. This waiting cost is minimized in a continuous price mechanism.

For the daily market, we found that several arguments seemed to favor a continued use of a period, order-driven trading mechanism. Firstly, since orders are stored for a single market clearing, the thickness of the market is larger in periodic than in continuous auctions. As such the order volume underlying the equilibrium market clearing price will be larger, and to a larger extent represent the ‘true’ underlying value of the commodity. In the daily market where a larger trading volume was to be desired, a perceived problem of the continuous mechanism was the potentially thinner liquidity associated with a virtually more segmented market. Secondly, there are several indications that a continuous mechanism might incur larger transaction costs in the context of the daily market. It is probable that the exchange’s costs of administering auctions and communicating information may rise if the trading mechanism is made continuous. With a starting point of a relatively low liquidity, this would imply higher average exchange-based trading costs[48]. It is also probable that the participant-generated transaction costs of bidding might be larger given a continuous mechanism. This cost may be substantial, as trading in this physical market normally is on a daily basis, often in all the daily contracts. Thirdly, there is the question of how the methods of congestion management should be implemented in such a market setting.

Given the current liquidity and maturity of the market, the periodic, order-driven auction was still recommended used in the daily market at the time. Due to the nature of the regulation power market, where bids by nature have to be stored for merit order selection in real-time, the continued use of the periodic order-driven auction was also recommended.

Timing and Frequency of Auctions

Given the continued reliance on a periodic order-based trading mechanism for both the short-term markets, the question still remained to what degree the missing opportunity to adjust spot obligations from the time of market clearing to delivery, contributed to larger deviations than necessary. The daily market was in the current setting cleared once a day, while the regulation power market was cleared during delivery on the basis of the bids submitted the previous day. The question here is whether changes in the timing of auctions, as well as in the number of auctions per day could contribute to greater efficiency.

[48] In the case of a quote-driven mechanism it was also likely that the large inherent risks for market makers in the low-volume daily market would imply high bid-ask spreads.

A simple alternative would be to move the market clearing of the daily market nearer the time of delivery. Consider a scenario where the flow of new information is high and volatile in this 12-36 hour period from market clearing to delivery. If market clearing were closer to delivery, prices and the resulting allocation of power would then be based on more updated information, and possibly contribute to minimize deviations and thus the trading volume of the regulation power market. If the information flow in the period is low, however, the effect of this change would be rather minimal. This closer time span between spot market clearing and delivery might in principle also contribute to reducing possibilities for strategic bidding based on differences in market admittance in the two markets. Smaller traders might, however, in this setting be prevented from trading, for example if the new market clearing was to occur after normal office hours, and thus give rise to extra costs in following up contract obligations. To conclude on this alternative, however, a closer analysis of benefits, as well as of the consequences for the individual groups of traders in the markets was called for.

To offer larger flexibility, we also considered the alternative of introducing a second market clearing in the daily market, with the second auction nearer the time of delivery. This would enable participants both to enter into contracts in normal trading hours as before, and to be able to act upon any new information closer to the time of delivery, and by this possibly reduce deviations and the volume of the regulation power market. This advantage must, however, be balanced against increased exchange costs due to a second auction, as well as increased costs of providing an information system filling the requirements of the near-delivery market clearing. Other possible effects must also be considered: With a starting point of rather low trading volume, there was the danger that both auctions would exhibit even thinner liquidity. The splitting of the daily market into two auctions could also potentially open for strategic bidding, especially if some groups of traders for some reason confined their trading to only one of the two auctions. The method of congestion management is also an open question in this setting. Based on the preliminary investigation and discussion of the subject, it was concluded that it was doubtful that the potential benefits of trading nearer the delivery period at the time with an initial poor liquidity would outweigh the increased costs and disadvantages a more segmented market.

Sealed or Open Auction

Information on the market’s valuation of power is in general an important input to market participants in deciding on their own consumption and production strategies. In a hydro power system the market valuation of power is a particularly essential input for the producer in determining the ‘water value’ of his resources. Better information on market values may in principle improve inter-temporal production strategies. The issue here was whether adjustments that eased the trader’s access to the information generated in the trading process, could contribute to a more efficient market.

As we have seen, better information from the trading process may for example follow implicitly from the choice of trading mechanism. Confer here the above discussion on the continuous mechanism, or on increasing the number of auctions, where both these alternatives would have offered more information from the trading process. Our focus here is, however, on alternatives related to the order book. The market mechanism of the daily market was sealed and provided no information on the underlying orders. In principle an open auction format could be employed. In an open auction the depth of the order book is revealed. By observing the valuation of others, traders obtain information relevant for their own valuation, in principle enabling better forecasts of the resulting price and volume, and making it possible for the participants to adjust their bids to new information. Further, as open auctions tend to reduce the informed traders’ ability to hide among the uninformed, the exploitation of less informed traders by the better informed traders may be avoided.

The disadvantages of opening the auction might in principle also be severe, as, for example, if it makes the market more vulnerable to collusion. With a market structure characterized by partly dominating market actors and a potential threat of collusions, this argument seemed relevant for the electricity market. Administrative exchange costs may also increase, as open auctions may warrant an improved information system for conveying the information to the users. A further argument in favor of sealed auctions is to avoid the possibility of higher volatility following excessive revaluation based on the endogenously generated information derived from the order flow in the open auction.

A further issue to be considered was whether the potential benefit of this information actually will be significant in the electricity market. Compared to other markets, where bids often are single price-quantity bids, the implemented bid structure of the short-term electricity markets is special. Here the participants submit their bids through demand and supply curves, i.e. a vector of quantity bids that are contingent on the equilibrium price. This ensures that the cleared contracts are in accordance with their preferences, whatever the price may be, and thus contributes to avoiding costly errors in predicting the prices. Moreover, as the markets are cleared daily, and a large part of the inherent uncertainty of the market is related to the longer-term aspects of the seasonal and annual supply of water, the daily published market data might in fact to a large extent provide the necessary input to forming consumption and production strategies.

The recommendations from our preliminary discussions thus leaned towards the continued use of a sealed auction, supplemented by the publishing of aggregated market data after market clearing. This conclusion applied to both the daily market and the regulation power market. However, as several of the above noted advantages as well as disadvantages of an open order book in the daily market seemed relevant, a further investigation was deemed necessary for a final conclusion.

Integrated or Unbundled Short-Term Markets

A completely different organization of the short-term markets is given by the integrated electricity market systems. The Norwegian short-term markets for power represent an unbundled market system where energy and real-time adjustment capacity are unbundled into separately priced commodities in a sequence of separate markets. This is in contrast to integrated systems, which may be characterized as a 'smart' market based on the overall optimization of gains from trade, subject to system constraints[49]. The typical design includes a day-ahead optimization of generation, transmission and reserves resulting in indicative plans to be re-optimized hour-ahead and in real-time operations. Pricing and settlements are, however, based on the real-time, system-wide nodal differentiated opportunity costs (shadow prices). As an integrated market virtually is cleared real-time, there are no deviations, and thus no need for a special market for handling deviations.

For the Norwegian electricity market, this system was, however, not considered a relevant alternative. The unbundled market setting had proved a viable alternative for the Norwegian system which was founded a more dispersed market structure, and on completely different underlying production technologies, with a large ability and capacity for momentary adjustment. A basic feature of the Norwegian unbundled system is an emphasis on the efficiency gains from the incentives inherent in a competitive market, where demand as well as supply compete and adapt to competitively formed market prices. The Norwegian system also allowed physical bilateral trading alongside the organized markets, giving the participants flexibility as to the means of trading, and the power exchanges competiton and incentives for providing an efficient market place.

[49] See Wilson (2002) for a discussion on integrated versus unbundled electricity systems.

With an adherence to the principle of an unbundled market, the principle organization of the short-term market in the separate markets, with a day-ahead spot market and a deviation market, stood firm. In the above sections we have considered alternative trading mechanisms within this basic unbundled framework in the current short-term markets. Equally important are the implied incentives of the trading mechanisms and pricing principles. This will be the topic of the following section, as well as in chapter 4 which considers the alternative of a two-tier price structure for adjustment capacity.

However, before proceeding to these topics, let us conclude this section by briefly considering an earlier sketch of the possibility of a more integrated, but still unbundled structure of the spot and the deviation market. The basic idea was here to integrate spot market clearing and adjustment capacity bidding. An example of such a structure would be a setting where the bidding procedure of the regulation power market was replaced by, and implemented, as a closing procedure within the day-ahead auction. The idea was that the marginal bids of technically real-time flexible traders in the day-ahead market could provide the basis for regulation power market offers, which could follow either directly, or by a functional transformation[50]. As for the real-time selection procedure by the grid control, priority based on merit order could be employed as before. Benefits of the more integrated and less segmented short-term market would potentially follow from lower costs of trading, less room for strategic bidding due to market power or asymmetric information, and from higher liquidity in general. The efficiency of such a market system is on one hand highly contingent on the extent of such benefits. On the other hand, the efficiency of this structure rests on whether the marginal spot market day-ahead bids (by real-time flexible participants) are relevant estimates (or in general related by a given functional transformation) of the real-time adjustment costs. We, however, left this as an open question which was not further analyzed.

2.4 The Incentives of the Regulation Power Market and the Daily Market

Our focus in this chapter has been on the efficiency in the short-term markets for electricity. While the previous section focused on efficiency issues of the trading mechanism in the market clearing and price discovery process, an equally important issue for the efficiency of short-term markets was whether the chosen mechanisms provided the necessary incentives to

[50] Requirements would also be that electricity traders who signal participation are able to follow technical requirements.

induce efficiency in the short-run, as well as in the long-run. Further discussed issues related to the incentives of the short-term markets were the following:

- *Incentives for trade in the daily market versus trade in the regulation power market*: In principle the two short term markets fill different functions, with the daily market (i.e. the day-ahead spot market) as the core short-term market, and the regulation power market as a residual market for handling and pricing deviations. An important question in relation to the interaction of the two markets, is whether the chosen market design in fact induces incentives to avoid deviations, and to use the daily market as the main organized market for the short-term planned allocation of electricity.
- *Incentives for offering demand-based adjustment capacity in the regulation power market:* Efficient real-time balancing is contingent on that the market induces the use of the most efficient adjustment capacity. In principle, both real-time adjustable production capacity, as well as real-time adjustable consumption, may be employed to meet forthcoming imbalances. At the time only generation plants were represented as active adjustment capacity in the regulation power market. To ensure overall efficiency, this calls for arrangements where all technically capable adjustment capacity, including consumption, is represented in the market.
- *Incentives for offering sufficient adjustment capacity in the regulation power market:* On offering capacity to the regulation power market, the capacity holders submit bids stating the required compensation for activating capacity. In submitting the bid, the capacity holder is required to hold capacity ready for load adjustment during the period in question. The capacity holder receives the regulation power price if called upon to adjust his load. If not activated, the capacity holder receives no remuneration. As such the regulation power market employs a one-tier price. In relation to short-term efficiency, an important question is whether this compensation and price structure succeeds in attracting sufficient amounts of adjustment capacity to the market (and in particular the capacity which is most efficient in use). And, though the implied incentives may suffice under normal circumstances, a further question is whether the structure induces the adequate amount in times of scarcity.
- *Incentives for investments in adjustment capacity:* Closely related to the above issue is the question of whether the one-tier price system employed in the regulation power

market provides sufficient incentives for investments in this form of secondary regulation capacity, as well as in other ancillary services.

In the remaining part of this section our focus will be the first of these issues, that is, on market incentives with respect to the resulting interaction of the daily market and the regulation power market. An issue here is whether the market structure gives the participants incentives to truthfully bid their planned delivery in the daily market, or whether there are incentives for an implicit trade in the regulation power market. The two last-mentioned issues above covering alternative price structures and ancillary services are covered in chapter 4.

2.4.1 The Incentives of Trade in the Daily Market versus the Regulation Power Market

A liquid and well-functioning spot market for day-ahead planned transactions is crucial to the overall efficiency in the unbundled electricity market. The prime function of the regulation power market is limited to supply capacity on short notice to handle deviations between the actual and contractual power load. The incentives implied by the market structure should thus support this division of functions. Properly organized, traders should not expect to gain by consciously holding back volume from the daily market in order to trade in the regulation power market. Deviations should be caused solely by errors e.g. due to production defaults and unforeseen changes in the need for electricity.

To discuss the incentives and interaction of the two markets, we need to distinguish between two sets of participants in the regulation power market. On one side, we have the capacity holders who place their capacity at the disposal of the coordinator, and who on doing so may be ordered to activate their capacity, on which they earn the adjustment price. On the other side, we have the remaining market participants who unannounced induce deviations for which they have to pay the costs of adjustment. A given producer or consumer of electricity may thus virtually sell or buy his electricity either in the spot market or in the adjustment market. The latter is done simply by deviating from the contractual load by consuming or producing a different amount of electricity than agreed upon in the spot and forward markets. In principle, a producer or consumer would sell (buy) electricity in either the spot market or the adjustment market, depending upon which market exhibits the highest (lowest) market price. If the adjustment price is expected to be higher (lower) than the spot price, a risk neutral market participant would therefore prefer to sell (buy) electricity in the adjustment market, rather than in the spot market.

The description and numerical example of the clearing mechanism of the regulation power market in section 2.1.2 indicates that the regulation price will be *higher* than the spot price if the realized load is *higher* than the contractual load[51]. Correspondingly, if the realized load is *lower* than the contractual load, the regulation price will be *lower* than the spot price. The consequence is that the consumer or producer that contributes to the general trend of deviation, would have been better off is he had undertaken the corresponding transaction in the spot market. For example, a consumer, who enhances the overall net deviation by contributing to a higher realized load than contracted, will have to pay the regulation price for the extra quantity, where the regulation price is higher than the spot price. In accordance with this incentive structure of the two markets, we thus would expect that electricity traders do not plan to trade in the regulation power market.

There may, however, be situations where consumers and producers would have the incentive to trade actively in the regulation power market:

- *Different cost structure:* If the effective trading commission in the spot market is higher than the commission in the regulation power market, then it could be optimal to consciously trade some electricity in the regulation power market.
- *Private information:* Traders who can predict the total deviation better than the market might exploit their knowledge by arbitrage between the day-ahead and the deviation market. For example, let us assume that a producer knows that the actual net consumption will be higher than the contracted net consumption, and thus that the system operator will have to increase production real-time. In this case the producer will wish to offer less production in the day-ahead market, and produce more than the contractual load in the regulation power market, there by contributing to a lower deviation, and being paid the regulation power price which is higher than the spot price.
- *Different competitive pressure:* Differences in the competitive pressure of the two markets might lead to situations where, for instance, producers prefer to trade in the regulation power market.

[51] These conclusions are contingent on the strategy of bidding used in the example. This employed strategy implies that the active capacity holders require a price higher than the announced equilibrium spot price to sell in the regulation power market, and a price lower than the announced spot price to buy in the regulation power market. We find this assumption to be reasonable. See also the empirical test referred to below.

In conclusion, the day-ahead and deviation market auctions should be constructed so that traders have an incentive to trade in the day-ahead market rather than in the deviation market. In the absence of special asymmetries, such as differences in trading costs, differential information, or different competitive pressure, the above trading mechanism appears to adhere to this principle.

An empirical study on the subject is carried out in Knivsflå and Rud (1997) where the relationship between the then Norwegian-Swedish spot market for electricity and the Norwegian regulation power market is studied. Based on a large set of observations from Nord Pool, the integrated Norwegian-Swedish electricity market[52], strong support was found for the following relationship: A positive load deviation, i.e. where the actual load is higher than the total contracted load, as expected led to an adjustment price that was higher than the spot price. Symmetrically, a negative load deviation, i.e. where the actual load is lower than the total contracted load, led to an adjustment price that was lower than the spot price. The implication is that it is not cheaper for ordinary traders to buy and sell in the adjustment market rather than in the spot market. The traders who enhance the net deviation thus had to pay a premium relative to the spot price, by buying (selling) the adjustment load at a higher (lower) price than the spot price. This premium is the compensation to the capacity holders who have to adjust the power load on short notice, thus giving compensation for the energy supplied, the extra costs of adjusting rapidly, as well as for holding spare capacity to accomplish the adjustments.

It was also hypothesized that the spot electricity price, if properly anticipated, was an unbiased estimate of the adjustment price of electricity at the time of the spot auction. By simply comparing the means of the two price series, this hypothesis could not be rejected. This means that the adjustment price premium was considered as noise at the time of the spot auction. In analyzing the volatility of spot and adjustment prices, it was also found that the volatility of the spot price was less than the volatility of the adjustment price at the time of delivery.

As for the adjustment load, we expected that the total actual load would equal the contractual load plus a noise term with zero expectation and some variance. This means that the adjustment load is considered as noise at the time of the spot auction, i.e. as unexpected

[52] The data set consisted of hourly observations of the spot price, the contractual spot load, the adjustment price, the adjustment load, and the total load of the period of January 1 1996 to February 28 1997, in total 10200 hours. In all the series except the total load, there are missing observations, so that the actual sample varies from 8406 to 10200 observations.

changes in the expected load from the time of the spot auction to the time of realization. On the basis of the empirical tests, this was confirmed as it was found that the mean of the adjustment load was not significantly different from zero.

2.5 The Short-Term Physical Markets of Today

Today the physical day-ahead spot market, Elspot, may be said to have developed into a cornerstone of Nordic power trading. Elspot covers Norway, Sweden, Finland, Denmark, and in addition also the KONTEK area[53]. Almost 70 % of the total power consumption in the Nordic countries was in 2007 traded on Elspot, and its system price is the underlying reference price for Nord Pool's financial power market, as well as for the bilateral wholesale market in the Nordic region. The regulation power market, operated by Statnett, is still a market for handling real-time deviations. While the framework of the spot market and the regulation power market to a large extent still is in accordance with the original framework, the markets have continually been refined. In addition the Elbas market now provides the possibility of intra-day trading 24 hours a day, in the other Nordic countries and Germany, and is planned to open for trade in Norway in 2009. Below we will briefly account for some of the changes in market design in the short-term physical markets[54].

Elspot

As a licensed exchange, the Elspot market has continuously improved e.g. rules of trading, membership structure, informational services, market conduct rules, market surveillance, implemented electronic trading, etc. With our focus on the central trading mechanism, we find that the market still is cleared by means of the same sealed, period auction, where the equilibrium prices follow from the aggregate bids and offers from all market participants. Hourly contracts are traded daily for physical delivery in the next day's 24-hour period. As such the trading horizon is as before, i.e. 12-36 hours ahead. In an integrated Nordic market, now comprising a large variety of production technologies, allowable bidding formats have, however, been extended. In addition to the normal hourly bids, the bidding formats of block bids and flexible hourly bids have been introduced:

– *Hourly bid*: The hourly bid is the basic type of Elspot market order, similar to the original bidding format. Each participant selects the range of price steps for the bid individually.

[53] The KONTEK area is the Vattenfall Europe transmission control area in Germany.
[54] Sources: The information on the current (per summer 2008) system is mainly found on the websites of Statnett (www.statnett.no), and Nord Pool (www.nordpool.com).

The bid may consist of up to 62 price steps in addition to the ceiling and floor price limits set by Nord Pool Spot. As before, Nord Pool Spot makes a linear interpolation of volumes between each adjacent pair of submitted price steps.

- *Block bid*: The block bid format was introduced in 1999, and is meant to be useful in cases where the cost of starting and stopping power production is high, or in the case of inflexible production and consumption. The block bid gives the participant the opportunity to set an 'all or nothing' condition for all the hours within the block. The block bid is an aggregated bid for several hours, with a fixed price and volume throughout these hours. The block bid must be accepted in its entirety, if accepted at all. In this, the block bid price is compared to the average Elspot price for the hours to which the block bid applies. The block bid is accepted if the bid price of a sales (purchase) block is lower (higher) than the average Elspot area price. It is also possible to define links between block bids, meaning that the evaluation and acceptance of one block bid is dependent upon the acceptance of another block bid.
- *Flexible hourly bid*: The flexible hourly bid is a sales bid for a single hour with a fixed price and volume. The hour is not specified, but instead the bid will be accepted in the hour with the highest price, given that the price is higher than the limit set in the bid. This type of bid is for example meant to give companies with power intensive consumption the ability to sell back power to the spot market, by closing down industrial process for the hour in question. As such, the flexible hourly bid is meant to help the market in strained power load situations leading to very high prices.

It should be noted, that in principle, the handling of these different bid formats in clearing a congested market is not straightforward, and raise a long range of efficiency issues as to different specific methods of clearing the market based on these bids.

Regulation Power Market

The regulation power market is a tool to keep the balance between total generation and consumption of power in real time. Up till today, the main principles of the trading mechanism for clearing the regulation power market are mainly similar to those described above. Active participants submit preliminary bids by 2000 for the following day (24 hour period), and may adjust or submit new bids up to 45 minutes before the delivery hour in question. In real time the capacity is accepted on the basis of merit order, though technical or geographical considerations may, as before, override the selection.

In our early work referred to above, we noted several issues of concern as to confining the regulation market as a market for handling imbalances. We here note several measures of the current market organization seem to some extent to relieve this concern. Note for example the following:

Incentives for Trade in Elspot versus Deviations over the Regulation Power Market:
As to the direct incentives to avoid deviations, several measures seem to have strengthened these incentives:

- First of all, to trade at Elspot the participant has to sign the Participant Agreement, as well as the Balance Agreement with Statnett (Norwegian participants). Both compel the participant to implement production, consumption and trading plans that balance, and to use the bidding at Nord Pool to achieve balance between energy inflow and energy demand in each Elspot area.
- Secondly, the bidding rules require that the regulation power bids represent a premium in relation to the Elspot price: The lowest allowable bidding price for sales of regulated power (increased production or reduced consumepetion) is set to 5 NOK above the area price of Elspot, while the highest allowable bidding price for the purchase of regulated power (reduced production or increased consumption) is set to 5 NOK below the relevant area price of Elspot.
- Thirdly, even stronger incentives for balance may follow by the planned implementation (January 1 2009) of new rules for a closer harmonization of Nordic pricing and settlement of imbalances[55]. The balance account of all balance responsible entities will now in short be separated into two separate balance accounts: the *production balance* comprising all production the entity is responsible for, and which represents the producer's passive imbalance, and the *consumption balance* which comprises consumption and trade. Consumption imbalances will, as previously in Norway, be settled by the regulation power price (RP-price) only, i.e. in a one-price system. Production imbalances will be settled by a two-price system, with the RP-price for imbalances that enhance the system's imbalance, and the Elspot price for imbalances that have alleviated the system imbalance. The settlement rules are summarized in table 5. In short, with the two-price system for production imbalances,

[55] See the implementation document on www.Statnett.no: Statnett (2008): 'Nordisk harmonisert balanseavregning. Endelig Implementeringsbeskrivelse', Statnett 13. juni 2008. The web site also contains several papers on the subject.

a passive imbalance is priced to the less profitable price of the RP-price and the Elspot price, meaning to give the producers an even stronger incentive to plan their dispositions in balance.

	Imbalance	Up-regulation hours	Down-regulation hours
Production balance[56] (= actual production – planned production + actively activated regulation power)	Negative	RP-price (up)	Elspot price
	Positive	Elspot price	RP-price (down)
Consumption balance[57] (= planned production + actual consumption + actual trade + actively activated regulation power)	Negative	RP-price (up)	RP-price (down)
	Positive	RP-price (up)	RP-price (down)

Table 5 Nordic harmonized balance settlement

Incentives for Offering Adjustment Capacity in the Regulation Power Market in the Short and Long Run:

The above measures also to some extent contribute to cementing a premium for the active regulation power market participant. However, as will be discussed in chapter 4, this may not suffice to attract sufficient capacity. In the concluding section of chapter 4, we will see that further measures have been instituted that give more incentives for the short as well as the long run supply of such adjustment capacity.

Incentives and Opportunities for the Use of Real-Time Demand-Based Adjustment Capacity:
Flexible demand offering short-term adjustments is now able to participate as an active participant in the regulation power market.

[56] Note that in up-regulation hours we have RP-price > Elspot price. In down-regulation hours we have RP-price < Elspot price. If the producer's imbalance is negative, this means that his net production has been too low, causing him passively to buy regulation power to meet his obligations. If this is the case, and if the regulation power market as a whole is up-regulated, he has to pay the RP-price, which is higher than the Elspot price. If the regulation power market as a whole is down-regulated, he has to pay the Elspot price which now is higher than the RP-price. (Likewise, if the producer's imbalance is positive, this means that his net production has been too high, causing him passively to sell regulation power to meet his obligations. If this is the case, and if the regulation power market as a whole is up-regulated, he receives the Elspot price, which is lower than the RP-price. If the regulation power market as a whole is down-regulated, he receives the RP-price which is lower than the Elspot price)

[57] As a convention, consumption and sales are denoted as negative numbers, while production and purchases are denoted as positive numbers. Purchases and sales are moreover defined from the point of view of the balance entity. A negative imbalance implies that the balance entity buys regulation power, and a positive imbalance implies that the balance entity sells regulation power.

Elbas

While the Elspot market closes 12-36 hours prior to delivery, the participants, however, continually receive new information in this period which affects optimal, as well as possible plans. As we noted above in section 2.2, this time span may in itself cause further imbalances. The above measures, including the new two-price balance settlement for production, give participants even stronger incentives to plan their dispositions in balance, stating a need of a planned adjustment market.

Such an intra-day market, Elbas, was launched as a separate market for power balance adjustment in Finland and Sweden 1999, and operated by the wholly owned Nord Pool Spot subsidiary, Nord Pool Finland OY. It opened for trade in Denmark in 2004 and 2007, and is planned to be opened for trade by Norwegian participants in 2009.

Elbas is a physical (pre-delivery) balance adjustment market. The Elbas market is an intra-day market where you can buy or sell your imbalances up until one hour before the delivery hour. The Elbas market thus enables trade with contracts that lead to physical delivery for the hours that have been traded on the Elspot market, and that are more than one hour from delivery. The Elbas Market is open around the clock every day of the year. For each and every hour of the day, one power hour contract is quoted. Elbas operates with a continuous trading mechanism, with market makers to enhance the liquidity of the market.

3 A Futures Market for Electricity

Inflexible and non-market based trading systems were identified as a major source of inefficiency in the former market regime. The recommendation of the SAF reform project in 1988/89 was to establish formal organized power markets, with a spot market for short-term allocation, and a futures market to provide instruments for risk management. The introduction of a futures market for electricity was at the time a novel proposal in the electricity industry. Parallel to the market reform development in Norway, there were taken initiatives for establishing a futures market in electricity in England. The first attempt by London Fox in 1990 was for various reasons not successful and was rejected. In October 1991, an OTC market, the market for 'Electricity Forward Agreements' was established, with GNI Ltd. selected as the initial broker for the EFA market[58]. The first Norwegian organized forward market, the weekly market, started to operate in November 1992. With no clearing house, mark-to-market or margin payment arrangements, the market was a market for standardized bilateral forward contracts, and in this respect more a forerunner for the full-fledged futures market to come.

The focus of this chapter is on the early development of the futures market in Norway. In the first stage towards a Norwegian futures market for electricity we were firstly concerned with issue of the suitability of electricity as a commodity for futures trading. This is the topic of section 3.1. Secondly, focus was on issues of how a futures market for electricity should be organized as with respect to institutional set-up, microstructure design, and contract design. This is the topic of section 3.2 which also includes a description of the first organized forward market, the weekly market. These sections are based on several papers written prior to the establishment of the Norwegian electricity market, i.e. Rud (1989b), Rud (1990a), Rud (1990b), Rud (1990c), Hope, Rud and Singh (1992), as well as on several unpublished notes on the subject in the years of 1990-1994. Section 3.3 follows with an appraisal of the initial 1992 market organization, and is based on Rud (1992), Hope, Rud and Singh (1993b), Knivsflå and Rud (1995a), and Knivsflå and Rud (1995b). Section 3.4 concludes by providing a brief description of the current organized futures and forward market for electricity.

[58] Though this also was an organized forward market for electricity, it must be noted that there were several differences between the Norwegian and the English market that had implications for a fundamentally different design of hedging markets in Norway. This was related both to the overall organization of the electricity markets, the market structure of participants, and in differences in fundamentals affecting prices and price volatility.

3.1 The Rationale for a Futures Market

While a core issue of the SAF reform project was to establish a spot market and a futures market as the main organized markets for trading energy, the electricity futures market represented a rather exotic idea. There were at the time no other Norwegian commodity futures markets[59], nor were there any futures markets for electricity in any other country. The Norwegian electricity sector thus had limited experience on the functions and working of a futures market. As such one role of our initial work on the subject vis-à-vis the industry was also to motivate and describe how a futures market for electricity presumably could be a tool for risk management. A more fundamental issue was related to the potential of achieving a viable futures market for electricity in the Norway, given the characteristics of the commodity, the inherent risk and uncertainty of the spot market, and the new market structure. Section 3.1.1 briefly comments on the fundamentals of risk in the Norwegian electricity market, the need of efficient risk management, and the role of a futures market for electricity. Section 3.1.2 discusses the suitability of electricity as a commodity for futures trading

3.1.1 Risk in the Electricity Market, and the Role of an Electricity Futures Market

The transition to market-based trade in electricity with an organized spot market contributed to a more pronounced visualization of the inherent price uncertainty of the market. Electricity spot prices vary stochastically within the day, within the week, within seasons, and from year to year, however, with recognizable daily, weekly, and seasonal patterns. The uncertainty of spot prices follows from the stochastics of the underlying fundamental demand and supply factors. Important fundamentals of demand are, for example, temperature, the price of other fuels, business cycles, etc. As to the supply of electricity, the most important factor in a pure hydro-power system is the availability of energy, i.e. the supply of water[60]. At any point of

---

[59] The listed futures and options market in Norway started in 1990, in which the first traded derivatives where stock options, and options on the OBX stock index. Trading in financial futures was, however, at the time prohibited. In May 1992, the Ministry of Finance granted permission to commence futures trading at the exchange. Futures on the OBX were then introduced in September 1992, and Government bond futures were introduced in June 1993. (Source: Battley, N. ed. (1993) *The World's Futures & Options Markets*, Probus Publishing Company, Cambridge, England.)

[60] The importance of the water supply was early accentuated as in Hveding (1968): 'In a predominantly hydro power system it is still necessary for the planning to provide sufficient total plant capacity to meet the peak load, just as in the thermal system. But in addition, the hydro system presents also an energy problem, which tends to be the overriding one, thus making planning considerably more complex than that of the thermal system. Whereas fuel for thermal plants can be bought as required, water supply to a hydro plant cannot be controlled except by storage, carrying water over from the surplus period to deficit periods. Once drawn down, the storage cannot be replenished except by the run-off that may or may not occur. The larger the storage volume is, the

time the available amount of water is given by the reservoir levels and the inflow of water, the latter which follows from precipitation and snow-melting. The inflow of water, which depends on temperature and weather, is stochastic, and varies greatly from year to year, as well as within each year. An important challenge of the hydro-power producer is to plan the inter-temporal allocation of water. This involves estimating when the value of using the available water is the highest. Here the market expectations of future inflow and future demand conditions are important factors that affect spot prices.

For the participants in the electricity market uncertainty is manifested in price risk and quantity risk:

- *Price risk* refers to the volatility and uncertainty of spot prices facing the producers, traders, and end consumers in the future. For hydro-power producers price risk refers to uncertain selling prices for their production. For end-users price risk refers to uncertain buying prices for electricity. The price risk exposure of traders, who are middlemen, follows from the uncertainty of the price margins of their net trading position.
- *Quantity risk* arises due to the hydro-power producer’s stochastic inflow of water, and the stochastic demand for electricity.

In a highly volatile market risk-averse participants would consider to hedge against uncertainty in accordance with their risk preferences. A futures market offers the possibility of hedging *price* risk. Price hedgers, such as buyers and sellers of the commodity, will thus be the main group of traders in the futures market. Normally, however, there will be an imbalance between hedging motivated purchases and sales. Here, trading by risk-taking participants without positions in the underlying commodity play an important role in balancing the market and contributing to an efficient allocation and pricing of risk. The main motives of participants in the futures market may thus be categorized either as

> i) *hedging motives,* where producers or consumers of electricity enter futures positions to reduce the price risk associated with the future price movements of their underlying positions, or as
>
> *ii) speculative motives,* where the main motive is to profit from future movements in price, and where the futures position does not match any underlying commodity position. The speculator is thus a trader who enters the futures market in pursuit of profit, but who as such has to accept risk in the endeavor.

---

better are the chances that the problem will be mastered at all times, but there is always a certain risk that deficiencies may occur’.

In addition to providing an efficient instrument for hedging price risk, it should also be acknowledged that a liquid and well-functioning futures market may contribute to the overall efficiency of a market by fulfilling several roles:

- *Risk allocation:* The futures market provides hedging facilities. Relative to a market purely based on bilateral contracts, a well-functioning futures market provides flexibility and enables a dynamic risk management. Even more important is that the futures market provides a market-based pricing of risk, and efficiently allocates risks in accordance with the risk preferences of the market participants.
- *Informational role:* In contrast to a bilaterally negotiated price, which only reflects information held by the negotiating parties, the market-based futures price reflects an aggregate of the information held in the market[61].
- *Co-ordinational role:* Closely linked to the informational role is the implied co-ordinational role of the futures market in the hydro-power electricity market. In a hydro-power dominated market, where the variable production costs to a large extent follow from the alternative cost of using water, the market participants' expectations of future spot prices play an important role in their inter-temporal allocation of resources. In this, futures prices contribute to align expectations, and thus to aid the co-ordination of production decisions in the market.

With a highly volatile electricity spot market and with risk averse participants, the need of an efficient instrument for risk-allocation was evident. The establishment of a futures market represented an obvious candidate for organizing this function. A basic question was, however, whether electricity would prove a suitable commodity for futures trading.

3.1.2 Is Electricity a Suitable Commodity for Futures Trading?

Although efforts were underway in other countries, the development of futures markets for electricity was at the time still a new phenomenon, and there was limited experience to draw on. A basic question was whether electricity as a commodity would be suitable for futures trading. That is, whether it was possible, and probable, that a liquid well-functioning futures market for electricity could evolve. Experience with commodities successfully traded on other futures markets indicated that the following characteristics were common to these

[61] The degree of the informational contribution of the futures market also depends upon the chosen trading mechanisms. For example, if based on an open auction format, the futures market may in principle to a larger extent contribute to increased transparency and dissemination of market information, than if based on a closed auction format.

commodities: i) product homogeneity, ii) storability, iii) deliverability, iv) the existence of a viable spot market, v) price variability, and vi) the existence of speculators to take up the balance of open positions. Though these characteristics are by no means neither necessary nor sufficient conditions for the success of a commodity for futures trading[62], they provided a certain guideline for assessing the suitability of electricity for futures trading. Electricity is a commodity which is physically homogeneous, and where reasonable requirements of standardization may be met. It is subject to delivery and indirectly storable through the storage of water or fuels. Furthermore, price uncertainty as manifested in stochastic spot prices, was considered a major concern for market participants. Thus given that a well-functioning spot market was in the process of developing, the establishment of a futures market for electricity seemed a viable proposition. However, while these characteristics at best can give an indication of the suitability of a commodity for futures trading, the ultimate success depends on a number of other factors. Here futures contract design and the institutional arrangements for trading the contract are of crucial importance.

3.2 Building Blocks of a Futures Market for Electricity and a Description of the Weekly Market

The main effort in launching a new futures contract is normally that of designing the futures contract. With no existing commodity futures markets in Norway, and thus no appropriate existing futures exchange system to build upon, the establishment of a futures market for electricity also called for the design of a suitable institutional set-up, exchange security system, and market microstructure. In this section we review our discussion on basic elements of setting up a futures market for electricity, following the discussion of each with a description of solutions chosen in the first organized forward market, the weekly market. Section 3.2.1 comments upon issues of contract design, while section 3.2.2 comments upon design issues related to the exchange system for futures trading.

3.2.1 The Futures Contract

A well-designed futures contract is vital to the success of the futures market. The futures contract is a standardized contract which gives the holder the right as well as the obligation to buy or sell a specified grade and quantity of a commodity at an agreed price (the futures price) at a given future point of time (contract maturity). The futures contract is standardized in all

[62] For example, the characteristics do not explain the development of futures markets for non-storable commodities such as live-beef, fresh eggs, etc., nor non-deliverable futures such as interest rate options or trading in share price indices.

details except the price. The majority of participants in a futures market for electricity will be producers, trading companies, and consumers for which hedging is the main motive for trading. This has important implications for contract design: Different alternatives of contract design have different implications for the resulting correlation between the futures price and the price of the underlying spot asset, and thus the hedging effectiveness of the contract. Furthermore, the success of the market also depends on the resulting liquidity in the contract and the extent to which the contract contributes to attract hedgers as well as speculators to the market. Below we will look into fundamental design issues of the electricity futures contract such as the choice of settlement procedure, the delivery period and length of the contract, quality and size.

Contract Settlement Procedures

Only a small percentage of all futures contracts are normally held to maturity. This implies that the holder of a futures contract normally offsets his position in the contract prior to maturity. If held to maturity, however, contract obligations are to be settled as specified in the settlement procedure. It is this mechanism which links the futures price to spot prices, and as such plays an important role in causing the prices quoted in the futures market to converge to the spot prices when approaching maturity. The settlement specification thus has important implications for the potential hedging effectiveness of the futures contract. There are two main alternatives for contract settlement; physical delivery and cash settlement.

- *Physical delivery:* Under this settlement arrangement the holder of a contract at maturity has to obligation to deliver (take delivery of) the commodity if he has a short (long) position in the futures contract. The commodity is then delivered as specified in the delivery conditions of the futures contract. Contract design here refers to the specification of delivery provisions, including for instance time of delivery, place of delivery, payment of transportation fees, quality and possible quality substitutions. Existing procedures used in executing spot contracts would then provide a good starting point for specifying delivery procedures in the case of physical settlement of electricity futures contracts.

- *Cash settlement:* Under a cash settlement procedure the holder of the contract at maturity pays (or receives) the difference in value between the contracted futures price and the spot price at maturity[63].

[63] Note that in both cases the contract must also specify the period of delivery/settlement. The spot price for electricity normally refers to a given hour (or at the time, a bundle of hours - the price section). This is in many respects too small a period on which to define the underlying commodity,

The hedging effectiveness of the futures contact is dependent upon the price correlation with the value of the underlying endowment of the hedger. The main role of the settlement procedure is to link the futures price to the spot price and provide an adequate correlation of spot and futures prices. It is here the *possibility* of delivery at maturity, or equivalently the *possibility* of cash settlement at maturity which gives incentives for making arbitrage-motivated transactions, and thus causes futures prices to converge to spot prices. Liquid markets in turn make such transactions possible. In a liquid market, the efficiency of a given hedge is in particular not dependent upon whether the contract is held to maturity, or whether it is reversed prior to maturity. Moreover, if properly designed, both settlement procedures should basically be equally viable. The efficiency of both settlement procedures are, however, contingent on a liquid well-functioning and non-manipulative spot market, which in turn is dependent upon, among other things, market design.

We considered both cash settlement and physical delivery as viable candidates on which to base the settlement procedures for an electricity futures contract. Turning to literature it was found that the ability for physical delivery had been important in the early stages of most commodity futures market, while the markets, however, subsequently seemed to evolve into a more mature, and financial market[64]. The increased use of cash settlement in futures markets, however, indicated cash settlement as an interesting alternative. Ex ante it was difficult to conclude which method of settlement would provide the best correlation and hedging effectiveness for the electricity futures contract. As such the success of any of the two settlement forms also would follow from the specific choices of settlement design, delivery conditions, or construction of index.

In favor of physical settlement, it was noted that the majority of participants in the new futures market probably would be hedgers, which traditionally were used to forward contracts with physical delivery, implying that a physical mode of settlement might induce more

implying that the delivery/settlement period should cover a larger period. For the alternative of physical delivery, this refers to which hours the electricity is to be supplied if held to maturity. For the alternative of cash settlement, this refers to which hourly prices the futures price is settled against. The issue of the delivery/settlement period is further discussed below.

[64] Confer e.g. Errera (1984): ‘The ability to make and take delivery on a futures contract appears to be crucial to the early success of such contracts. In the initial stage of development of a futures contract, delivery is a much more common occurrence. During this initial phase, futures contracts are used only as forward pricing mechanisms, and market participants look to futures markets as another source of supply or as another outlet for the sale of unwanted product. This is one of the most attractive features of futures markets to those unfamiliar with their operation. As markets mature, delivery occurs less frequently, and the markets become more financial in nature. A fully mature futures market will typically experience delivery on less than 3 % of the contracts outstanding on a particular commodity’.

confidence in the market in the initial phase. A potential problem, however, arises if a large group of participants actually use the futures contract for delivery. This might in turn cause lower liquidity in the spot market as well as in the futures market. A further problem of this settlement form was also related to the suitability and attractiveness of the futures market for pure speculators and hedgers from the non-electricity energy sectors. With physical delivery as the sole means of settlement at maturity, these participants would in principle be forced to use the spot market for delivery, or offset their position prior to maturity. Failure to attract speculators to the market may be detrimental to the liquidity and efficiency of the futures market, and in turn affect the costs of hedging. We were also concerned whether strategic bidding in the spot market in the hours of which the futures contract was settled, could be a problem. This problem, however, related to both settlement forms, and represented an aspect for consideration in forming the specifics of the settlement procedures.

In conclusion we considered both settlement procedures to be viable for the electricity futures market. As the idea of a futures market was quite novel in the electricity industry, our preliminary recommendation was that settlement by physical delivery seemed to be the most viable settlement procedure to manifest trust in the futures market. This recommendation also followed experiences gained in other futures markets. For the electricity market, a main aspect would also be to promote the futures market as a market for hedging, and *not* as a market for regular physical delivery. The ultimate decision on settlement procedure, as well as on the other details of contract design, should, however, follow from a thorough analysis of the implications of the overall contract design for trading incentives and the potential hedging effectiveness.

*

The initial set-up of the market, i.e. the weekly market, was based on a settlement procedure of physical delivery.

Contract Delivery/Settlement Period and Contract Length

The underlying commodity of the futures contract is defined by the futures contract, normally in terms of quality, quantity and delivery date, etc. As a commodity, electricity is a largely homogeneous, continuously supplied, and not directly storable good. These characteristics make the contract specification of the time of delivery/settlement a crucial factor in defining the underlying commodity. The spot commodity is defined per hour (formerly price section), thus exhibiting hourly fluctuating spot prices. Considering the nature of the good, an hour is, for many reasons, too short a time on which to define the electricity a futures contract. Rather, we believed the underlying commodity of the electricity futures contract should refer to energy delivered within a specific *period*, thus defined as a bundle of hourly spot deliverances. The value of the underlying commodity at maturity thus follows from the aggregate spot market value of electricity delivered in the specified delivery period. In terms of a per unit price, the per unit settlement price may be interpreted as the average spot price of the delivery period[65].

The question encountered in futures design was therefore how the delivery period should be specified to promote an efficient and liquid market. In our discussion we will separately discuss and make a distinction between the delivery/settlement period on one hand, and the choice of weekly load profile on the other hand:

- The *delivery/settlement period* refers to the time of year covered by the contract, as for example a given week, month, or season, etc. In a hydro-power system price fluctuations and uncertainty across these periods are largely related to fundamental factors that affect the net supply of energy.
- In the choice of the intra-weekly *load-profile* focus is, however, on intra-weekly price variations and uncertainty. This variation and uncertainty is in contrast largely related to intra-daily and -weekly variations in demand, and in the costs of meeting load variations and peak-load.

This distinction in the choice of spot hours thus relates to different dimensions of the price risk facing the participants. The focus within this section will be on the delivery/settlement period, as well as issues related to when trading in the futures contract begins (i.e. the length of the futures contract). The aspect of the intra-weekly price risk and load-profile specifications is discussed below as a separate topic.

The choice of the delivery/settlement period is defined as the choice of which time periods within the year that are to be covered by each contract, and largely refers to uncertainty related to the value of energy. Before proceeding with issues of contract design let us first provide a brief background insight as to main patterns of energy price uncertainty in the pure hydro-power market over time.

[65] In the case of physical settlement on maturity, the delivery period refers to the period in which the energy amount is to be supplied in accordance with the specified load-profile. Similarly, for cash settlement at maturity, the settlement period refers to the period defining which spot prices cash settlement is based on. For simplicity, in the continuing we will use the terms ‘delivery period’ or ‘settlement period’ to refer to both cases.

- In the very long term of more than a year, the hedger faces price risk associated with the level of electricity prices in the coming years. In a hydro power system, this largely follows from the uncertain future supply of energy due to future precipitation, as well as weather and temperature conditions which affect the demand side. This year-to-year price risk may be important for hydro-power producers which have a possibility of inter-annual storage, and for example for power intensive industries that enter into long term supply contracts.
- Within the time span of a year, the hedger also faces highly stochastic spot prices following from the uncertainty of future net energy supply. Within the year there are periods in which the (stochastic) flow of new information is of special importance as to price formation. For example, during the autumn months with autumn rain, important information on the amount of precipitation and thus energy supply for the winter season is revealed. During the winter months, it is gradually revealed how cold the winter will be, and thus how winter demand for electricity will be. And, a further crucial time within the year is when spring breaks out and snow-melting and spring floods occur. This intra-annual price risk is important for all hedgers concerned with electricity costs or the uncertainty of revenues.
- The hedger also faces price risk on a short-term with price fluctuations from week to week. This uncertainty to a large extent reflects weather-related fluctuations in demand.

An objective in designing the futures contract is to construct a contract which by and large provides an efficient instrument for hedging the income or cost of underlying endowment of the hedging participants. This endowment may be the participant's expected consumption or production of electricity in a future given time period. The size of the endowment will normally vary from time period to time period, for example with an expected larger consumption or production in December rather than in July. By the nature of electricity, also the endowment is continuous. The main considerations in contract design are related to the resulting potential hedging effectiveness of the contract for major groups of participants.

The discussion on the choice of delivery/settlement period is in effect a question of how to segment the year in contracts. Questions encountered here not only referred to the specific length of time period to be covered by each contract, but also the questions of for example, whether all time periods should be covered, and whether longer period contracts for distant futures could be combined with shorter-period contracts for contracts closer to maturity. In addition is the question of how far from maturity trading in contracts should start. Our analysis was at the time of a qualitative nature, and it was noted that a closer study was warranted to determine the general hedging requirements of the market. Let us review some main lines of the discussion:

- *Delivery/Settlement Period with Fixed-Length Periods, and Consecutive Contracts*

In our first approach we assumed fixed length contracts. Basically this implies that trading in the futures contract is stopped when reaching maturity, i.e. at the beginning of the delivery period, and that delivery/settlement procedures then are effectuated. For example, if the delivery period is a month, trading in the futures is stopped immediately prior to the beginning of the month in question, meaning that no further adjustments in the futures position are possible after this time. Note that given this arrangement, the final payoff on the electricity futures position is not known before the delivery period is ended. The total payoff is dependent upon the difference of the aggregate value implied by the futures contract, and the aggregate value implied by the resulting spot prices. An implication is also that any price fluctuations within the period are netted out, and that the payoff then can be characterized as dependent upon the average price of the delivery period. As such, the specification of the delivery/settlement period plays an important role as to which price risk the futures contract is able to hedge, and thus has consequences for the resulting correlation and potential hedging effectiveness of the futures contract.

Given this scenario, several delivery time periods were considered. A delivery period of a year was initially considered in the debate on the Norwegian market. The idea was that it would allow participants to hedge (very) long time risk, securing a fixed average price and thus a given price level for the year. However, with a yearly fixed length contract, participants would neither be able to explicitly hedge intra-year risk, nor to monitor the size of their hedge vis-à-vis variations in hedging requirements throughout the year. For the alternatives of shorter delivery periods, for example seasonal contracts rather than annual contracts, monthly contracts rather than seasonal contracts, or weekly rather than monthly contracts, the hedger has an increasingly better opportunity to tailor his hedging portfolio to variations in the size of the underlying endowment over time. With a focus on the risk of his revenue or costs of his underlying endowment within shorter periods of time, a probable benefit will also be that the futures price to a larger extent will be correlated with the value of the underlying endowment, thus providing a more efficient hedge. If the price risk within seasons or months is of significant importance to the general hedger of the market, these arguments thus favor shorter delivery periods.

The efficiency of the market is also highly related to the implied liquidity in the contracts, as well as transaction costs. If we assume that contracts were defined consecutively for all periods, the choice of shorter delivery periods implies more contracts. It is here reasonable to believe that transaction costs in trading for example 52 weekly futures contracts per year will be higher than trading 12 monthly futures contracts per year. Furthermore, a large number of short-term contracts potentially implies a more segmented market with lower liquidity in each contract. While shorter delivery periods might offer a better match of the futures position with the underlying endowment, a potential danger is thus that this choice also might infer low liquidity in each contract.

Our conclusion did, however, tend towards shorter delivery periods as of a month, or possibly a week in the short to medium term, though it was stressed that a closer analysis was necessary. This preliminary recommendation was based on the assumption that monthly or weekly defined futures contracts would enable the participants to tailor their portfolio, and probably induce a reasonable price correlation and thus hedging effectiveness in the short to medium term. As such, if analyses were to show that price risk within the month was pronounced and significant to the participants, and, if the flow of new information is high in the short to medium term indicating incentives for liquid trading in the contract, weekly contracts could be warranted. For hedging long-term risk, however, a major concern was that the alternative of weekly contracts would be too large a number for hedging more distant price risk. This would be the case if expectations of price development in consecutive weeks in the far future to a large extent reflected the same expectations as to the general supply of energy, and as such did not distinguish between future consecutive delivery weeks. These aspects imply that shorter delivery periods, with a large number of futures contracts for futures contracts maturing in a more distant future time, might segment the market and imply a lower liquidity for the more distant contracts.

- *Delivery/Settlement Period with Selected Fixed-Length Settlement Periods*

An alternative also considered was whether it was necessary to trade futures for every consecutive week or month. The question here was whether electricity futures design should parallel the design of other commodity futures, where the contract trades for selected delivery periods only, for example for wheat which has contracts for expiration only in March, May, July, September, and December. The argumentation behind this alternative is related to the nature of hedging: The efficiency of a futures position constructed to hedge the underlying endowment of the hedger depends upon the *correlation* between the futures contract and the value of the underlying endowment. If the futures contract covers the same time period as the underlying endowment, the futures contract may constitute a perfect hedge to the endowment, other things being equal. In actual hedging applications, it is, however, normal that the hedged commodity and the futures positions will differ in the time span covered, quantities, qualities, etc. Such a hedge is a *cross hedge*, i.e. where the characteristics of the spot and futures positions do not match perfectly. The cross-hedge may, however, offer adequate hedging if the spot value of the underlying endowment is sufficiently correlated with the futures price. Turning back to the electricity futures contract, if a futures contract defined on consecutive weeks offers nearly the same correlation with the value of the hedger's endowment in nearby weeks, fewer contracts might be traded, offering a less segmented market, with higher liquidity in each contract. Whether this is the case, however, is an empirical question to be analyzed.

▪ *Dynamic Definition of the Delivery/Settlement Period*

Another alternative considered was alternative arrangements of how the contract was traded in the time until maturity. In the above setting, we assumed that the delivery/settlement period of the contract was fixed throughout the life of the futures contract, and that trading in the contract was stopped at maturity. For delivery periods covering longer periods of time, this represented a problem, as trading was stopped, whilst there still was considerable unresolved price risk within the delivery period. As such this left no room for tailoring short-term hedging within the period, possibly posing a problem as to the correlation of the futures price with the value of the underlying endowment. However, while the alternative with shorter delivery periods to some extent would alleviate these problems, it also implied a large number of contracts traded for more distant periods, posing potential problems in terms of transaction costs, as well as the danger of an illiquid and segmented market.

Here another possible solution was introduced, which offered the combined advantages of the two alternatives. Basically, the alternative involved a dynamic definition of the delivery/settlement period of the contract:

- At the start of trading in the futures contract, the delivery/settlement period is defined upon a relatively long period, such as a season or a month. When trade in the contract starts, for example a year prior to the delivery period, general market concern, as well as available information, most likely refers to the price level of the overall period, and not to short-term variations within for example close weeks. During the trade of the contract based on the longer-term defined delivery/settlement period, focus is thus expected to be on the general level of price risk of this period. At the same time the number of contracts for distant periods is limited, avoiding the danger of unnecessarily segmenting the market.

- Later, in approaching maturity, the futures contract is split into several new futures contracts, each covering smaller segments of the original time period: To illustrate, let the starting point of the delivery/settlement period be defined to be a month (or a four-week period). At a given point of time, the contract is split into four separate futures contracts, each referring to one of the four weeks. These weekly contracts are then traded separately, giving the participants larger flexibility to adjust their hedge, and contributing to a larger degree of correlation between the spot- and the futures prices.

To sum up, the idea was that such an alternative might give the market a larger flexibility and ability to cover both longer and shorter term price risk, as well as improve the potential correlation and hedging effectiveness of the contracts.

▪ *Contract Length*

Closely related to the choice of delivery period is the decision when to start trading in the contract. The length of a futures contract is the length of the total trading period from the first trading day till the maturity of the contract. Normally trading in futures contracts starts 3-18 months before the date of maturity. In the early Norwegian debate on the establishment of the futures market for electricity, the electricity industry expressed a need for opening up for trade in contracts several years ahead of maturity. The major problem of starting trade in a contract a very long time ahead of maturity is, however, related to the prospect of a very thin liquidity in the initial life of the contract. Normally trade in a given futures contract increases towards maturity. This is partly related to the information flow in the market, as more information relevant to spot price formation at maturity is continuously received, and uncertainty is gradually dissolved towards maturity. Far from maturity the information flow may be low, possibly contributing to low liquidity in the contract. If, however, further market analyses were to indicate that futures trading in (very) distant contracts could be vital, an implication is that only a small number of contracts should be defined to not unnecessarily segment the market.

*

In the weekly market, a delivery period of a week was chosen as the standard time unit for contracts. From the beginning, trade in the contracts started half a year before maturity. While contracts close to maturity were traded as separate weekly contracts, however, contracts far from maturity were traded in groups of weekly contracts: Each contract in principle covered a given week, but were traded in bundles or blocks of four weeks if time to maturity exceeded seven weeks. Then, at seven weeks prior to maturity, the blocks were transformed into individual weekly contracts. The trading of blocks versus individual contracts took place according to the pattern described in table 6. For the fifth week we see that the first block had been split into 4 weekly contracts, and that another block has been added at the end.

	Number of individually traded contracts with maturity in the closest weeks:	Number of consecutive 4-week blocks traded after the weekly contracts:
Week 1	7	5
Week 2	6	5
Week 3	5	5
Week 4	4	5
Week 5	7	5

Table 6 Weekly and block traded forward contracts

In September 1993 the period of trade was extended to one year, with up to 12 block contracts in trade. From Fall 1994 the trade period for base-load contracts was extended one more year, with a segmentation of seasonal contracts for the second year[66].

Contract Quality Dimension: Load-profile

In some respects the above specification of the delivery/settlement period of the contract is the main attribute that defines the electricity futures commodity. As noted above, this attribute to a large extent is related to the level of energy prices in future periods of time. However, spot prices fluctuate also within the week. The futures contract specification also has to include a specification of the load-pattern of delivery, i.e. which delivery hours within the

[66] After addition of Fall 1994, the contracts traded at any time covered the following maturity periods. Weekly contracts (4-7 week contracts covering each of the next 4-7 weeks), Block contracts (9-11 block-contracts each covering the 9-11 following 4-week periods. In approaching 7 weeks before maturity, the block-contract is transformed into 4 weekly contracts to be traded separately), and Seasonal contracts (3 seasonal contracts covering the next year thereafter, i.e. the seasons covering the periods of week 1-16, week 17-40, and week 41-52, respectively. In approaching 11 x 4 weeks to maturity, the seasonal contract was transformed into the equivalent of block contracts to be traded in separately.)

week that are covered by the contract[67]. To some extent, spot prices follow distinct patterns which reflect the daily load patterns of consumption and the costs of meeting these load variations[68]. For example, peak-load hours are normally associated with high spot prices, and low-load periods with low spot prices. A futures contract with a load-pattern based on a flat delivery with a constant load profile throughout the week, implies a completely different value than a contract based on a standard load-profile that for example, mimics a normal average daily consumption pattern. In a sense, the choice of load pattern for the futures contract can be interpreted as a quality specification, where deliverance of power in peak-load hours can be said to represent a high-quality product, while the deliverance in low-load hours represents a low-quality delivery.

Note that the spot price development within the week, though exhibiting distinct patterns, is not completely deterministic and thus adds extra uncertainty to the spot prices and to the maturity value of the futures contract. A given futures contract based on a fixed load profile does not provide an opportunity to hedge the intra-weekly price risk. With the objective of constructing a futures market to meet the main price hedging requirements of the market, the question here was whether the design of futures contracts should accommodate also this aspect, and if so, how it should be done. An alternative considered in the market was to trade several futures contracts based on different load profiles, for example that two futures contracts would be traded for a given delivery/settlement period, one based on e.g. a flat load profile, and one based on peak load hours.

A basic question to ask in assessing such an alternative is whether the intra-weekly price risk is of significant importance to the main groups of hedgers in the market. An affirmative answer here would indicate that specialized contracts could be warranted. In principle, a thorough analysis of intra-weekly spot price variability would be necessary to conclude on this question. Tentatively it seemed, however, that the uncertainty of price variations within the week did not represent a severe source of uncertainty in the pure hydro power (and reservoir-based) system at the time. Abundance of load capacity at the time, with relatively low cost differences and uncertainty in meeting load variations, tentatively indicated that

[67]In the case of settlement by delivery, the load profile refers to how the energy is to be delivered throughout the week. In the case of cash settlement, it refers to the selection or index of spot prices that are the basis for calculating the profit or loss on the contract.

[68]Load variations within the week largely follow from patterns of consumption. In a hydro-power system, some generation plants, such as river-based plants, cannot regulate their load. For generation plants that can regulate their load, variable costs may be related to costs following from lower power efficiency if the turbine is not operated at the optimal point, or loss of water in starting and stopping turbines.

instruments for hedging intra-weekly price variation were not justified in the pure hydro power market[69].

Moreover, it is reasonable to assume that the trade of parallel contracts, only differentiated by different load profiles, could be detrimental to liquidity. In considering the potential hedging effectiveness of two contracts, an important indicator is the correlation of the price of the contract with the value of the endowment of the hedger. If the correlation properties are similar, both contracts would be potential hedging instruments for the same underlying endowment. In the extreme and hypothetical case, where the intra-weekly price variations are given by fixed factors in relation to an average energy price of the week, the contracts are in principle identical with respect to this correlation. And, if parallel futures contracts are traded which in essence have the same or similar correlation properties, the extra contract is in essence redundant. By trading in parallel similar (and redundant) contracts, the market will be segmented, causing a thinner liquidity in each contract, and thus a less efficient market.

Our tentative conclusion at the time was that only a single futures contract per delivery period should be issued, for example based on a flat base-load delivery profile[70]. It should however be stressed that more thorough analyses were needed to firmly conclude.

*

In the implementation of the weekly market, parallel contracts for each week were introduced. For each week, the base-load power contract and day-load power contract were traded, while the night-load power contract was introduced in 1994:

- Base-load power contract: This contract was for a 24-hour delivery throughout the week (168 hours per week) with a flat load profile.
- Day-load power contract: This contract was for delivery during daytime (7.00 am to 10.00 pm) from Monday to Friday (75 hours per week), with a flat load profile during these hours.

[69] This conclusion applied to the partly isolated pure hydro power system of Norway at the time. Stochastic properties in relation to intra-weekly price variations may, however, be profoundly different in different electricity systems, such as thermal systems versus the pure hydro-power system. Note also that different conclusions thus may apply to a more internationally integrated Norwegian power market.

[70] A further issue related to the load-profile is the specifications as to the relative energy amount to be delivered in each of the specified hours. Here a flat load-profile has been custom, i.e. with an equal amount of energy delivered within each hour.

– Night-load power contract: This contract was for flat delivery during night (10.00 pm to 07.00 am) weekdays, and 24-hour per day Saturday and Sunday (in total 93 hours per week).

Quality Dimension: Geographic Definition of the Spot Price

All electricity is delivered, or taken delivery of, by means of the common electricity grid. For a given producer or consumer this is always at their physical connection point to the grid. With principles of common carriage and a transportation tariff system based on point charges, the tariff paid by the individual buyer or seller is given by his connection point to the grid, regardless of the location of the other party to the transaction. Normal quality attributes of the delivered electricity, such as voltage, frequency, probability of interruptions, etc., are, moreover, first and foremost quality attributes of the individual grid connection (and total system operation), and not the energy sales transaction. This applies whether the electricity is contracted by spot sales, physical futures delivery, or bilateral contracts. In this respect electricity itself (i.e. apart from transportation) is a homogenous good[71].

For the futures contract, however, the geographical aspect may have implications for the resulting correlation of the futures price with the value of the underlying endowment of the hedger. This follows as spot prices are geographically differentiated in cases of network congestion. The price risk of the price hedger is first and foremost related to the volatility of the geographical price he faces, and is thus also affected by the geographical dispersion of supply and demand and its implications for congestion. A question of contract design is thus related to which price to base the contract upon. One might think that several futures contracts based on different underlying geographically differentiated prices would have the potential for a better price correlation and hedging effectiveness. However, with a limited market for each contract, and where the main uncertainty is related to energy, this alternative would in all likelihood result in segmented markets with severely low liquidity. The main implication for contract design, whether settlement by cash or delivery, thus is to base the standardized

[71] Different terms of interruption could in principle have been specified in the standardized contract. Note e.g. that bilateral contracts earlier included the specification of different terms of interruption, or even network priority. Interruptible contracts might in general contribute to a more flexible market and more explicit pricing of interruptions. These aspects were also considered with respect to futures contracts. However, as the main purpose of the futures contract is that of hedging energy price risk, and not handling the uncertainty of network congestion, the inclusion of special aspects of security of delivery thus were ruled out for the derivatives market.

contracts on a common energy related price, i.e. the system price, which is the equilibrium price of the spot market when grid capacities not are taken into account[72].

*

The issue of geographically differentiated contracts was at the time not a controversial issue in the national Norwegian market, and the choice of the system price as the main underlying spot price was a natural choice for the weekly market.

Contract Size

A futures contract is standardized in quantity. The denomination of electricity is in terms of energy, i.e. MWh. As the deliverance of energy is continuous, the quantity of energy on the contract follows from the specified length of the delivery period, the specified load-profile within the period, and the scaling of the contract. The length of the delivery period, as well as the load-profile is discussed above. The remaining issue related to contract size is that of the scaling of the contract. Given the concept of a flat load, the scaling addresses the standard power load. For example, if chosen to be 1MW, the energy quantity of the above base-load contract is 168 MWh, and for the above day-load contract, 75 MWh. In principle a lower scaling implies that the trader has to buy or sell more contracts to hedge a given underlying position, and in turn possibly larger transaction costs. A higher scaling implies larger contracts that might exclude potential traders. The scaling of 1MW seemed to represent a natural choice for energy futures.

3.2.2 Market Micro Structure and Exchange Security System

The establishment of a futures market for electricity called for the set-up of a completely new exchange. In principle the role of the exchange is to offer a trustworthy organized market place for futures contracts, with non-discriminatory, fair and just rules of trade. A basic objective is to strive to achieve a liquid and efficient market, which as such is an underlying objective in designing contracts, auctions, information policies, control and security systems, rules of legal and ethical conduct, etc. The institutional set-up of the main market functions was commented above in section 1.7.2. In this section we will briefly focus on the organizational set-up of the futures market for electricity, which we, however, at the time did not analyze in detail. Basically, we envisaged that the organization of the futures exchange to a large extent should resemble that of other future exchanges. As we will see, the chosen structure of the weekly market was, however, in many respects to diverge from that of normal

[72] See chapter 5 on congestion management.

futures market organization. Our objective here is merely to draw attention to some issues in the organization of trade and security systems in the weekly market, providing a basis for further discussion in the following section.

Market Admittance

The terms of admittance to the futures exchange should basically screen potential participants to secure that they are able to fulfill their obligations in the market. This first and foremost refers to financial obligations, but also their ability to carry out necessary arrangements in the event of physical settlement by delivery on maturity. To lay the basis for a liquid and efficient market, in principle all potential traders that are able to meet their obligations should be admitted to trade in the market, directly, or indirectly. Potential traders in the futures market thus include producers, consumers, and retail trading companies from the electricity market, as well as any other potential traders such as hedgers from other industries, pure speculators, etc. Any terms of admittance that hinder trade by participants that would be able to fulfill their commitments in the market, in principle only will contribute to a thinner and less efficient market.

*

In the initial market organization of Statnett Marked, which organized the daily and weekly market, all participants had to sign a contract of participation that covered all markets of Statnett Marked. All the markets of Statnett Marked were, however, defined as physical markets. This also related to the weekly market which was based on settlement by physical delivery only. On this basis all participants, even in the weekly market, were required to have the necessary technical equipment to be able to trade. A further issue was that there in fact were technical software restrictions on the number of market participants that the market could handle.

Trading mechanism

The trading mechanism is the mechanism by which the latent supply and demand bids of the traders are transformed into realized transactions with market prices and the corresponding allocation of resources. It specifies the form of demand and supply bids, and the market clearing process. Section 2.1.3 gave a brief characterization of main categories of trading mechanisms. Different choices of trading mechanisms may have different implications for the resulting efficiency and liquidity of the market. Section 2.3 discussed alternative trading mechanisms for the daily market. While a spot market and a futures market serve highly different purposes in the market, reflecting highly different motives for trading and highly different informational needs, the most efficient trading mechanism in the two markets may be highly different. Our focus is here on the futures market. Prior to the establishment of the market, we did but briefly sketch alternative auctions leaving the issue an open question for further analysis. On one hand, the periodic order-driven auction based on submitted demand and supply vectors was familiar in the electricity market. While this mechanism offered the thickness of a periodic auction, its poor dissemination of information in the trading process and infrequent trading represented a major drawback for a forward market. On the other hand, the most common futures market auction was a system of open outcry, where the traders meet on the trading floor to match bids and offers. This is a continuous order- or quote-driven auction, and represented years of experience in use in liquid futures markets. Though this mechanism in most futures market at the time was based on physical appearance in the market, the emergence of automated trading systems represented a promising alternative for the electricity futures market.

*

From the beginning the weekly market employed the same sealed-bid, periodic auction system used in the daily market, which in turn was based on the one previously used in the former occasional power market. Participants submitted their orders in the form of demand and supply curves, and the market was cleared once a week on Thursdays. The market price was set at the equilibrium price that balanced the aggregate supply and demand curves. The exchange published the market price, the total volume, and in the case of no trade, the highest buying price and the lowest selling price. Trade in the contract was then closed until the next trading day, one week after.

Exchange security system

Normally internal and external control systems and safeguards are established to secure the financial integrity and neutrality of the exchange, and to manifest the market's confidence in the exchange and the futures contracts. In this an associated clearing house normally plays a fundamental role. The main function of the clearing house is to guarantee that all of the traders in the futures market honor their obligations. The functions of the clearing house include; - the registration and splitting of the bilateral relationship of buyer and seller, in which the clearing house adopts the position of buyer to every seller, and seller to every buyer, thus eliminating credit risk; - handling systems of e.g. mark-to-market, margin payments[73], and other guarantee funds, thus securing the financial position and integrity of the clearing house; - and the organization of the settlement of contracts that are held to maturity. All these functions contribute to secure the fulfillment of contracts, and to maintain the integrity of the futures market. Here we envisaged the establishment of similar arrangements for the electricity futures market as used in other futures markets.

*

The weekly market was initially not associated with a clearing house. The exchange itself, Statnett Marked, entered as the formal counterparty to all contracts and in this way broke the bilateral relationship between the buyer and seller. However, the initial weekly market employed neither margin payments nor a mark-to-market system. Positions were secured by a given deposit. In the absence of the mark-to-market system, every contract was designated by the price at which the contract was entered into. In principle every individually entered contract was held to maturity. Economic profit and loss on all transactions were settled at maturity.

3.3 Efficiency Issues of the Weekly Market

The weekly market established in November 1992 by Statnett Marked was thus to be the first organized market for forward contracts in electricity. The weekly market suffered poor liquidity. There were many probable reasons for this. Explanations may on one hand be related to general market-related conditions, such as frictions in the transition away from the former reliance on long-term bilateral contracts, or issues related to the functioning of the spot market. On the other hand, we believed that the poor liquidity to a large extent was due to several aspects of the infant design of the market, where the chosen solutions to a large extent diverged from more common features of efficiently working futures markets. This related to e.g. the choice of auction format, exchange security system, market admittance, as well as contract specifications. Below we briefly comment upon some design-related issues.

Redundancy of contracts: The market offered a large range of contracts, both within the week (base-load, day-load and night-load contracts), and also for the near future (contracts for each consecutive week). A question to be asked was whether this large a number of contracts contributed to an unnecessarily segmented and illiquid market. More specifically, if several of

[73] The margin is essentially a deposit. The size of the required margin reflects the maximum allowed change in prices each market day, in principle large enough to cover a day's potential loss on the contract. At the end of each trading day, all contracts are marked to market. Potential gains or losses are registered, and the margin position is dynamically adjusted according to specified rules.

the contracts had the potential of covering the same underlying volatility with nearly the same correlation with the hedgers' endowments, some of the offered contacts might be redundant. While this called for a closer analysis, an obvious example was found in the 1994 contract set, where the new night-load contract actually could be duplicated by a base load contract less the day-load contract.

Inflexibility due to lack of margin and mark-to-market systems: While a main role of the margin payment and mark-to-market provisions is to secure the financial integrity of the system, these systems also implicitly contribute to a more liquid market. Active risk management may require rebalancing the futures portfolio to tune in on new information. This is possible in a well-functioning and liquid futures market. In the weekly market, though market clearing was carried out repeatedly (initially once a week) in the time up to maturity, and did as such provide the participants a possibility of dynamic portfolio adjustment, we believe the absence of margin and mark-to-market systems contributed to hinder this process of adjustment. In the weekly market every contract was designated by the price at which the contract was entered into, and the price followed the contract till maturity. Also, with the chosen guarantee provisions in the weekly market, and the absence of margin payments, economic profit or loss on the transactions in contracts was settled only at maturity. In particular, this was the case even when the participant had sold out so as to have a net zero position, thus implying that every individual entered contract actually was held to maturity. Though the participant might alter his net position by trade, however, the above issues implied a cumbersome and complex process to carry out a dynamic risk strategy, which we believe also contributed to a lower liquidity in the market.

Restricted market admittance: Market admittance was coupled to a physical participant agreement, implying that only participants that adhered to the specific technical requirements were allowed to trade. In this, several important groups of traders were virtually excluded to trade. This might have contributed to exclude not only important groups of speculators, but also pure hedgers as for example firms with hedging needs due to the importance of electricity as an input, but without the specific technical equipment. If the physical ability to deliver or take delivery actually were a requirement that followed from the choice of a settlement procedure based on physical delivery, this would be an important cause for reconsidering the choice of settlement. However, given a liquid underlying spot market, we did not see why such technical requirements of the trader in fact were necessary. An alternative would for instance have been that delivery procedures for this group of participants were coupled with an automatic spot market arrangement, for example by means of price-independent bids in the spot market. The main concern in specifying market admittance should be to secure the financial ability of the participant to meet his obligations in the market.

Emphasis on delivery rather than hedging: In our view several of the choices in contract design, as well as in the exchange framework, seemed to mirror a physical delivery orientation, rather than a price hedging orientation in the set-up of the market. We have seen that a settlement procedure of physical delivery was chosen. As experienced in several other commodity futures markets, this mechanism is in principle viable for a liquid futures market, and may as such support an efficient price hedging market. This choice, however, in the combination with several other market design choices seemed to indicate that too much weight was put on the idea of physical delivery, rather than handling price risk. Several design choices may be interpreted in this light. For example, - contracts were tailored with several delivery profiles possibly reflecting delivery needs rather than hedging needs, - market admittance was based on the ability for physical delivery, - and trading and guarantee systems seemed to favor that the trader held the contract till maturity, rather than facilitating dynamic hedging strategies. As such, it seemed as though it was expected that the standardized forward market was to handle both the role of risk allocation, *and* to be used as a means of role of handling and securing physical delivery[74]. In trying to emphasize both these objectives in designing the market, the result is likely to be that neither of the objectives can be reached.

Insufficient financial integrity of the exchange?: A further question to be raised was whether the implemented exchange structure sufficiently guaranteed the financial integrity of the exchange. In the place of a clearing house the exchange itself entered as counter-party to all contracts. Its net exposure to the credit risk of contract holders was thus contingent on the well-functioning of the exchange deposit and security system. On one hand we noted that normal means of securing the forward contracts, such as margin payments, were not in place. In contrast to the normal dynamically adjusted deposits of the margin system, where margin payments are tailored to the development in the trading position of the traders, a concern related to a system of more fixed deposits is that the deposit in the course of time may be too

[74] Early market suggestions (for example the AS IPT's proposal for design of standardized contracts) even seemed to reflect a notion of the standardized market as a market to also secure delivery.

low, and then expose the exchange to a financial risk[75]. On the other hand, a concern of financial integrity was also related to the close relationship between the grid company and the power exchange.

Informational and efficiency properties of the trading mechanism: The trading mechanism of the weekly market was a periodic, order-driven, and sealed mechanism, which was in contrast to the normally employed continuous quote- or order-driven trading mechanisms. The chosen mechanism of the weekly market, where only information on the resulting transactions was made public, represented a minimum of dissemination from the trading process. In a forward market, where information related to the market's assessment of future price development is important, it may be asserted that the information dissemination of the chosen mechanism was at the very minimum, and may have been a factor contributing to the poor liquidity and efficiency of the market. As such, we believed improvements in the trading mechanism, as e.g. using a continuous order- or quote-driven mechanism could improve information dissemination, and thus the liquidity and efficiency of the market.

*

Above we have pointed out several sources of inefficiency in the infant weekly market, accounting for the severely low liquidity. The market response to this was the establishment of several competing exchanges outside the sphere of Statnett Marked; Norsk Kraftmegling opened a market place for seasonal and annual contracts in April 1994, Norsk Kraftmarked opened for trade and the clearing of standardized annual and seasonal contracts in May 1994, and Markedskraft, originally based on the brokerage of bilateral contracts, also planned taking a step towards more standardized products. Amongst these competing markets were market design features that to a greater extent where in line with the common futures market design. Examples are the employment of continuous order- and quote-driven pricing mechanisms, the use of a clearing central, margin requirements, and weekly settlement of contracts. The futures market of today is a further development of the original weekly market, and partly the futures market initiated by Norsk Kraftmarked, which was taken over and integrated in Statnett Marked. As such all the emerging alternative market places probably played an important role in the development towards today's futures market for electricity.

[75] If the deposits are too high, they may, dependent upon the interest rules, etc., represent large trading costs for the participants and thus also deter liquidity.

3.4 The Futures Market Today

The financial market for electricity has undergone large transformations from a relatively illiquid physically oriented forward market, to a full-fledged financial market for electricity derivatives. The traded volume at Nord Pool's financial market has increased considerably since the first products where launched. The total volume of financial contracts traded at Nord Pool's financial markets in 2007 was 1060 TWh, and the total volume of trades cleared by Nord Pool Clearing in 2007 was 2369 TWh. In comparison, the volume of Elspot was 290.6 TWh, and the 2007 production in the Nordic area was 397.3 TWh. We will now briefly comment upon the exchange organization and traded derivatives[76].

Trade in financial contracts is organized by *Nord Pool ASA*. The exchange holds a license as a derivatives exchange, and is as such under the supervision of the Financial Supervisory Authority (FSAN) of Norway. Trade is also thoroughly regulated by internal rule books, such as 'Trading Rules for Financial Electricity Contracts and Certificate Contracts', and trading agreements such as the 'Exchange Membership Agreement' and the 'Market Maker Agreement'. Exchange trading may be carried out by exchange members. An entity that is not an exchange member may carry out exchange trades as a trading client, represented by a client representative which is an exchange member approved to represent trading clients. *Nord Pool Clearing* is licensed as a clearing house, and clears all contracts traded on the Nordic Power Exchange, and in this guarantees financial settlement. Nord Pool Clearing is under the supervision of FSAN. In order to carry out exchange trading at Nord Pool, the exchange member must be a clearing member with Nord Pool Clearing. Market participants wishing to clear its contracts through Nord Pool Clearing have to post and fulfill collateral requirements at all times.

To promote trading by new market participants and to stimulate greater liquidity, the financial contracts were changed from physical-delivery contracts to pure financial contracts. All exchange-traded derivates are now settled by cash settlement[77]. The reference price for all Nordic financial contracts is the system price of the total Nordic power market. The previous

[76] Sources: The information on the current (per summer 2008) system is mainly found on the websites of Nord Pool (www.nordpool.com). For an overview, see 'Trade at Nord Pool ASA's financial market', Feb 20 2008.

[77] Parallel to this development, it should also be noted that an increasing share of bilateral contracts also are based on financial settlement. Clearing of standardized financially settled contracts traded off the exchange was introduced in 1997 at Nord Pool Clearing, and in 1999 price and volume information was published for the first time for standardized financially cleared power contracts traded off the exchange and registered with Nord Pool for clearing.

periodic auction trading system was in 1994 replaced by a continuous trading system, and further developed as an electronic trading system from 1996 supporting a continuous quote-driven trading system with the use of market makers.

The product structure of the exchange has undergone several changes over many years. The exchange now lists futures contracts for the near-maturity contracts, and forward contracts for more distant maturities:

- *Futures contracts:* Futures contracts are traded with maturity for the nearest 5-6 weeks. Settlement of the future contracts involves both a daily mark-to-market settlement, and a final spot reference cash settlement. The mark-to-market settlement covers gains or losses from day-to-day changes in the market price during the trading period, and calls for the contract holder to deposit the necessary amount in cash accounts. The final spot reference cash settlement takes place during the delivery settlement period. Throughout the final settlement period which starts on the due date, the contract holder is credited/debited an amount equal to the difference between the spot market price and the futures contract's final closing price. Nord Pool now lists the following series of Nordic futures contracts:
  - The Base Load Day contracts (ENODddmm-yy) cover 24 hours each, and are listed for the remaining of the nearest week.
  - The Base Load Week contracts (ENOWww-yy) cover 7 days each, and are listed for the 6 consecutive weeks[78].
  - The Peak Load Week contracts (ENOPLWww-yy) cover the intra-weekly defined peak hours of 08-20 Monday to Friday, and are listed for the nearest 5 weeks.

- *Forward contracts:* The forward contracts are listed for up to six years ahead. The forward contract is settled by expiry market settlement with spot reference cash settlement only. As such, unlike the futures contracts, in the trading period prior to the due date for all these forward products, there is no mark-to-market settlement. The mark-to-market amount is accumulated daily as daily losses or profits, but is not realized throughout the trading period. Cash is required in the trader's cash account during the delivery period, starting at the due date. Previous to this, other forms of collateral are acceptable. Settlement throughout the delivery period is given by the difference between

[78] The Base Load Week futures contract is split into Base Load Day futures contracts at the start of each settlement period.

the contract price and the system price of each hour. The following series of Nordic forward contracts are now listed[79]:

- The Base Load Month contracts (ENOMmmm-yy) cover the month in question, and are listed on a 6 month continuous rolling basis (and are not subject to further splitting)
- The Base Load Quarter contracts (ENOQq-yy) cover the quarter (3 months) in question.
- The Base Load Year contracts (ENOYR-yy) cover the year (Jan 1 – Dec 31) in question.
- The Peak Load Month contracts (ENOPLMmmm-yy) cover the above-defined peak hours, for the nearest 2 months.
- The Peak Load Quarter contracts (ENOPLQq-yy) cover the above-defined peak hours, for 3 quarters.
- The Peak Load Year contracts (ENOPLYR-yy) covering the above defined peak hours, for 1 year.

Before leaving this market, let us also mention that in addition to forward and futures contracts, also other contracts have been introduced at Nord Pool, with options introduced in 1999, Contracts for Differences in 2000, European Union Allowances contracts in 2005, and Certified Emission Reduction contracts introduced in 2007. Though our focus here has been on futures and forward contracts for risk management, let us end this chapter by very briefly describing these later added contracts for risk management, i.e. option contracts, and contracts for differences:

- *Option Contracts:* An option is a right to buy or sell an underlying contract at a predetermined price at a predefined date in the future. The option contracts now traded at Nord Pool are European-style, i.e., they can only be exercised at the exercise date. The underlying contracts are quarter and year forward contracts.
- *Contracts for Differences (CfD):* A CfD is a forward contract with reference to the difference between an area price and the Nord Pool Spot System Price. Nord Pool provides trading in CfDs for the following areas; Norway (Oslo), Sweden, Finland, Denmark West, Denmark East and SYGER.

[79] Note that all quarter and year contracts are automatically cascaded, following specified procedures. The year contracts are basically split into quarter contracts, and the quarter contracts into month contracts. After the cascading day, the ‘new’ contracts follow the closing price of each separate product.

4 System Reliability and Ancillary Services in a Market-Based System

In this chapter we take a closer look on how systems and markets for handling the momentary system balance and system reliability are organized in the market-based Norwegian electricity system. While supervision and operation of system balance and system reliability is the responsibility of the system operator, the system operator itself does not own any such capacity. A variety of generating and power load capacity is necessary in order to balance the system and uphold system reliability. This capacity can be supplied by generators and in principle also flexible consumers. Furthermore, an important insight is that a large part of the capacity fit to carry out such services, has competing uses, for example in the supply of ordinary planned energy, versus holding the capacity as a reserve capacity for momentary balancing. In a market-based system, the main principle is that these decisions of allocating capacity are made by the individual participants in the market. Moreover, these decisions are affected by the incentives provided by the market system. An objective of market design is thus to design organized markets and other pricing and allocation systems, so as to induce efficiency both with regard to the efficient supply and use of energy in general, and also in ensuring an efficient and reliable momentary power balance.

The focus on momentary power capacity and ancillary services was accentuated during the mid 1990's by the process towards a more open and integrated foreign trade regime, and with the subsequent establishment of an integrated Nordic power exchange. In many respects this development represented an integration of the pure hydro-power Norwegian system with more thermally based systems. This integration in essence offered the opportunity to reap the potential benefits of trade, following from the comparative advantages of the different systems. At the same time, this development also represented new challenges for market design. The concern was now whether the existing market systems would be able to contribute to an efficient allocation and remuneration of power and energy resources in this scenario.

In the setting of the Norwegian unbundled market system, this chapter focuses on the incentives of the pricing and allocation systems of the market, both as to ancillary services in general, and as to the allocation of planned and unplanned power load capacity in particular.

The chapter is based on Rud (1995), Rud and Tjøtta (1996) and Rud and Singh (1997). Section 4.1 discusses benefits and challenges following from a more internationally integrated power system. Section 4.2 gives a simplified introduction to concepts of system reliability and ancillary services, thus focusing on the supply of such capacity and the provision of system reliability. Section 4.3 is concerned with the demand for system reliability, and in particular inherent aspects of system reliability which may be interpreted as a public good. Section 4.4 discusses how different pricing and market systems reflect different dimensions of the commodity electricity, also describing the systems of the Norwegian market unbundled market system at the time. Section 4.5 considers an alternative pricing structure for ancillary services. Section 4.6 concludes by briefly accounting for current market systems.

4.1 Benefits and Market Design Challenges of a More Internationally Integrated Norwegian Electricity Market

In many respects the hydro-power system has an unique ability to carry out planned load-following and handle real-time balancing of the power load. The development towards a more internationally integrated electricity market implied a larger interface with thermally based systems with other characteristics. This section considers principle aspects of benefits from this integration and the implied challenges for market design.

4.1.1 The Benefits of More Internationally Integrated Power Systems

The initial opening of foreign trade that followed the general market reform, was based on market negotiated contracts. In this it represented a small, but clear step away from the former political restrictiveness of foreign trade. This was followed by further measures of opening trade, and subsequently thoroughly manifested in 1995 when Nordic energy ministers agreed to expand Nordic electric power co-operation. In 1996 the joint Norwegian-Swedish power exchange was started for trade in power contracts, and was later also to include Finland and Denmark.

A basic driving force of this development was the expected benefits of trade, following from the comparative advantages of the different energy systems. Here Norway was a purely hydro-based system, Sweden was a part hydro and part thermal system, and the other Nordic countries, as well as other close European countries to a large extent were thermal systems. The comparative advantages of hydro-power versus thermal systems are mainly related to issues of the supply of energy on one hand, and the value of power load capacity and ancillary services on the other hand:

- Energy supply in the hydro-power system is given by the nature-given precipitation, and is thus characterized by severe risks as to the available supply of energy. Energy generation in thermal systems is based on a variety of fuels which may be bought in markets. In principle, trade has thus the potential of offering a higher security of energy supply to the hydro-based system.
- Capacity for ancillary and load following services is necessary to ensure the momentary balance of the power flow and maintain the operating quality of the power system. In the thermal system, the ability and costs of providing different kinds of ancillary services, including load following, varies from plant type to plant type, but is in general relatively costly, with higher start/stop costs and higher variable costs. In the reservoir-based hydro-power system, the costs of supplying such services, as well as peak-load delivery, is in comparison relatively low. In this the hydro-power system offers cost advantages relative to the thermal system. This to a large extent follows from the underlying technology of the hydro-power system, where high investments in power load capacity to some extent are necessary to efficiently utilize the supply of water. Variable costs are also relatively low, with low start/stop costs, and low variable costs. For example, costs of momentary adjustments are largely related to the deviation from optimal turbine operation and the corresponding loss of water. Traditionally the Norwegian electricity system has thus had an abundant load capacity.

From a Norwegian point of view, the closer integration with other power systems offered the possibility of a better pay-off on the Norwegian power resources, especially with respect to peak-load supplies, load-following, and various ancillary services. The possibility to import energy also offered a higher security of supply, especially in the event of dry years in the hydro-power system. The realization of such benefits of integration, however, on one hand was contingent on the size of the actual cost differences in supplying power-load related services, and on the actual capacity of foreign trade, including the capacity of new investments in international cables. On the other hand, the realized benefit also was dependent on the extent to which the pricing systems of the market were able to reflect these values.

4.1.2 Challenges for Norwegian Market Design

In the anticipated closer integration with thermal systems, it was expected that the Norwegian market to a larger extent would face demand that represented a larger inherent willingness to pay for such services. This also implied a transition to a scenario of a greater relative scarcity of such capacities, and as such an implied higher inherent market value of e.g. peak-load and load-following abilities, capacities for momentary balancing and other ancillary services.

The concern was now whether the existing market systems would induce an efficient allocation and remuneration of these resources in the short, as well as in the long run. A basic insight is here that the actually resulting market prices, and thus the resulting remuneration, to some extent depends upon the chosen market design. Note that the commodity in this respect is a heterogeneous good, for example with different abilities for planned versus unplanned adjustments and load following, different response times, and different abilities with respect to a long range of other ancillary services.

The current pricing systems, developed in the initial stages of the market reform, did to some extent reflect different dimensions of the peak-load and momentary balancing attributes of the system[80]. The market reform had, however, taken place in a partly pure national market. The relatively isolated Norwegian power system in many ways represented a scenario of abundant capacity for the supply of balancing and ancillary services. From the abundance of capacity, it followed that the relative inherent market value of such services was low. Moreover, it seemed probable to assume that elaborate mechanisms in this scenario had not been needed to ensure an efficient allocation of these resources, and that this was reflected in the chosen design of the market system.

The question was now whether the market system, in a future state of a relatively larger scarcity of such capacity and services, would be able to bring forth market prices that would signal the real market value of different power load capacity and ancillary services. This has important implications for the short-run as well as long-run efficiency of the market. For example, the inherent value of specific services and capacity may be high, e.g. due to high demand and high willingness to pay for such services. However, if the inherent value is high, but the pay-off in market prices due to the explicit market design and price structure is low, the resulting incentives of the market system will result in inefficient resource allocation in the short-run, and may imply lower than optimal investments for the capacity in question. In this way the development of a more internationally integrated market also represented new challenges for market design in the Norwegian electricity system.

A further challenge of market design is related to the demand for such services, and for ancillary services in particular. For a large range of power generating capacity, there are in

[80] See section 4.3 for a further discussion of the pricing system at the time.

principle several competing uses, for example ranging from normal energy production to the provision of different forms of ancillary services. We have in our exposition already encountered such alternatives, i.e. where we argued that producers had the choice of offering their capacity in the spot market, or alternatively, in the regulation power market. In principle the optimal allocation of resources to different appliances is to follow from the decisions of the individual participants, on the supply side and on the demand side. In the presence of market imperfections, however, reliance on only individually based decisions will not necessarily induce optimality. In this respect we will argue that the system reliability in effect is a public good, in which special measures are required to ensure a sufficient level of system reliability in the power system.

4.2 About System Reliability and Ancillary Services

In this section we will dwell on the continuous nature of electricity, and seek to provide a minimum background insight into the concept of system reliability which is the quality of the system, and into the concept of ancillary services which are the services necessary to maintain this quality.

The commodity electricity is traded, registered and settled in terms of energy. Electricity is, however, a continuously supplied commodity. The actual production or consumption of electricity within a given time period may be represented by a continuous power load function. The term *power* may simplified be described as the momentary level of production / consumption. *Energy* is the integral of the power load function within a given time period. The connection between power and energy is illustrated in figure 7. The power load at a given instant of time is measured in Watt, while energy is measured in Watt hours[81]. An energy amount of 1 kWh thus refers to the energy equivalent of a constant power load function of 1 kW during an hour.

[81] Normal denominations are kWh, MWh, GWh or TWh. Here we have that 1 kWh = 1000 Wh, 1 MWh = 1000 kWh, 1 GWh = 1000 MWh, and 1 TWh = 1000 GWh.

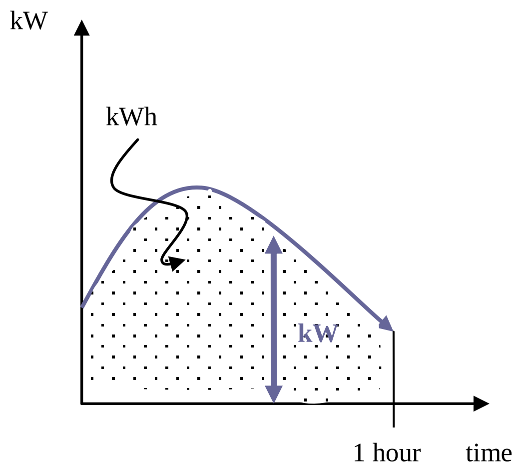


Figure 7: Power load (kW) and energy (kWh)

We can, somewhat simplified, say that electricity production and consumption (including losses) must balance at all times. Essentially this applies to the momentary level of power supply, as well as to the energy level. On an aggregate level, variations in the power loads of the individual producers and consumers will to some extent even out. Still, following the rhythm of daily life in the society, we find identifiable patterns of the aggregate load profile, for example with lower energy use at night, higher loads during the day-time, and with identifiable peak-load hours. It is, however, not to be expected that the aggregate production and consumption load will balance at all times by itself. This means that any implied imbalances in the aggregate net load supply, rapidly must be met to secure the momentary balance and quality of the system.

In maintaining this balance and system quality we can say that the system operator has to maintain an adequate *system reliability.* While our above discussion gave an ad hoc idea of the system quality in the notion of a 'system balance', the quality dimension is in essence more elaborate. We can, still simplified, summarize the aspect of system quality, in the dimensions of stability of frequency, quality of voltage, and security of delivery:

- *Frequency Stability:* The requirement of a stable frequency in the system refers to that frequency and time deviations are held within specified limits.

- *Voltage Quality:* This refers to the stability of the voltage level, with specified limits and rules as to the occurrence of voltage irregularities.

- *Reliability of Delivery:* This term refers to the risk of short-term interruptions, among other factors due to failure in the network. A further dimension is how fast the system can be restored after such interruptions.

On one hand, central supervision and coordination of the system by the system operator is an essential factor in the provision of system reliability. This is necessary as control and measures for adjustment are dependent upon the overall condition of the system. On the other hand, various forms of generation capacity are required to carry out these necessary adjustments of the system. These capacity and services we will term *ancillary services*.

Most ancillary services are provided by the general production plant owner, though also flexible consumers may represent a valuable source of some types of ancillary services[82]. In principle different production facilities with special attributes and abilities may be required to carry out the various services needed to balance the various quality attributes of the system, as well as to ensure a sufficient ex ante preparedness of the system to meet unforeseen momentary situations. Ancillary services thus include a large range of power-related functions and capacities that are necessary to keep the system working and reliable. While our focus will be on the economic system for procuring this capacity in the short and long term, we will not go into any kind of technical detail. However, to give an idea of the diversity of such capacity, let us briefly and simplified outline main categories of ancillary services. Our exposition illustrates the classification of the main categories of ancillary services at the time of study, as described in Wangensteen, Grande and Bakken (1996). Within each category we also see that the ancillary services can be further differentiated along different dimensions.

- *Primary regulation reserves:* This is capacity for instantaneous adjustment of power to deal with real-time momentary variations due to surges or reductions in aggregate demand. This includes reserves for handling momentary deviations in frequency, as well as other acute disturbances in the system. The reserves in this category are activated automatically and continuously, and thus represent the ancillary services that have the shortest response time. Important dimensions in this category are the level of deviation in frequency before the capacity is activated (e.g. deviations of 0.1 Hz, 0.15 Hz, 0.5 Hz), the response time (e.g. 5 seconds, 30 seconds), and the duration which indicates the time until the activation is to be ended.

[82] The regulation power market is a market for the supply of one form of ancillary services, namely secondary regulation reserves with a response time of 15 minutes. Note that flexible consumers may also supply this form for regulation given that they meet the requirements of a technical nature, as well as required response times.

- *Secondary regulation reserves:* These reserves are typically activated after the primary regulation reserves. The generator, or in principle also a flexible consumer, supplies regulation power, by offering to adjust power upward or downward within given response times. Relevant dimensions here are automatic versus manual activation, response time (e.g. 30 seconds, 15 minutes, 4 hours[83]), and duration which is the time period until which the activation is ended, and the next reserve category is to be activated.

- *System safeguard:* This is automatically activated measures to avoid breakdown of the power system. This category is activated in extraordinary situations after the reserves within the previous mentioned categories. An example is the pre-defined shut-down of production or consumption to handle serious and irregular disturbances in the system, such as if critical lines fall out, surcharge of lines, over-frequency, etc.

- *Reactive power:* Both the running reactive production and reactive reserves represent ancillary services that are activated due to deviations in voltage levels. The capacity may be differentiated along different dimensions, for example as to the size of deviations to occur for the capacity to be activated.

The specific definition and specification of ancillary services varies from country to country, and will in principle reflect the needs and characteristics of the underlying production technologies in the specific electricity system. A borderline case of ancillary services and normal deliverance of power are the services of so-called *load-following*, which deals with variations that take place systematically over a given, but short period, e.g. adjusting generation to adapt to predictable hour-to-hour and daily variations in demand. The load-following services thus address the planned following of a load curve, in contrast to the above-mentioned categories of ancillary services which address unforeseen instances. In the hydro-power system, where such costs are low, these services had not been a major issue. In contrast, the costs of load-following in thermal systems may be high, which implicates that greater attention is directed towards these services in thermal systems.

[83] Secondary regulation reserves with even longer response times may be termed tertiary reserves, etc. Also note that in thermal systems, the secondary regulation reserves may comprise generation by different production technologies. They may also classified by different response and ramping times, for example spinning reserves (capacity from generators already in operation that can be called on very shortly, e.g. 10 minutes), quick-start generators, and non-spinning reserves (that can be ramped within an hour).

Some of the above-mentioned ancillary services may require specialized equipment. For others, the production of the specified ancillary services is carried out with generation capacity which can be used in alternative means of production. An example is where the production capacity can be allocated to general planned energy production, as well as to reserves for providing ancillary services. An important objective of the pricing and resource allocation system of the market is here to ensure an efficient allocation of the production capacity with respect to its various competing uses.

4.3 Market Imperfections of System Reliability

Before proceeding to our discussion of pricing systems, let us also first dwell on the value of system reliability, and on the mechanisms for ensuring adequate system reliability in the market. We first discuss attributes of system reliability, and argue that it bears the characteristics of a *public good*. Secondly, we look into measures to deal with this market imperfection.

4.3.1 System Reliability as a Public Good

System reliability refers to the overall quality of the system, in terms of frequency and voltage stability, and the general reliability of delivery. These system qualities at the same time represent quality attributes of the delivered commodity itself. As such, it is quality dimensions that are experienced by all users and producers of the good. We will here argue that system reliability to a large extent adheres to characteristics of a public good.

Economists use two conditions to characterize a pure public good, that is, a public good is non-rivaling and non-excludable in use:

– *Non-rivaling in use:* A good is non-rivaling in use when the consumption of the good by one person does not reduce the quantity of the good available for others. This is in contrast to private goods where the benefits of a particular commodity or service accrue to a particular consumer, and to that consumer alone.

– *Non-excludable in use*: A good is non-excludable in use if it is impossible, or extremely costly, to deny others the use the good. In contrast, in the case of a private good, others can be excluded if they do not pay the price of the commodity.

In our opinion, system reliability is a public good. To see this, let us first note that both power producers and power consumers benefit from the supplied system reliability. It is quite self-evident that consumers prefer a stable frequency and voltage level, as well as low probabilities of interruption. Power producers also benefit from high system reliability, as it ensures a stable production environment, and since electricity prices in a system with higher system reliability, in all likelihood will be higher than in a system with lower system reliability. For both consumers and producers, there is also a value associated with the preparedness of the system, i.e. the fact that ancillary services at all times are ready to be used, even if they are not to be activated. This represents insurance as to the working and quality of the system, which all users connected to the system benefit from. In sum, we find it reasonable to assert that all producers and consumers benefit from high system reliability.

Let us now argue that system reliability is a non-rivaling good, as well as a non-excludable good: If one consumer benefits from a high level of system reliability, this does not hinder others from benefiting from the same system reliability. This implies that system reliability is a non-rivaling good. It is also reasonable to characterize system reliability as a good from which it is not possible to exclude others from using it. If a producer or consumer is connected to the grid, it is in practice not possible to exclude him from benefiting from the system reliability. The only possibility for exclusion is to deter him from using the network at all.

We do, however, in one instance find that part of system reliability is excludable in use, in the meaning that it in this instance is possible to register and collect payment for its use. This concerns part of the secondary regulation. Here it is possible to charge each participant for the deviation between their realized power load measured as energy within the hour, and their commitments of delivery. By charging for these deviations, the market participants face incentives to avoid deviations as far as possible. This may in turn partly reduce the overall need of secondary regulation reserves. Note, however, that deviations and balancing needs *within* the hour are non-excludable in use. This follows as it within our hourly settled energy system, is not possible to track the intra-hourly balancing needs to the single participant. For the other aspects of system reliability, our conclusion is nonetheless that the system reliability is a public good.

4.3.2 Measures for Ensuring System Reliability

From economic theory we know that special measures are necessary to attain an optimal and efficient level of the public good in the market. To see this, let us first consider a market for the normal competitive private good. In this market, efficient market allocation follows from market clearing based on the supply and demand bids of the individual participants. Further, the participants know that they only will be able to obtain the good in the desired amount, if they state their true willingness to pay. However, if the same normal market framework is instituted for the public good, the market equilibrium will not be efficient. In particular, less than the optimal level of the public good will be provided. Due to its non-excludable and non-rivaling nature, the individual market participants do not have incentives to bid their true willingness to pay. The result is that the supply of the public good will be too small, or to the extreme not provided at all.

Thus, a market based on normal market mechanisms, i.e. where individual supply and demand alone form the base of the equilibrium market allocation, is not able to provide an efficient level of the good system reliability. Total wealth of the society can, however, be increased, if the supply of the public good is increased. To achieve this, public policy measures are called for. The main tasks of providing the public good system reliability include: i) the optimal total level of system reliability has to be found, ii) the production of this level has to be financed, and iii) the optimal level of the good has to be supplied in a cost-efficient manner.

i) Optimal level of system reliability: Economic theory states that the optimal level of a public good is the level at which the marginal cost of providing the public good, equals the *marginal public benefit* of the good. The marginal public benefit is the *sum* of the benefits of the private individuals. For system reliability, the implication is that public authorities estimate and set the required level of system reliability. In the Norwegian electricity system the regulatory office, the Norwegian Water Resources and Power Board (NVE), has delegated this task to Statnett SF. Within a closer specified framework, Statnett SF is to specify and make operative the requirements of system reliability, covering both the generation and the network aspects. In principle, the level of system reliability is to be set so that the marginal cost of increasing system reliability, corresponds with the marginal public value of improved system reliability. A non-trivial aspect here is to assess the public value of system reliability, a value which is not reflected in any markets. A method commonly used for assessing such values includes the uses of surveys[84].

ii) Financing: The provision of system reliability also has to be financed. An objective in the choice of financing system is to find a financing form that minimizes the costs to society, both

[84] As an example see Haugland, Meyer, Rud and Singh (1989) for the design of a survey related to the value of the quality of delivery to consumers. This survey was carried out in 1990/91 and also in a modified version in 2001. It was among other things used to evaluate the value of different quality dimensions in the electricity system.

as to explicit transaction costs and as to implicit costs following from imposed deviations from the optimal market solution. In this question we argued that a financing by grid tariffs could be a highly relevant financing form, which also reflects the close connection of the grid and system reliability. There are though several challenges as to how such charges should be implemented.

iii) Cost-efficient provision of ancillary services: The third issue is related to the cost-efficient provision of system reliability. The system operator supervises the operation of the power system, and is responsible for the general system reliability. In relation to ancillary services, the system operator ex ante has to procure reserves for ancillary services, ensuring that the necessary capacities are available real-time, and in accordance with necessary specifications. During delivery the system operator has the responsibility for real-time running supervision and adjustment of the power flow to maintain system reliability. In this, reserves are activated when necessary to balance the system and maintain the quality of the system. Challenges of market design are here to construct mechanisms that contribute, on one hand, to the efficient allocation of resources to reserves/ancillary services, versus to alternative uses of the capacity in the market system. On the other hand, the mechanism should ensure that it is the most efficient of the available reserves that is activated. These issues will be the topic of the following two sections.

4.4 About Pricing Systems for Power, Energy, and Ancillary Services

We now turn to the pricing systems for power and energy in the unbundled electricity market. In many respects the core markets of the competitive electricity market are the spot and futures markets. These are markets in which the pricing and allocation of *energy* is central. However, energy markets alone do not suffice to ensure an efficiently working electricity market. This is due to the multifaceted requirements related to the supply of a reliable continuous flow of power. As we saw above, coordination and control of real-time power system operation is carried out by the system operator, while the necessary equipment is owned by generators, and at their disposal. This implies that market signals and incentives from the pricing systems of the market play a crucial role in the allocation of these resources in the short and long run. The objective of market design is to construct pricing systems that provide incentives that will contribute to an efficient short-term allocation, as well as signals for efficient long-term investment strategies. Following the closer integration with other power systems, it was also important from a Norwegian welfare point of view, whether the pricing systems sufficiently would reflect the value of the flexible hydro-power generating capacity, and thus provide the producers with a remuneration that reflects the intrinsic value of these resources.

Our focus in this section is to consider the extent to which the existing pricing systems reflected the different dimensions of power and energy capacities, and the resulting implications for efficiency in the short and long run. Note that we here use the term *pricing system* quite broadly, by referring to all markets and arrangements that set the framework for trade in the electricity system, and thus covering organized markets as the spot and regulation power market, contracts and payoff systems organized by the system operator, and network tariffs. Section 4.4.1 first discusses the continuous nature of electricity in relation to pricing systems, illustrating this aspect with a hypothetical continuous pricing scheme relating to the continuous power flow. While this representation in principle may capture all aspects of the intrinsic value of the good, it is, however, not an operable system. Section 4.4.2 illustrates that a set of several pricing systems is necessary within an unbundled electricity system. In effect each pricing system catches different dimensions of the attributes of the underlying power capacity. This section also briefly reviews and discusses the then current pricing systems of the Norwegian electricity system in relation to the pricing of power, energy and ancillary services.

4.4.1 About the Continuous Nature of Electricity

Electricity is a continuously supplied and produced commodity. The production or consumption of electricity may thus be fully characterized by the corresponding continuous electricity power load function. Let us assume that there is a true momentary intrinsic value associated with the momentary power load, following from the underlying benefits of consumption and costs of production. Further, let us envisage a pricing system in which the momentary dimension of the load profile is accurately priced. A candidate for such a hypothetical market system would be one in which the momentary power load itself was priced continuously in real-time, in other words a system operating with a continuous real-time market price function. This represents a continuous integrated real-time market. Such a continuous market system might also be visualized in an unbundled market. Here, the market system would be disintegrated into a planned, e.g. day-ahead, market, and a real-time market. The planned market equilibrium might be represented by specified continuous planned load-profiles for the period in question, together with the associated continuous price function. In real-time, deviations would be handled by the reserve capacity, and for example rewarded on the activation of the capacity by means of a continuous price function. This would in turn also represent the payments to be made by the participants who brought about the deviation.

These are both hypothetical scenarios, where the value of power load in both systems may be interpreted as the value of momentary supplied electricity, and represented by continuous market price functions. Such a pricing system, even if it were technologically possible, would in all likelihood be extremely costly and represent a severely complex system. As such it would, with today's technology, probably not constitute a cost-, nor an incentive-efficient market system, and also undermine the efficiency of market signals compared to a less accurate system.

Still, the nature of electricity is in fact continuous. While the continuous pricing system is not possible (nor desirable) in real-world electricity markets, a general approach to the continuous nature of electricity is thus to quote electricity in terms of energy, where the commodity is traded, registered and settled in terms of energy. Note that pricing systems based on the energy concept represent a discrete approximation: The commitment of energy to be supplied within each time period is a discrete approximation of the continuous load profile, while the energy prices which refer to given consecutive time periods may be interpreted as a discrete approximation to the (hypothetical) continuous price function.

Note, however, that even though the product in trade is dealt with in terms of energy, all production and consumption of energy is nevertheless in fact associated with a given continuous load-profile. To ensure a smooth functioning of the system, in principle, it would have been desirable to specify the load profile of every energy commitment. For example, an hourly energy commitment could be specified as e.g. a flat load-profile within the hour. Since all production and consumption, is registered and settled by the hourly amounts of energy, this, however, means that any deviations from the specified load profile cannot be controlled nor priced. The basic insight here is that the chosen time disaggregation of the market system in fact defines the finest possible measurement of the load profile, and thus represents the minimum amount of energy that can be controlled.

The choice of the minimum time period on which the markets and pricing systems are based upon, is thus a fundamental choice of the market design. Shorter settlement periods represent a finer approximation to the continuous load profile, and to the continuous price function. In principle, by continuously reducing the chosen standard settlement of the market, for example from hourly to half-hourly prices, from half-hourly to quarterly prices, or prices per five minutes, the energy market itself provides a closer approximation to the continuous load-profile of electricity, but at the same time represents an increasingly more costly market system.

A further implication is that any discrete approximation to the continuous commodity, alone is not sufficient to cover all crucial dimensions which are important to the efficient allocation of electricity resources. The market has to provide signals for an efficient allocation also in relation to the handling of the momentary balance of the system. Thus, the continuous commodity of electricity in real-world electricity markets has to be handled in a number of different pricing systems, each with different focuses, for example on planned vs. unplanned aspects, energy-related vs. momentary power-related aspects, etc. The chosen systems may be interpreted as systems that in different ways price the ex ante and the real-time values of the power load. This also illustrates that the normal comprehension of what is considered as the pricing and value of power load and energy dimensions, in fact is a product of how the market's pricing system is defined. Below we will discuss how different aspects of the Norwegian market system contribute to visualize the value of different dimensions of load-adjustment, power load, and energy.

4.4.2 Pricing Power, Energy and Ancillary Services in the Reformed Norwegian Electricity Market

Our focus here is on pricing systems within an unbundled electricity system, and in particular on how the value of power load adjustment properties and ancillary services, was reflected in the unbundled Norwegian system. We first consider planned aspects of the supply and demand of electricity, and proceed with pricing and allocation systems that handle momentary and stochastic aspects of the supply and demand of electricity.

– *Markets for Planned Energy:* Markets for planned energy deliverances are central to the unbundled electricity system. In principle these markets comprise organized spot markets, as well as organized and unorganized forward markets. The planned spot market is in effect a 'very-short-term' forward market, but is by convention termed a spot market. In Norway, the spot market is a day-ahead market, as described in chapter 2. The former delivery-oriented weekly market could be described as an organized forward market, while the trade of bilateral contracts is an unorganized forward market. Note also that the

Norwegian market is settled by the hour, implying that all trade is measured in terms of energy within the hour[85].

The spot market plays an important role in signaling significant aspects of the value of planned power load capacity. For example, prices are increasing in peak-load hours as increasingly more costly marginal production capacity is accepted in the market equilibrium of the day-ahead market. Likewise, prices are lower in periods with low power loads. As such the variation of spot prices from hour to hour also implies a power load related pricing, and the hourly based prices may be interpreted as a discrete approximation to the true value of the planned continuous load.

A question related to the spot market, as well as to the Norwegian energy market in general, concerns whether a larger integration with thermal systems will have implications for the optimal disaggregation of time in the Norwegian market. A change in this respect was already to be observed in the transition from the relatively isolated Norwegian market, to the integrated Norwegian-Swedish market. While the Norwegian market originally operated with spot contracts covering price sections, i.e. a given bundle of hours[86], from 1996, following the establishment of Nord Pool and the integrated market, the market was to be cleared on hourly contracts. Following a further integration with more thermally based systems, shorter periods for spot contracts, as well as for general physical settlement, could in principle prove more efficient. It might be that a finer approximation to the continuous power load, e.g. half-hourly prices, more accurately could signal the underlying variations of costs and benefits of the varying power load. In general the benefits of a closer approximation must be held against the costs of a closer approximation. For example, it might be that shorter settlement periods would be more beneficial in a thermal system than in a hydro-power based system, due to the higher costs in adjusting load in the thermal system, e.g. regarding both marginal costs, ramping costs, and other start/stop costs,. Awaiting the future development in the market, this was, however, left an open question at the time.

Before proceeding to other dimensions of the pricing system, let us also dwell a moment to discuss the implications for spot price formation in the Norwegian market of a closer

[85] Until 1996, however, the spot market, i.e. the daily market, operated with spot contracts that covered so-called ‘price section’, which consisted of one or several hours. The market price of the price section was thus the market price for each hour within the price section.

[86] Even though spot contracts covered several hours, note that the basic time unit of physical settlement was one hour.

integration with thermal systems. In principle, the cost of meeting planned peak-load in the thermal system tends to be more costly than in the Norwegian hydro power system. It would thus be expected that a more integrated market would enhance the prices in peak-load hours, thus contributing to a larger remuneration of power capacity also within the hydro-power system. Likewise, it may also be expected that higher intra-day variations may follow from a closer integration with thermal systems. As such, the energy markets by themselves provide a means of remunerating the power load capacity of the hydro-power system.

– *Planned Load Following:* Commitments of spot contracts and physical forward contracts specify the hourly committed production or consumption of energy. Due to the hourly specification of energy, the energy commitments are represented by discrete energy amounts for each hour. Consecutive hours with steadily increasing energy commitments, however, imply a steadily rising consumption. Though the commitments are specified as discrete jumps in energy, the real nature of demand will be a continuously rising energy output. To a large extent these variations take place systematically, and may be anticipated. It will then be efficient to plan production so that it follows the anticipated consumption. While the general delivery contract is specified as a standard load profile, arrangements may be made so that some generators supply load following services by following the anticipated rate of change in the power load. This might, for example, be specified as a commitment to supply the agreed energy amount, not as a flat load profile, but for example as a continuously rising load profile within the hour.

  Within thermal systems load-following may be relatively costly, while the costs of this service within the pure-hydro system in comparison are moderate. In power systems where the cost of load following is low, such as in a hydro-power system, it had not be custom to define predictable load following as an ancillary service, but rather more as an attribute of how the energy commitment is to be supplied. To our knowledge there only to a smaller extent were systems for pricing or remunerating these services. In other systems where load following is a costly service, it was more usual to include load following in the category of ancillary services. As such, planned load following services in principle represent an intersection between normal energy deliverances and ancillary services.

– *Secondary Regulation Reserves: The Regulation Power Market:* Aside from the anticipated power load variations, there will, however, still be unpredictable momentary variations in the produced and consumed power load. To balance the system, these variations have to be handled by means of ancillary services. The Norwegian regulation power market here plays an important role in handling *unplanned* load variations, including both deviations from plans and intra-hourly load variations. The supplied service in this market is defined as the supply of *secondary regulation capacity*, further specified as a capacity for load adjustment within a response time of maximum 15 minutes. The price mechanism was described in detail in chapter 2. To sum up the main issues, note that the supplier of such capacity places his bids in the market after the day-ahead market is cleared. The bids state the price at which the capacity can be activated. During the course of delivery the capacity is activated based on a merit order (with special rules in the case of congestion or other geographical considerations). The capacity holder is remunerated if activated, at the price representing the cost of the marginal capacity used, but does not receive anything if not activated. As such, the regulation power market also represents a market where power load capacity is priced, with a premium payment associated with the unanticipated and near-momentary nature of supply.

– *Other Ancillary Services:* At the time of our research, a greater part of ancillary services in Norway were procured by order of the system operator, and was in many instances only offered minor compensation. NVE (1994) specified the details of the responsibility for system reliability. Though Statnett as a system operator has the overall responsibility to supervise the system and to ensure smooth running operations, each plant owner also had a defined responsibility of contributing to a reliable system, with specified requirements as to e.g. technical equipment, voltage, active and reactive power, and energy production. These requirements were enforced through the grid-connection contracts with Statnett. The main principle was that the plant owner in general had the obligation to contribute with ancillary services, such as for example for frequency regulation and reactive power, only restricted by the technical specifications of their equipment. Furthermore, plant owners were only compensated to the extent that their contribution exceeded specified limits, and also if the plant owner incurred special costs or losses due to their obligation.

We have above provided a brief overview of pricing systems that at the time in different ways reflected the value of power-load related capacities and services. The Norwegian market reform took place at a time in which there was a relatively large abundance of such capacity. On one hand this implicitly implied that elaborate pricing mechanisms had previously not been necessary to ensure an efficient allocation of these resources. On the other hand, the market system, regardless of the abundance of the resources, still had required systems for handling the continuous aspects of electricity. We also saw that the power dimension in many ways in fact was reflected in the price structures of the market; - through varying intra-day hourly spot prices, - through the regulation power market inducing higher prices for regulation power than for ordinary planned spot commitments, - and through any ancillary service related contracts. Later on, the transition to a more internationally integrated market was, however, expected to imply a new market situation, with a greater scarcity of such resources due to a higher demand and willingness to pay for these services. It was to be expected that this changed market situation to some extent would be reflected in market prices, which partly would provide a larger remuneration of such services. An example of this would be a development with higher peak-load prices and higher regulation power prices. However, our concern was whether these market structures that were developed in a time of abundance, also in the new scenario would suffice to provide sufficient incentives for an efficient and reliable working electricity system in the short run, and to promote efficient investment strategies in the long run. This is the topic of the following section, where our focus is on alternative pricing structures for ancillary services.

4.5 An Alternative Price Structure for Secondary Regulation Power and Other Ancillary Services

As we have seen, the unbundled electricity market system consists of several different market and pricing systems, so also in the Norwegian market. The question now was whether the existing pricing systems would be able to contribute to an efficient resource allocation in a more power oriented market with a greater scarcity in the capacity for these services. In its full extent, the answer to this question would require an extended analysis of the market systems looking into the explicit and implicit treatment of the power dimension and covering questions of, for example, optimal time disaggregation, and the efficiency of trading mechanisms and price structures for energy, network capacities and ancillary services[87]. Here we limit our scope to a discussion on the pricing structure related to ancillary services in general, and to the supply of secondary regulation power reserve in particular. Section 4.5.1 discusses possible inefficiencies related to the pricing systems of ancillary services. Section

[87] This reflects that the efficiency, as well as the need for ancillary services, also follows from the interaction and efficiency of the other parts of the market system. For example, if the day-ahead market gives incentives for more accurate plans, and to adhere to these plans, the result may be lower need of secondary regulation capacity.

4.5.2 proposes a two-tier price structure for ancillary services. Section 4.5.3 discusses a representation of such a scoring rule as proposed by Robert Wilson. Section 4.5.4 discusses this structure in relation to the Norwegian hydro-power based electricity market.

4.5.1 Inefficiencies of the Current Price Structure of Ancillary Services

The provision of adequate system reliability is not possible to achieve by the market alone: Central control of the momentary balance and system quality is necessitated. As system reliability also is a public good, this implies that the optimal level of system reliability would not emerge by itself in a normal market setting. Here the system operator has the responsibility to estimate the optimal level, and to ensure that this level is adequately and efficiently supplied. To serve this role, the system operator has to procure the necessary services in a manner that contributes to the short-run cost-efficient provision of system reliability, while at the same time giving incentives for efficient investment strategies. Moreover, as this is a market with only one buyer, a basic prerequisite is also that efficient regulatory measures are in place. This requirement follows from its inherent characteristics of a natural monopoly. In the following we will assume that this framework is in place, and turn our attention to the mechanisms for the allocation of these ancillary services. As a first step, let us review a discussion by Rud and Tjøtta (1996) where the focus was on possible sources of inefficiency related to the price structure of the regulation power market in particular, and the pricing systems for ancillary services in general:

– *Short-term resource allocation of the regulation power market:* The short-term allocation of secondary regulation power is handled by the regulation power market, which is described in detail in chapter 2. Note that it is up to the supplier of reserve capacity to decide whether to offer his capacity in the regulation power market or not. By bidding and being accepted as a capacity holder, he is remunerated by the regulation market price if activated, but receives nothing if not activated. As argued in chapter 2, this per unit activated energy remuneration is normally higher than the per unit remuneration received in the spot market.

  If there is scarcity of capacity, the generator will offer his capacity in the regulation power market if this alternative represents the most profitable usage. The most likely alternative is to employ the capacity in the spot market. This means that a rational participant only will offer reserve capacity in the regulation power market if the expected profit of this market exceeds the profit of expected spot sales, and where also differences in the uncertainty of profit have to be considered; For spot sales, the expected profit follows from the expected spot price and cost of production, both of which are clearly defined. Likewise, for sales in the regulation power market, the expected profit of offering capacity in the market also follows from the expected remuneration less expected costs. However, the remuneration is here highly more uncertain and complicated to estimate: It is on one hand dependent upon the probability of being activated, which in turn is contingent on the probability that deviations will occur, the size of deviations, as well as the generator's bid and its implied place in the merit order list. On the other hand the remuneration is also dependent upon the activation price, which also depends on e.g. the size of deviation and the marginal accepted bid. The expected costs consist of the costs of holding the reserve in the required state of preparedness, and direct variable costs of activating the reserve.

As such, the participant does have incentives to channel capacity to the market which offers the highest expected profit. In a market with abundance of such capacity, we believe that it is probable that adequate capacity probably will be offered in this market. Likewise, it is reasonable to believe that normal competitive market forces in a tighter market to some extent will contribute to raising the premium of the regulation power market. However, given that there are costs of holding reserves, the one-tier price of the regulation power market must cover both activation and reserve holding. As the capacity is not always activated, this price structure may in some respects be likened with a lottery as to the remuneration of capacity holding. As the uncertainty of the regulation power profit in this sense is higher than the corresponding uncertainty of profit in the spot market, this risk of non-payment has to be taken into consideration in bidding, as well as in the allocation of resources amongst different usages. A basic question is here whether this aspect could constitute problems for the sufficient supply of capacity to this market in the short run.

– *Long-term resource allocation of the regulation power market*: Note that in making the capacity available, the participant supplies a service which represents a benefit for the system in terms of ensuring system reliability. At the same time, however, this represents a cost for the participant. If there are problems related to the sufficient supply to this market in the short run due to the uncertainty of remuneration, a likely implication is that the mechanism neither will induce sufficient investments in the long run. It should, however, be noted that investments in capacity with dual appliances, are triggered by price expectations in both uses.

– *Supply of non-energy oriented ancillary services:* At the time of our research, a greater part of ancillary services were procured by order of the system operator, in many instances offering none or only minor compensation. Given that system reliability is a public good, the requirements of the capacity holders to supply ancillary services can be interpreted as an allocation of the costs of system reliability, since the costs of normal ancillary services are charged the plant owners. However, this form of 'financing' may have adverse implications for investments in ancillary services, as the investor in assessing future investments will consider the supply of ancillary services partly as a cost, while it from a system point of view actually represents a benefit. This argument in effect implies that other financing forms should be considered, along with other systems for remunerating the supply of ancillary services.

These issues may indicate potential inefficiencies, both in the short-run with respect to the allocation to ancillary services versus alternative usages, and in the long-run with respect to the incentives for investments. For secondary regulation reserves, the regulation power market constitutes an important foundation and starting point for providing market-based incentives for ancillary services. The question raised was whether efficiency could be improved by employing price structures and pricing mechanisms that to a larger extent reflected the inherent nature of the supply of ancillary services.

4.5.2 Capacity Options: A Two-Tier Price Structure for Ancillary Services

In the quest for a more efficient price structure, let us turn to the underlying cost attributes of ancillary services. A common characteristic of ancillary services is that the capacity has to be held in preparedness, awaiting the need for the capacity to be activated real time. These characteristics are reflected in the costs of supplying these services:

- *Reservation costs:* In putting capacity at the disposal of the system operator, the generator provides necessary reserves to ensure the system reliability. In holding the reserve in the required state of preparedness, the plant owner incurs costs, such as for example extra energy costs, costs of non-optimal operation of the plants, extra workforce costs, and also costs in terms of the foregone profit from alternative usage of the capacity. The size of reservation costs, however, greatly vary for different production technologies, and are in general higher for thermally based technologies, than for hydro-power based generation.

- *Activation costs:* There are also direct costs of activating the reserves, as for example energy costs, extra costs due to non-optimal plant operation, extra wear and tear of the system, and start/stop costs. Also in this respect, costs vary greatly across different production technologies.

This clearly differentiated cost structure is important in defining a mechanism, where the objective is both to ensure an efficient allocation of reserve capacity, and an efficient handling of disturbances. Also note that the same differentiated structure can be identified in the benefits of the system reliability, where there are values associated with the preparedness of the system, as well as with the direct handling of disturbances.

This two-tier cost structure was neither reflected in the price structure of the current regulation power market, nor in the pricing systems for other ancillary services: The existing regulation power market employed a one-tier price, where the remuneration of capacity was related to the activation of capacity only. In a normal market, there normally is a clear correspondence between the price and the marginal cost. With a one-tier price, the activation price not only has to cover variable costs of activation, but also costs of capacity reservation. Under the one-tier price structure, it is difficult to accurately price these two important dimensions of ancillary services. For other ancillary services, we saw that in many respects, neither components of the cost structure were remunerated.

Following this, Rud and Tjøtta (1996) proposed that the contract for ancillary services took the format of a physical capacity option. The capacity option is based on the use of a two-tier price structure and mechanism for ancillary services:

- *A reservation price:* The reservation price is paid for the reservation of capacity, regardless of whether the capacity is activated or not.
- *An activation price:* The activation price is paid on the activation of the reserve capacity.

This two-tier price structure reflects the underlying structures of costs and benefits. The reservation price corresponds to the costs of holding the reserve in preparedness. The activation price corresponds to the costs of activating the reserve. Note that in this, the market prices also provide the system operator with better information on the costs of reserves and of activation, which is important information in planning levels and methods of achieving optimal system reliability.

The proposed contract for ancillary services may be termed a *capacity option*. This follows as the system operator in requiring the capacity for ancillary services, buys the right to activate the capacity as specified, but does not have any obligation to activate the capacity. The seller of the capacity option, however, has the obligation to be in preparedness for activation. In relation to the option framework, we see that the reservation price can be likened to the option price, and the activation price to the strike price.

In the spirit of the market-based electricity system, the basic idea is to implement this two-tier price structure in a market-based mechanism to promote an efficient supply of ancillary services. We here use the term ancillary service market in a broad sense. A market can to one extreme be a formalized organized market, such as the regulation power market, and to the other extreme a market with only a few suppliers, and implemented by more negotiated solutions. As such, the specific market design must be adapted to the characteristics of the specific ancillary service, the inherent competitive situation for the specific ancillary service, and the transaction costs of different systems. As aforementioned, there is one (regulated) buyer and in general several potential sellers. The main issue of markets for ancillary service is to bring competition upon the suppliers of these services, and by this provide a driving force both towards cost-efficiency, and towards an efficient allocation to the alternative uses of the capacities and resources.

In the following section our focus will be on the two-tier price structure, and the rules of bidding, pricing and selecting capacity for reserves and activation. Before proceeding to this, let us, however, first comment on some general aspects related to the organization of such a market[88].

- *Differentiation of contracts:* In principle the required ancillary services are highly differentiated. This implies that contracts should be differentiated by type of capacity, response time, duration and energy requirements, time of use, priority, and in some instances also its geographical location..
- *Contract period:* It is also highly likely that the contracts will be differentiated as to the contract period. For the regulation power capacity, which is a capacity that can be

[88] Note also that the resulting efficiency in the supply of ancillary services follows not only from the pricing mechanisms of ancillary services alone. It also follows from the interaction and efficiency of the other parts of the market system. For example, if the day-ahead market gives incentives for more accurate production and consumption plans, and incentives to adhere to these plans, the result may be lower need of secondary regulation power capacity. Such aspects have in part been discussed with respect to trading mechanisms and the timing of the spot and the regulation power market.

put to different uses, short contracts have been used, i.e. referring to specific hours of the following day[89]. For other ancillary services, it might be efficient to operate with more long-term contracts, for example referring to a year.

- *Control and sanctions:* Since the capacity options not necessarily are used, there are inherent incentives to wrongfully use capacity that should have been held as a prepared reserve. This implies that the contracts have to include specifications of control and sanctions.
- *Tailored price structure for different ancillary services:* The general form of the two-tier price structure in principle applies to all ancillary services. In relation to the individual type of capacity, it might, however, be that the weight given to the reservation part and the activation part of the price structure should vary. For example, for some forms of ancillary services, the costs of activation are negligible or difficult to measure or register, which thus implies that the reservation price will be the main factor. Here, it may be reasonable to implement the structure as a one-tier reservation-based price. For ancillary services where activation can be measured and registered, the full two-tier price structure might contribute to a better allocation, in the short-run as well as in the long run.
- *Trading mechanism:* Also the choice of trading mechanism should be tailored to different ancillary services, as the number of suppliers and potential of competition to a large extent varies across different categories of ancillary services. For example, where there are few potential suppliers, a market based on negotiations, or a system of putting the contracts out for tender might provide the most efficient solution. Where there are a large number of potential suppliers, such as in the case of regulation power, more formalized trading mechanisms could be efficient. Important elements of such trading mechanisms include bidding formats and market clearing rules both for the ex ante selection of reserves, and the real-time selection of which reserves to activate. A main objective in the design of this trading mechanism is to induce incentives that promote short- and long-run efficiency. An important issue in considering the efficiency of the trading rules is, moreover, the implied interaction with other markets, in particular the spot market.

[89] In principle also other contract periods might be considered for parts of the required secondary regulation power reserves.

In the next sections we focus on the two-tier pricing mechanism with respect to scoring and settlement rules, and implications of such a rule in a hydro-based power system.

4.5.3 The Chao and Wilson (2002) Scoring Rule

We have seen that the system operator requires a large range of different ancillary services covering a large range of time periods. In practice there may thus be a large number of simultaneous or consecutive markets for different ancillary services and for different time periods. The specifications of the ancillary service and specific auction or trading mechanism, will in principle have to be tailored to the characteristics of the given product, special geographical considerations, and the potential degree of competition. In addressing the issue of trading mechanisms, our focus will, however, here be on the similar features of ancillary services, rather than on the dissimilarities. In the following we will therefore consider a single general and stylistic ancillary service, in a competitive setting, and in this instance without regard to any special network considerations.

We have seen that the supply of ancillary services is a two-stage operation: The first stage is *capacity reservation*, in which the supplier of the ancillary service is committed to keep his capacity available during a specified time period. The second stage is *activation*, in which the capacity holder is required to activate his capacity by order of the system operator. This is in turn contingent on the running conditions of the electricity system. This two-part structure is mirrored in the proposed two-tier price structure, with a market price R^* for reserve availability, and a market price p_r^* for activation. Basically, the bid of the capacity holder *i* will also be a two part bid, (R^i, p_r^i), where R^i is the required compensation for holding the capacity available, and p_r^i is the per unit required compensation for supply of energy when activated, given that the capacity in question has been selected as a reserve capacity.

There are two fundamental elements of designing the procurement auction; the scoring rule and the settlement rule. The *scoring rules* specify, respectively, which bids should be selected for holding reserve capacity, and which of the selected reserves should be activated during the course of delivery period. The *settlement rules* specify how the selected capacities are remunerated, i.e. how the market prices of R^* and p_r^* are established. Both rules have important implications for the incentives of bid-making, including incentives to represent the true underlying costs, versus e.g. incentives of gaming. As such, the scoring and settlement rules are crucial to the efficiency of the market.

During the time of this research project, Robert Wilson visited the Norwegian School of Economics and Business Administration, and argued in favor of a set of simple, but incentive-compatible scoring and settlement rules. In table 7 we have summarized the scoring and settlement rule for the case of an incremental reserve, for example an incremental secondary regulation power service. The reader is referred to the later published article of Chao and Wilson (2002) for a more general representation, with details and proofs on the efficiency and incentive compatibility of this rule.

	Scoring Rule	Settlement Rule
Capacity Reserves	Selection of capacity reserves is based on a merit order ranking of the capacity reserve bids R^i.	Every accepted capacity reserve is paid the same price R^*, which is given by the lowest rejected capacity bid.
Activation	During delivery reserves are activated on a merit order ranking of their activation bids p_r^i.	Supplied energy through activation is paid the uniform per unit spot price p_r^* given by the bid of the marginally activated reserve.

Table 7 A incentive-compatible scoring and settlement rule for an incremental capacity reserve.

Candidates for the scoring rule would in principle be all combinations of the two parts of the bid. A main issue of the scoring rule of Chao and Wilson (2002) is the entire separation of the two parts of the bid. The selection of capacities that are to be held available, as well as the capacity settlement price, is based on the reserve capacity bids R^i only. Given the available pre-selected capacities during delivery, the activation of capacities to supply incremental power is likewise based on the activation bids p_r^i only, which also provide the basis for the settlement price p_r^*. A basic feature of these rules is that they encourage the supplier to bid according to his actual marginal costs. As such the rules contribute to avoiding the common fallacy of combining the two parts in the same scoring rule, which has shown to have important counter-efficiency incentive implications for bidding.

Note also that the main difference between the above procurement auction and that of the regulation power market, is the two-part nature of the procurement rule: The regulation power market is based on the energy bid and settlement only, though for this part, the scoring and settlement rules are basically the same.

4.5.4 A Discussion on the Chao and Wilson (2002) Scoring Rule in a Pure Hydro-Based Power System

The two-tier price structure based on the above scoring and settlement rules explicitly rewards the service of holding the capacity available on one hand, and the service of activating capacity on the other hand. Realized prices under such a pricing scheme will, however, depend on the fundamentals of the underlying technologies, as well as the specific market conditions. Let us now briefly consider factors which might affect price formation under the two-tier price scheme in a pure hydro power system. More specifically, we consider the ancillary services of secondary power regulation, illustrated in the case of capacity for increasing the momentary load. To focus on key aspects of the hydro power system, price formation is discussed in a setting of several simplifying assumptions, i.e. that both the spot market and the regulation power market are fully competitive, that there are no network constraints, and that the capacity holder in the short run perspective is risk neutral.

The participant has two alternative uses of his capacity, i.e. either to offer power in the spot (day-ahead) market, or to offer his capacity in the regulation power (real-time) market. For a given delivery hour the capacity can be sold in only one market. In principle the rational participant will wish to sell his capacity in the market which offers the highest profit. As such the bidding in each market will depend upon the sequence of the two markets. A stringent analysis of the bidding process and price formation thus would have to take this into consideration. Our intention here is, however, only to draw attention to more fundamental factors that might affect the value of reserve holding and activation. In this we take the simplifying hypothetical assumption that the markets are cleared simultaneously.

Let us now draw our attention to the reserve capacity and activation bid (R^i, p_r^i). It is reasonable to assume that in this setting, that the bid will reflect that the capacity holder requires at least the same profit as in the alternative of sales in the spot market. Furthermore, an important implication of the above scoring and settlement rules is that the participants may set both elements of the bid independently. This follows as the selection and pricing of capacity reservation only depends upon the submitted bids R^i, and likewise, the selection and pricing of activation only depends upon the submitted bids p_r^i. Let now us consider which factors affect each part of the bid, given the above assumptions.

The Activation Bid p_r^i

If the capacity has been selected as a reserve, the capacity holder receives R^*, regardless of whether the capacity is to be activated or not. In addition, the capacity holder receives p_r^* if the capacity is activated. The selection for capacity activation (from the reserved capacities), as well as the market activation price, is furthermore based on the activation bids p_r^i only. As such, the activation bid p_r^i is, according to the scoring and settlement rules, only relevant if the capacity has been selected as a reserve. Given that the capacity has been selected for reserve holding, the net profit of activation for a given participant, π_r^i, is $\pi_r^i = p_r^* - v_r^i - c_r^i$, where p_r^* is the market price of activation, found by the above scoring rule, v_r^i is the marginal energy cost of activation, and c_r^i are other marginal costs of activation. The assumption of a competitive market implies that the individual bid will not affect the equilibrium price. The net profit will thus be positive if $p_r^* > v_r^i + c_r^i$. Thus, under these assumptions, we have that the activation bid for capacity holder *i* will be $p_r^i = v_r^i + c_r^i$, implying a required non-negative profit to be activated. The question is thus what affects the two factors v_r^i and c_r^i for the hydro-power producers:

- *Marginal energy costs* v_r^i: In a hydro power system the marginal energy cost represents the value of the extra amount of water used in the activation of capacity for the supply of extra energy. Normally a hydro power generator will operate his turbines at an optimal level, which is at an operating rate less than at full capacity. At this optimal level, the amount of energy derived from each liter is maximized. In being activated, marginal energy costs will be the same as for spot sales if the producer still is able to operate on an optimal level. If the generator has to deviate from this optimal level, the energy produced per liter water is reduced, and the marginal energy cost of activation may thus be quite large, as the change in output intensity applies to the whole production during the time of activation[90].

- *Other marginal costs* c_r^i: Other marginal costs in relation to activation of the reserve are related to for example start costs. These costs are, however, expected to be relatively low in the hydro power system.

[90] Note, however, that if, as a starting point, the generator operates aside the best point, the implication is that marginal costs can be low, and even negative, if the activation causes the plant to operate at a more efficient level.

Though in principle the marginal costs of activation may be lower than in a thermal system, marginal costs of the hydro power generation plant may be higher than in normal planned production where the capacity is operated at its optimal state. In relation to the one-tier price structure of the regulation power market, with bidding after spot market clearing, we would expect that the same factors affect activation price bidding under the two price schemes. A main difference is that the capacity holder under the two-tier price scheme, if selected, is obliged to offer his capacity and submit bids.

The Reservation Bid R^i

The participant may offer his capacity in the spot market or as reserve capacity. It is reasonable to assume that the capacity holder will set his bid so that, if accepted, the expected profit is at least as high as in the alternative of spot sales. The expected profit of offering his capacity in the spot market may be represented as $E(\pi_s^i) = E\left[\max(p_s^* - v_s^i - c_s^i, 0)\right]$, where p_s^* is the spot price, $\mathrm{v}_\mathrm{s}^\mathrm{i}$ are variable energy costs in spot sales, and $\mathrm{c}_\mathrm{s}^\mathrm{i}$ are other variable costs, and where we have included the condition that the spot market sales only will occur for non-negative profits. The expected profit of offering the capacity in the regulation power market is $\mathrm{E}(\pi_\mathrm{R}^\mathrm{i}) = \mathrm{R}^* - \mathrm{C}_\mathrm{r}^\mathrm{i} + \mathrm{E}\left[(\tilde{\mathrm{p}}_\mathrm{r}^* - \mathrm{v}_\mathrm{r}^\mathrm{i} - \mathrm{c}_\mathrm{r}^\mathrm{i})\tilde{\mathrm{I}}\right]$, where R^* is the market reserve price, $\mathrm{C}_\mathrm{r}^\mathrm{i}$ are costs of holding the capacity in a state of preparedness, and $E\left[(\tilde{p}_r^* - v_r^i - c_r^i)\tilde{I}\right]$ is the expected activation profit, where $\tilde{I}$ is a stochastic 0/1 indicator function of whether the capacity is activated or not. The capacity holder will wish to offer his capacity in the regulation power market if $\mathrm{E}(\pi_\mathrm{R}^\mathrm{i}) > \mathrm{E}(\pi_\mathrm{s}^\mathrm{i})$. Under the above assumptions, this indicates a reserve capacity bid of: $\mathrm{R}^\mathrm{i} = \mathrm{C}_\mathrm{r}^\mathrm{i} + \mathrm{E}\left[\max(\mathrm{p}_\mathrm{s}^* - \mathrm{v}_\mathrm{s}^\mathrm{i} - \mathrm{c}_\mathrm{s}^\mathrm{i}, 0)\right] - \mathrm{E}\left[(\tilde{\mathrm{p}}_\mathrm{r}^* - \mathrm{v}_\mathrm{r}^\mathrm{i} - \mathrm{c}_\mathrm{r}^\mathrm{i})\tilde{\mathrm{I}}\right]$. There are three main factors of the reserve bid, i.e. $\mathrm{C}_\mathrm{r}^\mathrm{i}$, $\mathrm{E}\left[\max(\mathrm{p}_\mathrm{s}^* - \mathrm{v}_\mathrm{s}^\mathrm{i} - \mathrm{c}_\mathrm{s}^\mathrm{i}, 0)\right]$, and $\mathrm{E}\left[(\tilde{\mathrm{p}}_\mathrm{r}^* - \mathrm{v}_\mathrm{r}^\mathrm{i} - \mathrm{c}_\mathrm{r}^\mathrm{i})\tilde{\mathrm{I}}\right]$:

- *Reserve holding costs* $\mathrm{C}_\mathrm{r}^\mathrm{i}$: These are costs of holding the capacity in preparedness in accordance with the technical specifications of the given service. On one hand, there are payable costs, for example related to extra workforce needed. These costs are in general very low. On the other hand, the act of holding the reserve in a state of preparedness may induce energy costs. The level of these costs is highly dependent upon the imposed deviations from the optimal operation. If offering reserves can be done without diverging from the optimal rate of production, energy costs of holding reserves may be zero. In this case the plant is operated at its best point, which is less than 100% utilization of the available capacity, and reserves are available without any extra energy costs. However, if the holding of reserves means that the plant has to be operated at a non-optimal rate, energy costs of holding reserves available may be high. As such, this factor is highly dependent upon market conditions.

- *Alternative cost* $\mathrm{E}\left[\max(\mathrm{p}_s^* - \mathrm{v}_s^i - \mathrm{c}_s^i, 0)\right]$: If the participant has an alternative outlet for his capacity, the marginal alternative cost may be positive. If, however, optimal production plans state the same activity in the spot market regardless of offers in the regulation power market, this alternative cost is zero.
- *Expected profit of activation* $\mathrm{E}\left[(\tilde{\mathrm{p}}_r^* - \mathrm{v}_r^i - \mathrm{c}_r^i)\tilde{\mathrm{I}}\right]$: This is the profit from activation. Above we saw the bid is set so that this factor of expected profit is non-negative.

We see that the bid to a large extent depends upon the market situation, whether it is a period with abundant capacity, or a high-load period with relative scarce capacity. Using a stylized diagram, let us conclude this section by illustrating possible scenarios of market price formation in the situations of low and (very) high demand in the spot market. Figure 8 illustrates supply and demand in the spot market for a given hour. The spot market supply curve is given by S. Demand curves are depicted for two situations, D_L with low demand and thus abundance of capacity, and $\mathrm{D_H}$ shows demand in a situation of high demand and thus a relative scarcity of capacity. The figure depicts the market solution before the reserve requirement is imposed in the spot market, where the spot price, without any reserve requirements would respectively have been p_{sL}^* and p_{sH}^* given by the intersection of supply and demand. Let us now assume that a required reserve is imposed by Q^*, and consider the implications in both scenarios.

In the situation with low demand D_L, there is an abundance of capacity in the market, and the required reserve can be procured without any adjustments in the spot market. In this scenario it is reasonable to believe that the producers in supplying their normal spot and forward obligations will be able to operate their plants at an optimal level, thus implying that the reserve holding costs are low or even zero. If the expected profits of activation are high, this might even drive the reserve holding capacity bid to be zero, or even to be negative[91].

[91] If bidding for activation were restricted to the selected reserve capacities, an expectation of a positive activation profit could in effect imply that the reserve holding bid could be negative, as the selection also implied an exclusive right to bid. If bidding for activation is open for all participants, in

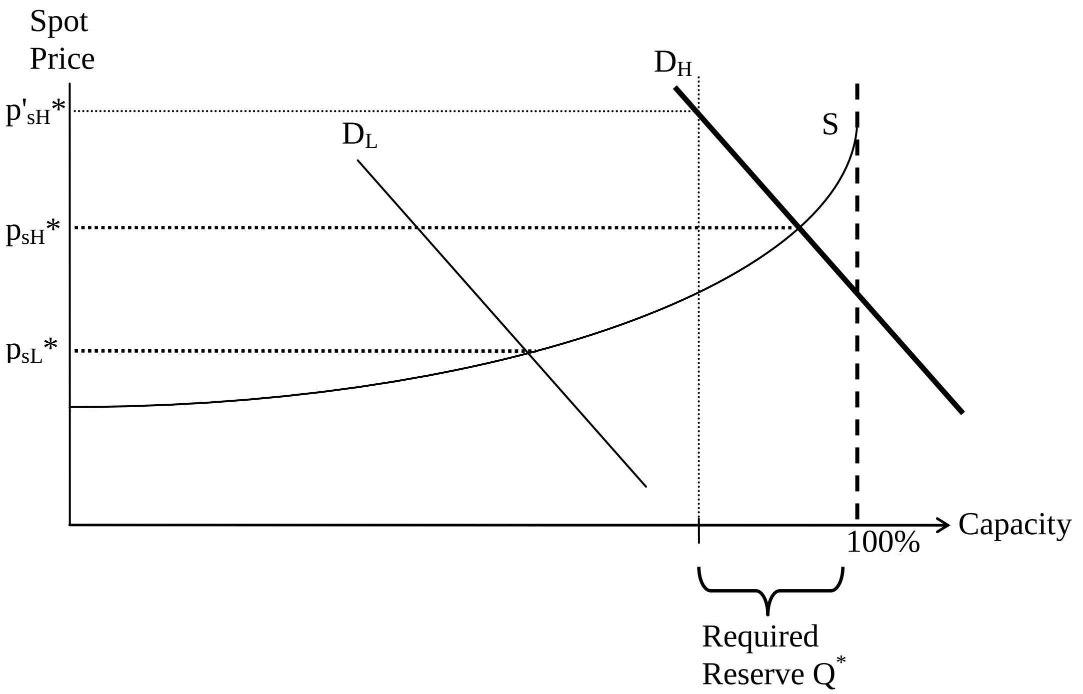


Figure 8: Spot market with reserve requirement and high and low demand

A possible low demand scenario is illustrated in figure 9 with the aggregate supply bids R_L and p_{sL}. The example shows a resulting market price R_L^* close to zero. The outcome is thus in effect the same as in the one-tier price structure regulation power market which has an implicit reserve price equal to 0. The main differences due to the two-tier price scheme thus are that the two-tier price structure explicitly visualizes that the zero price is a result of the abundance of capacity in the market, in addition to the fact that this scheme also obliges the selected capacities to hold their capacity in preparedness and submit bids for activation. As for the activation price, we have, somewhat arbitrarily, illustrated the bids as starting at the prior spot price, reflecting a required activation price premium relative to the spot market, for example due to higher costs of activation. This indicates an activation price p_{rL}^*. Different scenarios may be visualized for the activation supply curve, which we, however, believe to a large extent will reflect the same fundamentals of bidding as in the one-tier price structure scheme of the regulation power market.

addition to the participants that due to the selection were obliged to bid, the lower bound for the reserve capacity holding bid is zero.

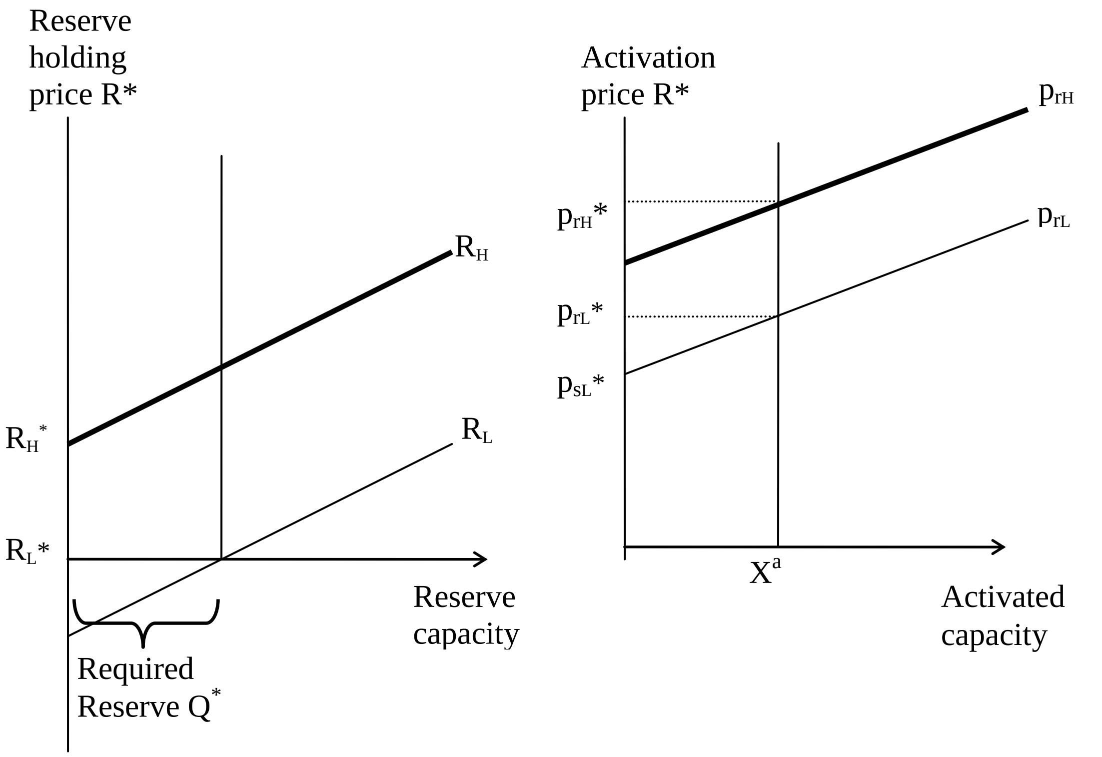


Figure 9: Regulation power market with reserve holding price and activation price, given high and low capacity utilization.

Now consider the situation of a market where capacity is relatively scarce. This is illustrated by the situation where demand for a given hour is represented by the high demand curve $\mathrm{D_H}$, while the underlying spot supply curve is as before. Note that the required reserve is not available in the form of production capacity if the spot market first is cleared at the indicated intersection of the demand curve $\mathrm{D_H}$ and the supply curve *S*. The unrestricted spot market clearing does as such not support the necessary reserve requirement[92]. Under our simplified assumption of simultaneous clearing of the two markets, the spot market has to be cleared under the restriction that sufficient capacity has to be withheld as a reserve capacity for momentary adjustments[93]. For example, if *S* represents the true marginal cost, this would indicate a spot price of p'^{*}_{sH}.

[92] An alternative or supplement would also be to include other forms of reserves, such as flexible consumption that can be reduced on a short notice.

[93] Though we will not discuss the topic of the sequential market clearing, let us briefly note that any chosen structure must secure the necessary reserve capacity. That is, if capacity is scarce, i.e. in high-load scenarios, it is necessary to clear the market for reserve capacity prior to the ordinary spot market, in order to obtain the sufficient amount of reserve capacity. This in particular applies to the reserve capacity holding part of the auction. However, given that the necessary capacity is withheld from the

Turning to the bidding for reserve capacity, we note that in this scenario of scarcity, the costs of holding reserves in a hydro power system can be high: The alternative costs of spot sales are high, as reflected by the high spot prices and scarce capacity. And, if power stations have to be operated at non-optimal levels to supply the necessary reserves, these factors may imply high costs of reserve holding. This is illustrated by the aggregate bid curve of R_H, with a resulting high reservation price R^*_H. Note, however, that this is but possible effects. A full analysis of price formation requires a more detailed knowledge of cost structures and market situations[94]. Turning to activation costs, if activation entails non-optimal operating levels, the costs and resulting activation price p^*_{rH} may, as illustrated in the figure be higher than in the in the scenario of abundance. Also here we, however, expect that the activation bidding to a large extent would reflect factors affecting the bidding within the current regulation power market[95].

Altogether, if a two-tier price structure were to be implemented, the main difference is related to the reservation of capacity and the explicit capacity reservation prices. We would expect these prices to reflect the available supply of reserve capacity, as well as the level demanded by the system operator. Higher requirements of system reliability, as well as an increased export of ancillary related services, may imply higher reservation prices. As such, a two-tier price structure would explicitly visualize the value of holding reserves, a value which in principle can be high in a more strained market, which might follow in a more integrated market. It is also expected to trigger the necessary supply of capacity reserves in the short run, also with incentives for investments[96].

spot market, the sequence of the spot market and the activation part of the reserve capacity auction may in principle be similar to the spot-regulation market sequence normally employed.

[94] For example, if by reducing spot market sales, reserve holding enables the producer to operate on more optimal levels, this would lower the reserve holding costs.

[95] For example, if the activation occurs at non-optimal levels, causing power stations to be operated on a more optimal level, the costs and resulting price p^*_{rH} may actually be lower. We have here not taken into account the measures like the later imposed requirement in the regulation power market that the incremental regulation power price bids have to exceed the corresponding spot price

[96] If reserve holding implies frequent operation of the system on non-optimal levels, and triggers further investments in power capacity in the power plants, a possible consequence may also be that normally sufficient reserves are available when running the system on optimal levels, with low costs and a low reserve capacity price, and where a one-tier price structure might be sufficient.

4.6 Current Market Systems for Ancillary Services[97]

System security and reliability is in Norway the responsibility of Statnett, the national transmission system operator. In this respect Statnett has the responsibility to ensure momentary generation/demand balance with a satisfactory quality of supply in the power system at all times. The Norwegian Water Resources and Power Board (NVE) enforces regulations related to power system operation[98]. As such, the system operator has to have adequate operational reserves at its disposal at all times. Here it is stated that the system operator shall to the greatest possible extent make use of measures which are based on market principles. Today several systems and agreements are in place to ensure the momentary balance of the power system:

For secondary power regulation, the regulation power market has been supplemented with various forms of capacity options which contribute to ensuring that sufficient capacity is available, as well as providing incentives for investments:

- *Regulation power market:* The regulation power market is utilized by the common Nordic market, and basically operates along the same basic principles as previously. See section 2.5 for an update on current organization, and the plans of a Nordic harmonization in the settlement of imbalances.
- *Regulation power capacity option market[99]:* This reserves option market was launched in 2000, in response to the concern that, due to the limited generation margin, there was a risk that all Norwegian generating capacity on winter week days might be sold in the day-ahead Elspot market to cover Norwegian demand and to be exported. The capacity options require that the reserves are held in preparedness, and offered in the regulation power market. This is a market for incremental power in the form of fast operating reserves. It includes generation capacity, as well as consumers that can

[97] The description of current (per summer 2008) systems is mainly based on material found on www.statnett.no.
[98] See NVE (2002) 'Forskrift om systemansvaret i kraftsystemet' (Regulations relating to Power System Operation) , Oslo, 17 May 2002, www.nve.no.
[99] In Norwegian: 'Regulerkraftopsjonsmarked'. Market rules are stated in Statnett (Oct 19 2005): 'Vilkår for tilbud, aksept og bruk av regulerkraftopsjoner i produksjon/forbruk' (Terms and Conditions for Offering, Acceptance and Use of Regulating Capacity Options for Production and Consumption). Descriptions and comments on the system may for example be found in Borgen, H. and B. Walther (2006): 'Effektregulering i det norske kraftsystemet', Elektro, no. 4, April 2006, and in Walther, B. and I.H. Vognlid (2004): 'Statnett's Option Market for Fast Operating Reserves', ieadsm Demand Response Dispatcher, Vol. 1 (6), April 2005.

reduce their consumption within the 15 minute required response time. As such, the market has also resulted in a substantial volume of flexible demand to compete with generation.

The market is operated on a weekly basis only in wintertime (November – March)[100]. Prior to the Thursday deadline of bidding, participants place bids stating the required capacity option price and the power capacity they are willing to offer at this price. Statnett systemizes the offers and analyzes the amount to be purchased in each area. The cheapest offers are chosen up to the desired volume. Other factors influencing the selection may be network conditions and the different regulating qualities of generation and demand. The price of the marginally (most expensive) accepted offer in the area defines the price of all contracts in the area. Thursday at 1400 Statnett posts information about the concluded purchase for the next week. Accepted capacity is thus rewarded the regulation capacity option price. The following week the contracted reserves have to be offered in the regulation power market on all weekdays between 05 and 23, with hourly activation price bids determined by the capacity owner. The regulation power market is thus cleared as usual, based on these bids alongside bids from other participants.

- *Longer-term capacity options:* In addition to the weekly capacity option market described above, Statnett also has entered into bilateral agreements of 5-10 years duration with generators. An underlying intention is to contribute to the rehabilitation of old units and increased size of new units to be installed.

For other ancillary services, remuneration is on one hand based on negotiations between the Norwegian Electricity Industry Association (EBL) and Statnett, specifying the rate of remuneration for different ancillary services such as e.g. load following, reactive power, etc[101]. On the other hand, more market-based systems have been developed:

- *Market for frequency-regulated reserves[102]:* This market started operating January 2008. The products exchanged on this market are power reserves which are

[100] From the beginning the contracts referred to a 3-month or 1-year basis, subsequently on a 1-month basis, and from Fall 2004 on a weekly basis. The main reason for a shorter contract period was to reduce the uncertainty of participants when offering the capacity, and to enable Statnett better to estimate and tailor the purchase to the need of capacity.

[101] See e.g. 'Protokoll fra forhandlinger mellom Statnett SF (Statnett) og Energibedriftenes landsforening (EBL) om godtgjørelse for systemtjenester', available on www.statnett.no.

[102] In Norwegian: 'Marked for frekvensstyrte reserver'. See 'Vilkår for tilbud, aksept, rapportering og avregning i Marked for frekvensstyrte reserver', Statnett Jan 21 2008, available on www.statnett.no.

automatically activated on specified variations in frequency. The contracts consist of a minimum base delivery, and a market-based delivery. Participants submit bids in the market, which is operated both on a weekly basis with contracts for specific time periods of the coming week (i.e. weekday-day, weekday-night, weekend-day, weekend-night), and on a daily basis with 24 hourly contracts for the next day. The main market clearing criteria is price, and the market price is set to be the highest accepted bid within each area.

- *Market for spinning reserves*[103]*:* The objective of this market is to secure the operator access to spinning reserves in the 15-minute period prior to planned changes in production levels in the system. More specifically, these reserves are spinning reserves that may be activated on a 1 minute response time at any point of time within 15 minutes before the generator has planned to increase his production. On signaling their interest to offer these reserves, the suppliers are remunerated according to specified pricing formulas if activated.

The above services refer to ancillary services for handling the ordinary momentary balance of the system. As a system operator Statnett is also responsible for the security of supply of the system, and has to develop measures for handling severely scarce energy situations (SAKS), i.e. in the advent that the market alone is not able to handle the situation[104]. Implemented measures in this respect are:

- *Consumption energy options:* The objective is to lower the probability of rationing, by reducing consumption in a SAKS situation. Participants submit two-part bids, where the option price bid states the necessary annual compensation to offer this flexibility, and the activation price bid states the price on which the option is to be activated, i.e. when the consumer has to reduce his consumption. Activation may only occur if the probability of rationing is higher than 50%.

- *Reserve generation power:* Further, Statnett has purchased reserve generation power, which is to be used only if the probability for rationing is severely high.

[103] In Norwegian: 'Marked for innfasingsreserve'. See 'Vilkår for tilbud, aksept, rapportering og avregning for Innfasingsreserve', Statnett, Nov 9 2007, available on www.statnett.no.

[104] The framework for establishing measures for such situations ('Svært anstrengte kraftsituasjoner (SAKS)') is given by St.meld.nr,18 (2003-2004) on the security of supply. Measures are to be approved by NVE.

5 Pricing Network Capacity in a Market-Based Electricity System

Above we have covered three of the main pricing/market systems that constitute the fundamental building blocks of the unbundled electricity market, and where each of them have their special function: The spot market is a day-ahead market with a focus on planned energy supply and demand. The main focus of the regulation power market and other pricing systems for ancillary services is on handling the continuous and momentary stochastic nature of supply and demand. The futures market has its focus on risk allocation. With the reliance on a common capacitated grid network, the overall efficiency of the electricity market also depends upon the efficient pricing of network facilities. The fourth main building block of the electricity market, thus comprises the pricing and allocation systems of network capacity, and will be the topic of this section.

Due to the common reliance on the electricity grid for transportation, the efficiency in the allocation of electricity production and consumption is inextricably intertwined with the efficient management of the grid capacity. Our main focus will mainly be on congestion management. It is here necessary that electricity markets and grid management interact, as any congestion only can be relieved by altering production and/or consumption plans. This chapter mainly serves as a brief introduction to the market design aspects of pricing systems for network capacity, and contains a brief sample of the issues of congestion management. The chapter is based on Jørnsten, Rud and Singh (1997). An important issue in infant stages was to understand and integrate network aspects in the economic modeling of the electricity market. First, however, section 5.1 briefly comments on the general tariff structure. The main part of this chapter is section 5.2 which provides an introduction to the modeling of how market dispositions are affected by the network. The section also presents a first basic view on efficiency implications of congestion management by zonal pricing. Section 5.3 concludes by briefly commenting on current Nordic methods of handling congestion.

5.1 General Tariff Structure

This section briefly reviews main principles of the pricing of network services and the main structure of network tariffs. In effect, network tariffs specify how the transportation of electricity is priced. The underlying pricing principles for electricity transportation are in effect crucially different from normal freight pricing, which often may be based on a per-unit cost reflecting the marginal cost of transportation, and where the price often is related to the transport distance. In contrast electricity exhibits several special features, with special implications for transportation pricing in this sector:

- *Sub-additive cost function:* The cost function of transmission and distribution may be characterized as a sub-additive function, with marginal costs lower than average costs.
- *Loop-flow features:* Power injected to in the network may broadly be characterized as a pool of energy, where the actual power flow in the grid follows from physical laws and the overall monitoring of system balance. These physical relations are important in defining tariffs as well as handling congestion, and are closer described in section 5.2. Here we note that the path of a given transaction is not determinable, and that the resulting power flow follows from the entire power flow of the grid, rather than the attributes of the individual power transaction.
- *Capacity constraints and congestion:* The network has a given capacity. This capacity cannot be exceeded, as any 'overload' would result in an immediate blackout. The term network congestion refers to the scenario of which the planned net demand implies a power flow for a given transmission line which exceeds its actual capacity. Measures must be taken so that an actual 'overload' does not occur.

To sum up, and without going into detail, the main implications of these features for constructing tariffs for transmission and distribution are:

- *Two-part tariff:* With a per-unit cost equaling marginal costs, revenues would not cover total costs of transmission. Furthermore, a per-unit price which is higher than marginal costs, might unduly discourage use of the grid and is thus inefficient from an economic standpoint. The implication of this is the use of a two-part tariff; a fixed fee with the main function of covering costs, and a per-unit fee to cover marginal costs such as losses.
- *Point tariff:* Due to the loop-flow features, the marginal cost of a given transaction between a given seller and a given buyer is not related to the distance or specific connection between these two parties. The implication is here that the per-unit part of the tariff should be independent of distance.
- *Congestion fee:* An objective of congestion management is to ensure an efficient electricity allocation which is feasible within the given network. Demand for transmission is derived from the underlying energy demands and supplies. Congestion must thus be handled by adjusting demand or supply. An efficient handling of

congestion thus implies that changes in demand and supply must be in accordance with the underlying willingness to pay or marginal production costs. In broad terms, the transmission tariff thus has to incorporate some form of a congestion fee. A further discussion of congestion management is contained in section 5.2.

Following the market reform, there was a thorough revision of the old tariff system presenting a new structure based on new pricing principles. The new tariff structure reflected these three main ingredients. The main change was that the former ‘channel’ charges were replaced by so-called *point charges*. Here the charges solely depend on the volume and place where power is being fed into or tapped from a connection point in the grid, and not on the (artificial) notion of the path taken. Each agent is charged for the transportation tariff by the owner of the grid to which he is connected, and the grid owner only pays the connection charge to the grid to which he in turn is connected. For example, a distribution company pays for its connection to the regional grid, and a regional grid owner to the main grid. The transmission tariff after the market reform (1992) consisted of the following parts:

a) *Energy charge:* The energy charge is a variable charge and consists of two parts: The *marginal loss rate* for feeding and tapping the central grid was in principle differentiated according to low-load and peak-load periods of the week, and summer and winter periods across the year. The *congestion charge* was placed on all power in bottleneck situations. This was a market-based positive (negative) charge on demand (supply) on the deficit side of the bottleneck, and a negative (positive) charge on demand (supply) on the surplus side.

b) *Connection charge:* This is a fixed charge, which is not dependent upon the amount of energy delivered or supplied on the net. The basis for calculating the charge was the sum total of generating capacity and demand in a situation of a national peak load behind every connection point in the central grid system. This charge also applied to grid owners in other levels, where the grid owner pays the connection charge only to the grid to which he is connected.

c) *Power charge:* This too was a fixed charge, not dependent upon the varying energy amounts transferred. The basis for calculating this charge was the measured power exchange with the central grid in a situation of a potential peak load, adjusted for available unused generating capacity. As for the connection charge, this also applies to grid owners of regional or distribution networks connecting to the transmission grid.

5.2 Congestion Management

The transmission grid connects the various regions of the entire country. The capacity of the grid is in the short-run fixed. In periods of high demand the capacity may constrain the power flow as to what is possible to deliver. Prior to the market reform, priority was connected to the type of contract underlying the transaction. Firm power contracts had priority in the network, while the delivery of power purchased on the occasional market was stopped in the case of congestion. In this respect, the reorganization of the power markets also changed this principle. Spot market transactions were given full priority, though as we shall see below, the market equilibrium was adapted with geographically differentiated prices to create a feasible power flow. Further, the fulfillment of bilateral contracts in the case of congestion had to be met through spot market transactions. For a holder of a bilateral contract with obligations in an area behind a capacity constraint, any obligation had to be met by conducting transactions in the spot market.

In principle, a basic understanding of the underlying physics of the power flow, as well as the interaction of power trade and grid management, is essential to designing an overall efficient electricity market system. This applies to the design of all market systems for the pricing and allocation of power, as well as to the design regulation policies that affect short and long run strategies related to grid management, production, and consumption. Moreover, it is important to incorporate such effects in economic models of the electricity market to fully assess the implications of different policy measures. Here it became custom to model an approximation of the true power flow by means of a simplified 'DC' network. This method captures the main effects due to the loop-flow features of the network, while it at the same time renders a manageable model where it is possible to model and analyze the main economic features of market design and policy analysis.

This section provides a sample of such modeling, based on Jørnsten, Rud and Singh (1996), which represents one of our first steps in incorporating the features of electricity transmission in market design. With the example related below in section 5.2.2, we seek to give an insight into the intrinsic features of congestion management. First, however, the following section briefly summarizes the basics of the power flow within the simplified 'DC' network.

5.2.1 Power Flow Equations

Normal power networks are alternating current (AC) networks. A direct representation of power flows within this network involves a rather complex modeling. The direct current (DC) network is a special instance of the AC-network, but still involves a rather complex representation. A further simplification is the simplified 'DC' network, which allows us to work with power variables directly, and provides a more manageable model[105]. It is further convenient to assume that all network lines have identical technical characteristics, with impedances equal to 1 and with no losses. This is also the basis for our modeling of the power flow. We will now review the power flow equations within this simplified 'DC' network.

To illustrate, let us now consider the following classic three-node network, with production in node 1 and 2, and consumption in node 3. All nodes are connected, thus there are three lines in the network which is illustrated in figure 10.

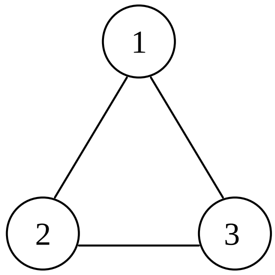


Figure 10   Three-node electricity grid

Let x_i be the net injection of power in node *i*, where net production is denoted by $x_i > 0$, and net consumption by $x_i < 0$. The power flow on line *ij* from node *i* to node *j* is denoted x_{ij}. If positive, the flow is in the direction from *i* to *j*. If negative, the flow is in the direction from *j* to *i*. Thus, we have that $x_{ij} = -x_{ji}$. The power flow along each line is given by three physical laws. The equations for our three node network, with a somewhat simplified explanation are:

- *Kirchhoff's junction rule* states that the input to and output from each node has to balance. In our example with three nodes this implies that[106] $x_1 = x_{12} + x_{13}$, $x_2 = x_{21} + x_{23}$, and $x_3 = x_{31} + x_{32}$.
- *Kirchhoff's loop rule* states that the power flow along all paths between two nodes is equal. In our simple grid, there are only two possible paths. We can find these paths by considering the flow of energy from node 1 to node 3: The power can either flow from node 1 to node 3 directly, or it can alternatively flow from node 1 via node 2 to node

[105] The interested reader is referred to Bjørndal (2000) which gives a simplified overview of AC, DC and 'DC'' power flow analysis, also giving further references to read.
[106] Equivalently, $x_1 = x_{12} + x_{13}$, $x_2 = -x_{12} + x_{23}$, and $x_3 = -x_{13} - x_{23}$.

3. With our assumption of impedance equal to 1, the equation for this rule is given by $x_{13} = x_{12} + x_{23}$.

- *The Law of conservation of energy* states, in our example with no losses, that no energy disappears. Thus consumption equals total production, or $x_1 + x_2 + x_3 = 0$.

By solving these equations, the power flow along each line is found, and may be expressed in terms of the net injection to each node: $x_{12} = \frac{1}{3}x_1 - \frac{1}{3}x_2$, $x_{13} = \frac{2}{3}x_1 + \frac{1}{3}x_2$, and $x_{23} = \frac{1}{3}x_1 + \frac{2}{3}x_2$. For example, with all demand in node 3, a production of 90 in node 1, and no production in node 2, the flow along the lines are $x_{12} = 30$, $x_{13} = 60$, and $x_{23} = 30$.

5.2.2 Congestion Management

Congestion is in principle alleviated by reducing the net demand in the deficit area, and increasing net demand in the surplus area. The challenge is to create a system that induces an optimal allocation of power. Several methods have been used, or proposed, such as nodal pricing, zonal pricing, financial transmission rights, flowgate rights, counter purchases, etc. Not all methods will achieve the most efficient market solution. A method which is theoretically optimal, but not necessarily efficiently operable, is that of *nodal pricing*. In the pure nodal pricing method prices are differentiated by location and represent the shadow price of injection or withdrawal in each node. A higher (lower) price in a node reduces (increases) demand and increases (decreases) production, thus relieving congestion.

In the Norwegian system prices are geographically differentiated in the spot market to obtain a planned power flow which is feasible within the capacities of the network. The congestion fee arrangement implemented in the daily market is a simplified version of pure nodal pricing, in the sense that prices are differentiated, not by the node, but by zones which are an aggregation of several nodes. The system was taken over from the former Norwegian Pool for Occasional Power, and applied to all power traded on the organized markets of Statnett Marked.

An important development in the economic understanding of the interaction of the spot market and the grid in the treatment of congestion, was to include simplified 'DC' models of the network in economic electricity market models. One of our first approaches is represented by Jørnsten, Rud and Singh (1996). Among other issues, this working paper takes a first look at the method of zonal pricing used in the Norwegian market. In the below example, we first illustrate that the efficiency of the zonal pricing method may depend upon how the zones in fact are defined. Furthermore, the example also illustrates the non-optimality of the zonal pricing method. Inspired by the work of Wu and Varaiya (1995), the example shows how the efficiency of the market may be improved by letting participants enter into multilateral contracts after the zonal pricing solution is known. In this, the example also highlights fundamental aspects in the integration of market and network considerations.

Unconstrained Three Node Market Model

We now consider a simple three node market, where all nodes are connected by lines, as illustrated in figure 10 above. In each node there is supply as well as demand, all of which are represented as linear demand and supply curves. The supply curves of node 1, 2 and 3 are given respectively by $p_1 = \frac{1}{10} x_1^s$, $p_2 = \frac{2}{10} x_2^s$, and $p_3 = \frac{4}{10} x_3^s$, where p_i is the price in node *i*, and x_i^s is the supply in node *i*. The demand curves in all nodes are the same, and given by $p_i = 20 - \frac{5}{100} x_i^d$. In the non-capacitated market, market equilibrium is given by equating aggregate supply and demand, and clearing the market on a common price. Table 8 shows the market solution for this example[107]. In the absence of network constraints, the price is the same in all nodes. The demand and supply surplus is given by normal economic definitions[108], and line flows follow from the above equations following from the laws of Kirchhoff.

	Price	Supply	Demand	Net Injection	Producer Surplus	Consumer Surplus	Social Surplus	Line	Flow
Node 1	15.48	154.8	90.3	64.5	1198.75	203.95	1402.71	1-2	25.8
Node 2	15.48	77.4	90.3	-12.9	599.38	203.95	803.33	1-3	38.7
Node 3	15.48	38.7	90.3	-51.6	299.69	203.95	503.64	2-3	12.9
Total		271.0	271.0	0.0	2097.81	611.86	2709.68		

Table 8 Market solution with no line constraints

[107] All numbers of the following tables are based on the accurate prices found in solving the market equilibrium. The represented numbers in the tables are abbreviations of the full numbers, implying that the represented numbers and sums of the table do not add directly. It may, though be checked that the represented numbers are consistent with the following prices: Table 8 and table 11 are based on $p_1 = p_2 = p_3 = 15.48387097$, table 9 on $p_1 = 15.410344828$ and $p_2 = p_3 = 15.72413793$, table 10 on $p_1 = p_3 = 15.27272727$ and $p_2 = 15.92727273$, and table 12 on $p_1 = 15.18584071$, $p_2 = 15.82300885$, and $p_3 = 15.50442478$.

[108] Producer surplus is given by the sales revenue, less the cost of production. For producer *i* this is $PS_i = \frac{1}{2} p_i x_i^s$. Consumer surplus is defined by the area under the demand curve, less the cost of purchase. For consumer *i* this is $CS_i = \frac{1}{2}(a_i - p_i) x_i^d$, where a_i is the intersection of the demand curve. Further, in the case of congestion, there is a grid surplus given by $GS = \sum p_i x_i^d - \sum p_i x_i^s$. If there is no congestion, this grid surplus is 0. Social surplus is found as the total of all producer, consumer and grid surplus, i.e. $SS = \sum_i PS_i + \sum_i CS_i + GS$.

Constrained Capacity: Zonal Pricing

Let us now assume that the maximum capacity of line 1-2 is 20.0. The above market solution which implies a flow of 25.8 on this line, is thus not feasible. The implemented Norwegian method of handling congestion is that of *zonal pricing*. In this method the market is split into zones that separate the market on the two sides of the capacitated line. In our three-node network, there are three possible ways to split the market, as illustrated in figure 11. With zonal division A, node 1 constitutes a zone alone, while nodes 2 and 3 constitute the other zone where both share the same market price. With zonal division B, nodes 1 and 3 constitute one zone with a common market price, while the other zone comprises node 2 alone. Note that the illustrated zonal division C does not represent a workable division of the market, as it does not give any control with the capacitated line.

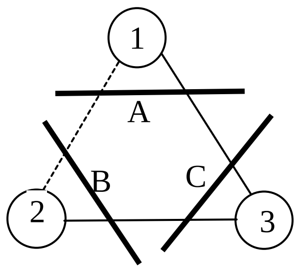


Figure 11   Zonal divisions of the network

The resulting market solutions for zonal divisions A and B are shown in tables 9 and 10, and are found as the allocation that balances the market under the restrictions that the line flow over line 1-2 at the maximum is 20, and that the prices of nodes within the same zone are equal. Note that when prices in different nodes differ, the payment paid by consumers exceeds the payment received by the producers. The difference is the grid surplus.

	Price	Supply	Demand	Net Injection	Producer Surplus	Consumer Surplus	Grid Surplus	Social Surplus	Line	Flow
Node 1	15.10	151.0	97.9	53.1	1140.57	239.76	20.20	1400.53	1-2	20.0
Node 2	15.72	78.6	85.5	-6.9	618.12	182.83	1.66	802.61	1-3	33.1
Node 3	15.72	39.3	85.5	-46.2	309.06	182.82	11.10	502.99	2-3	13.1
Total		269.0	269.0		2067.75	605.42	32.96	2706.14		

Table 9   Market solution with zonal division A, when line 1-2 is constrained to 20

	Price	Supply	Demand	Net Injection	Producer Surplus	Consumer Surplus	Grid Surplus	Social Surplus	Line	Flow
Node 1	15.27	152.7	94.5	58.2	1166.28	223.47	12.28	1402.04	1-2	20.0
Node 2	15.93	79.6	81.5	-1.8	634.20	165.87	0.81	800.87	1-3	38.2
Node 3	15.27	38.2	94.5	-56.4	291.57	223.47	-11.90	503.14	2-3	18.2
Total		270.5	270.5		2092.05	612.81	1.19	2706.05		

Table 10   Market solution with zonal division B, when line 1-2 is constrained to 20

The example shows that both zonal divisions in fact are able to handle the problem of the restricted capacity of 20 on line 1-2. However, the social surplus differs in the two alternatives, where the social surplus of zonal division A is marginally larger than the social surplus of zonal division B. This illustrates that the efficiency of the zonal pricing congestion method is dependent upon how the zones are defined. Also note that, though alternative A is the most efficient solution of the two, i.e. with the highest social surplus, the two alternatives have different allocation effects. For example, we see that producers in node 1 and 2, and consumers in node 3, would be better off if the zonal division B had been chosen (assuming no further allocational measures).

Constrained Capacity: Zonal Pricing + Multilateral Contract

Though the capacity of line 1-2 is fully utilized under the zonal pricing scheme, let us now argue that the social surplus of the market can in fact be increased without violating the capacity constraint. We will motivate why and how an additional multilateral trade between the participants, can increase the efficiency of the zonal market solution. At the same time, this also illustrates that the congestion method of zonal pricing does not render the optimal market solution.

To see this, consider zonal division A, which was the above zonal alternative with the highest social surplus. Let us now consider how the social surplus can be increased by allowing multilateral trade after the market is cleared by zonal pricing. The reason why this is possible follows from two features:

- First, note that the above zonal solutions can be said to represent an imbalance in the total market, due to the capacity constraint: In the resulting supply and demand plans in node 1, we have an equilibrium where marginal costs equal marginal benefits which in turn equal the market price of 15.10. However, producers and consumers in nodes 2 and 3 have adapted to a market price of 15.72, which implies that marginal costs and benefits are 15.72. As the marginal benefit in nodes 2 and 3 is higher than the marginal cost in node 1, this indicates that additional trade is desirable.

- Secondly, note that it in fact is possible to increase the transport along the network, while still observing the restrictions implied by the laws of Kirchhoff: The line flow in alternative A is given by $x_{12} = 20.0$, $x_{13} = 33.1$, and $x_{23} = 13.1$, where the capacity of line 1-2 is fully utilized. Let us now see what happens if we produce an unit extra in node 1, to be consumed in node 3. Unfortunately, this would result in an extra flow on line 1-2, i.e. $\Delta x_{12} = \frac{1}{3}$, which is not possible[109]. However, if we in addition produce an extra unit in node 2, this induces a counter flow on line 1-2, i.e. $\Delta x_{21} = \frac{1}{3}$, or equivalently $\Delta x_{12} = -\frac{1}{3}$[110]. Together the two flows net each other out, thus without violating the capacity constraint. This implies that production and consumption in the market can be increased, and transported in the network with the given constraints. For example, an increase of production with one unit in node 1 and one unit in node 2, together an increased consumption of 2 units in node 3, would in our example give the following net nodal injections of $\mathrm{x}_1 = 54.1$, $\mathrm{x}_2 = -5.9$, and $\mathrm{x}_3 = -48.21$. The corresponding line flows are $x_{12} = 20.0$, $x_{13} = 34.1$, and $x_{23} = 14.1$.

Let us now assume that the consumer in node 3 desires to buy an extra unit from the producer in node 1. The transaction is likely to be profitable as the marginal benefit of 15.72 exceeds the marginal cost of 15.10. The catch is, however, that this transaction alone would violate Kirchoff's law and is therefore not feasible. This bilateral contract can thus not be executed. However, as we have seen, more power can be transported in the network, in this case if the production in node 2 is increased at the same time: It *is* feasible to transport the combined extra production and consumption package of $\Delta x_1 = 1$, $\Delta x_2 = 1$, and $\Delta x_3 = -2$, i.e. where producers in node 1 and 2 both increase their production with one unit, and the consumer in node 3 consumes two extra units. Neither of the two transactions alone are feasible to transport, nor is the transaction between node 2 and 3 profitable alone, since the marginal cost for an extra unit exceeds the marginal benefit of this unit. The solution is to initiate a multilateral contract, i.e. a tri-lateral contract between the three parties. Let us now consider whether such a transaction is profitable.

Table 11 sums up the effect of the extra trilateral agreement between the producers in node 1 and 2, and consumers in node 3. The obligations of the original market clearing with respect

[109] For the additional transaction we have $\Delta x_1 = 1$ and $\Delta x_3 = -1$. The Kirchhoff equations (omitting one superfluous equation, e.g. $\Delta x_2 = 0 = -\Delta x_{12} + \Delta x_{23}$) for the additional flow are $\Delta x_1 = 1 = \Delta x_{12} + \Delta x_{13}$, $\Delta x_3 = -1 = -\Delta x_{13} - \Delta x_{23}$, and $\Delta x_{13} = \Delta x_{12} + \Delta x_{23}$. The solution is the resulting flow of $\Delta x_{12} = \frac{1}{3}$, $\Delta x_{13} = \frac{2}{3}$, and $\Delta x_{23} = \frac{1}{3}$.

[110] For this transaction we have $\Delta x_2 = 1$ and $\Delta x_3 = -1$. The Kirchhoff equations (omitting one superfluous equation, e.g. $\Delta x_1 = 0 = \Delta x_{12} + \Delta x_{13}$) for the additional flow are $\Delta x_2 = 1 = -\Delta x_{12} + \Delta x_{23}$, $\Delta x_3 = -1 = -\Delta x_{13} - \Delta x_{23}$, and $\Delta x_{13} = \Delta x_{12} + \Delta x_{23}$, with the resulting flow of $\Delta x_{12} = -\frac{1}{3}$, $\Delta x_{13} = \frac{1}{3}$, and $\Delta x_{23} = \frac{2}{3}$.

to both prices and quantities stand firm, and are the same as shown above in table 9. The tri-lateral agreement we are considering is a package of $\Delta x_1 = 1$, $\Delta x_2 = 1$, and $\Delta x_3 = -2$. Here the contract price for the individual three parties, i.e. P_1, P_2 and P_3 has to be negotiated. The table below shows the resulting efficiency gain. To be explicit upon the effect on the consumer and producer surplus, we have quoted the income/benefit and costs for each group of participants separately[111]. The surpluses for participants not involved in the transaction are the same as in table 9, while the net profits of the tri-lateral agreement for each of the parties in the contract depend upon the negotiated prices.

	Market clearing Zonal Division A Quantities			Additional agreement Quantities			Net per node	Total producer surplus for all contracts			Total consumer surplus for all contracts			Contract
Node	Price	Sell	Buy	Price	Sell	Buy	Net inj.	Producer Income	Producer Cost	Producer Surplus	Consumer Benefit	Consumer Cost	Consumer surplus	Profit of contract
1	15.10	151.0	97.9	P_1	1		54.1	2281.14 + P_1	1155.72	1125.42 + P_1	1718.86	1479.10	239.76	P_1 -15.15
2	15.72	78.6	85.5	P_2	1		-5.9	1236.24 + P_2	633.95	602.30 + P_2	1527.51	1344.68	182.83	P_2 -15.83
3	15.72	39.3	85.5	P_3		2	-48.2	618.12	309.06	309.06	1558.86	1344.68 + P_3	214.18 - P_3	31.35 -P_3

Table 11 Zonal market clearing with an additional tri-lateral contract

The total social surplus of this scenario (including the surplus of the additional contract) is now 2706.505. This is found by adding the producer and consumer surpluses of all participants, and also including the grid surplus of 32.96 from the zonal market clearing[112]. The increase in social surplus is in total 0.37. This increased surplus equals the total profit on the contract, i.e. note that 0.37=31.35 - 15.15 - 15.83[113]. The allocation of this extra surplus is given by the price negotiations of the three parties.

---

[111] Note that the total cost of the producers after including the extra contract, is given by $\frac{1}{2} c_i x_i^{p2}$, where c_i is the slope of the supply curve, i.e. 0.1, 0.2 and 0.4, respectively, and x_i^p is the total volume. Likewise, the total benefit of the consumers after including the extra contract is given by the area under the demand curve, i.e. by $a_i x_i^d - \frac{1}{2} b_i x_i^{d2}$, where a_i is the intersection of the supply curve, i.e. 20 for all nodes, b_i is the slope of the demand curve, i.e. 0.05 in all nodes, and x_i^c is the total volume of the node.

[112] The grid surplus is not altered due to the additional contract. This is because we find it reasonable to assume that the contract holders agree upon prices so that $P_1 + P_2 = P_3$. In other words, we assume that the producers receive the full amount paid by the consumer. In contrast, this is not the case with the zonal solution. Here, in the event of congestion, total payment by consumers does not equal the amount received by the producers.

[113] The costs of producer 1 have increased with 15.15 from 1140.57 to 1155,72, the costs of producer 2 have increased with 15.82 from 618,12 to 633,94, and the benefit of consumer 3 has increased with 31.35 from 1527.51 to 1558.86.

We have thus seen that the zonal pricing scheme did not render the optimal and most efficient market solution. This was illustrated by the construction of a trilateral contract, which in effect represented a Pareto improvement. It is in principle possible to construct such profitable tri-lateral contracts, until the optimal market solution as given by the optimal nodal pricing congestion method is found.

Constrained Capacity: Nodal Pricing

Let us now conclude this section by reviewing the optimal solution of nodal pricing. Here prices are differentiated in all nodes. The optimal nodal pricing allocation is found by maximizing social surplus, under the restrictions that the power flow follows the laws of Kirchhoff, and that the power flow of line 1-2 is less or equal to its maximum capacity. For our example we have[114]:

$$\begin{aligned} &Maximize \quad Social\,Surplus = \sum_{i=1}^{3}\left[\tfrac{1}{2}(a-p_i)x_i^d - \tfrac{1}{2}p_i x_i^s\right] \\ &\qquad\qquad st. \qquad x_{12} \le 20 \\ &\qquad\qquad\quad x_1 + x_2 + x_3 = 0 \end{aligned}$$

The resulting optimal solution for our example is shown in table 12.

	Price	Supply	Demand	Net Injection	Producer Surplus	Consumer Surplus	Grid Surplus	Social Surplus	Line	Flow
Node 1	15.19	151.9	96.3	55.6	1153.05	231.76	-843.96	540.85	1-2	20.0
Node 2	15.82	79.1	83.5	-4.4	625.92	174.47	70.01	870.40	1-3	35.6
Node 3	15.50	38.8	89.9	-51.2	300.48	202.10	793.06	1295.64	2-3	15.6
Total		269.7	269.7	0.0	2079.45	608.34	19.12	2706.90		

Table 12 Market solution nodal pricing

There are a variety of methods used for handling congestion. The efficiency of the established operable methods may, however, imply large deviations from the efficiency of other near-optimal solutions. The reader is referred to Bjørndal (2000) which provides an overview of different congestion pricing methods, along with further analyses of congestion management and pricing issues in the Norwegian electricity market. The choice of congestion methods may have important implications from a national, as well as from an international perspective, and represent a vast area of unresolved questions.

[114] The first part of the social surplus is the area under the demand curve, the second part is the costs of production. The flow of line 1-2 follows from the above quoted laws of Kirchhoff, i.e. $x_{12} = \frac{1}{3}x_1 - \frac{1}{3}x_2$. The last restriction specifies that sum production has to equal sum consumption.

5.3 Network Pricing Today

The general tariff structure still reflects the above-mentioned principles of point charges, a two-part tariff with fixed and variable fees, and an implicit congestion fee through the geographically differentiated area prices in the case of congestion:

The variable fee is dependent upon the amount of energy input/output to the grid. This part of the tariff is meant to reflect the marginal losses of the input or output to the grid, and the marginal loss rates are recalculated weekly.

The fixed fees are based on various formulas based on previous years' production or consumption. Reduced fees are offered for example for new production capacity in areas where new production is beneficial from a network point of view, and for consumption on interruptible power agreements.

Bottlenecks between Elspot bidding areas are handled by zonal pricing. For the other Nordic countries the bidding areas are consistent with the geographical area of the transmission system operators, i.e. Sweden, Finland, KONTEK, Denmark-Jutland and Denmark-Zealand (the two latter grids are not physically interconnected). In addition there are the Norwegian bidding areas. Norway is as such the only country which uses the zonal pricing scheme for handling internal bottlenecks, though only for large and durable bottlenecks. The Norwegian grid is usually divided into two bidding areas, but three or more areas within Norway is possible. It is the Norwegian transmission operator Statnett who determines how to split the Norwegian grid into bidding areas based on physical conditions.

Internal bottlenecks within the bidding areas are handled by different systems. Within Norwegian bidding areas bottlenecks are normally to be dealt with by use of the regulation power market. Within Sweden, Finland, and Denmark, grid congestion management is managed by counter-trade purchases based on bids from generators. In this case the transmission operator pays for a downward regulation in the surplus area and for upward regulation in deficit areas.

6 Concluding Remarks on Market Reform and Market Efficiency

We have now reviewed the early phase of the electricity market reform in Norway, with a particular emphasis on wholesale electricity market design. After nearly two decades, the Norwegian market, and subsequently the integrated Nordic market, is basically organized along the same fundamental principles, while details of market design and regulatory policies continually have been refined. Here we will briefly comment upon the efficiency of the Nordic market after the reform, with an emphasis on the Norwegian part of the market.

Before proceeding, let us as an indication of the viability of this pioneer market organization note that several markets worldwide have adopted similar structures. Furthermore, observing what Joskow (2006) later has referred to as a 'textbook architecture for restructuring and competition', we find that these principles largely coincide with the reform steps of the Norwegian market in 1991[115]. Also, as noted by Bye and Hope (2006), in a comprehensive EU-financed research project on European electricity reforms (SESSA), the Nordic electric power model was suggested as a potential benchmark for market organization and the efficient functioning of electric power markets[116].

Drawing on recent studies of the Norwegian electricity market, we will in the following look into issues of short and long run efficiency in the current market. As an overall assessment, we find a broad consensus that the basic structure and implementation of the market system have proved viable for promoting an efficient electricity sector, however, still with a potential of further improving market performance and meeting future challenges. Section 6.1 comments upon issues of short-run efficiency. Long-run efficiency and driving forces for

[115] Joskow (2006) defines ten key components of a 'textbook' architecture for restructuring, regulatory reform and the development of competitive markets for power: 1. Privatization of state-owned utilities; 2. Vertical separation of competitive segments from regulated segments; 3. Horizontal restructuring of the generation segment to create an adequate number of participants; 4. Horizontal integration of transmission and network operations; 5. Voluntary public wholesale spot energy and operating reserve market institutions; 6. Regulatory rules and supporting network institutions to promote efficient access to the network; 7. Unbundling retail tariffs; 8. Measures for retail competition; 9. Independent regulatory agencies; and 10. Transition mechanisms for the reform. Except for components 1 which was not politically accepted, and 3 which was not necessary due the large number of generators, we find that the mentioned components all are issues which were central in the Norwegian electricity reform.

[116] See the SESSA webpage, www.sessa.eu.com, and also Hope and Singh (2006).

investments are the topics of section 6.2. Section 6.3 comments upon the market-based system as a framework for implementing environmental measures. Section 6.4 concludes the chapter.

6.1 Short-Run Economic Efficiency of the Market

Let us now consider the efficiency of the market in the short-run, i.e. within the given capacity of the market. Short-run efficiency is, simply put, related to the extent to which the market succeeds in allocating power in accordance with the marginal benefits and costs of power. With rapidly changing and volatile market conditions a crucial issue is also the flexibility of the market in adapting to these changing conditions. In principle, the efficiency gain brought forth by the market reform, can be measured as the gain in net social surplus following from this market organization, as compared to the pre-reform organization of the market. While it is a vast, if not impossible task, to accurately estimate this gain, in practice studies on the efficiency of the reform have focused on more isolated measures that may indicate the degree of efficiency in the market. This will also be our focus.

For the competitive part of the market, efficiency is to a large extent related to market prices; How well do market prices reflect the relative scarcity and true marginal costs and benefits of power, on wholesale and retail levels, and towards different producer and consumer groups? Moreover, how well do producers and consumers respond to changing market prices? The resulting degree of efficiency is furthermore a product of several factors, for example related to the efficiency of organized markets and regulatory measures, as well as to the general market structure and degree of competition. For the non-competitive part, i.e. transmission and distribution, efficiency is on one hand related to whether the regulatory framework induces cost efficiency in the supply of these services and a socially optimal level and quality of services. On the other hand, efficiency in the short run is also related to whether the tariff structure and levels give incentives for an optimal use of these services.

Our purpose here is not to provide a full analysis, but rather to point to aspects of the market which have contributed to efficiency, and also to indicate areas for further improvement. Subsection 6.1.1 reviews some evidence on market performance, while subsection 6.1.2 seeks to relate this to aspects of market organization.

6.1.1 Evidence of Short-Run Power Market Performance

The general picture is that the relative scarcity of power has been reflected in market prices, both in periods of abundance and of scarcity. As noted by Bye and Hope (2006), the market reform was implemented in a scenario of excess capacity, with highly differentiated consumer prices, and large waste in the form of large spillovers of water. The market rather successfully managed the transition from a situation with politically set prices (based on long-term marginal costs), to competitively set prices. In the first years after the reform, the market responded by driving prices down, thus reflecting the excess capacity. Efficiency gains following from deregulation and competition may also have contributed to a downward pressure on prices. As demand increased and Norwegian capacity was restricted, prices then started to rise. With the deregulation of Sweden, Finland, and Denmark, excess capacity in these countries, combined with Norwegian imports, contributed to keep prices low. In later years, with a further increasing demand, and a tightened energy balance, the market has again responded with rising energy prices.

On the demand side, even from the immediate beginning of the new market-based system, it was found that the falling wholesale prices were mirrored in retail prices: Bye and Hope (2006) noted that the market reorganization seemed to reduce price differentials between consumers, as well as to narrow the gap between end-user prices and the (wholesale) equilibrium prices. As of the current situation, several studies, as e.g. Econ Pöyry (2007a) and Amundsen et al. (2006) point out that changes in wholesale prices are rapidly reflected in Norwegian retail prices, and that increases in end-user prices furthermore have a considerable and relatively rapid impact upon demand. With a focus on energy efficiency, Econ Pöyry (2007a) further concludes that the development in energy efficiency after the market reform has been favorable, and possibly better than in other comparable countries, with similar as well as other energy systems.

On the supply side, a major question of efficiency is whether the market reform induces a socially optimal management of the Norwegian hydro-power system. Wolfgang et al. (2007) find that reservoir management has changed after the market reform, with changes in all seasons of the year, and an average reduction in reservoir levels of 4.6 %. The largest change is in autumn just before the winter period, while changes in reservoir levels at end winter are non-significant. Simulations, however, indicate that there have been no noticeable changes in the handling of reservoirs under dry year scenarios. There are several factors that may account for the change: Some changes follow directly from the new market framework. For example, the pre-reform local responsibility of security of supply implied that production decisions were based on a rather high (pre-stated) rationing price, which induced producers to operate with higher reservoir levels due to the high water values at low reservoir levels. Market reform, with the abandonment of local security of supply responsibility and with more inter-regional and international trade, is thus likely to have induced part of the change in reservoir management. Other non-market system issues may also have affected reservoir management, such as better models and data-processing capacity, better reservoir and weather forecasts, and expectations of larger precipitation due to climate changes; all which indicate that it is profitable to tap the reservoirs to lower levels than before. It is however interesting to note that Wolfgang et al. (2007) conclude that there is no evidence that the current reservoir management renders too low a reservoir level compared with optimal solutions based on optimizing models. Multiconsult (2007) which considers different environmental impacts of the changes in reservoir management, however, raises the question whether these environmental costs to a larger degree should be incorporated in decisions guiding reservoir management.

For the power market, a further crucial performance aspect is the ability of the market to cope with periods of severe scarcity of energy. In a system vulnerable to large nature-given variations in supply, such as the Norwegian hydro-power dominated system, this is of increasing importance as tighter market conditions emerge due to growing demand and a slower pace of investments. The market has here been put to several tests, most notably the situation of the winter 2002/03. Starting in the spring of 2002, inflow exceeded normal levels and implied a higher water level than normal. This situation consequently increased production and lowered prices. In Autumn, however, rains failed and resulted in an extremely dry 2002/03 winter, with extremely low reservoir levels, and an impending danger of rationing. Several studies, e.g. Amundsen et al. (2006), Bye and Hope (2006), Bye et al. (2003), find that the Nordic market in fact was able to manage the serious shortage of energy. Prices adjusted rapidly with rocketing prices, and with a fairly rapid demand and supply response on all levels of the market, rationing was avoided. Rainy January days and an early spring, however, also highly contributed to bringing the situation under control. Most important, this incident indicated that the market was robust enough to withstand even quite extreme energy supply shocks. While the market did pass this test of extreme market conditions, and offered proof of a flexible and robust market, this incidence, however, also highlighted the vulnerability of the system.

6.1.2 Market Organization and Short-Term Performance

Market performance seems hitherto to be indicative of a fairly good short-term efficiency, where market prices to a large extent reflect underlying fundamental market conditions, such as precipitation and reservoir filling, development of fuel prices, and patterns of consumption.

Extreme market situations have also revealed that the market is fairly robust and flexible. Market performance is in many ways a product of the specific chosen market organization. In this section we will briefly offer some comments upon aspects of market structure and organization that we believe have contributed to an efficient, flexible and robust market, but we will also point to areas for potential improvement.

- *Organized Spot and Hedging Markets:* With the continued development of the organized markets, the Nordic organized power markets may now be characterized as a highly developed and well-functioning market in relation to size, liquidity and transparency. In this we believe that the organized markets of Nord Pool play a crucial role for the short-run efficiency of the market. In our opinion, there has also been demonstrated an overall will and ability to tailor market design to achieve liquidity in meeting the new challenges of the market, as for example in coping with the international integration of markets, in the development of hedging instruments, and in incorporating environmental instruments. We, however, also believe that a continued dynamic view on future market design is important for future liquidity and efficiency.

  We have, though, a specific comment on current spot market design. The spot market has developed from a market with pre-dominantly hydro-power producers, to a Nordic market representing several different generation technologies. To accommodate technologies with e.g. inflexible production and high start-up or shut-down costs, the block bid format was introduced, giving the participants the opportunity to set an all-or-nothing condition for all the hours within the block. The use of block bidding, however, in effect turns the spot electricity auction into a combinatorial exchange, where specific block bid formats and equilibrium market clearing methods have implications for the efficiency of the market[117]. There are here a number of unsolved questions regarding the implied efficiency and equilibrium properties of these bidding formats. As further market integration and future investments in generation capacity may further shift the underlying technology mix, a concern is that inappropriate block bidding procedures may detrimentally affect the efficiency of the spot auction. Related to this is work on how to price non-convexities like fixed start-up costs[118].

- *Network Pricing and Congestion:* The handling of congestion is closely related to the spot auction for planned day-ahead schedules. We believe, for many reasons, that also

[117] See for example Meeus et al. (2008).
[118] See e.g. Hogan and Ring (2006), Bjørndal and Jörnsten (2008), and O'Neill et al. (2005).

congestion management represents an important area for future improvement. Firstly, further insight is needed as to the efficiency of different congestion methods, also taking into account the interaction with the spot auctions, where new bidding formats such as block bids, may have important implications also for congestion management. Secondly, there are efficiency aspects related to the fact that different methods for congestion handling coexist in the Nordic market; This is on the one hand within countries, for example in Norway, where zonal pricing is used for the largest and most frequent bottlenecks, while other congestion situations are handled by special regulation based on regulation power bids. On the other hand, different countries also employ different methods, where for example Sweden in contrast to Norway uses counter purchases for internal bottlenecks. There are here major unresolved questions as to the efficiency and allocational effects of different co-existing methods, confer Bjørndal et al. (2006). Thirdly, a further aspect, as noted by Bye and Hope (2006), is related to the efficiency effects of the coexistence of different regulatory regimes employed for networks in the Nordic countries, which may have implications for the ability to achieve efficient and harmonized methods for congestion handling.

- *System Operation and Ancillary Services:* The above discussion on spot market organization and congestion handling refers to market systems for short-term planned energy. A crucial task of the electricity system is, however, also to attend to the momentary balance of the system, and uphold a sufficient quality and reliability of the electricity system. As such, this is also an important aspect of the short-run efficiency of the market. In the pure hydro-power system, we have argued that power generation capacity for such tasks in general is abundant and relatively cheap compared to systems based on other technologies. We have, however, experienced a tightening power balance due to growing demand, low investment rates (see below), and increasing demand for the export of services related to e.g. peak load, load-following and momentary balancing. Tighter market conditions thus mean that system operation becomes more challenging. While this has not been a problem yet for the Norwegian system, it may be in the near future.

  The responsibility for operation of the national network and system operation lies with the national transmission company. Here we have seen that in response to the tightened power balance, market design has continually been adapted. For example, to allocate sufficient capacity to ancillary services, the regulation power market has been supplemented with a regulation power capacity option market, and a market arrangement for longer-term capacity options. Market arrangements for other ancillary services have also been developed, e.g. a market for frequency-related reserves, and a market for spinning reserves. In this we see that the basic framework of the market also has proved a viable framework for introducing and developing systems or market-based arrangements to meet new challenges in this area. Here we also expect to see further developments in the future. Furthermore, with respect to the increasingly more integrated Nordic market, the Nordic transmission network system, however, remains decentralized in the sense that national transmission companies are responsible for system operation in their national networks. Though cooperation between transmission companies takes place on a voluntary basis through NORDEL, Bye and Hope (2006) point out that there may be potential inefficiencies of the current system, and thus room for improvement with respect to common system operation.

- *Security of Energy Supply in the Short Run:* A further aspect of the short-run security of supply is to ensure a sufficient supply of energy. In a market-based system, the task of balancing the demand and supply of energy in the short run is first and foremost a task carried out by market forces. For example, in relation to the Norwegian system, where the supply of energy is highly dependent upon precipitation, the market in times of scarcity balances by raising prices, which in turn depresses demand, increases imports, and encourages an economizing of water and carrying of storage in the advent of periods of even greater scarcity. There has, however, been a large degree of political concern about the ability of the market in handling such situations. Seeing that the market under the 2002/03 shortage handled the supply shock rather well, with rapidly adjusted prices combined with rapid demand and supply response, Amundsen et al. (2006) note that fears regarding supply security and adequacy seem to be unfounded. Still, the effect of a potential crisis of supply shortage is of considerable social economic cost, which in principle might not be adequately reflected in prices. If this is the case, further security measures may be necessary. Here we note that Statnett is responsible for the security of supply of the system, and thus has incentives to develop measures for handling severely scarce energy situations, in the event that the market alone would not be able to handle the situation. Implemented measures in this respect are consumption energy options, and the purchase of reserve generation power. It is an open question whether further measures are necessitated.

- *Network Efficiency – Regulation:* A further integral aspect of the overall short-term efficiency of the electricity market is related to the network: Firstly, market efficiency is contingent upon efficient network pricing, where the design of price structure, as well as the resulting price level both are important for the short-term efficiency of the market. Secondly, network efficiency is related to supplying an efficient level and quality of network services, i.e. the level which in principle balances the marginal cost of supply against the marginal benefit for different levels and qualities. Thirdly, network efficiency is related to achieving cost-efficiency in supplying these services. All these aspects of efficiency, however, depend upon the incentives provided by the regulatory regime. Different regulatory regimes have been employed. With the introduction of the new Energy Act, rate-of-return regulation was introduced for network companies. In 1997, this regulation was replaced by income regulation, described by Bye and Hope (2006) as essentially a revenue-cap incentive mechanism, which also contains elements of rate-of-return, price-cap and yardstick regulation. The revenue cap is based on the total cost coverage of network activities, including a stipulated rate of return on invested capital. An efficiency improvement factor was defined for each network owner, based on a data envelopment analysis (DEA) of the efficiency improvement potential for each company. There is no doubt still room for further improvement of the regulation mechanisms for network companies. Bye and Hope (2006) also note that the need for harmonization of regulation within the integrated Nordic electric power market has gradually become more pressing.
- *Market Power and Market Dominance:* Investigations by competition authorities and research studies do not seem to have documented any major instances of unilaterally or collectively exercise of market power that is to the detriment of competition in the Nordic power market. Several studies, as e.g. Bye and Hope (2006), Hope (2005), Bye et al. (2003), Econ Pöyry (2007b), all seem to conclude that there is an overall satisfactory level of competition in the market, with a large number of market participants, and a low degree of market concentration, especially as to the integrated Nordic market. A considerable challenge for the Competition Authority, however, lies in maintaining a competitive market in the future, both with respect to vertical integration and strategically motivated actions. Bye and Hope (2006) note that as the electric power market has become more concentrated through mergers and acquisitions, the Norwegian Competition Authority has tightened its market surveillance and enforcement policy to prevent the abuse of market

power. A further aspect is the implications of transmission capacity constraints for market competition, where challenges are related to the special market properties of an electric power system. For example, any plant on the margin in a restricted price area, even a small firm, may in principle be in a position to abuse market power. Hope (2005) also raises the question whether an ex ante system of market monitoring and regulatory oversight of the electric power markets, should be developed as an integral part, or at least as a supplement, to the ex post based competition and enforcement system.

- *Retail Markets:* A general impression is that price signals of the wholesale market are fairly rapidly conveyed to end users in the Norwegian retail market, and that price differences between different end user groups have been diminishing. There may be several explanatory factors for this development. For example, Econ Pöyry (2007a) points out that the degree of competition between power suppliers in Norway contributes to lower margins in the retail sector. Also, as changes in spot prices are rapidly channeled to end-users that have spot-based contracts, this indicates that the large degree of spot-based contracts has been important. Furthermore, there are several specific Norwegian measures that have contributed to promoting an efficient retail market. For example, as pointed out by Bye and Hope (2006): A load-profile demand measurement was introduced in 1995 to facilitate trade in the retail market by avoiding investment in expensive metering equipment for retail customers. A couple of years later, fees for consumer switching were eliminated to stimulate consumer switching and market competition, and time allowed for consumer switching was reduced to one week. In 1998 the Norwegian Competition Authority also introduced a price information system for retail prices from power suppliers to improve market transparency.

  An issue of concern relating to consumer response in the Nordic market as a whole, is however, that there seems to be large differences within the Nordic market. For example, as noted by Amundsen et al. (2006), while the wholesale market appears to be strongly integrated, with prices in different areas diverging for shorter periods only, this is not so with respect to the retail prices in different Nordic countries. There may be several causes for this. For example, a larger part of Norwegian retail customers, than in the other Nordic countries have spot price based contracts. Other explanations may be related to differences in retail competition and differences in the degree of vertical integration. Further, there are differences in regulatory measures. As such, the retail market seems to represent an area for further efficiency gains, especially with respect to the Nordic market, but also with possible efficiency gains for the Norwegian retail market.

6.2 Long-run Economic Efficiency of the Market

Long-run efficiency is related to whether the market system induces optimal investments, both with respect to investments in generation capacity and network capacity, and investments related to capacity for future consumption as well. Investments in these different areas are, however, guided by completely different frameworks and incentive mechanisms. Investments in generation capacity, as well as consumer devices, are basically guided by the profitability of the investments, where future price development is crucial. Investments in network capacity, however, are highly dependent upon the incentives provided by the prevailing regulatory regime. In terms of long-run efficiency, a major question is whether the prices and incentives of these systems encourage optimal investments, including the right mix of generation, consumption and network capacities. For example, an expected future energy shortage may be managed by investments in generation capacity, by investments in network capacities that alleviate bottlenecks and enable increased imports to the area of shortage, or by investments in power-saving appliances. On top of this, there are several and partly diverging goals of energy policy, including energy efficiency, security of supply and environmental considerations. These are the challenges of long-run efficiency in the power sector. As commented Bye and Hope (2006) the issue of investments is high on the political agenda, with a widespread skepticism as to the ability of the market to provide optimal investments. In this section we will briefly review some studies on investment activity in the deregulated market, and comment upon the framework and incentives for optimal investments. Sections 6.2.1 and 6.2.2 review studies related to investments in respectively generation and network capacity.

6.2.1 Generation Capacity

The market reform brought forth a completely new framework for investment decisions. In the pre-reform era investments were to a large degree guided by the compulsory requirements of a local security of supply based on firm power contracts, and where investment levels were more or less made to match rather optimistic forecasts of a rising and perceivably inelastic demand. This resulted in excess generation capacity. Following the reform, the power market became a fully accepted means of procuring power, and local security of supply became a matter of the market. Investment initiatives are now in principle to be taken by the individual generators, with the incentives of profitability as the main driving force. From a market perspective, it should also be added that the society point of view on security of supply has gradually shifted from a focus on the national balance of demand and production, to that of a national power balance including international trade; If power imports are readily available at a lower cost, it is rational to base power supply on imports rather than on the more expensive national investments.

Bjørndalen et al. (2007) give an overview of investments in generation capacity. Following a beginning excess capacity, investments in new production capacity in fact began to fall as early as in the eighties, long before deregulation. After the market reform, electricity prices first fell due to competition and excess capacity, and signaled that investments were not profitable. Thus, investment levels continued to fall. With a growing demand, the falling market prices were subsequently replaced by rising market prices. Bjørndalen et al. (2007) note that investments made after the reform account for a total 4 % increase in (momentary) power capacity in the period 1990-2004, while generation capacity in terms of energy has increased with about 7.5 TWh since 1991. This is, however, only half the demand growth in the same period. In sum, with low investment rates and an increasing demand, the gap between demand and production capacity has throughout the post-reform period continually narrowed, moving from a national surplus of generation capacity, to a national power deficit. There seem to be several factors which may contribute to explain these investment rates:

- *The acceptance of international trade as a source of supply:* The cost of procuring energy by new investments must be weighed against the costs of imported electricity. Here, Econ Pöyry (2007b) points out that the transmission capacity between Nordic countries, as well as to other neighboring countries has been extended[119]. The implication is that if import is the most profitable means of procuring power, this may imply that several generation investment projects in comparison may have been unprofitable.
- *Higher costs of available new production capacity:* Both investment costs as well as future variable costs have been rising in this period. In accordance with this, Econ Pöyry (2007b) notes that it is not surprising that there have been lower investments than historically normal, and that the investments that have been made, mostly have

[119] Transmission capacity between Norway and both Denmark and Sweden has been extended recent years, and the NorNed cable to Netherland has been built. From a Nordic perspective, there is also a relatively large exchange capacity to Germany, Russia, Poland, and Estonia. There is, however, a clear tendency that also the Nordic power balance has been weakened since the turn of the century.

been in hydro power generation capacity. It is first in recent years that the expected market prices indicate that investments in e.g. gas-fired plants and other technologies for which costs have been escalating, may be profitable.

- *Stricter regulatory policies, arising mostly from environmental concerns:* Investments are subject to a granted concession. Here different environmental concerns, for hydro-power plants, as well as for other technologies (e.g. gas-fired generation plants), have played an increasingly larger role with respect to the level and nature of investments that have been granted a concession. For environmental desirable technologies, such as e.g. wind power plants, case studies by Econ Pöyry (2007b) further indicate that aid schemes are of particular significance.
- *Political uncertainty and concession risk:* The electricity market is, as many other commodity markets, a market of inherent price and quantity uncertainty. Bjørndalen et al. (2007), however, point out that additional political uncertainty and uncertainty and costs related to concessions, also may contribute to hindering investments.

In relation to long-run efficiency the main question is whether the market-based electricity system has provided a framework for economically and environmentally optimal investments in generation capacity. We have seen that there are several explanatory factors. If the low investment rate reflects considerations that are justified from an economic or environmental point of view, even low investment rates may in principle be in line with long-run economic or environmental efficiency. If, however, low investment rates to a large extent reflect other issues that represent market imperfections, such issues thus represent important potential sources of improving long-run efficiency. Let us review some comments on investment performance:

Several studies indicate that the low investment rates largely follow from political and/or environmental considerations, rather than from a lacking willingness of generators to invest. For example, Bye and Hope (2006) comment that politicians seem unanimous in blocking new investments in large hydropower plants, gas-fired plants and nuclear and other thermal plant technologies, so that the only feasible alternatives seem to be renewable technologies based on e.g. wind, biomass, solar energy and wave power. And, since these technologies are costly, the market prices are still not sufficiently high to stimulate investment in the absence of strong financial support. Also Amundsen et al. (2006) note that generation investment in the Nordic market is not so much a question of commercial, as of political will.

Comments on the role of the market framework as to giving incentives for achieving long-run efficient investments, however, seem to indicate that the market itself to a large extent does give adequate signals for investment based on economic considerations alone. A main conclusion of Econ Pöyry (2007b) is that the Energy Act seems to provide a better framework for socially rational and efficient investments, than other realistic alternatives: Market-based prices reflect the value of new power, and not only provides an important benchmark for assessing the profitability of new investments, but also induces an important flexibility on the supply side as well as the demand side, and also with respect to international trade. Econ Pöyry (2007b) notes that projects with a low environmental impact in fact seem to be carried out without undue delay if they are profitable, while the most important explanation if such investments have not been carried out, is that of lacking profitability. Also Amundsen et al. (2006) conclude that there is little support for the view that generators have not seized on profitable investment opportunities, thus indicating that prices are sufficient to attract investor interest. As to the current status of the security of supply Econ Pöyry (2007b) notes that the Nordic market may be characterized as a market with a satisfactory balance both with respect to power capacity and energy. Though the Norwegian power balance has deteriorated, new transmission capacity to other countries has been built, also adding flexibility to the market There is, however, a problem of regional energy imbalances in parts of Norway, due to a combination of increased (industrial) demand, constrained network capacity, and low investments in generation capacity and/or network capacity to the areas.

Above, we have discussed generation investments in general terms. Before proceeding to the issue of network investments in the next section, let us, however, comment upon two more specific issues of investments:

- First, there is a major challenge in coordinating generation capacity investments and network investments to achieve an efficient mix of these different capacities. Due to the externalities and loop-flow features of the network, the production capacity decisions based on market prices alone might not be the overall optimal investment strategies. This may follow from several aspects; - price signals are imperfect due to coarse congestion handling methods; - capacity investments may affect future nodal prices, making it difficult for the producer to evaluate the true profit of the investment; - and the socially optimal investment strategy may involve a mixed investment strategy of network and generation capacity investments[120]. These issues represent important challenges for regulatory policies related to investments in grid capacity as well as in generation capacity.

- Secondly, there is an important heterogeneity in power supply with respect to the different attributes for handling e.g. base load versus peaks and momentary power balance requirements. While energy supply has been scarce, until recently the momentary power capacity has in contrast been abundant in the Norwegian system. As the Norwegian market becomes increasingly integrated with electricity systems where such capacity is more costly, this scenario of abundance is changing. An important aspect of long-run efficiency is thus whether the market system also provides sufficient incentives for investments in this area, also handling potential market imperfections. While this represents an important future challenge, we do, however, find several recent developments of the market system that seem to bring forth important investment incentives, e.g. the market for regulation power capacity options and for longer-term capacity options, and markets for frequency-regulated reserves and for spinning reserves. As such, this development demonstrates a strength of the system, i.e. the ability to adapt the framework to meet future challenges.

6.2.2 Transmission and Distribution Capacity

During the period from the early 1970's till the late 1980's, large infrastructure projects were carried out, extending transmission networks from west to east in Norway, and to a certain extent from north to south. Once these large infrastructure projects had been completed, investments in network infrastructure capacity decreased, with a sharp decrease starting in the late 1980's. As pointed out by Bye and Hope (2006) this decrease in investments also coincided with the debate that took place before and during regulation of the Norwegian electricity market in 1991. After the reform, incentives for investment in network capacity follow from the incentives provided by the network regulatory regime, however, where investments also are subject to a concession grant where environmental and other societal concerns are taken into consideration. With strongly falling investment rates, a main concern of long-run efficiency is whether the regulatory regime is able to induce optimal network investments.

A new regulatory regime was introduced for transmission and distribution network companies in 1991, where the new regime of rate-of-return regulation and 'yardstick competition'

[120] See Lund and Rud (2008).

contributed to reduce network tariffs. Bye and Hope (2006) note that this yardstick competition was meant to secure cost efficiency, and an improved social investment strategy. The initial rate-of-return regulation was, however, in 1997 replaced by an income regulation regime where the efficiency rate plays an important role. The efficiency rate is specific to each distribution network, and is meant to reduce the annual allowable network-specific income, and through this provide incentives for the distribution network to reduce costs, which is an important issue of short-term efficiency. It may, however, be questioned whether the new regulatory regime can succeed in encouraging optimal investments in the network, both with respect to the level of investments, the nature of investments as to quality of supply, and to hinder that lacking network investments will contribute to bottlenecks that pave the way for less competition in restricted areas.

Amundsen et al. (2006) point out that in the case of Statnett, there appears to be no lack of will to invest, and that, were it not for the more restrictive view taken by Norwegian regulatory authorities, more transmission capacity would have been built. However, network investments in other levels have been severely small. As pointed out by Bye and Hope (2006), it is an open question whether the regulatory regime provides sufficient incentives for optimal network investments that enable the market-based electricity system to function efficiently. We believe that there is room for large improvements in this area. There are several challenges. Firstly, an important aspect relates to the inherent incentives of the regulating mechanism, as there are major challenges in combining incentives both for cost-efficiency, efficient pricing, and optimal investments. Secondly, as noted by Amundsen et al. (2006), a further challenge is to solve co-ordination problems associated with different national TSOs and investments in interconnector capacity. Thirdly, long-run efficiency depends upon how the market system handles the interaction between investment in network and generation capacities, which in effect are highly integrated, and where the socially optimal investment strategy may be a mixed investment strategy.

6.3 Environmental Challenges

To the extent that the electricity market conveys the scarcity of electricity, the market per se contributes to economizing with the scarce resource, and gives signals to lower electricity consumption. The question with respect to future environmental efficiency is, however, to an even greater extent whether the market system is an instrument fit to carry out policies for handling environmental externalities, in the short run and in the long run. Econ Pöyry (2007b) here conclude that the Energy Act has provided a flexible framework that also may contribute to reaching environmental goals. This may e.g. be goals of reducing greenhouse gas emissions (e.g. by using different forms of market-based allowances or quotas), and goals of promoting investments in renewables (e.g. by direct investment aid or differentiated net tariffs). While the deregulated market system has retained the possibility to evoke necessary regulatory measures, the system has thus also provided an important framework for implementing market-based measures which give appropriate incentives to the participants of the market. An important tool is also the concession system which enables authorities to invoke measures to incorporate specific environmental considerations, locally as well as nationally. However, though the market system seems to provide an adequate framework for implementing environmental measures, the main future challenge still lays in designing the appropriate incentive mechanisms within this framework.

6.4 Concluding Remarks

An unanimous verdict seems to be that the market system and overall framework provided by the Energy Act is a flexible system for dealing with changes in market conditions, as well as environmental challenges. The market system has proved to be robust in handling several extreme market conditions, and changes due to e.g. international market integration. However, there is still a need for continuous improvement of details of the framework, the organized markets, and the regulatory mechanisms including the plan- and concession system, to achieve an environmentally and economically efficient power sector.

References

Amundsen, E. S. (1989): 'Grunnrente, effektivitet og inntektsfordeling', SAF-arbeidsnotat nr. 12/1989. (Translation of title: Economic Rent, Efficiency and Allocation)

Amundsen, E. S., L. Bergman, and N.-H. M. von der Fehr (2006): 'The Nordic Electricity Market: Robust by Design?', in F. P. Sioshansi and W. Pfaffenberger (2006), *Electricity Market Reform. An International Perspective*, Elsevier, 2006.

Battley, N. ed. (1993) *The World's Futures & Options Markets*, Probus Publishing Company, Cambridge, England.

Berg, M. (1989): 'Prioriterte leveranser', SAF-arbeidsnotat nr. 9/1989. (Translation of title: Priority Deliveries)

Bjørndal, M. (2000): 'Topics on Electricity Transmision Pricing', Doctoral Thesis, Norwegian School of Economics and Business Administration, May 2000. (Also: SNF-rapport no. 8/2000).

Bjørndal, M., E. Hope, and H. O. Husum (1994): 'Sentralnettets utstrekning. En prinsipiell analyse', SNF-rapport nr. 62/94.

Bjørndal, M., and K. Jörnsten (2008): 'Equilibrium prices supported by dual price functions in markets with non-convexities', *European Journal of Operational Research*, 190, 768-789.

Bjørndal, M., K. Jörnsten, and L.M. Nonås (2006): 'Flaskehalshåndtering i Norden. Samfunnsøkonomiske vurderinger', Forthcoming SNF-Rapport.

Bjørndalen, J., J. Bugge og T. Olsen (2007) 'Investeringer i kraftproduksjon - Kapasitetsutvikling 1975-2006', Rapport ECgroup, 6/6-2007.

Bjørndalen, J., E. Hope, E. Tandberg, og B. Tennbakk (1989): 'Markedsbasert kraftomsetning i Norge', SAF-rapport nr. 7/1989. (Translation of title: Market-based Electricity Trade in Norway).

Borgen, H. and B. Walther (2006): 'Effektregulering i det norske kraftsystemet', Elektro, no. 4, April 2006.

Bye, T. and E. Hope (2006) 'Electricity Market Reform – The Norwegian Experience', in L. Sørgard (ed.): *Competition and Welfare: The Norwegian Experience*, Konkurransetilsynet (Norwegian Competition Authority), Bergen, Norway.

Bye, T., N.H.M. von der Fher, C. Riis, and L. Sørgaard (2003): 'Kraft og makt – en analyse av konkurranseforholdene i kraftmarkedet', Report of an Expert Group appointed by the Norwegian Ministry of Labour and Administration, 2003.

Chao, H.-P. and R. Wilson (2002): 'Multi-Dimensional Procurement Auctions for Power Reserves: Robust Incentive-Compatible Scoring and Settlement Rules', *Journal of Regulatory Economics*, 22:2, 161-183.

Econ Pöyry (2007a) 'Energiloven og energieffektivisering', Econ Pöyry rapport no. 2007-077.

Econ Pöyry (2007b) 'Vilkår for ny kraftproduksjon', Econ Pöyry rapport no. 2007-097.

Errera, S. (1984) 'Exchanges and their contracts', in Treat, J. E. (ed.) *Energy Futures. Trading opportunities for the 1980s*, PennWell Publishing Company, Tulsa, Oklahoma

Hammer, U. (2007) 'Investeringer i kraftproduksjon og nett. En rettslig studie'

Haugland, S.A., C.B. Meyer, L. Rud og B. Singh (1989): 'Leveringskvalitet på elektrisk kraft. Del 1', SAF-rapport nr. 30/1989. (Translation of title: Quality of Delivery of Electricity)

Hogan, W.W. and B.J. Ring (2003): 'On minimum-uplift pricing for electricity markets', Working Paper, John F. Kennedy School of Government, Harvard University.

Hope, E. (2005): 'Market dominance and market power in electric power markets: A competition policy perspective', Report, Stockholm: Swedish Competition Authority.

Hope, E., L. Rud and B. Singh (1992): 'Markets for electricity in Norway', in E. Hope and S. Strøm (eds.) (1992): 'Energy Markets and Environmental Issues: A European Perspective', Scandinavian University Press.

Hope, E. L. Rud og B. Singh (1993a): 'Märkte für Elektrizität: die ökonomische Reform des norwegischen Elektrizitätssektors', Zeitschrift für Energiewirtschaft, Januar 1993. (Engelsk versjon: 'Markets for Electricity: Economic Reform of the Norwegian Electricity Industry' SNF arbeidsnotat 12/1993).

Hope, E., L. Rud og B. Singh (1993b): 'Det norske kraftmarkedet', SNF-rapport nr. 65/1993. (Translation of title: The Norwegian Electricity Market)

Hope, E. and B. Singh (2006) 'Harmonising an Effective Regulation', in M. Glachant and F. Lévêque (eds.) *Achieving European Union electricity markets in 2009? Assessment and proposals.* EU SESSA.

Hveding, V. (1968): 'Digital simulation techniques in power system planning', *Economics of Planning,* 8 (1-2)

Joskow, P. L. (2006) 'Introduction to Electricity Sector Liberalization: Lessons Learned from Cross-Country Studies', in F. P. Sioshansi and W. Pfaffenberger eds. (2006) *Electricity Market Reform. An International Perspective,* Elsevier 2006.

Jørnsten, K., L. Rud og B. Singh (1997): 'Samspillet mellom nettariffering og markeder for effekt og energi', SNF-arbeidsnotat nr. 55/1997 (EnFO publ. 318-1998). (Translation of title: Interaction of Grid Tariffs and Markets for Power and Energy)

Knivsflå, K.H. og L. Rud (1995a): 'Markets for Electricity: Structure, Pricing, and Performance', Pacific and Asian Journal of Energy, vol. 5, no. 2, pp. 261-284. (SNF working paper no. 115/1994)

Knivsflå, K.H., og L. Rud (1995b): 'Organisering av omsetningsfunksjonen i kraftmarkedet', SNF rapport nr. 11/1995. (Translation of title: Organizing the Exchange Function of the Electricity Market'

Knivsflå, K.H. og L. Rud (1997): 'Price and Volume Relationships between the Electricity Spot Market and the Ancillary Market', SNF-working paper, no. 57/1997 (EnFO publ. 317-1998).

Lommerud, K. E. (1989): 'Regulering av naturlige monopoler', SAF-arbeidsnotat nr. 7/1989. (Translation of title: Regulation of Natural Monopolies)

Lov 29. juni 1990 nr 50 om produksjon, omforming, overføring, omsetning, fordeling og bruk av energi (energiloven). (Energy Act June 1990, Law of production, transmission, sale and distribution of energy)

Lund, A.-C. and L. Rud (2008): 'Investment Evaluation in a Constrained Electricity Network with Stochastic Nodal Price Processes', see essay 4 in this thesis.

Meeus, L., K. Verhagen, R. Belmans (2008) ‘Block Order Restrictions in Combinatorial Electric Energy Auctions’, *European Journal of Operational Research,* Article in Press.

Multiconsult (2007): ‘Rapport om rammer for magasindisponering’, Rapport Multiconsult i samarbeid med Advokatfirmaet Mageli ANS og NINA, 9. Oktober 2007.

Nordpool (2008): ‘Trade at Nord Pool ASA’s financial market’, Feb 20 2008.

NVE(1994) ‘Retningslinjer for systemansvaret i kraftsystemet’.

NVE (2002) ‘Forskrift om systemansvaret i kraftsystemet’ (Regulations relating to Power System Operation) , Oslo, 17 May 2002,

O’Neill, R.P., P.M. Sotkiewicz, B.F. Hobbs, M.H. Rothkopf, and W.R. Stewart Jr. (2005): ‘Efficient market-clearing prices in markets with nonconvexities’, European Journal of Operational Research, 164, pp. 269-285.

Ot.prp. nr. 43 (1989-90)

Rud, L. (1989a): ‘Økonomistyring i produksjons- og distribusjonsselskap: Regnskapet som informasjonsgrunnlag for beslutninger’, SAF-arbeidsnotat nr. 4/1989. (Translation of title: Economic Management of Production and Distribution Companies: Using Accounting Information in Decision Making)

Rud, L. (1989b): ‘Organisering av markeder: Nye former for risikodeling’, SAF-arbeidsnotat nr. 10/1989. ((Translation of title: Organizing Markets: New Forms for Risk-Sharing)

Rud, L. (1990a): ‘Futures markeder i kraftnæringen: Oppbygging og virkemåte av et futures marked’. SAF-arbeidsnotat nr. 50/1990. (Translation of title: Futures Markets in the Electricity Sector: The Construction and Working of a Futures Market)

Rud, L. (1990b): ‘Noen aspekter ved utviklingen av kraftmarkedet’, SAF-rapport nr. 65/1990. (Translation of title: Some Aspects of the Development of the Power Market)

Rud, L. (1990c): ‘Organisering av markeder for elektrisk kraft’, SAF-rapport nr. 39/1990. (Translation of title: Organizing Markets for Electric Power)

Rud, L. (1992): ‘Omsetningssystem for standardkontrakter’, Note accompanying letter to Samkjøringen av kraftverkene i Norge, March 1992. (Translation of title: Exchange System for Standard Contracts)

Rud, L. (1993): ‘Markedsbasert Kraftomsetning - Implikasjoner for organisering og styring av everk’, Nordisk Seminar- og Arbejdsrapport 1993:522. (Translation of title: Market-based Electricity Exchange – Implications for Organization and Management of Power Companies)

Rud (1995), L.: ‘Prising av effekt i et markedsbasert kraftomsetningssystem. Et forprosjekt’, SNF-arbeidsnotat nr. 54/1995. (Translation of title: Pricing Power in a Market-Based Electricity Exchange System)

Rud, L. (1996): ‘Effektiv kraftomsetning: mellomleddenes rolle’, SNF-rapport nr. 13/1996. ((Translation of title: Efficient Power Exchange: The Role of Brokers and Traders)

Rud, L. og B. Singh (1997): ‘Markeder for systemtjenester: Grunnstruktur, incentiver, og samspillet med det øvrige kraftmarked’, SNF-arbeidsnotat nr. 54/1997 (EnFO publ. 316-1998). (Translation of title: Markets for Ancillary Services: Structure, Incentives, and the Interaction with other Parts of the Energy Market)

Rud, L. og S. Tjøtta (1989): ‘Integrasjon i kraftnæringen’, SAF-arbeidsnotat nr. 5/1989. (Translation of title: Integration in the Power Sector)

Rud, L. og S. Tjøtta (1996): ‘Systempålitelighet i et markedsbasert omsetningssystem for kraft’, SNF-arbeidsnotat nr. 59/1996. (Translation of title: System Reliability in a Market-Based Exchange System for Electricity)

Salvanes, K. og Tjøtta, S. (1990): ‘Cost differences in electricity distribution: economies of scale and economies of density in the Norwegian electricity distribution industry’, SAF-working paper No. 30/1990.

Singh, B. (1989): ‘Futures markets for non-oil energy. The hydro power sector in Norway.’ SAF-working paper no. 11/1989.

Statnett (2005): ‘Vilkår for tilbud, aksept og bruk av regulerkraftopsjoner i produksjon/forbruk’, Oct 19 2005, (Terms and Conditions for Offering, Acceptance and Use of Regulating Capacity Options for Production and Consumption).

Statnett (2007): ‘Vilkår for tilbud, aksept, rapportering og avregning for Innfasingsreserve’, Nov 9 2007

Statnett (2008): ‘Vilkår for tilbud, aksept, rapportering og avregning i Marked for frekvensstyrte reserver’, Jan 21 2008.

Statnett (2008): ‘Nordisk harmonisert balanseavregning. Endelig Implementerings-beskrivelse’, Statnett 13. juni 2008.

Storting Melding Nr. 18 2003/2004

Storting Report No. 11 1995/1996.

Strandenes, S. P. (1989): ‘Organisering av nettselskap i kraftforsyningen’, SAF-arbeidsnotat nr. 6/1989. (Translation of title: Organizing Grid Companies in the Electricity Industry)

Tandberg, E. og B. Tennbakk (1989): ‘Kraftomsetningssystemene i Sverige, Storbritannia, New Zealand og USA’, SAF-arbeidsnotat nr. 13/1989. (Translation of title: Electricity Market Systems in Sweden, Great Britain, New Zealand, and USA)

Vaage, K. (1989): ‘Etterspørselen etter elektrisitet’, SAF-arbeidsnotat nr. 8/1989. (Translation of title: Demand for Electric Power)

Walther, B. and I.H. Vognlid (2004): ‘Statnett’s Option Market for Fast Operating Reserves’, ieadsm Demand Response Dispatcher, Vol. 1 (6), April 2005.

Wangensteen, I., O.S. Grande and B.H. Bakken (1996) ‘Systemytelser’, EFI TRA4441

Wilson, R. (1985): ‘Incentive efficiency of double auctions’, *Econometrica*, 53(5): 1101-1115.

Wilson, R. (1987): ‘On the equilibria of bid-ask markets’, in Feiwel, G.R. (ed.) *Arrow and the Ascent of Modern Economic Theory*. London: MacMillan Press, 375-414.

Wilson, R. (2002): ‘Architecture of Power Markets’, *Econometrica*, 70(4): 1299-1340.

Wolfgang, O., A. Haugstad, B. Mo, I. Wangensteen, og G. Doorman (2007) ‘Magasindisponering før og etter energiloven’, Sintef Energiforskning AS, Teknisk rapport TR A6569, Nov 2007

Wu, F.F., and P. Varaiya (1995) ‘Coordinated Multilateral Trades for Electric Power Networks: Theory and Implemantion’, Department of Electrical Engineering and Computer Sciences, University of California.

ESSAY 2

Capacity Charges: A Price Adjustment Process for Managing Congestion in Electricity Transmission Networks*

Mette Bjørndal
Kurt Jörnsten
Linda Rud

Abstract

In this paper we suggest a procedure based on capacity charges for managing transmission constraints in electricity networks. The system operator states nodal capacity charges for transmission prior to market clearing. Market clearing brings forth a single market price for electricity. For optimal capacity charges the market equilibrium coincides with that of optimal nodal pricing. Capacity charges are based on technical distribution factors and estimates of the shadow prices of network constraints. Estimates can be based on market information from similar congestion-situations, and then brought near the optimal values through an iterative process.

* Forthcoming in Bjørndal, E., M. Bjørndal, P. Pardalos, and M. Rönnqvist (eds.), *Energy, Natural Resource and Environmental Economics*, Springer 2009.

1 Introduction

The goal of deregulating the electricity market has been to achieve efficiency through competition between supply and demand. A special feature of electricity is the reliance on a common network for transmission, where network constraints have special implications for optimal economic dispatch, due to the externalities created by the loop flow. Highly different market designs have been chosen for handling network constraints, with different implications for efficiency. Our objective is to combine several of the suggested approaches, and see whether it is possible to find a good approximation to optimal nodal prices by using system operator announced capacity charges.

In the capacity charge approach capacity constraints are handled by issuing nodal capacity charges. Market clearing brings forth a single market price for energy common to the entire pool. The net nodal price thus equals the common market price less the nodal capacity charges. For positive (negative) capacity charges, the net nodal price is lower (higher) than the market price. Optimally set capacity charges allow the market to reach optimal dispatch as net nodal prices equal optimal nodal prices. Capacity charges are issued by the system operator, and are based partly on technical load factors, and partly on the shadow prices of congested lines. Implementation of the capacity charge approach can be on an ex post or an ex ante basis.

If capacity charges are announced ex post, i.e. *after* bidding and market clearing, the approach is merely a different representation of the nodal pricing method, now with the aggregate effect of capacity constraints in the grid priced explicitly for each node. With ex post announcement of capacity charges, the system operator has full information of shadow prices, and the calculation of capacity charges is straightforward.

On an ex ante basis capacity charges are issued *prior to* bidding and market clearing, and market participants respond to the issued capacity charges in submitting their supply and demand curves. Without full information, as with the ex ante announcement, however, the shadow prices of congested lines have to be estimated. With the ex ante announcement of capacity charges, the implementation of the approach further depends upon the overall design of the electricity market. Organization of competitive electricity markets varies with respect to how real-time balancing is organized, ranging from solutions with electricity spot markets which in effect are real-time markets, to solutions with a day-ahead scheduling market, and where real-time balancing is handled in a separate real-time balancing market.

Within the market framework of a separate day-ahead scheduling market and a real-time balancing market, we assume that the function of the day-ahead market is to reach an efficient schedule which is feasible with respect to expected capacity constraints. The market participants state their bids after the announcement of capacity charges. Thus, the market clearing brings forth a market equilibrium consistent with the estimated shadow prices of expected capacity constraints. The efficiency of the market equilibrium can be improved through an iterative adjustment process to reach an optimal and feasible market solution.

Within the framework of a real-time spot market, the ex ante announcement of capacity charges may affect demand and supply bids. These estimated capacity charges will, however, not be able to clear the market alone, and there is no room for a direct iterative process. This requires using e.g. ex post nodal prices together with the pre-announced capacity charges.

The capacity charge approach offers several advantages. A main issue is the role of capacity charges as important market signals for demand and supply, signaling geographical differences in the nodal cost of aggregate network constraints. Compared to approaches such as zonal pricing, the method also incorporates the advantages of nodal pricing, as capacity charges may be nodally differentiated. Further, as capacity charges apply to all contracts of physical delivery, the method enables spot and bilateral contracts to coexist. Also, as capacity charges are issued by the system operator, and are based on technical information and estimates of shadow prices, this may enable a clearer distinction between the role of the exchange and the system operator, and might facilitate coordination in areas where there for some reason exist multiple exchanges. As for implementing ex ante announced capacity charges in a market based on a real-time spot market only, the combination of ex post nodal prices with pre-announced capacity charges may be a source of enhancing market efficiency. In a market which is cleared real-time, we note that the producer or consumer has to be able to respond automatically to prices in order to submit price-elastic bids. If not, only price-inelastic bids can be submitted. The pre-announcement of capacity charges enables this group to respond on the signals conveyed by the capacity charges, reducing or increasing planned demand according to the signals of expected congestion cost conveyed by the capacity charges.

The rest of the paper is organized as follows: Section 2 discusses the approach in relation to other methods for handling congestion. Section 3 presents the foundation for the capacity charge approach, and shows its relation to optimal dispatch by nodal prices, using a model with a ‘DC’-approximated network. Section 4 exemplifies the approach of optimal capacity charges in a six-node model. Section 5 discusses iterative approaches for implementing the capacity charge method, illustrating with a standard gradient method. Section 6 discusses a heuristic approach in obtaining feasible flows. Section 7 concludes the paper, and future research is discussed.

2 Literature Review

The concept of nodal prices is discussed by Schweppe et al. (1988). Optimal nodal prices are produced by the solution of the welfare maximization problem as the dual prices of the power flow equations, and are interpreted as the value of power in each node (cf. Wu et al. (1996)). A mechanism enforcing optimal nodal prices, where generators and consumers adapt to the local (nodal) market price when deciding on output, ensures social optimum in the short run. Wu et al. (1996), however, point to several counter-intuitive and possibly troublesome characteristics of implementing the nodal pricing approach. For instance, for the system operator to calculate the optimal economic dispatch and implement it, suppliers and consumers must truthfully reveal cost and demand functions, and they may not be willing to give away such strategic information.

On the other hand, the price system suggested by Chao and Peck (1996) represents a system for ‘explicit congestion pricing’, where, instead of providing locational energy prices as nodal prices do, the use of scarce transmission resources is priced. This is accomplished through the design of a trading rule, based on load factors or distribution factors, specifying the transmission capacity rights that traders must acquire in order to complete an electricity transaction. In optimum, Chao-Peck prices are consistent with optimal nodal prices, and in accordance with the shadow prices of the transmission constraints of the optimal dispatch problem. A slight modification of Chao-Peck prices is suggested by Stoft (1998), where a ‘hub’ price is determined by allowing energy bids at any given node (or ‘hub’) in the network. Both mechanisms rest upon a market bringing forward the prices of transmission rights on the links, and the number of prices these systems have to derive is usually far less than the number of nodes in the network. However, although Chao and Peck (1996) and Stoft (1998) give some indications, the specific design of a market mechanism to determine the Chao-Peck prices is still regarded an unresolved problem.

The coordinated multilateral trade model suggested by Wu and Varaiya (1995) is intended to attain optimal dispatch without requiring the system operator to collect private information, i.e. supply and demand curve bids. Instead, brokers carry out profitable multilateral trades under feasibility constraints. More specifically, central coordination is achieved through an iterative process, where, in the case of transmission constraints, loading vectors are announced by the system operator, based on which brokers must evaluate the feasibility of the trades in question. Consequently, the decision mechanisms regarding economics and the feasibility of system operation are separated. Economic decisions are carried out by private multilateral trades among generators and consumers, while the function of ensuring feasibility is coordinated through the system operator who provides publicly accessible data, based upon which generators and consumers can determine profitable trades that meet the secure transmission loading limits.

In relation to the optimal dispatch problem, the coordination models can be interpreted as different relaxation schemes, with competitive players in generation and consumption and the system operator solving different sub-problems, and information is exchanged back and forth. The decompositions corresponding to nodal pricing and Chao-Peck pricing are price-driven. In the case of nodal prices, the system operator hands out the optimal nodal prices of energy obtained after solving the optimal dispatch problem, and optimal dispatch is achieved as producers and consumers adapt to their local prices. For Chao-Peck prices, a market is supposed to bring forth the competitive prices of transmission rights, while the system operator provides information on how trades affect every single link. When traders adapt to the transmission charges of the links imposed by the prices of the transmission rights, the overall problem is solved. The coordinated multilateral trade model can be interpreted as a Benders decomposition, where the market players maximize net profit, and quantities are communicated to the system operator, which checks feasibility and generates constraints. The new constraints must be taken into consideration when additional trades are placed and the process continues. Due to the complexity of electric networks, each method has its limitations in practical use.

In this paper, we combine several of the above approaches. Our objective is to find good approximations of the optimal nodal prices based on an uncongested system price and the loading vectors of congested lines. This approach may be interpreted as a Chao-Peck pricing approach including a ‘hub’, as suggested by Stoft (1998), where we estimate/guess the shadow prices of congested lines. Our approach is also similar to the coordinated multilateral trade model of Wu and Varaiya (1995) in that we need not rely on the disclosure of private information. Compared to Wu and Varaiya (1995), instead of announcing the constraints through the publication of the loading vector, the grid operator announces a set of nodal capacity charges that is based on an estimate/guess of the shadow price of the constraint in question, and the loading vector. The approach is also similar to that of Glavitch and Alvarado (1997) who use market information to estimate cost parameters.

3 The Capacity Charge Approach

The optimal market equilibrium is the market solution which gives the maximum social surplus attainable within the constraints of the system, i.e. the equilibrium which replicates the solution to the optimal economic dispatch problem. In this section we compare the market equilibrium of the capacity charge approach with the optimal economic dispatch.

Consider an electricity market where supply and demand are located in *n* nodes which are interconnected by a constrained transmission grid. Demand in node *i*, q_i^d, depends upon the market price p_i in the node. The demand curve is specified as the general function $p_i^d(q_i^d)$ which is assumed to be non-increasing in q_i^d. Note that for a given price p_i, the corresponding demand q_i^d is found as the inverse of the function. Likewise, supply in node *i*, q_i^s, also depends on the market price p_i of the node, and the supply curve is specified as the general non-decreasing function $p_i^s(q_i^s)$. The supply q_i^s in node *i* for a given nodal price p_i is the inverse of the function.

The social surplus, Π_{ss} is defined as total willingness to pay, less total cost of production, as shown in (3.1). Note that the total willingness to pay is given as the area under the demand curve, while the total cost of production is the area under the supply curve.

$$\Pi_{ss} \equiv \sum_{i=1}^{n} \left[\int_0^{q_i^d} p_i^d(q)dq - \int_0^{q_i^s} p_i^s(q)dq \right] \tag{3.1}$$

where q_i^d and q_i^s are the quantities demanded and supplied in node *i*. Social surplus may be decomposed into demand surplus, supply surplus and grid revenue, the latter due to congestion. These are shown respectively as the three terms of (3.1').

$$\begin{aligned}\Pi_{ss} \equiv & \sum_{i=1}^{n} \left[\int_0^{q_i^d} \left[p_i^d(q) - p_i^d(q_i^d)\right]dq \right] + \sum_{i=1}^{n} \left[\int_0^{q_i^s} \left[p_i^s(q_i^s) - p_i^s(q)\right]dq \right] \\ & + \sum_{i=1}^{n} \left[p_i^d(q_i^d)q_i^d - p_i^s(q_i^s)q_i^s\right]\end{aligned} \tag{3.1'}$$

The transmission grid consists of several lines connecting the nodes. To illustrate the nodal capacity charge approach, we consider real power using the lossless linear 'DC' approximation of power flow equations, with reactances equal to 1 on every link[121]. Each line *ij* of the network is defined by the two nodes *i* and *j* which it interconnects. Let q_{ij} be the flow along line *ij*. If positive, the flow is in the direction from node *i* to node *j*. If negative, the flow is in the direction from node *j* to node *i*. Under the lossless 'DC' approximation we have $q_{ij} = -q_{ji}$.

Production in node *i* is q_i^s, while consumption in the node is q_i^d. The net injection in node *i* is defined by:

$$q_i \equiv q_i^s - q_i^d \tag{3.2}$$

The power flow on each line is determined by Kirchhoff's junction rule, Kirchoff's loop rule, and the Law of conservation of energy.

$$q_i = \sum_{j \neq i} q_{ij} \qquad i = 1, \ldots, n-1 \tag{3.3}$$

$$\sum_{ij \in L_\ell} q_{ij} = 0 \qquad \ell = 1, \ldots, m-n+1 \tag{3.4}$$

$$\sum_i q_i = 0 \tag{3.5}$$

Kirchhoff's junction rule (3.3) states that the current flowing into any node is equal to the current flowing out of it. There are *n* nodes, and there are *n*-1 independent equations. Equation (3.4) follows from Kirchhoff's loop rule that states that the algebraic sum of the potential differences across all components around any loop is zero. The number of independent loops is given by $m-n+1$, where *m* is the number of lines in the grid. $(\bar{L}) = (L_1, \ldots, L_{m-n+1})$ is the set of independent loops[122] and L_ℓ is the set of directed arcs *ij* in a path going through loop ℓ. The law of conservation of energy (3.5), states that, in the absence of losses, total generation is equal total consumption.

In general, for a given network and load, the power flows may be summarized by a matrix of load factors. Each load factor β_{ij}^{lm} shows the fraction of an injection in node *l* with withdrawal in node *m* that flows along line *ij*. Note that $\beta_{ji}^{lm} = -\beta_{ij}^{lm}$ and $\beta_{ij}^{ml} = -\beta_{ij}^{lm}$. Under the 'DC'

[121] The 'DC' approximation is the customary approximation used in economic literature when dealing with the management of transmission constraints. Under these assumptions, and with well-behaving cost and benefit functions, the optimal dispatch problem is convex. For the specifics of the 'DC' approximation, see for instance Wu and Varaiya (1995), Chao and Peck (1996) or Wu et al. (1996). In the 'DC' approximation both losses and reactive power are left out.

[122] See Dolan & Aldous (1993).

approximation the load factors are constants, i.e. independent of load[123]. By introducing a reference point r for withdrawals, the load factors may be represented by a loading vector $\overline{\beta}_{ij}(r) \equiv (\beta_{ij}^{1r} \quad \beta_{ij}^{2r} \quad \dots \quad \beta_{ij}^{nr}) \equiv (\beta_{ij}^{1} \quad \beta_{ij}^{2} \quad \dots \quad \beta_{ij}^{n})$ for each link ij. Element k of loading vector $\overline{\beta}_{ij}(r)$ shows the flow along line ij if 1 MW is injected into node k and withdrawn in the reference point r. A general trade between node l and m, may be viewed as a combined trade between node l and r and between r and m. Thus, we have $\beta_{ij}^{lm} = \beta_{ij}^{lr} + \beta_{ij}^{rm} = \beta_{ij}^{lr} - \beta_{ij}^{mr} = \beta_{ij}^{l} - \beta_{ij}^{m}$. With net injections given, the line flow along line ij is consequently:

$$q_{ij} = \sum_i \beta_{ij}^i q_i \tag{3.6}$$

Capacity constraints $CAP_{ij} \geq 0$ and $CAP_{ji} \geq 0$ on line ij require that $q_{ij} \leq CAP_{ij}$ and $q_{ji} \leq CAP_{ji}$. The capacity constraints may thus be stated as[124]:

$$\sum_i \beta_{kl}^i q_i \leq CAP_{kl} \qquad k = 1,\dots,n, \quad l = 1,\dots,n, \quad k \neq l \tag{3.7}$$

Under the 'DC' approximation, optimal economic dispatch is then given by the following convex optimization problem:

$$\begin{aligned}
&\textit{maximize} && \Pi_{ss} \equiv \sum_{i=1}^{n} \left[\int_0^{q_i^d} p_i^d(q)dq - \int_0^{q_i^s} p_i^s(q)dq \right] && \\
&\textit{subject to} && q_i = q_i^s - q_i^d && i = 1,\dots,n \\
& && q_i = \sum_{i \neq j} q_{ij} && i = 1,\dots,n-1 \\
& && \sum_{ij \in L_\ell} q_{ij} = 0 && \ell = 1,\dots,m-n+1 \\
& && \sum_i q_i = 0 && \\
& && \sum_i \beta_{kl}^i q_i \leq CAP_{kl} && k = 1,\dots,n, \quad l = 1,\dots,n, \quad k \neq l
\end{aligned} \tag{3.8}$$

In the unconstrained case, where neither of the capacity constraints of (3.7) are binding, there will be a uniform price in the market. For the capacity constrained case, where at least one

[123] In general AC systems the load factors depend on the distribution of loads over the network. Our method applies also for general AC systems, however, requiring recalculations of the load factors according to the load.

[124] Note that for a non-existing line ij we have $\beta_{ij}^{lm} = \beta_{ji}^{lm} = 0$, and $CAP_{ij} = CAP_{ji} = 0$.

capacity constraint is binding, nodal prices will differ and may be different for all nodes[125]. If the constraint $q_{kl} \le CAP_{kl}$ is binding, we have $q_{kl} \ge CAP_{kl}$ and thus $q_{kl} \ge 0$. As $q_{lk} = -q_{kl}$, the corresponding constraint $q_{lk} \le CAP_{lk}$ is not binding. Define the shadow prices of (3.7) as $\mu_{kl} \ge 0$. Thus, if $\mu_{kl} > 0$, we have $\mu_{lk} = 0$.

Under the capacity charge approach, the system operator first provides nodal capacity charges, cc_i. On receiving this information, the participants determine supply and demand bids. Market clearing, results in an equilibrium energy price, p, which is common to the entire pool. The capacity charges may be positive or negative. A positive capacity charge, $cc_i > 0$, is defined as the amount cc_i the suppliers in the node pay per unit supplied, or equivalently the amount cc_i the consumers receive per unit consumed. For a negative charge, consumers pay, while producers are compensated. The net nodal price thus equals $p_i = p - cc_i$.

> **Proposition:** *The market equilibrium of the capacity charge approach is in accordance with optimal economic dispatch when capacity charges are optimally defined.*

Proof. If we relax the capacity constraints in (3.8), we obtain the Lagrangian function:

$$L(\bar{\mu}) = \sum_i \left[\int_0^{q_i^d} p_i^d(q)dq - \int_0^{q_i^s} p_i^s(q)dq \right] + \sum_k \sum_l \mu_{kl} \left[CAP_{kl} - \sum_i \beta_{kl}^i (q_i^s - q_i^d) \right] \tag{3.9}$$

For a given vector $\bar{\mu}$ consisting of shadow prices for all lines of the network, the relaxed problem $h(\bar{\mu}) = \{\max L(\bar{\mu}) \text{ s.t.} (3.2)\text{-}(3.5)\}$ provides an upper bound on the objective function value of (3.8). This follows from weak duality. Because of strong duality, solving the dual problem $\min_{\bar{\mu}} h(\bar{\mu})$ also provides the solution to our original problem (3.8). Considering the objective function of the dual problem and rearranging terms, we get:

$$\begin{aligned} L(\bar{\mu}) &= \sum_i \left[\int_0^{q_i^d} p_i^d(q)dq + \sum_k \sum_l \mu_{kl} \beta_{kl}^i q_i^d \right] - \sum_i \left[\int_0^{q_i^s} p_i^s(q)dq + \sum_k \sum_l \mu_{kl} \beta_{kl}^i q_i^s \right] \\ &\quad + \sum_k \sum_l \mu_{kl} CAP_{kl} \\ &= \sum_i \int_0^{q_i^d} \hat{p}_i^d(q)dq - \sum_i \int_0^{q_i^s} \hat{p}_i^s(q)dq + \sum_k \sum_l \mu_{kl} CAP_{kl} \end{aligned} \tag{3.10}$$

[125] Refer to Wu et al. (1996) for the characteristics of optimal nodal prices.

The rearranged Lagrangian function is quite similar to the original (3.1), however, with two alterations. First, the original supply and demand functions are perturbed, and have been shifted by the term $\sum_k \sum_l \mu_{kl} \beta^i_{kl}$, as shown in (3.11).

$$\begin{aligned} \hat{p}^d_i(q) &\equiv p^d_i(q) + \sum_k \sum_l \mu_{kl} \beta^i_{kl} \equiv p^d_i(q) + cc_i \\ \hat{p}^s_i(q) &\equiv p^s_i(q) + \sum_k \sum_l \mu_{kl} \beta^i_{kl} \equiv p^s_i(q) + cc_i \end{aligned} \tag{3.11}$$

This perturbation is equivalent to the shift in supply and demand curves resulting from the capacity charge approach, where suppliers and consumers in node *i* face a capacity charge

$$cc_i = \sum_k \sum_l \mu_{kl} \beta^i_{kl} \tag{3.12}$$

Secondly, we have the addition of the last term $\sum_k \sum_l \mu_{kl} CAP_{kl}$. For a given shadow price vector μ, this term is a constant. For optimal shadow prices, μ^*_{kl}, the term is equal to the Merchandising surplus (cf. Wu et al. (1996)), which is equivalent to our definition of grid revenue in (3.1'). Thus, social optimum is achieved by the system operator issuing optimal capacity charges $cc^*_i = \sum_k \sum_l \mu^*_{kl} \beta^i_{kl}$, and subsequently solving the unconstrained optimal dispatch problem by clearing the market according to the perturbed supply and demand functions of (3.11). ■

In our approach capacity constraints are managed by means of nodal capacity charges, which cause shifts in the supply and demand curves. Thus, constraints are implicitly taken care of, and market equilibrium results from clearing the market on a common energy price, *p*, i.e. the system price. In optimum, the net prices of each node, $p_i = p - cc_i$, are equivalent to the optimal nodal prices of the nodal pricing approach. Note, however, that although optimal capacity charges cc_i are given by (3.12), they are not uniquely defined, but are associated with the load factors. When using load factors associated to a reference point, both the level of the capacity charges and the system price are affected by the chosen reference point. Using optimal shadow prices will however always ensure the same optimal net nodal prices, regardless of which reference point is chosen. As market participants optimally adapt to the net price, i.e. the market energy price corrected by the capacity charge, market equilibrium is not affected by the choice of reference point.

4 Capacity Charges: An Example

To illustrate the nodal capacity charge approach, we construct an electricity market of six nodes, each with production and consumption. We will here as a benchmark show the outcome of the optimal unconstrained and constrained dispatch, the latter using nodal prices. With optimally defined capacity charges, we show that the outcome of the capacity charge approach is identical to that of nodal pricing.

Model and Parameters

Generators are assumed to have quadratic cost functions, with a profit function π^s of the general form $\pi^s = (p - cc_i)q_i^s - \frac{1}{2}c_i {q_i^s}^2$, which gives us linear supply curves. We also assume linear demand functions. Supply and demand curves, including capacity charges, are shown as:

$$p = c_i q_i^s + cc_i \tag{4.1}$$

$$p = a_i + cc_i - b_i q_i^d \tag{4.2}$$

where $a_i > 0$, $b_i > 0$ and $c_i > 0$ are positive parameters. The parameters of our example are shown in table 4.1.

Table 4.1 Parameters

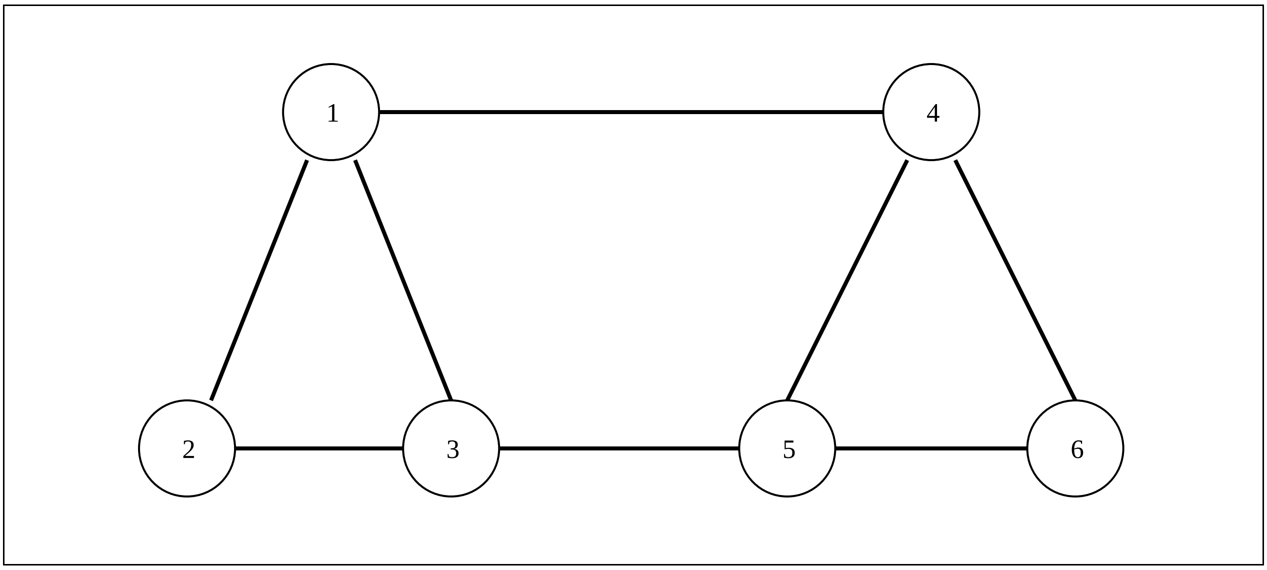
	a_i	b_i	c_i
Node 1	20	0.05	0.2
Node 2	20	0.05	0.1
Node 3	30	0.10	0.7
Node 4	20	0.05	0.2
Node 5	30	0.10	0.7
Node 6	30	0.10	0.1

Social surplus, decomposed into the surpluses of suppliers and consumers, and grid revenue due to congestion, is given by (4.3).

$$\Pi_{ss} \equiv \sum_{i=1}^{6} \tfrac{1}{2}(a_i - p + cc_i)q_i^d + \sum_{i=1}^{6} \tfrac{1}{2}(p - cc_i)q_i^s + \sum_{i=1}^{6} cc_i(q_i^s - q_i^d) \tag{4.3}$$

The network connecting the six nodes is shown in figure 4.1.

Figure 4.1 Network



We apply the lossless linear ‘DC’ approximation of the power flow equations, with reactances equal to 1 on every link. For given net injections, power flows are determined according to (3.3), (3.4), and (3.5), here stated as follows:

$$\begin{aligned} q_1 &= q_{12} + q_{13} + q_{14} \\ q_2 &= -q_{12} + q_{23} \\ q_3 &= -q_{23} - q_{13} + q_{35} \\ q_4 &= -q_{14} + q_{45} + q_{46} \\ q_5 &= -q_{35} - q_{45} + q_{56} \end{aligned} \tag{3.3'}$$

$$\begin{aligned} q_{13} &= q_{12} + q_{23} \\ q_{13} &= q_{14} + q_{45} - q_{35} \\ q_{13} &= q_{14} + q_{46} - q_{56} - q_{35} \end{aligned} \tag{3.4'}$$

$$q_1 + q_2 + q_3 + q_4 + q_5 + q_6 = 0 \tag{3.5'}$$

The load factors β_{ij}^{lm} show the power flow on line *ij* following from an injection of one unit in node *l* and withdrawing it in node *m*. By solving (3.3’), (3.4’) and (3.5’) with $q_l = 1$ and $q_m = -1$, we can represent the power flow of network by a matrix of load factors. The matrix of load factors for our example network is shown in table 4.2.

Table 4.2 Load factors

β_{ij}^{lm}		**TRADES *(lm)***														
		12	**13**	**14**	**15**	**16**	**23**	**24**	**25**	**26**	**34**	**35**	**36**	**45**	**46**	**56**
	12	19/30	4/15	1/10	1/6	2/15	- 11/30	- 8/15	- 7/15	- 1/2	- 1/6	- 1/10	- 2/15	1/15	1/30	- 1/30
L	**13**	4/15	8/15	1/5	1/3	4/15	4/15	- 1/15	1/15	0	- 1/3	- 1/5	- 4/15	2/15	1/15	- 1/15
I	**14**	1/10	1/5	7/10	1/2	3/5	1/10	3/5	2/5	1/2	1/2	3/10	2/5	- 1/5	- 1/10	1/10
N	**23**	- 11/30	4/15	1/10	1/6	2/15	19/30	7/15	8/15	1/2	- 1/6	- 1/10	- 2/15	1/15	1/30	- 1/30
E	**35**	- 1/10	- 1/5	3/10	1/2	2/5	- 1/10	2/5	3/5	1/2	1/2	7/10	3/5	1/5	1/10	- 1/10
S	**45**	1/15	2/15	- 1/5	1/3	1/15	1/15	- 4/15	4/15	0	- 1/3	1/5	- 1/15	8/15	4/15	- 4/15
(i-j)	**46**	1/30	1/15	- 1/10	1/6	8/15	1/30	- 2/15	2/15	1/2	- 1/6	1/10	7/15	4/15	19/30	11/30
	56	- 1/30	- 1/15	1/10	- 1/6	7/15	- 1/30	2/15	- 2/15	1/2	1/6	- 1/10	8/15	- 4/15	11/30	19/30

For example, a trade consisting of injecting 1 MW in node 1 and withdrawing it in node 3, gives a flow over line 12 equal to $\frac{4}{15}$ in direction from 1 to 2. Likewise, injecting 1 MW in node 2 and withdrawing it in node 3, results in a flow over line 12 of $-\frac{11}{30}$, i.e. a flow of $\frac{11}{30}$ from 2 to 1. If combined, the flows over line 12 from the two trades partly cancel, resulting in a net flow over line 12 equal to $q_{12} = \frac{4}{15} + (-\frac{11}{30}) = \frac{1}{10}$.

Alternatively, by introducing a reference point r for withdrawals, the load factors may be represented by the loading vectors $\overline{\beta}_{\mathrm{ij}}(r)$ for each link *ij*. The load factors using reference point 3 are for example given by β_{ij}^{l3} derived from the five columns for trades with node 3 in table 4.2, (columns 13, 23, 34, 35 and 36), and noting that $\beta_{ij}^{ml} = -\beta_{ij}^{lm}$ and that $\beta_{ij}^{33} = 0$ by definition. This loading vector shows line flows for trades from any injection point to the reference point only. Note that all information of table 4.2 is contained in these five columns. A general trade between e.g. node 1 and 6, may be viewed as a combined trade between node 1 and 3 and between 3 and 6. For example, the flow on line 12 resulting from this trade is $\beta_{12}^{16} = \beta_{12}^{13} + \beta_{12}^{36} = \frac{4}{15} + (-\frac{2}{15}) = \frac{2}{15}$.

Unconstrained optimal dispatch

Assuming no congestion in the network, optimal dispatch and maximum social surplus results from aggregating supply and demand curves, and clearing the market so that the prices of all regions are the same, i.e. $p_1 = p_2 = p_3 = p_4 = p_5 = p_6$. Due to the absence of constraints, all resulting flows are feasible, and capacity charges and grid revenue due to congestion are thus equal to 0. In our example, the market price of energy in the scenario of zero capacity charges is 17.09. Table 4.3 shows the optimal unconstrained dispatch of our example. The table also displays total social surplus, 7552.33, and its allocation to production, consumption and the grid.

Table 4.3 Unconstrained dispatch

	Price	Supply	Demand	Net Injection	Supply Surplus	Demand Surplus	Grid Revenue	Total Surplus
Node 1	17.09	85.47	58.14	27.33	730.43	84.51	0.00	814.93
Node 2	17.09	170.93	58.14	112.79	1460.86	84.51	0.00	1545.36
Node 3	17.09	24.42	129.07	-104.65	208.69	832.95	0.00	1041.64
Node 4	17.09	85.47	58.14	27.33	730.43	84.51	0.00	814.93
Node 5	17.09	24.42	129.07	-104.65	208.69	832.95	0.00	1041.64
Node 6	17.09	170.93	129.07	41.86	1460.86	832.95	0.00	2293.81
Total		561.63	561.63	0.00	4799.96	2752.37	0.00	7552.33

Line	Flow
12	-34.40
13	43.99
14	17.73
23	78.39
35	17.73
45	43.99
46	1.07
56	-42.93

Nodal prices

Now, assume that the capacity of the lines are as shown in table 4.4. In the example we assume that $CAP_{ij} = CAP_{ji}$[126]. These capacity constraints make the unconstrained optimal dispatch infeasible, as the constraints of lines 23, 35 and 45 would be violated at this solution.

Table 4.4 Line Capacities

Line	Capacity
12	60
13	60
14	60
23	60
35	10
45	30
46	8
56	60

With the nodal pricing approach optimal nodal prices result in optimal dispatch. Table 4.5 displays optimal dispatch, nodal prices, and the allocation of social surplus. In optimal dispatch, we find that the capacity of lines 23, 45 and 46 are binding. Line 35, although expected, is not constrained, while the flow direction of line 46 has changed and the flow limit is binding in optimal dispatch. The table also shows optimal shadow prices for each line. Since $\mu_{ji} = 0$ if $\mu_{ij} > 0$, we have displayed only one shadow price per line. If positive, it indicates that the constraint CAP_{ij} is binding. If negative, it indicates that the constraint CAP_{ji} is binding, where the absolute value of μ_{ji} is the shadow price of the constraint CAP_{ji}.

[126] In reality, this may not be so, as the grid lines may be operated with different capacities depending on the direction of the flow over the interconnection.

Table 4.5 Optimal dispatch - nodal prices

	Price	**Supply**	**Demand**	**Net Injection**	**Supply Surplus**	**Demand Surplus**	**Grid Revenue**	**Total Surplus**
Node 1	17.05	85.23	59.08	26.15	726.42	87.26	-445.78	813.68
Node 2	16.15	161.47	77.05	84.42	1303.69	148.43	-1363.17	1452.11
Node 3	18.71	26.73	112.89	-86.17	250.06	637.26	1612.20	887.32
Node 4	16.28	81.40	74.39	7.01	662.62	138.36	-114.08	800.98
Node 5	19.48	27.82	105.24	-77.41	270.95	553.74	1507.73	824.69
Node 6	17.30	173.00	127.00	46.00	1496.45	806.45	-795.80	2302.90
Total		555.66	555.66	0.00	4710.18	2371.50	401.11	7482.79

Line	**Flow**	**Shadow price**
12	-24.42	0.00
13	35.58	0.00
14	14.99	0.00
23	60.00	3.46
35	9.41	0.00
45	30.00	6.14
46	-8.00	-1.16
56	-38.00	0.00

This nodal pricing approach follows the concept of nodal prices, as discussed by Schweppe et al. (1988) and Hogan (1992). In order to implement such a system of nodal prices, it is required that the system operator calculates optimal nodal prices on the basis of information of the network, supply and demand.

Optimal Capacity Charges

Under the capacity charge approach a positive or negative capacity charge cc_i is issued to each node, while the market is cleared at a single equilibrium price *p*. If capacity charges are announced prior to market clearing, consumers and producers will take the announced capacity charge into account when deciding supply and demand bids. Market response to optimally defined capacity charges, will result in a feasible and optimal market equilibrium. Table 4.6 shows the optimal capacity charges, when the system price is defined as the unconstrained energy price. Note that the net nodal prices, $p_i = p - cc_i$, equal the optimal nodal prices, and that the resulting market equilibrium and social surplus of the two methods coincide. Likewise, if capacity charges are announced *after* bidding and market clearing, we see that it is straightforward to represent nodal prices by a common market price and nodal capacity charges, so that $p_i = p - cc_i$.

Table 4.6 Capacity charges

	Price	Capacity Charge	Supply	Demand	Net Injection	Supply Surplus	Demand Surplus	Grid Revenue	Total Surplus	Line	Flow
										12	-24.42
Node 1	17.09	0.05	85.23	59.08	26.15	726.42	87.26	1.23	814.91	**13**	35.58
Node 2	17.09	0.95	161.47	77.05	84.42	1303.69	148.43	79.83	1531.95	**14**	14.99
Node 3	17.09	-1.62	26.73	112.89	-86.17	250.06	637.26	139.37	1026.69	**23**	60.00
Node 4	17.09	0.81	81.40	74.39	7.01	662.62	138.36	5.69	806.68	**35**	9.41
Node 5	17.09	-2.38	27.82	105.24	-77.41	270.95	553.74	184.50	1009.19	**45**	30.00
Node 6	17.09	-0.21	173.00	127.00	46.00	1496.45	806.45	-9.52	2293.38	**46**	-8.00
Total			555.66	555.66	0.00	4710.18	2371.50	401.11	7482.79	**56**	-38.00

Optimal capacity charges are defined by (3.12), using the optimal shadow prices from table 4.5, and load factors defined by the physical characteristics of the grid from table 4.2. Load factors are defined relatively to the chosen reference point. The level of both the system price and the capacity charges depend on this chosen point of reference. The net nodal price, $p_i = p - cc_i$, as well as the nodal differences between both net prices and between capacity charges, however, are the same, regardless of the chosen point of reference. Table 4.7 shows examples of optimal sets of energy price and capacity charges, depending on the chosen reference point.

Table 4.7 Optimal capacity charges and energy prices

		Basis for determining capacity charges						
	Node	**Unconstrained energy price**	**Reference Point 1**	**Reference point 2**	**Reference point 3**	**Reference point 4**	**Reference point 5**	**Reference point 6**
Energy Price		17.09	17.05	16.15	18.71	16.28	19.48	17.30
Capacity Charge	**1**	0.05	0.00	-0.90	1.66	-0.77	2.43	0.25
	2	0.95	0.90	0.00	2.56	0.13	3.33	1.15
	3	-1.62	-1.66	-2.56	0.00	-2.43	0.77	-1.41
	4	0.81	0.77	-0.13	2.43	0.00	3.20	1.02
	5	-2.38	-2.43	-3.33	-0.77	-3.20	0.00	-2.18
	6	-0.21	-0.25	-1.15	1.41	-1.02	2.18	0.00

As market participants face identical net nodal prices in all cases, their resulting supply and demand will be the same as in optimal dispatch, i.e. as shown in table 4.5. We will find the same production, consumption, line flows, social surplus and allocation of surplus, including identical grid revenues. Moreover, the grid revenue is equal to the merchandizing surplus under optimal nodal prices. However, although all market participants are equally well off in all cases, we may find that their perception of the situations may differ. For the individual market participant, it may be difficult to see the relation between the market price and the capacity charges. A market participant facing a capacity charge would thus be likely to think of his burden due to the constraint as the capacity charge cc_i he faces, with a total burden of $cc_i q_i^s$ for suppliers and $-cc_i q_i^d$ for consumers. Table 4.8 displays the *perceived* burdens of the consumers and producers due to the constraints.

Table 4.8 Perceived burdens of the constraint

		Basis for determining capacity charges						
	Node	Unconstrained energy price	Reference point 1	Reference point 2	Reference point 3	Reference point 4	Reference point 5	Reference point 6
Producer	1	4.00	0.00	-76.60	141.86	-65.27	207.13	21.64
Consumer		-2,77	0.00	53.09	-98.33	45.24	-143.58	-15.00
Producer	2	152.70	145.12	0.00	413.88	21.46	537.54	186.12
Consumer		-72,87	-69.25	0.00	-197.50	-10.24	-256.50	-88.81
Producer	3	-43.23	-44.49	-68.51	0.00	-64.96	20.47	-37.70
Consumer		182,61	187.91	289.37	0.00	274.36	-86.45	159.24
Producer	4	66.16	62.34	-10.82	197.83	0.00	260.16	83.01
Consumer		-60,46	-56.97	9.89	-180.80	0.00	-237.77	-75.86
Producer	5	-66.31	-67.62	-92.62	-21.31	-88.92	0.00	-60.55
Consumer		250,81	255.75	350.33	80.59	336.34	0.00	229.03
Producer	6	-35.81	-43.93	-199.41	244.02	-176.41	376.50	0.00
Consumer		26,29	32.25	146.39	-179.14	129.50	-276.39	0.00
Total		401.11	401.11	401.11	401.11	401.11	401.11	401.11

For instance, considering the producer in node 6, the choice of node 5 as the reference node leads to a total payment of 376.50, whereas the choice of node 2 as the reference point, induces a total compensation of 199.41. However, since the net prices are identical in all situations, the surpluses of each participant will be the same as shown in table 4.5. For example, the producer in node 6 has a supply surplus of 1496.45 in all cases.

5 An Iterative Adjustment Process

We found that optimally stated capacity charges lead to optimal dispatch. From equation (3.12) we see that the informational requirements for issuing optimal capacity charges are the loading vectors and the shadow prices of congested lines. Loading vectors are technical information, which we assume are readily available. Shadow prices are in principle found by solving the optimal dispatch problem, thus requiring that the system operator has information on cost and benefit functions. When capacity charges are issued prior to bidding and market clearing, shadow prices have to be estimated. Estimated shadow prices can be improved through an iterative process, making use of market responses to obtain good estimates of the shadow prices. Such an iterative approach is similar to that of Wu and Varaiya (1995), however, while they use feasibility and clever market agents (brokers) to arrange multilateral trades, we will use prices and market response to coordinate the nodal markets.

The problem of the system operator in this case is to state capacity charges, based on estimates of the shadow prices of the congested lines, and improved through an iterative process. The iterative approach may be interpreted and implemented in a direct or indirect manner. The direct approach involves a series of actual iterations in clearing the market, where participants after each market clearing receive adjusted capacity charges to which they respond with adjusted supply and demand curve bids. The final market price and capacity charges will be those of the last iteration. While this direct approach would contribute to ensuring 'correct' prices for each point in time, the transaction costs of several iterations for each market clearing could be quite large. Alternatively, the iterative approach may be implemented indirectly. On observing a pattern of congestion similar to earlier periods, the iteration comes about when the system operator uses information on earlier market responses to improve estimates. This may be a more cost efficient method, and justifiable if congestion situations last for a period of time. In this case, market responses to earlier capacity charges may be used to obtain better estimates. It is also possible to start with an estimate based on information obtained from earlier estimates and market observations, and improve capacity charges through a few iterations.

We will illustrate the iterative process by using a simple updating procedure in our example, assuming the direct interpretation, or alternatively, identical market conditions in consecutive time slots. In the example we start with capacity charges set to zero, implying that no lines are congested. This results in the unconstrained dispatch solution, which is not feasible. Alternatively, starting points may be based on forecasts of congested lines and shadow prices.

In each iteration shadow prices, and consequently capacity charges, are updated, to relieve the congested lines. The objective here is to illustrate the approach, rather than finding the most efficient rule in this case. There is a vast literature on algorithms for updating, see for example Minoux (1986). For illustration, we have employed a standard gradient method, where the shadow prices are updated on the general form:

$$\mu_{ij}^{t+1} = \mu_{ij}^{t} + \lambda_{ij}^{t} \frac{\gamma_{ij}^{t}}{\left\|\gamma_{ij}^{t}\right\|} \tag{5.1}$$

where μ_{ij}^{t} is the estimated shadow price of line *ij* at iteration *t*, γ_{ij}^{t} is a gradient of the objective function valued at iteration *t* of our objective function, $\left\|\gamma_{ij}^{t}\right\|$ is a normalization of the gradient, and λ_{ij}^{t} is the step chosen at time *t*. In the example, we have defined the terms as follows:

$$\gamma_{ij}^{t} = \frac{\partial L^{t}}{\partial \mu_{ij}} = CAP_{ij} - \sum_{k} \beta_{ij}^{k} q_{k}^{t} = CAP_{ij} - q_{ij}^{t} \tag{5.2}$$

$$\left\|\gamma_{ij}^{t}\right\| = CAP_{ij} \tag{5.3}$$

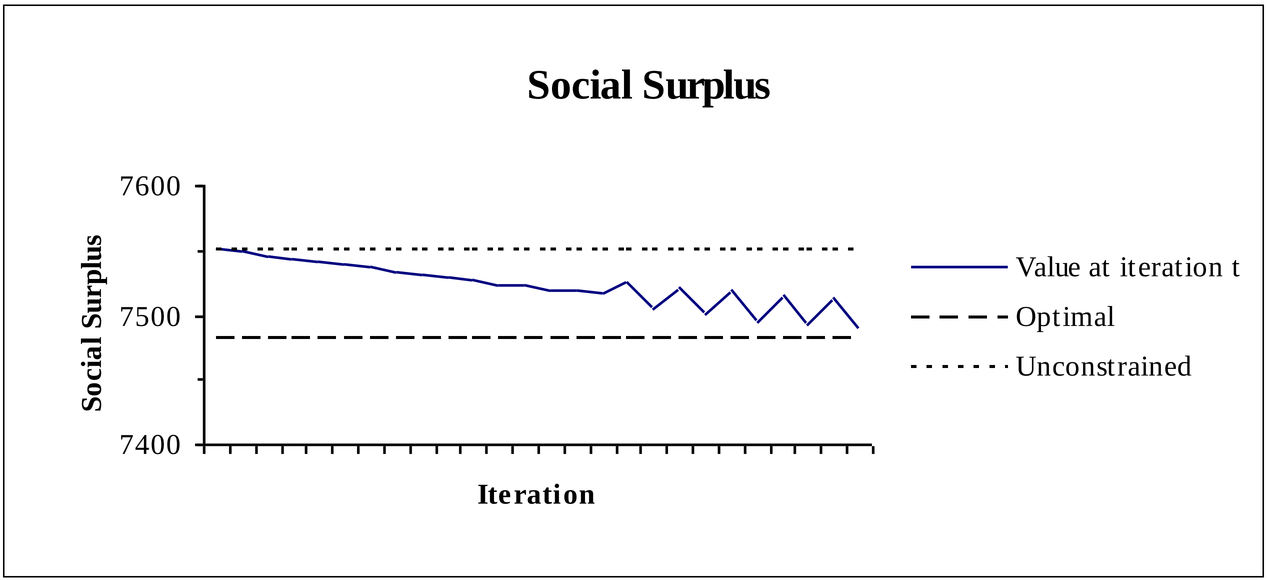
where we see that γ_{ij}^{t} is the under- or over-utilization of the line, and by normalizing by CAP_{ij}, we have the relative under- or over-utilization of the line.

If $CAP_{ij} - q_{ij} < 0$, the line is congested, requiring the shadow price estimate of line *ij* to be raised. If $CAP_{ij} - q_{ij} > 0$, the line is not congested and any shadow price estimate for the line has to be driven towards 0. Thus the step λ_{ij}^{t} is negative. The size of the step λ_{ij}^{t} determines the speed of change. In our example relatively small steps induce a slow convergence towards the optimal value, while larger, but still moderate steps give a faster convergence. However, steps which are large relative to the congestion of the line, cause an oscillation of capacity utilization and shadow price around the optimal values. A mixed definition related to the degree of capacity utilization, for example as shown in (5.4), may give a faster convergence when over-utilization is high, while reducing oscillation around the optimal value.

$$\lambda_{ij}^{t} = \begin{cases} -1 & \text{for} \quad \dfrac{\left|CAP_{ij} - q_{ij}^{t}\right|}{CAP_{ij}} > 0.1 \\[2ex] -0.1 & \text{for} \quad \dfrac{\left|CAP_{ij} - q_{ij}^{t}\right|}{CAP_{ij}} \leq 0.1 \end{cases} \tag{5.4}$$

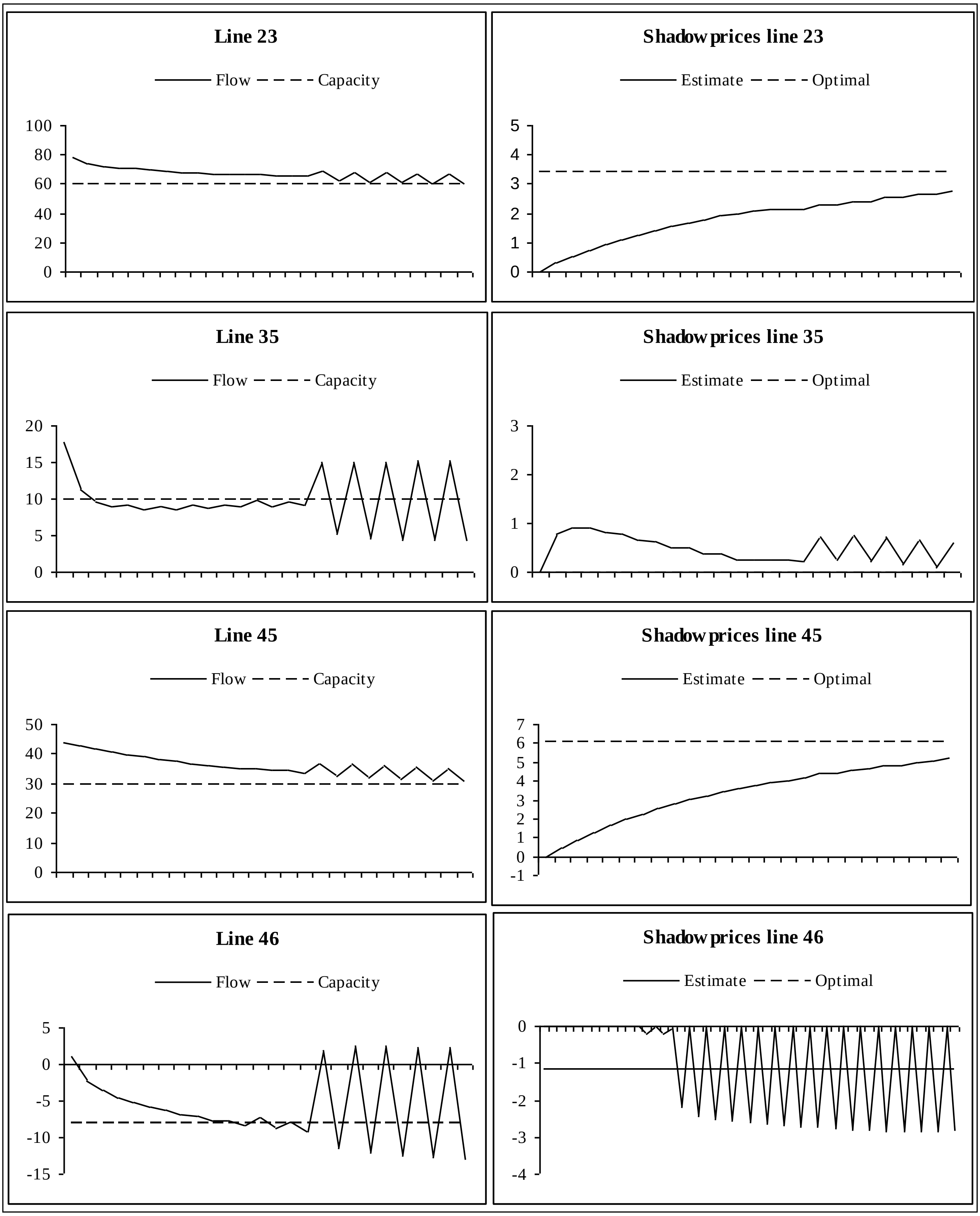
Figure 5.1 shows the resulting development of social surplus in 25 iterations defined by (5.1)-(5.4) and with non-negativity constraints on the shadow prices. Starting from the value connected to the infeasible unconstrained case, social surplus evolves towards the level of the optimal case, however, oscillating due to our rather crude definition of the iteration process.

Figure 5.1 Iterations: Social surplus



The corresponding development of capacity utilization and shadow prices is displayed in figure 5.2. The left hand side displays the capacity and capacity utilization of the lines that are constrained starting with zero capacity charges, and/or that are constrained in optimal dispatch. As in the above tables, capacities and flows of the lines *ij* are displayed as a positive number when the flow is in the direction *i* to *j*, and as a negative number when the flow is in the direction *j* to *i*. The right hand side shows the estimated shadow prices, where we equivalently have determined $\mu_{ij} > 0$ as the shadow price for the capacity in the direction of *i* to *j*, while $|\mu_{ij}|$ is the shadow price of μ_{ji} if $\mu_{ij} < 0$.

Figure 5.2 Iterations: Capacity Utilization and Shadow Prices



In the unconstrained market solution, we find that the flows over lines *23*, *35* and *45* exceed capacity. This indicates a positive shadow price for each of these lines, and the adjustment rule induces a rise in the shadow price estimate. Nodal capacity charges are issued based on these estimates, and the result in the second market clearing is a reduction of the actual flow over these lines. Iterations gradually increase the shadow price estimates of the lines, further reducing the flow over the capacitated lines towards a feasible solution.

It should be noted that the flow over all lines changes as a result of the capacity charges and the corresponding market equilibrium. Line *35* is initially infeasible in the unconstrained solution with no capacity charges, and the shadow price of this line is estimated to be positive. However, though initially infeasible, changes in the flow of other lines due to capacity charges, actually induce a lower flow over line *35*, making the resulting flow feasible in optimum. Lines *12*, *13*, *14*, *46* and *56* are initially feasible. New market solutions as a result of the stated capacity charges causes changes in lines *12*, *13*, *14*, and *56* with flows still within the line capacity. Line *46* starts with an initial feasible flow of 1.07 in the direction from node *4* to node *6*. Changes elsewhere in the network gives a reduced flow, and subsequently a change in the flow direction, from node *6* to node *4*, as illustrated in the figure by negative numbers. Further changes in the network implicate a required flow from *6* to *4* beyond the capacity of the line, thus inducing a positive shadow price estimate.

The above iteration procedure illustrates the use of the gradient as indicating the direction for updating the shadow price estimate. In our case example, we find that with a constant small step, e.g. 0.1, the line flow is driven asymptotically towards the capacity, albeit necessitating a large number of iterations. A higher step size would speed the process, when the shadow price is far from the optimal value, but result in an oscillation around the optimal value when near, as shown in the example. The engineering of more efficient algorithms may reduce the number of iterations called for. However, taking into account the costs of iterations, in order to obtain feasible flows within a small number iterations, the adjustment procedure has to be combined with some other mechanism as e.g. curtailment or counter purchases.

6 A Heuristic Procedure

Above we have illustrated how iterations based on a simple standard gradient method can bring the market solution towards optimal dispatch, and lower the line flows of constrained lines towards capacity limits. We see that this procedure may require a rather large number of iterations to reach the optimal solution. Comparing the cost of further iterations with the gain in social surplus, it may be optimal to terminate the iterative procedure before reaching the optimal solution. However, note that the illustrated procedure is an upper bounding procedure, where the line flows of constrained lines are driven to the capacity limit from above. By prior termination of the iteration procedure, the resulting flow would not be feasible. To find a feasible flow, an alternative is curtailment, however, a problem is to curtail such that the resulting quantities constitute a market equilibrium. An infeasible flow can also be corrected through a secondary market, for instance a market organizing counter-purchases. An alternative or supplement to such cut-off mechanisms is to ‘force’ the iteration itself to reach a feasible solution. This section discusses a heuristic approach for finding a feasible flow that is also a market equilibrium, and as we will see, also brings us near the optimal solution.

The proposed heuristic procedure is based on Everett (1963) and reviewed below. Let us first summarize the problem. Focusing on the capacity constraints, the optimal dispatch problem of (3.8) may be formulated as follows:

$$\begin{array}{lll} \textit{maximize} & \Pi_{ss}(\bar{q}) & \\ \textit{subject to} & q_{kl}(\bar{q}) \leq CAP_{kl} & \forall kl \end{array} \tag{3.8'}$$

where $\bar{q} = (q_1^s, \ldots, q_n^s, q_1^d, \ldots, q_n^d)$ is the vector of production and consumption in each node, and $q_{kl}(\bar{q})$ is the flow on line *kl*. The shadow price vector $\bar{\mu}$ gives us the shadow prices for all lines in both directions. The Lagrangian function is:

$$L(\bar{\mu}) = \Pi_{ss}(\bar{q}) + \sum_k \sum_l \mu_{kl} \left[CAP_{kl} - q_{kl}(\bar{q})\right] \tag{3.9'}$$

In this setting Everett’s theorem can be stated as follows:

1. Choose an arbitrary vector $\bar{\mu}$ of non-negative shadow prices for all lines.

2. Find a solution $\bar{q}^*$ which maximizes the unconstrained Lagrangian function $L(\bar{\mu})$.

3. Then, $\bar{q}^*$ is the solution to the constrained maximization problem with the same objective function as (3.8'), but with modified capacity limits $CAP_{kl}^{'}$, given by $CAP_{kl}^{'} = q_{kl}(\bar{q}^*)$.

The ex ante announcement of capacity charges is easily interpreted within this theorem. In order to set capacity charges according to (3.12), i.e. $cc_i(\bar{\mu})$, the system operator has to choose an arbitrary set of shadow prices $\bar{\mu}$. Market participants bid on the basis of the capacity charges. By clearing the market on a single spot price, we thus actually find the unconstrained maximum of the new Lagrangian function. The solution $\bar{q}^*$ is a true market equilibrium, and is both feasible and optimal with respect to the *modified* constraints. Our problem is to reach a feasible and optimal solution within the *real* constraints of the network, calling for an adjustment of shadow prices through an iterative process. Different sets of $\bar{\mu}$ lead to different flows. However, if the chosen $\bar{\mu}$ gives a feasible solution, and the flow actually equals the real constraint for binding constraints, the Everett theorem states that this solution also is the optimal solution to the original constrained problem.

Based on this insight, we have slightly modified the updating procedure. The main issue is to reach shadow prices (and thus the capacity charges) that will give rise to a flow that matches the constraints. As a heuristic we have tried to speed movement towards a feasible solution by simply exaggerating the over-utilization by redefining the capacity size used in calculating both the gradient and the step. Starting with a market equilibrium that causes an infeasible flow over line *ij*, the updating procedure is still given by (5.1) - (5.4), with one alteration. We have substituted the real capacity CAP_{ij} with $CAP_{ij}^{'} = \alpha_{ij} CAP_{ij}$, where $\left|\alpha_{ij}\right| < 1$. The relative over-utilization of the line will then be exaggerated, the normalized gradient will be larger, and with the specified step of (5.4), the step also might be higher. This causes a faster adjustment of the shadow prices towards a choice that gives a feasible flow. Note that not all lines become feasible within the same iteration. Thus, even though a given line becomes feasible within a given market equilibrium, further iterations may be called for to obtain feasible flows over other lines. Subsequent iterations however slightly alter equilibrium supply and demand, causing changes in line flow over all lines. There will thus still occur changes in the line flow over lines that have become feasible. The moment the flow of a given line becomes feasible with respect to the real capacity, any further adjustments in shadow prices have been made using the real capacity in (5.1) - (5.4).

To obtain a market equilibrium, supply and demand must balance according to the net prices of the nodes, and at the same time obtain a flow that is feasible. There exists many market equilibria, though they do not represent the optimal dispatch, cf. Wu et al. (1996). Market clearing ensures that supply and demand balances. By Everett's theorem we find that if the line flows defined by the market solution are feasible and equal the capacity limit for binding constraints, the resulting market solution is indeed the optimal dispatch.

Figures 6.1 and 6.2 illustrate a simple version of the heuristic, where $\alpha_{ij} = 0.7$ is constant and equal for all lines. In figure 6.1 we see that social surplus more quickly advances the optimal level of social surplus than is the case without the heuristic.

Figure 6.1 Iterations: Social surplus

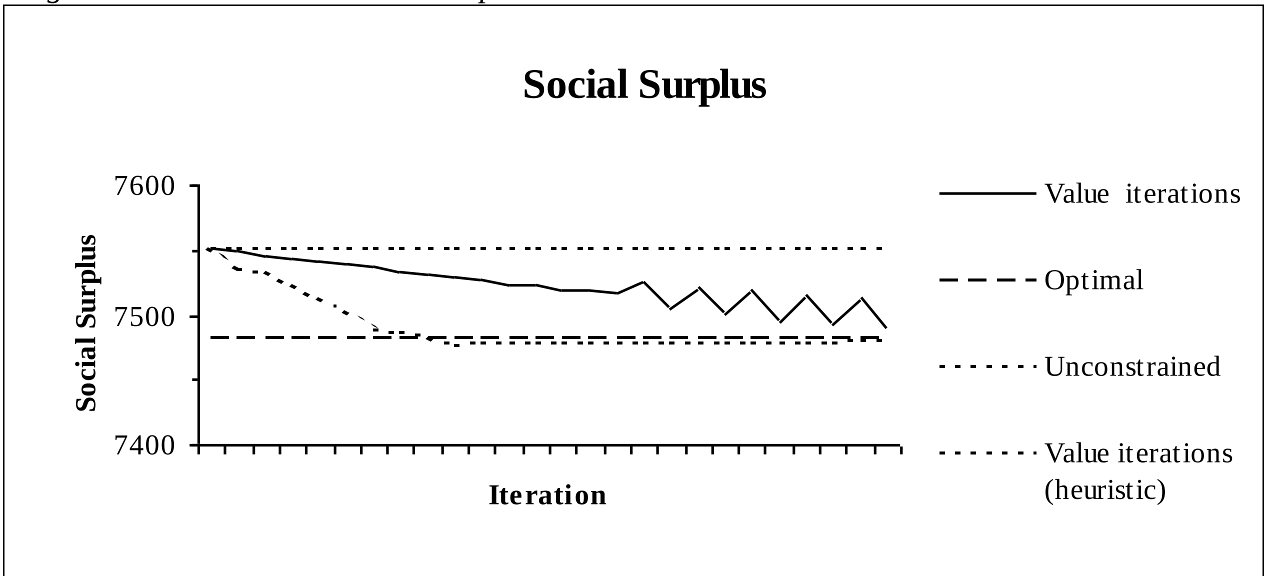
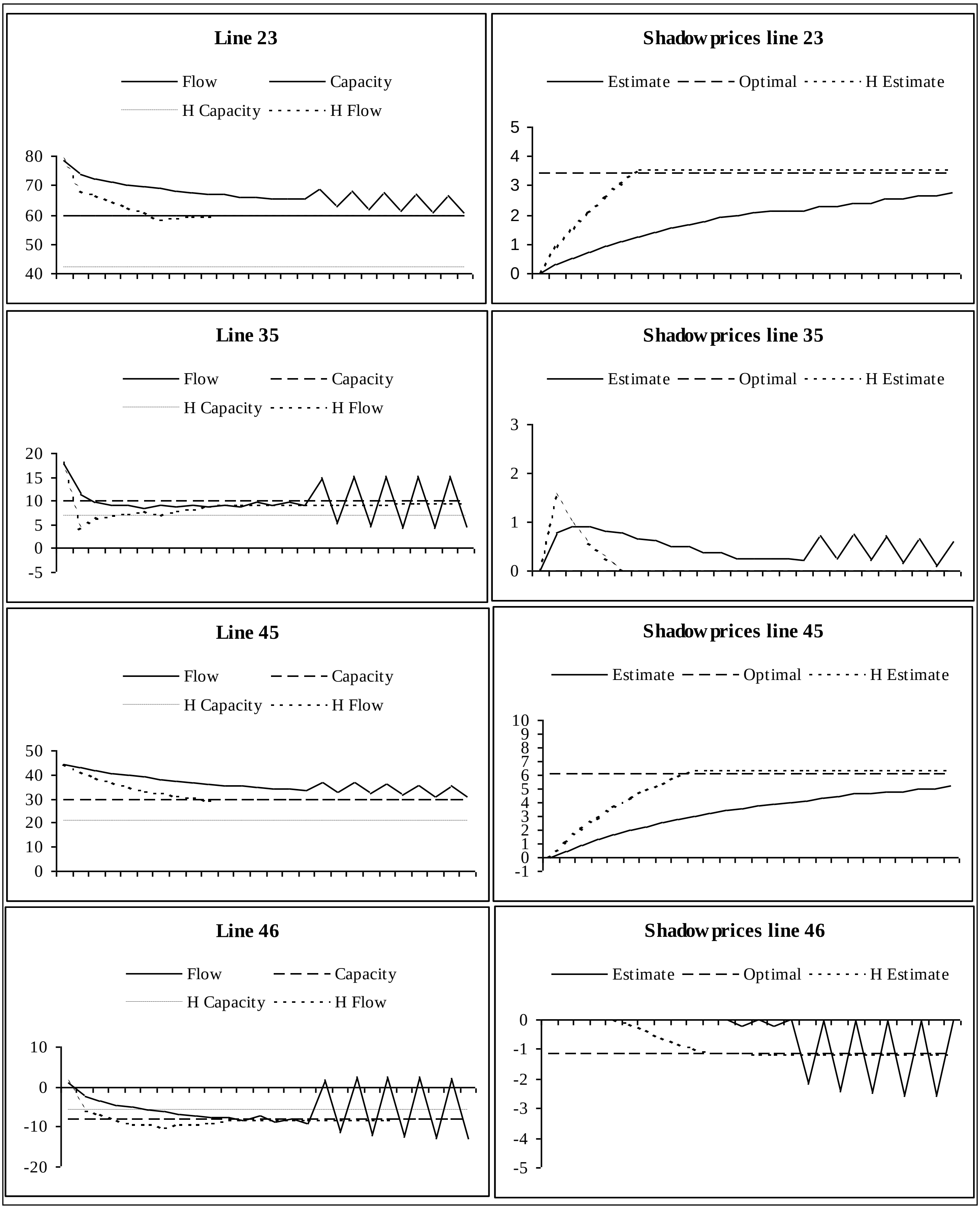


Figure 6.2 shows how the heuristic effects line flow and shadow price estimates. Even with this rather crude definition of the heuristic, using the same α_{ij} for all lines, we see that the line flows of the constrained lines become feasible in less iterations. Also note that the estimated shadow prices come close to the optimal shadow prices.

Figure 6.2 Iterations: Capacity Utilization and Shadow Prices



7 Summary and Topics for Future Research

Electricity markets have been reorganized introducing competition in supply and demand to achieve greater efficiency. Capacity constraints in electricity networks represent rather complex constraints on an efficient market solution, due to the externalities created by the loop flow. Several different market designs have been chosen for handling network constraints. Optimal nodal prices, as the solution of the welfare maximization problem, represent an efficient market equilibrium, where optimal nodal prices reflect both supply and demand conditions, together with the constraints of the network. In the capacity charge approach we combine several of the suggested approaches, while at the same time trying to withhold the optimal characteristics of nodal prices. An objective is to find good approximations of the optimal nodal prices based on a competitive market price and nodal capacity charges issued by the system operator.

A main issue of the approach is the role of capacity charges as important market signals for demand and supply. The capacity charge approach is not a flowgate based approach, but signals the nodal cost of aggregate congestion in the network. Compared to an approach such as zonal pricing, the method incorporates the advantages of nodal pricing. Capacity charges apply to all contracts of physical delivery, facilitating the co-existence of exchange traded and bilaterally traded contracts. The calculation of capacity charges is based on technical information, together with estimates of shadow prices. The capacity charges are issued by the system operator. This recognizes the coordination task performed by the system operator, as well as the system operator’s access to important information on physical dispatch. At the same time, the approach enables a clearer distinction between the role of the exchange and the system operator, and might facilitate coordination in areas with separate exchanges. The efficiency of the approach is, however, contingent on the quality of shadow price estimates. We show that estimated shadow prices can be improved through a (direct or indirect) iterative process, making use of market responses to obtain good estimates of the shadow prices. A potential source of inefficiency is related to the system operator’s incentives for stating capacity charges that boost their revenue. This is an issue for further investigation, and should be seen in connection with the regulation of the grid-company.

In principle, capacity charges may be announced before or after market bidding. If capacity charges are announced after market clearing, the approach is equal to the nodal pricing approach. When announced prior to bidding and market clearing, capacity charges give signals to the market as to expected congestion. Within the market framework of a separate day-ahead scheduling market, and a separate real-time balancing market, as the Norwegian system, market participants act upon this information in stating their supply and demand bids. The efficiency of the market equilibrium can be improved through an iterative adjustment process to reach an optimal and feasible solution. Within the framework of a real-time spot market, the ex ante announcement of capacity charges signals expected congestion prices, and may induce a shift in supply and demand which otherwise would be inelastic with respect to the real-time spot price.

We think the capacity charge approach may be a realistic procedure for managing transmission constraints in a competitive electricity market, for example in the Nordic scheduled day-ahead power market. The capacity charge approach bears a close relation to the system used today. The market operates with a general price referred to as the 'system price of energy' which is the price of unconstrained dispatch. Estimated congestions at the time of market clearing are mainly managed through zonal pricing. With the capacity charge approach, however, capacity charges are announced prior to market clearing, and for each node instead of only for a few zones. Also, the use of nodal capacity charges is similar to what is already done for marginal losses, in that loss factors are published for every connection point in the central high voltage grid. We also believe that the suggested procedure will easily facilitate bilateral trading to go alongside with the pool. As for the iterative process, we note that many congestion situations seem to last for a period of time, and that the zonal division only comes into practice when the congestion is expected to last for several days. In this case, market responses of initial capacity charge estimates might be used to obtain better estimates. To ensure real-time balance of the system, this is handled by through the regulating market, a market that is already present.

Further development of the capacity charge approach, however, still leaves a number of questions to be investigated. First, we would like to do numerical tests on larger networks, and the exact procedure used for adjusting shadow price estimates is of special interest. In this, we can rely on optimization theory, for instance to find an adjustment scheme producing near-optimal dual variables within a few iterations. In this setting, we should also consider how close to optimality we need to be in order for the system to perform satisfactorily. In this case the procedure has to be combined with some other mechanism to obtain feasible flows.

Moreover, it would be interesting to perform simulations, where market data are slightly perturbed in each step. This would simulate how grid revenue develops in an adjustment process where we use market information from similar congestion situations to obtain initial guesses on the shadow prices. Since the suggested procedure also involves a mechanism like for instance curtailment or counter purchases to obtain feasible flows, the specific design of the mechanism is of special interest, also taking into account how it affects the performance of e.g. separate real-time balancing power markets. Furthermore, gaming possibilities and more generally, regulatory issues should be examined.

Employing the 'DC' approximation of the power flow equations, we do not focus on transmission losses, although a complete system for transmission pricing should address losses as well. We believe that marginal losses can be readily taken care of in our approach by issuing nodal loss factors similar to what is already done in the present Norwegian system, but possibly with a 'hub' as a reference point when computing loss factors. If the pool is located at the 'hub', the clearing price of the pool could be used in pricing marginal losses. One of the other simplifications of the 'DC' model is that load factors are constants. The non-linear nature of the AC power flows implies that loss factors depend on loads, however our approach is still valid if the load factors we use are marginal load factors (Stoft (1998)). So, if we are to use the full AC power flow model, in principle, we have to recalculate the load factors whenever the load changes, in addition to solving non-linear systems.

References

Bjørndal, M. & K. Jørnsten (2001), 'Zonal Pricing in a Deregulated Electricity Market', The Energy Journal, 22, 1, 51-73.

Chao, H.-P., & S. Peck (1996), 'A Market Mechanism for Electric Power Transmission,' Journal of Regulatory Economics, 10, 25-59.

Dolan, A. & J. Aldous (1993), *Networks and Algorithms: An Introductory Approach*, John Wiley & Sons.

Everett, H., III, (1963), 'Generalized Lagrange Multiplier Method for Solving Problems of Optimum Allocation of Resources', *Operations Research*, 11, 399-417.

Glavitch, H., & F. Alvarado (1997), 'Management of Multiple Congested Conditions in Unbundled Operation of a Power System,' IEEE Transactions on Power Systems, 13, 374-380.

Hogan, W. W. (1992), 'Contract Networks for Electric Power Transmission', Journal of Regulatory Economics, 4, 211-242.

Minoux, M. (1986) *Mathematical Programming. Theory and Algorithms*, John Wiley and sons.

Schweppe, F. C., M. C. Caramanis, R. D. Tabors, & R. E. Bohn (1988), Spot Pricing of Electricity, Kluwer Academic Publishers.

Stoft, S. (1998), 'Congestion Pricing with Fewer Prices than Zones,' The Electricity Journal (May), 23-31.

Wu, F., & P. Varaiya (1995), 'Coordinated Multilateral Trades for Electric Power Networks: Theory and Implementation,' Department of Electrical Engineering and Computer Sciences, University of California.

Wu, F., P. Varaiya, P. Spiller & S. Oren (1996), 'Folk Theorems on Transmission Access: Proofs and Counterexamples', Journal of Regulatory Economics, 10, 5-23.

ESSAY 3

Understanding the Stochastics of Nodal Prices: Price Processes in a Constrained Network*

Arne-Christian Lund

Linda Rud

Abstract

Network congestion in competitive electricity markets may be managed by geographically differentiated nodal prices. The stochastics of an unconstrained equilibrium price reflect the underlying fundamentals of demand and supply. The stochastics of nodal prices in addition reflect the consequences of grid congestion. This paper demonstrates how a static three-node model may be combined with dynamic modeling of fundamental parameters, giving stochastic nodal price processes consistent with the underlying grid. These price processes may be employed in analyzing production, hedging, and investment decisions under uncertainty.

* The authors would like to thank Mette Bjørndal for important contributions. Any errors are the sole responsibility of the authors. The paper is also included in A.-C. Lund (2004): *Stochastic modeling approaches in economics & finance*, PhD Thesis, Norwegian School of Economics and Business Administration, November 2004.

1 Introduction

Electricity prices in competitive markets have proven to be highly volatile. A thorough understanding of the stochastic price processes is important in e.g. production planning, risk management and investment planning. The stochastics of equilibrium commodity prices in general reflect the stochastic nature of the underlying fundamentals of demand and supply. For electricity, however, a special factor is the effect of the constraints in the underlying grid. By simply clearing the market on a common equilibrium price, often denoted the *system price*, the resulting allocation of production and consumption may be infeasible due to grid congestion. In the case of congestion a feasible market solution may be obtained by clearing the market on differentiated nodal prices[127]. Optimal nodal prices may be characterized as the prices that optimize the aggregate social surplus implied by the bid curves, *within* the capacity constraints of the grid. For the individual consumers and producers, the relevant prices processes are those of the area/nodal specific prices, not the system price process. The objective of our paper is to gain insight into the impact of grid limitations on the stochastic nodal price processes.

In literature, there have been two main approaches to modeling electricity prices. Firstly, there are several contributions which focus on finding appropriate stochastic processes to model a given time series of electricity prices. The processes are in turn applied to e.g. valuing contingent claims. For example, Lucia and Schwartz (2002) study the use of different one and two factor models in modeling the system spot price on the Nord Pool spot exchange. These processes are fitted to data, and then used to value futures and forward contracts. Weron et al. (2004) formulates a jump diffusion model, and a regime switching model for the spot price process. The

[127] The concept of nodal electricity prices is first discussed in Schweppe et al. (1988).

parameters in their models are fitted to Nord Pool price data. Johnson and Barz (1999) discuss 8 diffusion and jump diffusion models. By maximum likelihood methods these models are calibrated to the spot prices at four different energy markets.

The focus of all these models is to find appropriate stochastic processes to describe the spot price. An advantage of this approach is related to the abundance of price data, allowing the processes to be fitted directly to the observed prices series. It does not, however, take into account that the market in periods may be separated due to transmission constraints arising from the specific geographical distribution of supply and demand. Nor does it model the relation between the prices of the separated markets. Thus, the resulting price processes are not directly relevant for the individual market participants for which the nodal price differences are important.

Secondly, there are partial equilibrium models based on models of the underlying market and its transmission constraints, where market clearing nodal/area prices are found, given the geographical dispersion of supply and demand[128]. The main contribution is related to understanding the effect of nodal prices, as well as different methods for handling congestion. These models have, however, been inherently static, or two-periodic at most. Though congestion issues are well described, these models do not capture the underlying dynamics of the stochastic price processes, and can at most be regarded as giving a ‘snap shot’ of the real world. In evaluating investments, production plans, etc., under uncertainty, these models do not provide a sufficient input to handle the challenges of future uncertainty.

In this paper, we seek to combine features of both these traditions, by introducing dynamics and stochastics into the static nodal price models, thus obtaining nodal price processes consistent with the physical laws of the transmission net. Demand and supply are represented by given price-elastic

[128] See for example, Bjørndal (2000) and the references therein.

functions. Uncertainty is introduced by specifying demand or supply parameters as stochastic processes. A related idea is represented in Barlow (2002), where the market price process is characterized on the basis of an inelastic demand market with a static functional form of the supply functions. We have in addition modeled an underlying three-node network, where the market is cleared by nodal prices. These nodal and system price processes may further be employed in analyzing e.g. strategies of production, or of investments in net and production capacity when the nodal price is the relevant price.

The paper is organized as follows: Section 2 presents our model. Equilibrium prices are derived both in the presence and the absence of binding capacity constraints, and the system price and nodal prices are characterized. In section 3 we assume that the above market situation occurs repetitively with consecutive and independent market equilibria. We introduce uncertainty by characterizing demand as a stochastic function. Different choices of stochastic functions may represent different assumptions as to e.g. daily and seasonal variations. Based on the choice of stochastic process for demand in section 3, section 4 looks at the characteristics of the resulting stochastic nodal price processes, as well as discussing other applications of the model. Section 5 discusses different choices of model specification, while section 6 concludes the paper.

2 Price Formation in a Three-Node Electricity Market

Consider a simple three-node electricity market, with generation in node 1 and 2, and demand in node 3. The Generators in node 1 and 2 have the quadratic profit functions

$$\pi_i^s = p_i q_i^s - \tfrac{1}{2} c_i {q_i^s}^2 . \tag{1}$$

Here p_i is the price, q_i^s is the quantity supplied, and c_i is the cost factor in node *i* = 1,2. This gives the linear supply functions

$$p_i = c_i q_i^s \tag{2}$$

where production in each node, $q_i^s = \frac{1}{c_i} p_i$, follows directly from (2). All demand is consumed in node 3, i.e. $Q_D \equiv q_3^d$. Assuming no losses, in equilibrium aggregate supply $Q_S \equiv q_1^s + q_2^s$ equals aggregate demand, i.e. $Q_S = Q_D$. The implicit benefit of demand, π_3^d, for a given point of time is assumed to be of the form

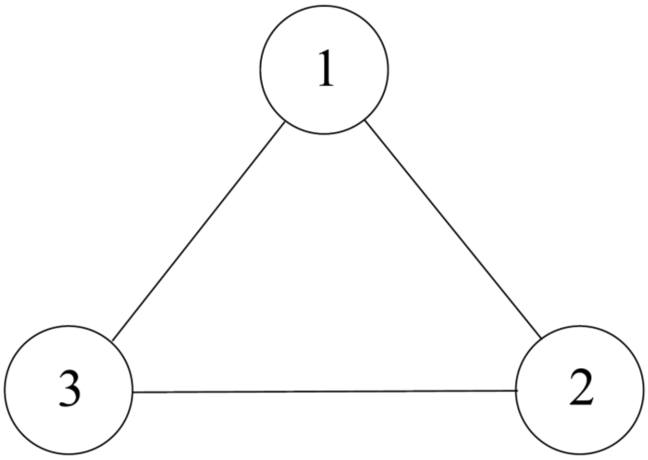
$$\pi_3^d = (a - p_3) Q_D - \tfrac{1}{2} b Q_D^2 , \tag{3}$$

giving the linear demand function

$$p_3 = a - b Q_D \tag{4}$$

where $a, b > 0$. The consumed quantity in node 3 is therefore $Q_D = \frac{1}{b}(a - p_3)$ when the nodal price is p_3.

The three nodes are interconnected by a simple three-line 'DC' network, as illustrated in figure 1.



Each line is assumed to have identical technical characteristics and impedances equal to 1. The flow over each line is solely determined by physical laws. In our simple network the line flows over lines *ij* in the direction from *i* to *j* resulting from production in node 1 and 2 are given by the equations of (5)[129]. For flows in the opposite direction we have $q_{ji} = -q_{ij}$.

$$\begin{aligned} q_{12} &= \tfrac{1}{3} q_1^s - \tfrac{1}{3} q_2^s \\ q_{13} &= \tfrac{2}{3} q_1^s + \tfrac{1}{3} q_2^s \\ q_{23} &= \tfrac{1}{3} q_1^s + \tfrac{2}{3} q_2^s \end{aligned} \tag{5}$$

[129] These equations follow from three physical laws. For this network we have Kirchhoff's Junction Rule, $q_i = \sum_{i \neq j} q_{ij} \ for\, i = 1,2$, that states that the total current out of the production node has to equal the total input to the production node. Following from Kirchhoff's Loop Rule we have $q_{13} = q_{12} + q_{23}$ which in the absence of losses states that the algebraic sum of flows over any path in a loop is equal. The Law of Conservation of Energy states that total generation equals total consumption, $q_3^d = q_1^s + q_2^s$.

2.1 No Capacity Limits

Let us first assume that there are no capacity limits in the network. The challenge of the market is then to match supply and demand. In this case the market will be cleared at a common price $p_s = p_1 = p_2 = p_3$. This equilibrium price of a non-capacitated market is also termed the system price. The aggregate supply curve as a function of the market price p_s is

$$Q_S = \frac{1}{c} p_s \tag{6}$$

where $c \equiv \frac{c_1 c_2}{c_1 + c_2}$ is the resulting cost factor of aggregate supply. The proportion of total generation generated in node i is now $\alpha_i \equiv \frac{q_i^s}{q_1^s + q_2^s}$. By inserting (2) we have the production weights

$$\alpha_i = \frac{c_{3-i}}{c_1 + c_2} \qquad for\, i = 1,2. \tag{7}$$

The market is cleared at the price which matches supply and demand, i.e. for $Q_S = Q_D$, where Q_D is given by (4) and Q_S by (6). The equilibrium price, the system price, is thus given by $p_s = \frac{ac}{b+c}$, giving

$$p_s = \frac{ac_1c_2}{bc_1 + bc_2 + c_1c_2}. \tag{8}$$

The corresponding total equilibrium quantity is given by $Q_S = Q_D = \frac{a}{b+c}$, i.e.

$$Q_S = Q_D = \frac{a(c_1 + c_2)}{bc_1 + bc_2 + c_1c_2} \tag{9}$$

In this unrestrained solution the proportions of the total production produced in node 1 and 2 are given by the least-cost production mix α_1 and α_2 of (7). Thus, nodal production is $q_i^s = \frac{\alpha_i a}{b+c}$, giving

$$q_i^s = \frac{ac_{3-i}}{bc_1 + bc_2 + c_1c_2} \quad for\, i = 1,2 \tag{10}$$

By inserting the above flow quantities into the line flow equations (5), we find the line flows in an unrestricted market clearing:

$$\begin{aligned} q_{12} &= \frac{a(c_2 - c_1)}{3(bc_1 + bc_2 + c_1c_2)} \\ q_{13} &= \frac{a(c_1 + 2c_2)}{3(bc_1 + bc_2 + c_1c_2)} \\ q_{23} &= \frac{a(2c_1 + c_2)}{3(bc_1 + bc_2 + c_1c_2)} \end{aligned} \tag{11}$$

2.2 Capacity Limit on Line 12

In the case of capacity limits in the network, the optimal market equilibrium is defined as the allocation that maximizes social surplus, given the constraints of the network. Now let us assume that the capacity of line 12 is restricted to $\hat{C}$ for flows in either direction, while the capacity of the other lines will not be binding. The capacity of line 12 will thus be binding for any combination of parameters that satisfies the inequality $|q_{12}| > \hat{C}$, i.e.

$$\left| \frac{a(c_2 - c_1)}{3(bc_1 + bc_2 + c_1c_2)} \right| > \hat{C} \tag{12}$$

In section 3 we will assume a repetitive market where the intersect *a* of the demand curve is stochastic. The slope of demand and supply curves, *b*, c_1, and c_2, are for simplicity chosen to be constants. In this setting it is therefore the level of *a* that determines whether the capacity limit of line 12 will be binding. To simplify, let us define the producer in node 1 as the least cost producer assuming that $c_1 < c_2$. This implies that $q_1^s > q_2^s$ with a constant direction of the line flow on line 12, i.e. $q_{12} > 0$. Clearing the market at a uniform price the constraint on line 12 becomes binding when the demand reaches the level which solves $q_{12}(\hat{a}) = \hat{C}$. This critical demand intersect level is

$$\hat{a} = 3\hat{C}\frac{bc_1 + bc_2 + c_1c_2}{c_2 - c_1} \tag{13}$$

By inserting $\hat{a}$ from (13) into (9) and (10) we find the corresponding levels of production in node 1 and 2, $\hat{q}_1^s$ and $\hat{q}_2^s$, and consumption, $\hat{Q}_D$;

$$\hat{q}_i^s = 3\hat{C}\frac{c_{3-i}}{c_2 - c_1} \quad for\, i = 1,2 \tag{14}$$

$$\hat{Q}_D = 3\hat{C}\frac{c_1 + c_2}{c_2 - c_1} \tag{15}$$

For a realization of $a \le \hat{a}$ the unrestrained solution resulting from a common market price p_s is feasible. At $a = \hat{a}$ the line q_{12} is fully utilized, and the unrestrained market solution with the above quantities of (14) and (15) is feasible. This quantity, $\hat{Q}_D = \hat{Q}_S$, is the maximum feasible aggregate production given the production mix α_1 and α_2 of the unrestrained solution, which also is the minimum-cost production mix.

A demand realization $a > \hat{a}$ calls for a higher total production and consumption. The unrestrained solution based on a uniform market price p_s is however now not feasible. Still, it is in fact possible to achieve higher production levels, i.e. $Q_S > \hat{Q}_S$ without violating the capacity constraints of line 12. The clue is to define differentiated nodal prices and change the production mix. This can be seen by studying the line flows resulting from Kirchhoff's law. For each extra unit produced in node 1, equation (5) states that $\frac{1}{3}$ of the quantity will flow on line 12 from node 1 to node 2. This action alone is not possible if the line is already congested. We however find that equation (5) also states that for each extra unit produced in node 2, $\frac{1}{3}$ of this quantity will flow on line 12 from node 2 to node 1. Additional quantities $(Q_S - \hat{Q}_S)$ are thus feasible by producing equal additional amounts in each node, i.e. with the production mix $\hat{\alpha}_1 = \hat{\alpha}_2 = \frac{1}{2}$.

We have a new kinked aggregate supply curve, which for quantities above $\hat{Q}_S$ reflect the marginal cost of additional production using the new production mix. For any aggregate production $Q_S > \hat{Q}_S$ (occurring when $a > \hat{a}$, the production in node i is given by $q_i^s = \alpha_i \hat{Q}_S + \hat{\alpha}_i (Q_S - \hat{Q}_S)$, i.e. with the least cost mix α_i for the critical level of non-congested quantity, and a mix of $\hat{\alpha}_i = \frac{1}{2}$ for any additional quantity. By substituting for $\hat{Q}_S$, α_i, and $\hat{\alpha}_i$, the feasible nodal production as a function of any total production $Q_S > \hat{Q}_S$ is

$$q_i^s = \tfrac{1}{2} Q_S + (3 - 2i) \tfrac{3}{2} \hat{C} \quad \text{for } i = 1{,}2 . \tag{16}$$

For any $Q_S > \hat{Q}_S$, the aggregate cost of production, $\Pi_C = \frac{1}{2} \sum_{i=1,2} c_i {q_i^s}^2$ is

$$\Pi_C = \tfrac{1}{2} \sum_{i=1,2} c_i \left[\alpha_i \hat{Q}_S + \hat{\alpha}_i (Q_S - \hat{Q}_S) \right]^2 \tag{17}$$

giving the marginal cost of

$$\frac{\partial \Pi_C}{\partial Q_S} = \tfrac{1}{4} (c_1 + c_2) Q_S + \tfrac{3}{4} (c_1 - c_2) \hat{C}_{12} \tag{18}$$

The aggregate benefit of consumption, Π_B, is given by omitting the payments in (3), i.e. $\Pi_B = a Q_D - \frac{1}{2} b {Q_D}^2$, giving the marginal benefit of consumption

$$\frac{\partial \Pi_B}{\partial Q_D} = a - b Q_D$$

By equating the marginal cost of aggregate production and the marginal benefit of consumption, we find the equilibrium amount $Q^* = Q_S = Q_D$ for $a > \hat{a}$ to be

$$Q^* = \frac{4a + 3\hat{C}(c_2 - c_1)}{4b + c_1 + c_2} \tag{19}$$

Given the equilibrium amount Q^* the quantities in each node are found by (4) and (16).

To achieve these quantities and balance the resulting regional markets, it is necessary to operate with nodal prices tailored to induce the required quantities. Prices in node 1 and 2 are given by the individual supply curves of (2), $p_i = c_i q_i^s$. For $a > \hat{a}$ prices as a function of the total quantity supplied are

$$p_1 = c_1(\tfrac{1}{2}Q_S + \tfrac{3}{2}\hat{C}) \tag{20}$$

$$p_2 = c_2(\tfrac{1}{2}Q_S - \tfrac{3}{2}\hat{C}) \tag{21}$$

The nodal price in node 3 is given by (4).

2.3 Nodal Prices as a Function of the Demand Parameter *a*

To sum up, the market price is dependent on the realization of the demand parameter *a*. The market price in a non-capacitated market, i.e. when $a \le \hat{a}$, is given by (8); $p_s = \frac{ac_1c_2}{bc_1+bc_2+c_1c_2}$. By inserting Q^* of (19) into (4), (20) and (21), we obtain the nodal prices of the capacitated market when $a > \hat{a}$. The nodal prices as a function of the demand parameter *a* can be summarized by

$$P_1 = \begin{cases} p_s = \dfrac{ac_1c_2}{bc_1+bc_2+c_1c_2} & for\, a \le \hat{a} \\ p_1 = \dfrac{c_1(2a+3\hat{C}(2b+c_2))}{4b+c_1+c_2} & for\, a > \hat{a} \end{cases} \tag{22}$$

$$P_2 = \begin{cases} p_s = \dfrac{ac_1c_2}{bc_1+bc_2+c_1c_2} & for\, a \le \hat{a} \\ p_2 = \dfrac{c_2(2a-3\hat{C}(2b+c_1))}{4b+c_1+c_2} & for\, a > \hat{a} \end{cases} \tag{23}$$

$$P_3 = \begin{cases} p_s = \dfrac{ac_1c_2}{bc_1+bc_2+c_1c_2} & for\, a \le \hat{a} \\ p_3 = \dfrac{a(c_1+c_2)+3\hat{C}b(c_1-c_2)}{4b+c_1+c_2} & for\, a > \hat{a} \end{cases} \tag{24}$$

where $\hat{a}$ is given by equation (13). The corresponding production and consumption in the three nodes are given by inserting prices into the individual supply and demand functions, i.e. $q_1^s = \frac{1}{c_1} p_1(a)$, $q_2^s = \frac{1}{c_2} p_2(a)$, and $Q_D = \frac{1}{b}(a - p_3(a))$.

In figure 2 the prices are plotted for different levels of *a* for a model where $b = 0.05$, $c_1 = 0.2$, $c_2 = 0.8$, and $\hat{C} = 120$, implying that $\hat{a} = 126$. Due to the simple model setup the functions are all piecewise linear in *a*. When the line capacity is not fully utilized, the market is cleared at the common market price, p_s. For higher demand levels, $a > \hat{a}$, prices in node 1 are set lower than the system price to curb production and reduce the flow in the direction from 1 to 2, while prices in node 2 are set higher than p_s to relieve capacity problems by inducing a greater counterflow on line 12.

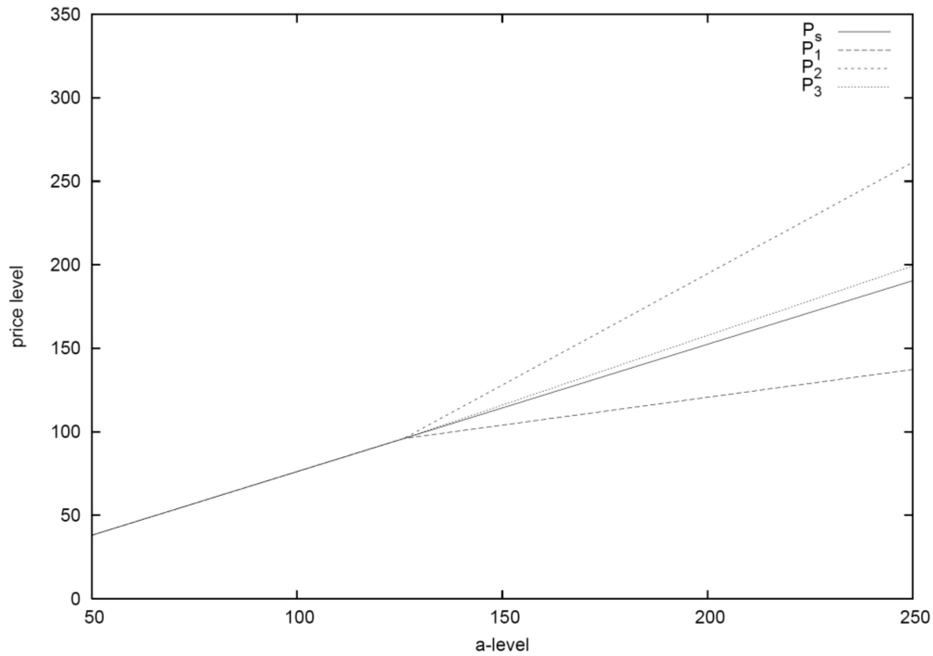


Figure 2: Plot of prices as a function of the *a*-parameter.

3 Introducing Dynamics and Uncertainty

Our aim is now to analyze the effect of limited grid capacity on nodal price processes. Prices at a given point of time are given by (22)-(24) above[130]. Basically, uncertainty in prices is driven by the uncertainty of underlying fundamentals related to demand and/or supply. To focus on the basic effect of grid constraints, we have introduced uncertainty in the demand function (4) only, keeping a transparent and controllable model. Uncertainty is implemented by defining *a* as a stochastic function. From the definition of the demand function it is clear that shifts in *a* represent parallel shifts in the demand function. In a market with electrical household heating, the most natural example could be the temperature. By defining a stochastic process for *a*, and assuming that the process *a* takes only admissible values (e.g. $a > 0$), equations (22)-(24) will define the stochastic processes for p_1, p_2, and p_3.

In order to derive the nodal price processes, we must first define a stochastic process for our fundamentals, in this case the parameter a_t. Finding a good representation of the real-world stochastic process is not a simple task. It is important to have a clear understanding of the effects we want the model to capture, and also how a reasonable process should evolve. Several questions arise in this context, for example; - Is the real world process explicitly time dependent? - What is the level of the process expected to be at a given point of time? - How volatile is the process? - Is the noise constant, or is it dependent on time and/or the level of the process? - Does the driving process display periodic patterns? - Is the driving process mean reverting, and if so, how fast on average?

[130] Note that the allocation of production and consumption at each instant of time is as in the static model described above, implying that none of the market participants in this model act strategically over time.

In our paper we do not, however, aim for ultimate realism in any particular market. For computational tractability we employ a continuous time framework. We have chosen to model the parameter a_t by a particular continuous time Itô diffusion, namely a time dependent mean reverting (Ornstein-Uhlenbeck) process of the form

$$da_t = \gamma_t(\eta_t + \frac{\eta_t'}{\gamma_t} - a_t)dt + \sigma_t dW_t \tag{25}$$

Here W_t is a Brownian motion, γ_t is related to the speed of mean reversion and η_t is the 'on average' seasonal process level. Variants of this process are used extensively in the literature, see e.g. Lucia and Schwartz (2002) or Johnson and Barz (1999), while the above formulation is based on Lund and Ollmar (2003). This process captures several important properties of a demand curve for electricity. The process is mean reverting, giving demand a tendency to normalize after some time, an attribute which is characteristic of demand. By choosing a rather high γ_t, the process would relatively rapidly be driven towards the average seasonal process level, a lower γ_t causes a slow drive towards the average level. As we will see below, this process formulation also opens for specifying cyclical patterns of demand. Note that the Ornstein-Uhlenbeck formulation above implies that volatility represented by $\sigma_t dW_t$ is independent of a_t. Though this process captures many important aspects of demand, it may attain negative values, requiring caution as to applications where this may be problematic.

For $s < t$ the explicit solution of equation (25) is

$$a_t = (a_s - \eta_s)e^{-\int_s^t \gamma_u du} + \eta_t + \int_s^t \sigma_u e^{-\int_u^t \gamma_r dr} dWu \tag{26}$$

We now simplify by letting $\gamma_t \equiv \gamma$ and $\sigma_t \equiv \sigma$, that is, constant volatility and a constant speed of mean reversion. Since σ is deterministic, the Itô integral is normally distributed and we can write a_t as

$$a_t = (a_s - \eta_s)e^{-\gamma(t-s)} + \eta_t + \sigma\left(\frac{1-e^{-2\gamma(t-s)}}{2\gamma}\right)^{1/2}\varepsilon \tag{27}$$

where ε is a standard normal distributed random variable. For a given a_s, the Gaussian process a_t has a conditional mean and variance[131] at time $t > s$ equal to

$$E[a_t | a_s] \equiv \mu_t = (a_s - \eta_s)e^{-\gamma(t-s)} + \eta_t \tag{28}$$

$$Var[a_t | a_s] \equiv \rho_t = \sigma^2\left(\frac{1-e^{-2\gamma(t-s)}}{2\gamma}\right) \tag{29}$$

Since the expected value of a_t is equal to η_t when $t \to \infty$, we can interpret this as the long run mean level for the process.

Demand clearly follows several cyclical patterns over time, reflecting patterns of nature as well as human activity that vary over the day, the week and across the year. In our model, price variations are induced by these variations in demand. These fluctuations may be modeled by including several trigonometric functions with different parameters. We have specified the average seasonal process level, η_t, as

$$\eta_t = A_0 + \sum_{j=1}^{k}\{A_j \cos(\omega_j t) + B_j \sin(\omega_j t)\}$$

where the parameters of A_j, B_j and the frequencies ω_j are chosen to specify the k different patterns we wish to model.

[131] To simplify notation we suppress the s, a_s dependence in μ, ρ. Still they must always be seen as functions of these variables.

In the paper Lund and Ollmar (2003) the process is calibrated to the spot price at the Nordic electricity marked. Here one unit of time corresponds to one hour, and the frequencies are set to

$$\omega_1 = 2\pi/8760 \quad \omega_2 = 2\omega_1 \text{ (year)}$$
$$\omega_3 = 2\pi/168 \quad \omega_4 = 2\omega_3 \text{ (week)}$$
$$\omega_5 = 2\pi/24 \quad 2\omega_6 = \omega_5 \text{ (day)}$$

to model yearly, weekly and daily variations in the spot price. Demand variation highly matches that of the spot prices. To illustrate how our model may be applied, we therefore use approximately the same parameters as Lund and Ollmar (2003), this time to model the demand process *a*[132].

The two lower panels of figure 3 show the resulting (time varying) long term mean level of the demand parameter a_t. In the first of these panels, covering five weeks, we are able to see the resulting daily and weekly variations in the mean, while in the second covering fifty weeks, we are able to see yearly variations in a_t. In the two upper panels of figure 3 we have plotted a realization of the process for $a_0 = 80.0$, $\sigma = 2.0$.

[132] The parameters chosen for *A* and *B* are calibrated to reflect the level estimated in Lund and Ollmar (2002): $A_1 = 32.2$, $A_2 = -8.4$, $A_3 = -3.1$, $A_4 = -0.2$, $A_5 = -5.3$, $A_6 = 1.6$, $B_1 = -3.7$, $B_2 = 12.3$, $B_3 = 4.1$, $B_4 = 2.8$, $B_5 = -4.8$, and $B_6 = 4.2$. Apart from this, the remaining parameters are chosen to fit our example, with the reversion parameter γ set to 0.01 and $A_0 = 100.0$.

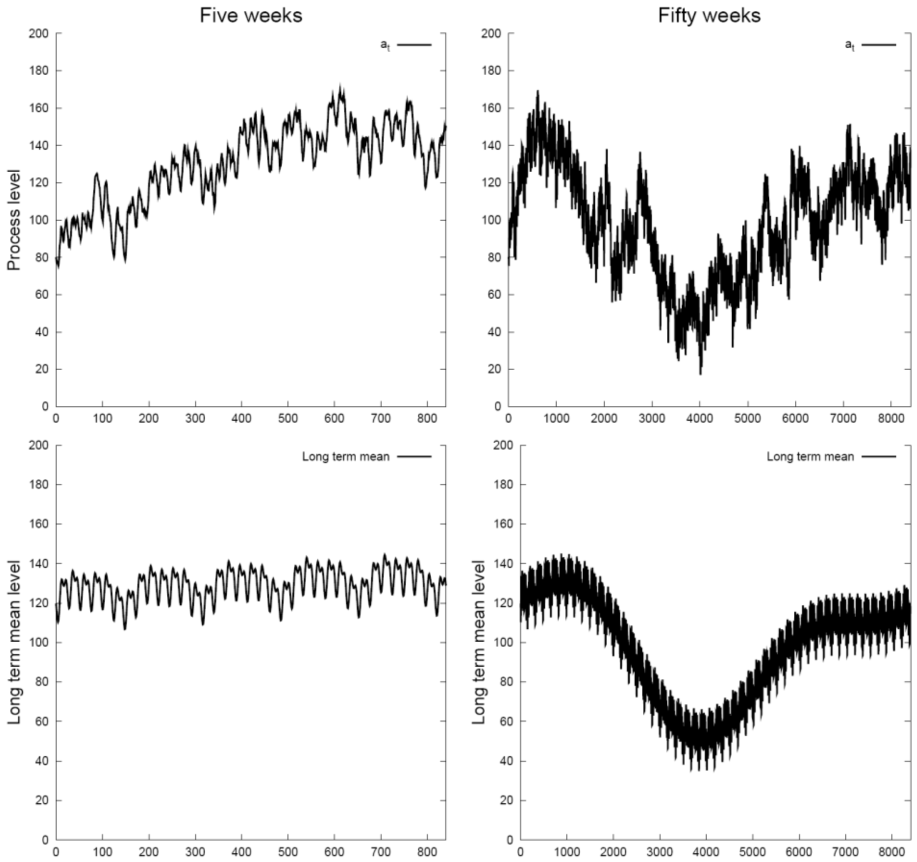


Figure 3: Plot of one replicate of the *a*-process, together with its long term mean level.

4 The Stochastics of Nodal Prices: Characteristics and Applications

The price process of our model is driven by the variations in the demand parameter *a*. The nodal price processes are thus completely defined by the price equations (22)-(24) together with the stochastic process of a_t given by equation (27). Based on this characterization of the nodal price processes, we have a tool useful for several purposes. Within a given model, we are able to study the stochastic properties of nodal prices, as well as performing numerical and analytical analyzes of different issues in the market. The purpose of this paper has been to gain insight into how grid limitations influence the underlying stochastic nodal price processes. In this section we will briefly sketch some applications.

- In figure 4 we have illustrated the realization of the price process for the above realization of a_t. This is but one possible replicate of the price processes. More interestingly, our characterization of the price processes allows us to calculate e.g. expected levels of nodal prices, as well as the estimated volatility of prices. This is information which is important input in evaluating power contracts, investment decisions or other applications for which future prices are important.

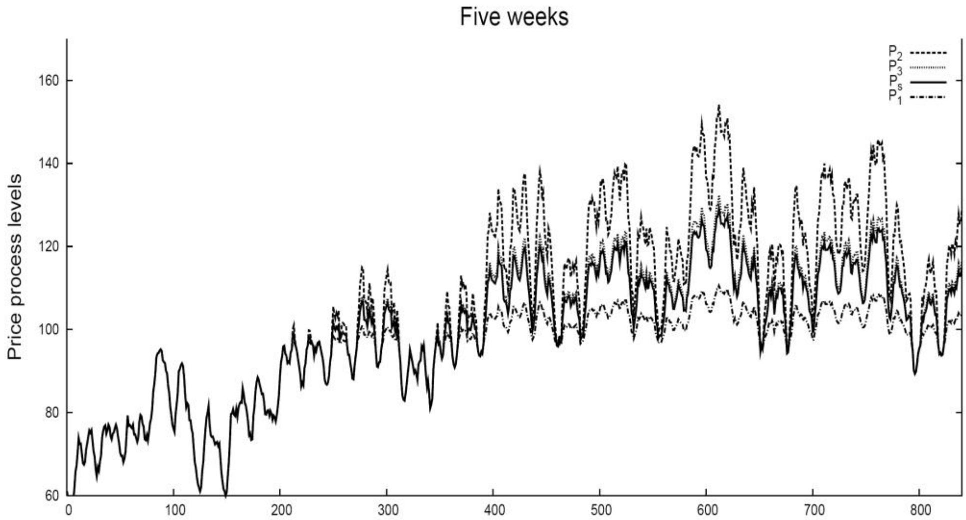


Figure 4: Price evolution for this a-replicate.

As an illustration, assume that we want to find the expected price levels in this model, given the current level of demand a_s. With the notation introduced in the appendix we can write price expectations as

$$E\left[p_s(t)\middle|a_s\right]=\tfrac{c_1c_2}{\Psi}\mu_t$$

$$\begin{aligned}E\left[p_1(t)\middle|a_s\right]&=E\left[p_s(t)I_{\{a_t\le\hat{a}\}}\middle|a_s\right]+\tfrac{2c_1}{\Phi}E\left[a_tI_{\{a_t>\hat{a}\}}\middle|a_s\right]\\&\quad+\tfrac{3c_1}{\Phi}\hat{C}(2b+c_2)E\left[I_{\{a_t>\hat{a}\}}\middle|a_s\right]\\&=\tfrac{c_1c_2}{\Psi}\overline{A}_t+\tfrac{2c_1}{\Phi}(\mu_t-\overline{A}_t)+\tfrac{3c_1}{\Phi}\hat{C}(2b+c_2)(1-G(\tfrac{\hat{a}-\mu_t}{\sqrt{\rho_t}})\end{aligned}$$

$$E\left[p_2(t)\middle|a_s\right]=\tfrac{c_1c_2}{\Psi}\overline{A}_t+\tfrac{2c_2}{\Phi}(\mu_t-\overline{A}_t)+\tfrac{3c_2}{\Phi}\hat{C}(2b+c_1)(1-G(\tfrac{\hat{a}-\mu_t}{\sqrt{\rho_t}})$$
$$E\left[p_3(t)\middle|a_s\right]=\tfrac{c_1c_2}{\Psi}\overline{A}_t+\tfrac{c_1+c_2}{\Phi}(\mu_t-\overline{A}_t)+\tfrac{3b}{\Phi}\hat{C}(c_1-c_2)(1-G(\tfrac{\hat{a}-\mu_t}{\sqrt{\rho_t}})$$

where $\Psi=bc_1+bc_2+c_1c_2$, $\Phi=4b+c_1+c_2$ and G denotes the cumulative standard normal distribution function. The price expectation functions are plotted in figure 5.

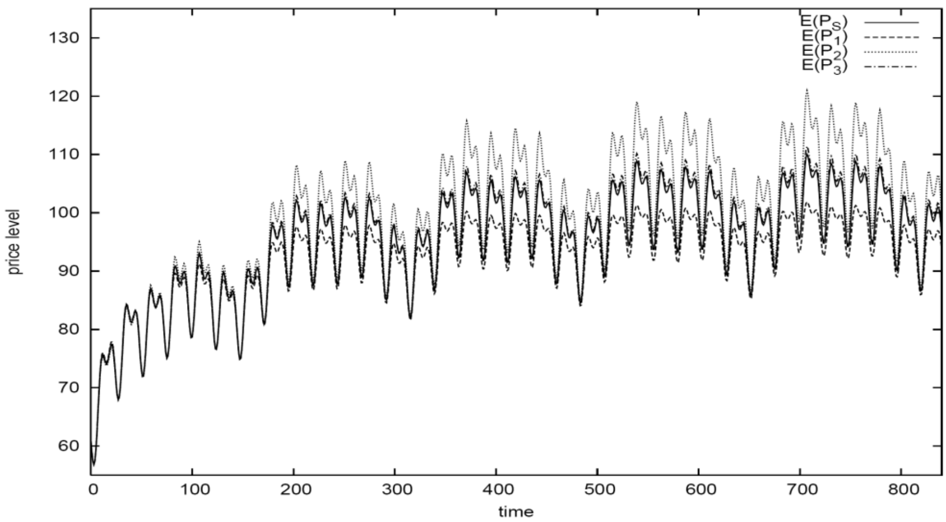


Figure 5: The expected nodal prices.

Likewise, the volatility of the price processes may be calculated in a similar manner. For example, again with notation from appendix A, we have, given the state a_s at time s that

$$Var[p_s(t)] = (\tfrac{c_1c_2}{\Psi})^2 Var[a(t)] = (\tfrac{c_1c_2}{\Psi})^2 \rho_t$$

$$\begin{aligned} Var[p_1(t)] &= E\left[(p_s(t) - E[p_s(t)])^2 I_{\{a_t \le \hat{a}\}}\right] + (\tfrac{2c_1}{\Phi})^2 E\left[(a_t - E[a_t])^2 I_{\{a_t > \hat{a}\}}\right] \\ &= (\tfrac{c_1c_2}{\Psi})^2 (\overline{C}_t - 2\overline{A}_t\mu_t + \overline{F}_t\mu_t^2) + (\tfrac{2c_1}{\Phi})^2 (\overline{D}_t - 2\overline{B}_t\mu_t + (1-\overline{F}_t)\mu_t^2) \end{aligned}$$

- Price differences are highly important for producers operating in several regions, as well as producers and consumers hedging their positions for example by derivative contracts written on the system price. Figure 6 shows the price differences between the nodal prices and the non-constrained system price. In this realization, we find an initial period where the demand has not led to binding network constraints, while the following time periods show differentiated nodal prices. Our model allows the price difference processes to be explicitly analyzed, for example by calculating expectation and variance functions, and applying these in evaluating contracts and market positions.

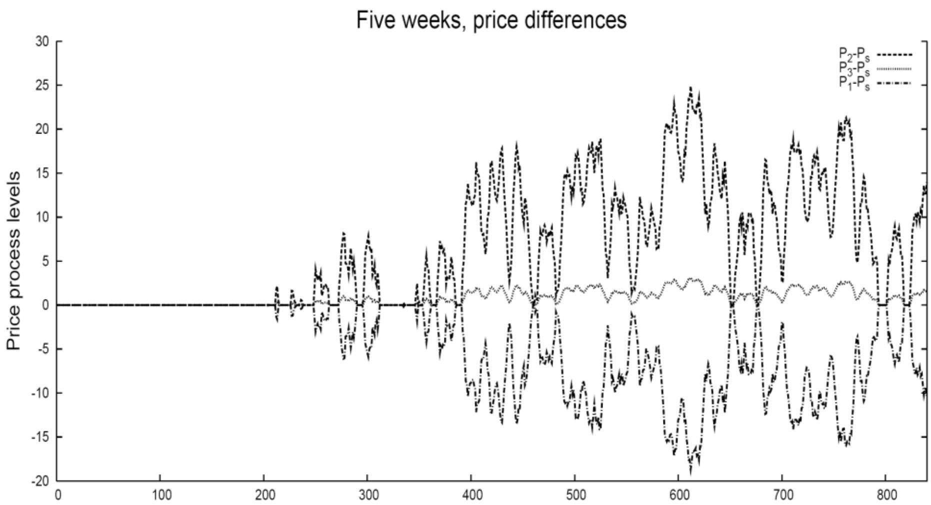


Figure 6: Evolution of price differences.

- Another useful application is to estimate the degree to which the grid will be congested in a given time period. Let $F_{\hat{a}}([\hat{S},T])$ be the expected fraction of hours with nodal prices in the future period $[\hat{S},T],\ \hat{S} > s$. Within our setup this can be calculated as follows:

$$
\begin{aligned}
F_{\hat{a}}([\hat{S},T]) &= \tfrac{1}{\text{T-}\hat{\text{S}}} E\left[\int_{\hat{S}}^{T} I_{\{a_t > \hat{a}\}} dt \middle| a_s \right] \\
&\overset{Fubini}{=} \tfrac{1}{\text{T-}\hat{\text{S}}} \int_{\hat{S}}^{T} E\left[I_{\{a_t > \hat{a}\}} \middle| a_s \right] dt \\
&= \tfrac{1}{\text{T-}\hat{\text{S}}} \int_{\hat{S}}^{T} (1 - G(\tfrac{\hat{a}-\mu_t}{\sqrt{\rho_t}})) dt \\
&= 1 - \tfrac{1}{\text{T-}\hat{\text{S}}} \left\{ \left[tG(\tfrac{\hat{a}-\mu_t}{\sqrt{\rho_t}}) \right]_{\hat{S}}^{T} - \int_{\hat{S}}^{T} t (\tfrac{\hat{a}-\mu_t}{\sqrt{\rho_t}})' \tfrac{1}{\sqrt{2\pi\rho_t}} e^{-\frac{(\hat{a}-\mu_t)^2}{2\rho_t}} dt \right\}
\end{aligned}
$$

  where μ_t, ρ_t are seen as functions of the initial process level a_s as in (28) and (29). With the parameter values from the previous example, we find that $F_{126}([0,336]) = 0.29$ by numerical integration. That is, we would expect nodal prices in 29% of the hours the first two weeks when $\hat{C} = 120$ (or alternatively, $\hat{a} = 126$).

- Other applications of the model may involve analyzing the sensitivity of changes in model parameters on price expectations, volatility, etc. This may be done by analytic or numerical studies of the consequences of changing the model parameters, as for example related to the demand slope or cost parameters, or changing the capacity of the grid. For example, in figure 7 we show how the expected fraction of nodal prices varies by varying the capacity $\hat{C}$ of the congested line[133].

[133] See Lund and Rud (2008) for an analysis within this framework related to investments in grid capacity.

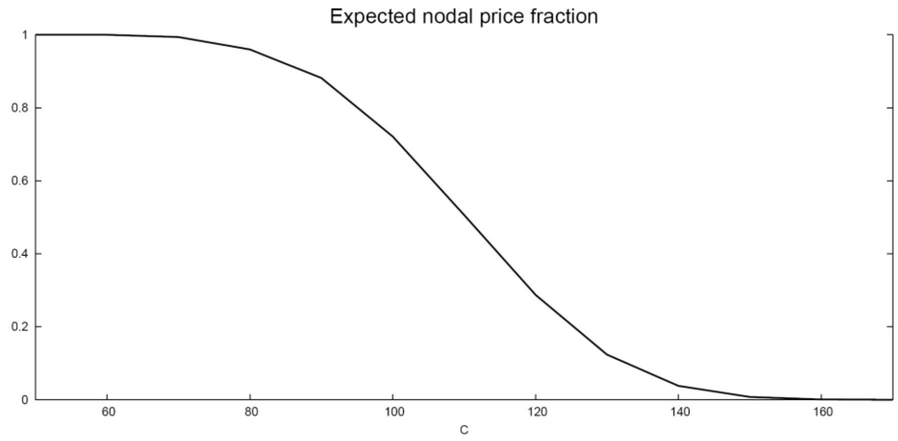


Figure 7: Expected fraction of the hours with nodal prices in the two first weeks, as a function of the line capacity.

5 Other Model Specifications

Above we have illustrated some of the main issues in modeling stochastic nodal prices. The model may be extended in several directions, for example with respect to the network, the representation of supply and demand, as well as the choice and modeling of stochastic parameters. In applying the principles of the model to analyze more complex real world problems, the challenge is to simplify, but yet capture the key features of the problem. It is then important to have a clear idea of which effects we want to analyze. Also, note that several extensions of the model will necessitate numerical simulation rather than analytical characteristics of the nodal price processes. Below we will shortly discuss possible model extensions.

Network: The 3-node network is simple. Still it may for example be seen as an approximation to the Norwegian/Swedish part of the Nordic market, where zonal aggregation often has been applied with two Norwegian zones, and one Swedish zone. By applying slightly more complex networks, we may similarly be able to pinpoint key features of the networks in other countries.

Demand and Supply Functions: By modeling demand and supply in the same node, problems of negative quantities will be avoided, as a negative net demand simply indicates net production. Further, other demand and supply functions may be implemented, though linear demand and supply functions facilitate a clear and simple model.

Stochastic processes: For the demand process a_t above, we have chosen a simple process. Clearly more complex process could be applied. a_t could also be a process defined by combinations of other (possibly correlated) processes. As an illustration suppose that demand consists of household and industrial consumption. Let us assume that the industrial consumption varies slowly due to changing market conditions, and this is modeled by a reverting process *X*. Household consumption, on the other hand, is exposed to temperature variations, and this could be modeled by a more volatile stochastic process *Y*, normalizing relatively quickly. It may then be appropriate to use a process e.g. of the form $a_t = \beta_1 X_t + \beta_2 Y_t$. Clearly these processes could also incorporate long-term trends.

In another model we could be particularly interested in the effects of demand variations throughout the day. We know that day and night variation in demand may be large in some periods and smaller in others. It is also often the case that the consumption during the night has low variance, while the peak consumption varies much more. This could be captured by introducing stochastic amplitudes in demand. As a simple illustration consider $Y_t = 50 + X_t(1.2 + \sin(2\pi \frac{t}{24}))$, where $X_t > 0$ is a process reverting at 50. This leads to periods with high prices at day time, with low fluctuations in the night as seen in figure 8.

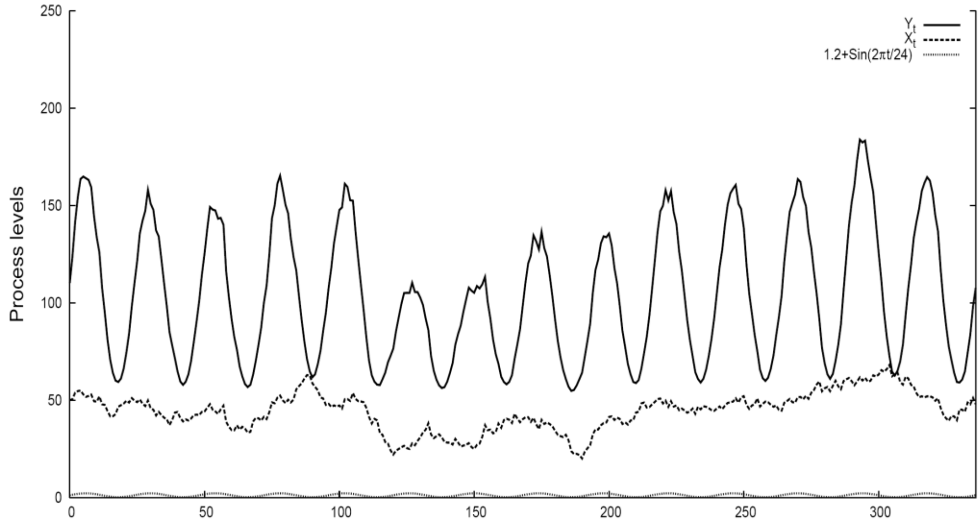


Figure 8: Plot of a process with large volatility in daytime.

In section 2.3 we focused on the nodal prices as functions of the parameter *a*. Clearly we could have considered the prices as functions of the other parameters, as for example the demand slope, or production parameters.

6 Conclusions

In this paper we have demonstrated how capacity limits in the grid influence the stochastic price processes. Based on an intuitive model for the demand dynamics, we have shown how our framework can help analyze the characteristics of the price processes. We have also indicated possible applications of the model.

This paper can be seen as a 'primer' for introduction of dynamics and uncertainty in a nodal network. In the paper we focused on examples to illustrate how the model could be applied. The main focus has been on simplicity, rather than realism. Still we have given indications of possible generalizations that could be important when real world problems are studied.

As the fundamental processes can be unobservable in practice, it may be difficult to calibrate our model to fit observed nodal prices. Compared to models based on observed nodal prices, our approach may be advantageous in several cases. As an illustration we point out that our model may give insight in problems where there are no observations, for instance for other potential grid specifications. It is also simple to include subjective beliefs about the future market situation. Clearly this is highly relevant for investment decisions. In addition note that a naïve calibration of the price process directly from data may give strange results, for instance price movements in neighboring nodes that are physically impossible. In our model, the dynamics will always be consistent with the underlying physical grid.

Appendix

In this section we compute some useful expectations. An important function is $I_{\{a_t \leq \hat{a}\}}$ which is 1 when $a_t \leq \hat{a}$, i.e. when there is no congestion and non-differentiated nodal prices, and 0 otherwise indicating a congested net. Assume $t > s$, and that we know a_s. When a is modeled by the Ornstein-Uhlenbeck process, we know that it is Gaussian with mean and variance given by equations (28) and (29). Therefore the probability that the net is un-congested at time t, given present state a_s, can be written as

$$
\begin{aligned}
\overline{F}_t &\equiv E\left[I_{\{a_t \leq \hat{a}\}} \middle| a_s\right] \\
&= \int_{-\infty}^{\hat{a}} \left(\frac{1}{\sqrt{2\pi\rho_t}} e^{-\frac{(x-\mu_t)^2}{2\rho_t}} \right) dx \\
&= G^{\mu_t,\rho_t}(\hat{a}) \\
&= G\left(\frac{\hat{a}-\mu_t}{\sqrt{\rho_t}}\right)
\end{aligned}
$$

where G denotes the cumulative standard normal distribution function. Thus, we also have that the probability that the net is congested at time t, given the state a_s, is

$$E\left[I_{\{a_t > \hat{a}\}} \middle| a_s\right] = 1 - \overline{F}_t$$
$$= 1 - G\left(\frac{\hat{a} - \mu_t}{\sqrt{\rho_t}}\right)$$

For calculating price expectations etc., we need several calculations. Recall that $\hat{a}$ is defined by (13), and that ρ_t and μ_t depend on s and a_s. We further calculate $E\left[a_t \middle| a_s, a_t \le \hat{a}\right] \equiv E\left[a_t I_{\{a_t \le \hat{a}\}} \middle| a_s\right]$:

$$\begin{aligned}
\overline{A}_t &\equiv E\left[a_t I_{\{a_t \le \hat{a}\}} \middle| a_s\right] \\
&= \frac{1}{\sqrt{2\pi\rho_t}} \int_{-\infty}^{\hat{a}} x e^{-\frac{(x-\mu_t)^2}{2\rho_t}} dx \\
&= \frac{1}{\sqrt{2\pi\rho_t}} \int_{-\infty}^{\hat{a}} (x - \mu_t) e^{-\frac{(x-\mu_t)^2}{2\rho_t}} dx + \mu_t \int_{-\infty}^{\hat{a}} \frac{1}{\sqrt{2\pi\rho_t}} e^{-\frac{(x-\mu_t)^2}{2\rho_t}} dx \\
&= -\frac{\sqrt{\rho_t}}{\sqrt{2\pi}} e^{-\frac{(\hat{a}-\mu_t)^2}{2\rho_t}} + \mu_t G\left(\frac{\hat{a} - \mu_t}{\sqrt{\rho_t}}\right)
\end{aligned}$$

This clearly implies that

$$\overline{B}_t \equiv E\left[a_t I_{\{a_t > \hat{a}\}} \middle| a_s\right] = E\left[a_t \middle| a_s\right] - \overline{A}_t = \mu_t - \overline{A}_t$$

In addition note that $E\left[a_t^2 \middle| a_s\right] = \rho_t + \mu_t^2$, and that

$$
\begin{aligned}
\overline{C_t} &\equiv E\left[a_t^2 I_{\{a_t \le \hat{a}\}} \middle| a_s\right] \\
&= \frac{1}{\sqrt{2\pi\rho_t}} \int_{-\infty}^{\hat{a}} (x-\mu_t)^2 e^{-\frac{(x-\mu_t)^2}{2\rho_t}} dx + \frac{2\mu_t}{\sqrt{2\pi\rho_t}} \int_{-\infty}^{\hat{a}} (x-\mu_t) e^{-\frac{(x-\mu_t)^2}{2\rho_t}} dx \\
&+ \mu_t^2 \int_{-\infty}^{\hat{a}} \frac{1}{\sqrt{2\pi\rho_t}} e^{-\frac{(x-\mu_t)^2}{2\rho_t}} dx \\
&= -\frac{\sqrt{\rho_t}}{\sqrt{2\pi}} (\hat{a}-\mu_t) e^{-\frac{(\hat{a}-\mu_t)^2}{2\rho_t}} + \rho_t G(\frac{\hat{a}-\mu_t}{\sqrt{\rho_t}}) \\
&- \frac{2\mu_t\sqrt{\rho_t}}{\sqrt{2\pi}} e^{-\frac{(\hat{a}-\mu_t)^2}{2\rho_t}} + \mu_t^2 G(\frac{\hat{a}-\mu_t}{\sqrt{\rho_t}}) \\
&= -\frac{\sqrt{\rho_t}}{\sqrt{2\pi}} (\hat{a}+\mu_t) e^{-\frac{(\hat{a}-\mu_t)^2}{2\rho_t}} + (\mu_t^2+\rho_t) G(\frac{\hat{a}-\mu_t}{\sqrt{\rho_t}})
\end{aligned}
$$

Finally also observe that

$$
\begin{aligned}
\overline{D_t} &\equiv E\left[a_t^2 I_{\{a_t > \hat{a}\}} \middle| a_s\right] \\
&= E\left[a_t^2 \middle| a_s\right] - E\left[a_t^2 I_{\{a_t \le \hat{a}\}} \middle| a_s\right].
\end{aligned}
$$

References

Barlow, M.T. (2002): ‘A diffusion model for electricity prices’, *Mathematical Finance*, 12(4):287-298.

Bjørndal, M. (2000): *Topics on Electricity Transmission Pricing*, PhD thesis, Norwegian School of Economics and Business Administration, March 2000.

Johnson, B. and G. Barz (1999): ‘Selecting stochastic processes for modelling electricity prices. In *Energy Modelling and the Management of Uncertainty*, pages 3-21, RISK Books, 1999.

Lucia, J.J. and E.S. Schwartz (2002): ‘Electricity prices and power derivatives: Evidence from the nordic power exchange. *Review of Derivatives Research*, 5(1):5-50.

Lund, A.-C. and F. Ollmar (2003): ‘Analyzing flexible load contracts in the energy market’, in *An Analysis of Derivatives Prices in the Nordic Power Market*, PhD thesis, Norwegian School of Economics and Business Administration, September 2003.

Lund, A.-C. and L. Rud (2008): ‘Investment Evaluation in a Constrained Electricity Network with Stochastic Nodal Price Processes’, see essay 4 in this thesis.

Schweppe, F.C., M.C. Caramanis, R.D. Tors, and R.E. Bohn (1988): *Spot Pricing of Electricity*, Kluwer Academic Publishers.

Weron, R., M. Bierbrauer, and S. Trück (2004): ‘Modeling electricity prices: jump diffusion and regime switching’, *Physica A*, (336):39-48.

ESSAY 4

Investment Evaluation in a Constrained Electricity Network with Stochastic Nodal Price Processes*

Arne-Christian Lund

Linda Rud

Abstract

The interconnectedness of supply and demand with a common capacitated network is a main feature of electricity markets. This paper studies the implication of this interaction for evaluating the profitability of investments in production and grid capacity under uncertainty. In doing so, we pinpoint several potential pitfalls, and discuss principle issues in the use of system prices versus nodal prices, how investments may affect future price processes, the mix of grid and production investments, and the issue of externalities. The issues are illustrated within a three node electricity market model, which is combined with dynamic modeling of underlying fundamental parameters, giving stochastic nodal price processes consistent with the underlying grid.

* Mette Bjørndal, Kurt Jörnsten, Per Manne, Kristian Miltersen, and Jøril Mæland have provided useful comments that are greatly appreciated. All errors are the authors' responsibility.

1 Introduction

The degree of long-run efficiency achieved in an electricity market is contingent upon the optimality of investment strategies induced by the market system. In many respects the electricity market shares common features and challenges of other markets: As in most commodity markets, the electricity market is organized with competition in supply and demand, and exhibits highly volatile prices due to the volatile nature of the underlying fundamentals, thus posing similar challenges for investments under uncertainty. The electricity market is also a network economy where all supply and demand is interconnected by a common capacitated network, thus posing similar challenges of optimal investment strategies in other network economies.

In the electricity market these features are, however, combined. Supply and demand are competitive activities, where investments in production capacity and energy-saving appliances follow from the incentives given by market prices. Incentives for investments in grid capacity, however, follow from the framework of the chosen regulation policy due to the natural monopoly characteristics of the network. The nature of interconnectedness in the grid is also further enhanced due to Kirchhoff's laws that determine the physical power flow. For example, while individual investments normally do not significantly alter the market equilibrium in other competitive commodity markets, even minor investments can in principle affect nodal electricity prices, even in electricity markets that are highly competitive. And, there may be positive or negative network externalities of investments in production capacity, so that the optimal investment plan for the market is related to achieving the right mix of investments in network and production/consumption capacities. These issues are due to the interconnectedness and loop-flow attributes of the common capacitated network. It is in the combination of such highly different regimes that the electricity market faces its challenges in achieving long-run efficiency.

In this paper we will study the implications of these interactions for evaluating the profitability of investments under uncertainty. In doing so, we pinpoint several potential pitfalls in evaluating grid and production investments in electricity markets. We look into the use of system prices versus nodal prices, the identification and calculation of correct nodal price processes, and the issue of externalities. As such, our results have implications for the individual investor in evaluating investments, as well as for authorities in designing market architecture and regulation policies for addressing market imperfections such as natural monopoly and externalities.

The issues are illustrated within a simple electricity market model, applying the modeling of stochastic nodal price processes from Lund and Rud (2004), where a static three-node model is combined with dynamic modeling of fundamental parameters, giving stochastic nodal price processes consistent with the underlying grid. Section 2 discusses investments in a network economy under uncertainty, and addresses potential pitfalls in evaluating investments in production and grid capacity. Section 3 reviews the electricity market model. Section 4 introduces welfare measures. Sections 5-7 exemplifies investment decisions concerning investments in grid and production capacity, and discusses principle issues of grid investments, showing how unfortunate use of time series based processes may fail to identify socially optimal investment strategies. Section 8 concludes the paper.

2 Investments in a Network Economy

The overall benefit of an investment may be defined as the net change in social surplus that the investment induces. In principle this implies comparing the market equilibrium before the investment with the resulting market equilibrium after the investment. In most well-functioning competitive commodity markets individual investments do not significantly alter market equilibrium, and the gain in private surplus is likely to closely equal the gain in social surplus. In electricity markets, however, even rather minor investments may potentially affect the general equilibrium due to the interconnectedness and loop-flow attributes of the common capacitated network. This has several important implications for evaluation of and incentives for investments, as we will discuss below.

Investments in Production Capacity: In a competitive market production capacity investments are in principle made by the individual market participants, driven by the incentives provided by the market system, i.e. the price signals. In a volatile and competitive network-based economy, there are several potential pitfalls in evaluating investments that may hinder socially optimal investment strategies. Some pitfalls are related to the following:

- *Problems in using the system price process for evaluation:* The future income of investments in production capacity depends upon future prices that are uncertain. Investment evaluation thus requires assessment of future prices and the stochastic attributes of the price processes. Looking to literature, most contributions in this area have focused on finding appropriate stochastic processes to model time series of system spot prices, see for example Lucia and Schwartz (2002), Weron et al. (2004), and Johnson and Barz (1999). However, if the proposed capacity investments are located in areas where nodal prices processes differ from system price processes, analyzes based on system price processes will misrepresent future profitability.

- *Problems in the use of time-series-based nodal price processes:* Following this approach to modeling electricity prices, a natural extension would be to model nodal price processes based on time series of nodal prices. By using correct techniques, the estimated nodal price processes may well describe the existing nodal price processes. However, in a network economy this also represents a potential pitfall. Although the system price of the electricity market might not be noticeably affected by investments, even rather minor investments may in principle affect the grid flow and state of congestion, and thus nodal prices. In this case time-series-deduced nodal price processes may give false premises for appraising the projects.

- *Problems in evaluating the effect of investments on future nodal prices:* If investments are assumed to affect grid flow and thus the distribution of nodal prices, investment evaluation calls for modeling future nodal prices. This requires an analysis of how production capacity investments affect market equilibrium, and thus future nodal prices. Many contributions in this area are based on partial equilibrium models of the underlying market and its transmission constraints, and analyze the effect on nodal prices. See e.g. Bjørndal (2000). However, most of these contributions represent 'snap shots', being inherently static, or two-periodic at the most. To provide input for investment analyzes, they have to be combined with stochastic models, to induce implications for nodal price processes. For more stylized electricity market models, as used in this paper, future nodal price processes may be explicitly calculated on the basis of the specified supply, demand and network, and specifications of underlying fundamentals. For more comprehensive models, simulations may be necessary to derive time series for deducing the attributes of future nodal price processes.

- *Problems of network externalities:* Production capacity investments evaluated on the basis of 'correct' nodal price processes in principle represent the true profitability of the investor. Due to loop flow features and the interconnectedness of the grid, the investments may, however, not only alter the surplus of the investor, but also the surplus of other participants: Investments that relieve congestion tend to lower the price in the previously 'congested' area, and raise the price in other areas. This may imply that investments that are unprofitable (profitable) for the investor, are profitable (unprofitable) for the society as a whole. In this instance investments in production and network capacity exhibit externalities, a feature which is important in the design of overall regulatory and incentive mechanisms in the electricity market.

Investment in Network Capacity: Congestion affects the maximum attainable social surplus. The costs of congestion arise both as the overall possible electricity production is constrained, and from the opportunity cost that results from out-of-merit-order dispatch due to congestion, so that the resulting production mix does not reflect the cheapest mix of production capacity available. In such a case, grid investments have the potential of being highly profitable and to raise total social surplus. However, also here, there are aspects that may deter optimal investment plans.

- *Incentives for investment:* Optimal network investments are highly contingent upon the design of market framework, regulatory policies and their implications for network management and investment incentives. While grid investments may relieve congestion and increase the net social welfare, the 'grid surplus', here defined as the revenue from capacity charges, of the transmission system operator may be reduced. Due to the natural monopoly attributes of the grid, the challenge therein lies in designing regulatory policies that induce a well-organized transmission company, with incentives for promoting investments that enhance social surplus.

- *Interaction of grid and production capacity investments:* Production capacity investments, as well as network capacity investments, may relieve grid congestion. Complicating the matter of grid investments is the fact that the overall optimal investment strategies in effect are decisions that involve finding the right mix grid *and* production capacity investments.

In the following sections, and based upon a simple three node electricity market model, we will look into both grid investments that relieve congestion directly, and production capacity investments that relieve congestion indirectly. Uncertainty is imposed by modeling a stochastic demand, making the benefit of investments highly uncertain. Investments may influence the occurrence of congestion, and thus also the future price processes. Within our model we analyze investment decisions based on an explicit evaluation of how price processes are affected, and through this also illustrate common pitfalls of evaluating investments in a network economy.

3 The Model

We apply the three-node electricity market model of Lund and Rud (2004), as illustrated in figure 1, with generation in nodes 1 and 2, and consumption in node 3. Uncertainty is imposed in the market by modeling the fundamentals of demand as a stochastic process. The network is modeled by the classic 'DC' approximation to the network, where each line has identical technical characteristics and impedances equal to 1.

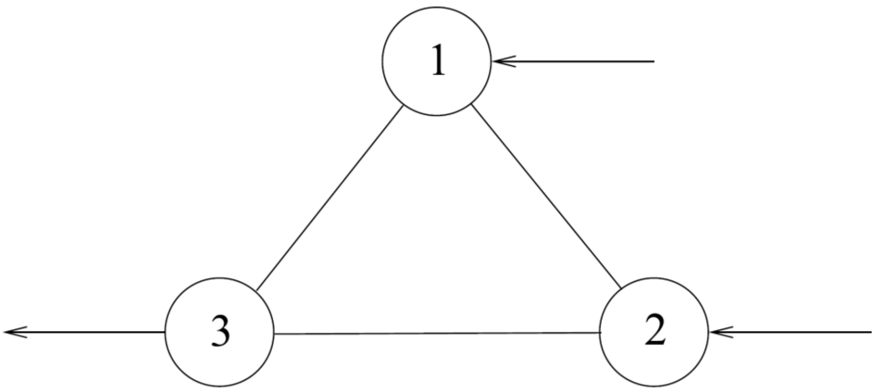


Figure 1: The network

Demand and Supply: Generation is located in node 1 and 2, and is represented by the simple linear supply curves of

$$p_i = c_i q_i^s \quad i = 1,2 \tag{1}$$

where p_i is the price, q_i^s is the quantity supplied, and c_i is the cost factor in node *i* = 1,2. We assume an unlimited capacity of the production in node 1 and 2, and furthermore that producer 1 is the low cost producer, i.e. $c_1 < c_2$. All demand is located in node 3, and is defined by the linear demand curve of

$$p_3 = a - bq_3^d \tag{2}$$

where q_3^d is the quantity demanded, and *a* and *b* are positive constants, $a,b > 0$. In our stylized model we assume no losses. This implies that in market equilibrium aggregate supply has to equal aggregate demand, i.e. $q_1^s + q_2^s = q_3^d$.

Network Flows: The flows over each line follows from Kirchhoff's laws. The line flow over line *ij*, in the direction from *i* to *j* is denoted q_{ij}. Note that if power flows from node *j* to *i*, we have $q_{ij} < 0$. In our simple 'DC' network, the flows over each line for production quantities q_1^s and q_2^s in node 1 and 2, and consumption q_3^d in node 3 are[134]

$$\begin{aligned} q_{12} &= \tfrac{1}{3} q_1^s - \tfrac{1}{3} q_2^s \\ q_{13} &= \tfrac{2}{3} q_1^s + \tfrac{1}{3} q_2^s \\ q_{23} &= \tfrac{1}{3} q_1^s + \tfrac{2}{3} q_2^s \end{aligned} \tag{3}$$

Below we will impose a capacity constraint on line 12. A simplifying assumption in this respect is that $c_1 < c_2$, implying that $q_1^s \geq q_2^s$. As a consequence, we will have $q_{12} \geq 0$, i.e. a fixed flow direction on line 12.

Market Clearing Prices: In the case of no capacity limits in the network, the market is cleared, i.e. $q_3^d = q_1^s + q_2^s$, on a common price $p_s = p_1 = p_2 = p_3$ where

$$p_s = \frac{a c_1 c_2}{b c_1 + b c_2 + c_1 c_2} \tag{4}$$

The price p_s is termed the system price. The equilibrium production and consumption quantities, q_1^s, q_2^s, and q_3^d may be found by inserting p_s into (1) and (2).

This is also the market equilibrium in a network with capacity limits, given that the resulting line flows do not exceed the line capacities. The resulting flow is found by inserting the quantities into equations (3). Also, note that

[134] These equations follow from three physical laws: Kirchhoff's Junction Rule, $q_i = \sum_{i \neq j} q_{ij}$ for $i = 1,2$, states that the input to nodes 1 and 2 has to equal to the current flowing out of it. Following from Kirchhoff's Loop Rule we have $q_{13} = q_{12} + q_{23}$ which, in the absence of losses, states that the algebraic sum of flows over any path in a loop is equal. The Law of Conservation of Energy states that total generation equals total consumption, $q_3^d = q_1^s + q_2^s$.

when the market is cleared on the common system price p_s, the relative production mix of q_1^s and q_2^s is the least cost production mix to bring forth the resulting aggregate quantity of supply.

If, however, the implied flows of the system price equilibrium exceed the capacity of restricted lines, the market has to be cleared on differentiated nodal prices. The optimal market equilibrium is defined as the allocation that maximizes social surplus, given the constraints of the network. In the model the capacity of line 12 is assumed to be restricted to $\hat{C}$ for flows in either direction, while we have assumed that the capacity of the other lines will not be binding.

By inserting q_1^s, q_2^s, and q_3^d in (3), we find the flow over line 12 to be $q_{12} = \frac{a(c_2 - c_1)}{3(bc_1 + bc_2 + c_1 c_2)}$. The capacity of line 12 will thus be binding for any combination of parameters that satisfies the inequality $|q_{12}| > \hat{C}$:

$$\left| \frac{a(c_2 - c_1)}{3(bc_1 + bc_2 + c_1 c_2)} \right| > \hat{C} \tag{5}$$

Below we will model the uncertainty of our electricity market model, by representing the demand factor *a* by a stochastic process. Whether the line capacity is binding or not, thus depends upon the realization of *a*. Using the least-cost production mix implied by the unconstrained problem, we find that the capacity of line 12 becomes binding if the demand factor a exceeds the level $\hat{a}$ which solves $q_{12}(a) = \hat{C}$. Thus, when the demand intersection is $a = \hat{a}$, where

$$\hat{a} = 3\hat{C} \frac{bc_1 + bc_2 + c_1 c_2}{c_2 - c_1}, \tag{6}$$

the capacity of line 12 is fully utilized, that is, given the least-cost production mix which is implied by the unconstrained market solution on the common market price p_s. If *a* is to exceed $\hat{a}$, the unconstrained market solution based on the least-cost production mix, is thus not feasible.

Following the equations of network flow in (3) we, however, find that it is possible to increase overall production by changing the production mix for the additional quantities above $q_i^s(p_s(\hat{a}))$, from the least cost production mix, to a production mix of equal amounts of q_1^s and q_2^s for the extra marginal quantity. Increased production based on this production mix is possible as equal additional production quantities in node 1 and 2 cancel out in the definition of the flow over line 12.

To achieve this, nodal electricity prices have to be differentiated for the market to clear. We refer to Lund and Rud (2004) to verify that the equilibrium prices as a function of the demand parameter *a* are:

$$P_1 = \begin{cases} p_s = \dfrac{ac_1c_2}{bc_1 + bc_2 + c_1c_2} & for\, a \le \hat{a} \\ p_1 = \dfrac{c_1(2a + 3\hat{C}(2b + c_2))}{4b + c_1 + c_2} & for\, a > \hat{a} \end{cases} \tag{7}$$

$$P_2 = \begin{cases} p_s = \dfrac{ac_1c_2}{bc_1 + bc_2 + c_1c_2} & for\, a \le \hat{a} \\ p_2 = \dfrac{c_2(2a - 3\hat{C}(2b + c_1))}{4b + c_1 + c_2} & for\, a > \hat{a} \end{cases} \tag{8}$$

$$P_3 = \begin{cases} p_s = \dfrac{ac_1c_2}{bc_1 + bc_2 + c_1c_2} & for\, a \le \hat{a} \\ p_3 = \dfrac{a(c_1 + c_2) + 3\hat{C}b(c_1 - c_2)}{4b + c_1 + c_2} & for\, a > \hat{a} \end{cases} \tag{9}$$

Uncertainty: Uncertainty is introduced by assuming that the intersection *a* of the demand curve follows a stochastic process, while *b*, c_1 and c_2 are deterministic constants. The parameter *a* represents the level of demand. Shifts in *a* give parallel shifts in the linear demand curve, for example due to differences in temperature. Though electricity demand fluctuates, it normally tends to move back towards normal conditions, a characteristic which may be represented by a mean reverting process. While we also have chosen not to incorporate jumps in our model, we here use the model of the demand parameter *a* employed in Lund and Rud (2004) where it is modeled as a mean reverting (Ornstein Uhlenbeck) process of the form:

$$da_t = \gamma(\eta_t + \frac{\eta_t'}{\gamma} - a_t)dt + \sigma dW_t. \tag{10}$$

Here W_t is a Brownian motion, σ is a constant time-independent volatility, γ is a constant speed of mean reversion, and η_t is the 'on average' seasonal process level which reflects that mean demand levels follow cyclical patterns over time. To model this effect, the average seasonal process level, η_t, is modeled as a sum of several trigonometric functions:

$$\eta_t = A_0 + \sum_{j=1}^{k}\left\{A_j \cos(\omega_j t) + B_j \sin(\omega_j t)\right\}.$$

One unit of time corresponds to one hour. The parameters may be calibrated to model normal yearly, weekly and daily variations of market demand. The frequencies are set to $\omega_1 = 2\pi/8760$ and $\omega_2 = 2\omega_1$ for yearly cycles, $\omega_3 = 2\pi/168$ and $\omega_4 = 2\omega_3$ for weekly cycles, and $\omega_5 = 2\pi/24$ and $2\omega_6 = \omega_5$ for daily cycles. The remaining parameters are somewhat arbitrarily[135] set to $A_0 = 100.0$, $A_1 = 32.2$, $A_2 = -8.4$, $A_3 = -3.1$, $A_4 = -0.2$, $A_5 = -5.3$, and $A_6 = 1.6$, and for the B-parameters: $B_1 = -3.7$, $B_2 = 12.3$, $B_3 = 4.1$, $B_4 = 2.8$, $B_5 = -4.8$, $B_6 = 4.2$. Further, the revision parameter γ is set to 0.01.

Let us now turn back to the process of a_t. Since σ is deterministic, the Itô integral is normally distributed, and the explicit solution can be written as

[135] As this is only an illustration, the parameters are not calibrated to real demand data. However, we find that demand variations to some extent can be likened to variations in system prices, with daily variations characterized by daily peak periods and low night levels, with weekly variations characterized by lower weekend levels, and with yearly seasonal variations with lower summer prices, and higher winter prices. We have therefore chosen to use an approximation of the parameters used for spot *price* process calibration in Lund and Ollmar (2002).

$$a_t = (a_s - \eta_s)e^{-\gamma(t-s)} + \eta_t + \sigma\left(\frac{1-e^{-2\gamma(t-s)}}{2\gamma}\right)^{1/2}\varepsilon \tag{11}$$

where $s < t$ and ε is a standard normal distributed random variable.

For a given a_s, the Gaussian process a_t has a conditional mean and variance

$$E\left[a_t \middle| a_s\right] \equiv \mu_t = (a_s - \eta_s)e^{-\gamma(t-s)} + \eta_t \tag{12}$$

$$Var\left[a_t \middle| a_s\right] \equiv \rho_t = \sigma^2\left(\frac{1-e^{-2\gamma(t-s)}}{2\gamma}\right) \tag{13}$$

Since the expected value of a_t is equal to η_t when $t \to \infty$, we can interpret η_t as the long run mean level for the process.

To illustrate, the two lower panels of figure 2 plot the long term mean level η_t for five and fifty weeks, respectively. Here clear daily, weekly, and seasonal patterns may be observed. The two upper panels of figure 2 show a specific realization of the process for five, and fifty weeks respectively, given the parameters of $a_0 = 80.0$, and an annual $\sigma = 2.0$. Figure 3 illustrates the nodal prices resulting from the this particular realization of a_t.

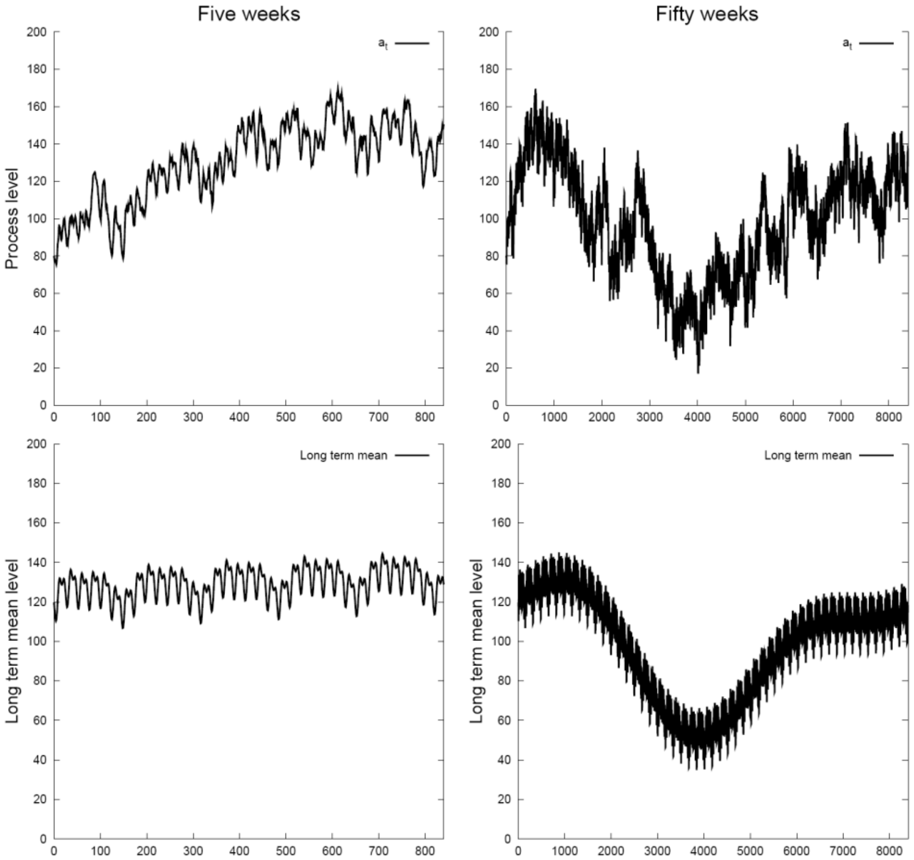


Figure 2: Plot of one replicate of the process level a_t (upper panels), and plot of long term mean level η_t (lower panels).

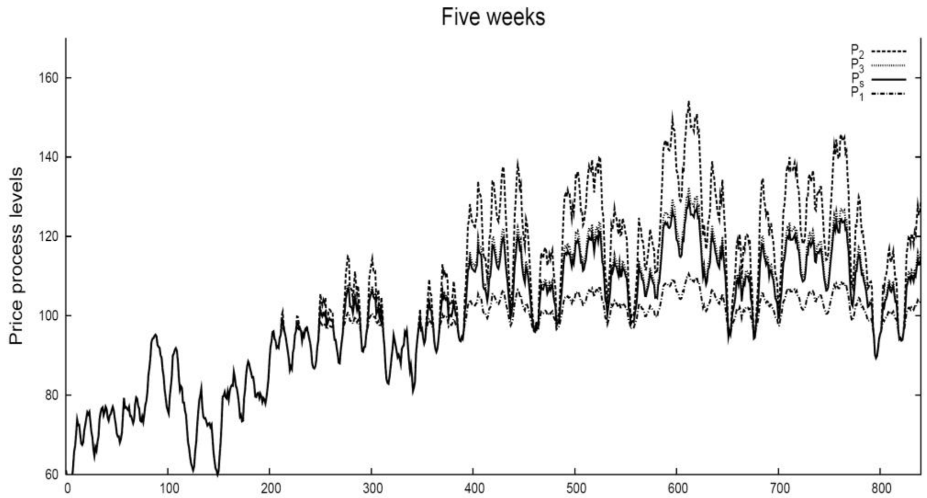


Figure 3: Price evolution for the a_t -replicate in figure 2.

4 Welfare Measures

Social surplus measures the overall efficiency of the market, while allocational issues may be studied by decomposing it into consumer surplus, producer surplus and grid revenue (the latter due to congestion and use of nodal prices).

Social Surplus: Assuming a utilitarian welfare function, social surplus is defined as the total willingness to pay less the total costs of production at the equilibrium quantities. In our model, for a given point of time[136], this corresponds to $SS = \frac{1}{2}(a + P_3)q_3^d - \frac{1}{2}P_1 q_1^s - \frac{1}{2}P_2 q_2^s$. The prices P_1, P_2, and P_3 are given by equations (7)-(9), and the quantities by inserting prices into (1) and (2). As the demand parameter *a* is the source of uncertainty within our electricity model, prices and quantities in each node at a given point of time depend upon the realization of *a*, implying that the specific equation for social surplus also depends on the level of *a*. Recall that $\hat{a} = 3\hat{C}\frac{bc_1 + bc_2 + c_1c_2}{c_2 - c_1}$, is the border level of *a* as to when the network is congested. For $a \le \hat{a}$ the network is uncongested and the market is cleared on the common system price p_s. For $a > \hat{a}$ the network is congested and the market is cleared on differentiated nodal prices p_1, p_2, and p_3. For a given point of time, the social surplus may thus be represented as:

$$SS(a) = \begin{cases} SS_{(a \le \hat{a})} \equiv \dfrac{a^2 c_1 c_2 - p_s^2 (bc_1 + bc_2 + c_1 c_2)}{2bc_1 c_2} & for\, a \le \hat{a} \\ SS_{(a > \hat{a})} \equiv \dfrac{a^2 c_1 c_2 - p_3^2 c_1 c_2 - p_1^2 bc_2 - p_2^2 bc_1}{2bc_1 c_2} & for\, a > \hat{a} \end{cases}$$

For further computations, we have sorted the terms with respect to the demand parameter *a*. Social surplus may thus be written as:

[136] This is the social surplus for a given point of time. The exposition is simplified by omitting the *t* subscript in *a*, in prices, and in quantities.

$$SS(a)=\begin{cases} SS_{(a\le\hat a)} \equiv \dfrac{[c_1+c_2]a^2}{2(bc_1+bc_2+c_1c_2)} & for\, a\le\hat a \\ SS_{(a>\hat a)} \equiv \dfrac{\left[-9\hat C^2(bc_1+bc_2+c_1c_2)\right]+\left[6(c_2-c_1)\hat C\right]a+[4]a^2}{2(4b+c_1+c_2)} & for\, a>\hat a \end{cases}$$

Investments in grid or production capacity alter future allocations in the electricity market, typically for several years ahead. Surpluses at a given future point of time are related to the realization of the demand factor *a*, and are thus inherently uncertain. Using a risk neutral probability measure, we compute the value of the aggregated social surplus, $\mathbf{SS}(a_0)\equiv E\left[\int_0^\infty e^{-rs}SS(a_s)ds\middle|a_0\right]$ to compare the different investment scenarios.

$$\begin{aligned}\mathbf{SS}(a_0) &\equiv \int_0^\infty e^{-rs}\left\{E\left[SS_{\{a_s\le\hat a\}}I_{\{a_s\le\hat a\}}\middle|a_0\right]+E\left[SS_{\{a_s>\hat a\}}I_{\{a_s>\hat a\}}\middle|a_0\right]\right\}ds \\ &= \int_0^\infty e^{-rs}\left\{\overline{\theta}^2_{SS}E\left[a_s^2 I_{\{a_s\le\hat a\}}\middle|a_0\right]+\theta^0_{SS}E\left[I_{\{a_s>\hat a\}}\middle|a_0\right]\right\}ds \\ &+\int_0^\infty e^{-rs}\left\{\theta^1_{SS}E\left[a_s I_{\{a_s>\hat a\}}\middle|a_0\right]+\theta^2_{SS}E\left[a_s^2 I_{\{a_s>\hat a\}}\middle|a_0\right]\right\}ds\end{aligned}$$

where *r* is the risk free interest rate, $\overline{\theta}^i$ denotes the coefficients of a^i for the surplus in question when there is no congestion, and θ^i are the coefficients of a^i when there is congestion. For social surplus, the non-zero coefficients are given by $\overline{\theta}^2_{SS}=\frac{c_1+c_2}{2(bc_1+bc_2+c_1c_2)}$, $\theta^0_{SS}=\frac{-9\hat C^2(bc_1+bc_2+c_1c_2)}{2(4b+c_1+c_2)}$, $\theta^1_{SS}=\frac{3\hat C(c_2-c_1)}{4b+c_1+c_2}$, $\theta^2_{SS}=\frac{2}{4b+c_1+c_2}$. Furthermore, to calculate this value, we have weighted the elements of the social surplus term by the indicator functions $I_{\{a_t\le\hat a\}}$ and $I_{\{a_t>\hat a\}}$. The function $I_{\{a_t\le\hat a\}}$ is 1 when $a_t\le\hat a$, and 0 otherwise. Likewise, the function $I_{\{a_t>\hat a\}}$ is 1 when $a_t>\hat a$, and 0 otherwise. The formulas for the values of $E\left[I_{\{a_t\le\hat a\}}\middle|a_s\right]$, $E\left[I_{\{a_t>\hat a\}}\middle|a_s\right]$, $E\left[a_t I_{\{a_t\le\hat a\}}\middle|a_s\right]$, $E\left[a_t I_{\{a_t>\hat a\}}\middle|a_s\right]$, $E\left[a_t^2 I_{\{a_t\le\hat a\}}\middle|a_s\right]$, and $E\left[a_t^2 I_{\{a_t>\hat a\}}\middle|a_s\right]$ can be explicitly calculated and are found in the appendix.

Consumer Surplus: The net surplus of the consumer group is given by the benefit of consumption defined by the area under the demand curve, less the cost of purchase at the given market price in the particular node. In our model we have $CS=\frac{1}{2}(a-P_3)q_3^d$. The consumer surplus $CS(a)$ for a given period with the realization *a* may in the model may be written as

$$CS(a)=\begin{cases} CS_{(a\le\hat{a})}\equiv\dfrac{\left[b(c_1+c_2)^2\right]a^2}{2(bc_1+bc_2+c_1c_2)^2} & for\, a\le\hat{a} \\ CS_{(a>\hat{a})}\equiv\dfrac{\left[\frac{9}{2}b\hat{C}^2(c_2-c_1)^2\right]+\left[12b\hat{C}(c_2-c_1)\right]a+\left[8b\right]a^2}{(4b+c_1+c_2)^2} & for\, a>\hat{a} \end{cases}$$

The value of the aggregated consumer surplus which is $\mathbf{CS}(a_0)\equiv E\left[\int_0^\infty e^{-rs}CS(a_s)ds\middle|a_0\right]$, is likewise given by

$$\begin{aligned}\mathbf{CS}(a_0)&\equiv\int_0^\infty e^{-rs}\left\{E\left[CS_{\{a_s\le\hat{a}\}}I_{\{a_s\le\hat{a}\}}\middle|a_0\right]+E\left[CS_{\{a_s>\hat{a}\}}I_{\{a_s>\hat{a}\}}\middle|a_0\right]\right\}ds\\ &=\int_0^\infty e^{-rs}\left\{\bar{\theta}_{CS}^2E\left[a_s^2I_{\{a_s\le\hat{a}\}}\middle|a_0\right]+\theta_{CS}^0E\left[I_{\{a_s>\hat{a}\}}\middle|a_0\right]\right\}ds\\ &+\int_0^\infty e^{-rs}\left\{\theta_{CS}^1E\left[a_sI_{\{a_s>\hat{a}\}}\middle|a_0\right]+\theta_{CS}^2E\left[a_s^2I_{\{a_s>\hat{a}\}}\middle|a_0\right]\right\}ds\end{aligned}$$

with the factors of $\bar{\theta}_{CS}^2=\frac{b(c_1+c_2)^2}{2(bc_1+bc_2+c_1c_2)^2}$, $\theta_{CS}^0=\frac{9b\hat{C}^2(c_2-c_1)^2}{2(4b+c_1+c_2)^2}$, $\theta_{CS}^1=\frac{12b\hat{C}(c_2-c_1)}{(4b+c_1+c_2)^2}$, and $\theta_{CS}^2=\frac{8b}{(4b+c_1+c_2)^2}$.

Producer Surplus: The net surplus of the producer is given by the sales revenue, less the cost of production, i.e. $PS_i=\frac{1}{2}P_iq_i^s, i=1,2$. Producer surpluses, $PS_1(a)$ and $PS_2(a)$ in node 1 and 2 for a given period with the realization of *a* are thus

$$PS_1(a)=\begin{cases} PS_{1(a\le\hat{a})}\equiv\dfrac{\left[c_1c_2^2\right]a^2}{2(bc_1+bc_2+c_1c_2)^2} & for\, a\le\hat{a} \\ PS_{1(a>\hat{a})}\equiv\dfrac{\left[9c_1\hat{C}^2(2b+c_2)^2\right]+\left[12c_1\hat{C}(2b+c_2)\right]a+\left[4c_1\right]a^2}{2(4b+c_1+c_2)^2} & for\, a>\hat{a} \end{cases}$$

$$PS_2(a)=\begin{cases} PS_{2(a\le\hat{a})}\equiv\dfrac{\left[c_1^2c_2\right]a^2}{2(bc_1+bc_2+c_1c_2)^2} & for\, a\le\hat{a} \\ PS_{2(a>\hat{a})}\equiv\dfrac{\left[9c_2\hat{C}^2(2b+c_1)^2\right]-\left[12c_2\hat{C}(2b+c_1)\right]a+\left[4c_2\right]a^2}{2(4b+c_1+c_2)^2} & for\, a>\hat{a} \end{cases}$$

The value of the aggregated producer surplus of producer *i*, $\mathbf{PS}_i(a_0)\equiv E\left[\int_0^\infty e^{-rs}PS_i(a_s)ds\middle|a_0\right]$, is given by:

$$\begin{aligned}\mathbf{PS}_i(a_0) &\equiv \int_0^{\infty} e^{-rs}\left\{E\left[PS_{i\{a_s\le\hat{a}\}}I_{\{a_s\le\hat{a}\}}\middle|a_0\right]+E\left[PS_{i\{a_s>\hat{a}\}}I_{\{a_s>\hat{a}\}}\middle|a_0\right]\right\}ds\\ &=\int_0^{\infty} e^{-rs}\left\{\overline{\theta}^2_{PS_i}E\left[a_s^2 I_{\{a_s\le\hat{a}\}}\middle|a_0\right]+\theta^0_{PS_i}E\left[I_{\{a_s>\hat{a}\}}\middle|a_0\right]\right\}ds\\ &+\int_0^{\infty} e^{-rs}\left\{\theta^1_{PS_i}E\left[a_s I_{\{a_s>\hat{a}\}}\middle|a_0\right]+\theta^2_{PS_i}E\left[a_s^2 I_{\{a_s>\hat{a}\}}\middle|a_0\right]\right\}ds\end{aligned}$$

where $\overline{\theta}^2_{PS_1}=\frac{c_1c_2^2}{2(bc_1+bc_2+c_1c_2)^2}$, $\theta^0_{PS_1}=\frac{9c_1\hat{C}^2(2b+c_2)^2}{2(4b+c_1+c_2)^2}$, $\theta^1_{PS_1}=\frac{6c_1\hat{C}(2b+c_2)}{(4b+c_1+c_2)^2}$, $\theta^2_{PS_1}=\frac{2c_1}{(4b+c_1+c_2)^2}$, $\overline{\theta}^2_{PS_2}=\frac{c_1^2c_2}{2(bc_1+bc_2+c_1c_2)^2}$, $\theta^0_{PS_2}=\frac{9c_2\hat{C}^2(2b+c_1)^2}{2(4b+c_1+c_2)^2}$, $\theta^1_{PS_2}=\frac{-6c_2\hat{C}(2b+c_1)}{(4b+c_1+c_2)^2}$, and $\theta^2_{PS_2}=\frac{2c_2}{(4b+c_1+c_2)^2}$.

Grid Surplus: For a given equilibrium, consumers pay $P_3q_3^d$, while producers in total receive $P_1q_1^s+P_2q_2^s$, where $q_3^d=q_1^s+q_2^s$. In the case of no congestion, nodal prices equal the system price, payments balance and this grid surplus is zero. With different nodal prices, the consumer payments may differ from the revenue received by the producers. This difference, $GS=p_3q_3^d-(p_1q_1^s+p_2q_2^s)$, is collected by the grid operator, and is termed the grid surplus. The grid surplus for a given period may be written as $GS(a)$;

$$GS(a)=\begin{cases}GS_{(a\le\hat{a})}\equiv 0 & for\, a\le\hat{a}\\ GS_{(a>\hat{a})}\equiv\frac{\left[-9c\hat{C}^2(bc_1+bc_2+c_1c_2)\right]+\left[3\hat{C}(c_2-c_1)\right]a}{4b+c_1+c_2} & for\, a>\hat{a}\end{cases}$$

and with the value of the aggregated grid surplus given by $\mathbf{GS}(a_0)\equiv E\left[\int_0^{\infty}e^{-rs}GS(a_s)ds\middle|a_0\right]$:

$$\begin{aligned}\mathbf{GS}(a_0)&\equiv\int_0^{\infty}e^{-rs}\left\{E\left[GS_{\{a_s>\hat{a}\}}I_{\{a_s>\hat{a}\}}\middle|a_0\right]\right\}ds\\ &=\int_0^{\infty}e^{-rs}\left\{\theta^0_{GS}E\left[I_{\{a_s>\hat{a}\}}\middle|a_0\right]+\theta^1_{GS}E\left[a_sI_{\{a_s>\hat{a}\}}\middle|a_0\right]\right\}ds\end{aligned}$$

where $\theta^0_{GS}=\frac{-9\hat{C}(bc_1+bc_2+c_1c_2)}{4b+c_1+c_2}$ and $\theta^1_{GS}=\frac{3\hat{C}(c_2-c_1)}{4b+c_1+c_2}$.

5 Network Investments: Increasing the Capacity of Line 12

If market equilibrium is restricted by the capacity of the network, investments in network capacity may potentially increase net social surplus. In the following we will, by means of our electricity market model, focus on principle issues of how network investments affect the social surplus, and the surpluses of different groups of participants. In the following, when referring to changes in social surplus, we will refer to changes in the gross social surplus as defined above. The net effect is, however, given by the gross change in social surplus less investment costs. While investment costs are relatively straightforward to estimate, the change in social surplus is more complex to estimate, as it involves an estimation of how the investment changes future production, consumption, congestion, and thus also market prices.

In our electricity market model, a network investment refers to increasing the capacity $\hat{C}$ of the line between node 1 and 2. For any future scenario, i.e for any a-replicate, a line investment that increases the capacity of this line will clearly reduce the expected fraction of time in which the network is congested. This in turn increases the fraction of periods in which it is possible to clear the market on a common system price, and consequently also the fraction of time in which the full potential of social surplus is realized. As such, the line investment will clearly increase the gross social surplus.

Our focus is now to gain a further understanding of how surpluses are affected. Note that social surplus is a function of production, consumption and realized market prices, all of which explicitly or implicitly are functions of the grid capacity $\hat{C}$. They also, however, depend upon the uncertainty of the demand parameter a, making the profitability of the investment dependent on the uncertainty of future market conditions.

In our simple market model, we have assumed that we know the true stochastic processes of market fundamentals. This enables us to calculate and study principle effects of the investments. Figure 4 illustrates the value of the aggregate surpluses, $\mathbf{SS}(\hat{C}, a_0)$, $\mathbf{CS}(\hat{C}, a_0)$, $\mathbf{PS}_i(\hat{C}, a_0)$, and $\mathbf{GS}(\hat{C}, a_0)$, as a function of different line capacities $\hat{C}$, given a life horizon of 10 years for the investment[137] [138].

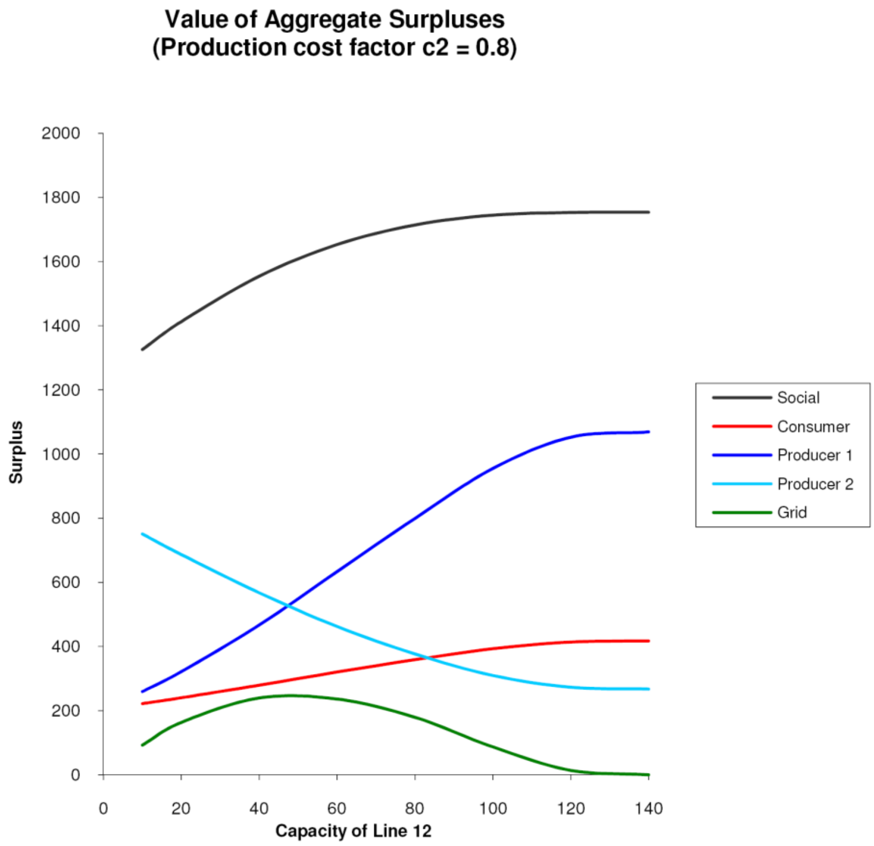


Figure 4: Value of aggregate surpluses as a function of the capacity $\hat{C}$ of line 12.

---

[137] Parameters of the example are given by $b = 0.05$, $c_1 = 0.2$, $c_2 = 0.8$, $a_0 = 80$, an hourly interest rate $r = 0.05/8760$, and an hourly $\sigma = 2/8760$.

[138] Note that we have assumed that the line capacities of line 13 and 23 will not be binding. This is a simplifying assumption. In general, by increasing the capacity in some parts of the grid, other lines that previously were not constrained, might reach their capacity limits.

First consider the effect on social surplus of increasing $\hat{C}$. We find that higher levels of capacity on line 12 steadily give higher levels of social surplus, but, as expected, at a decreasing rate of increase. Starting with a relatively low capacity on line 12, we see that a given positive line increment $\Delta\hat{C}$ substantially raises social surplus. Starting on a higher level of initial line capacity, the benefit of the same incremental investment is less, and near zero for capacity increments above a capacity of 120. This may be explained by considering the critical level $\hat{a}$ (defined in (6)), above which the grid is congested. By increasing line capacity with $\Delta\hat{C}$, the critical level $\hat{a}$ is raised. This has two effects. First, $\Delta\hat{C}$ lowers the probability of congestion, i.e. the probability that $a > \hat{a}$. This effect is more pronounced if, as a starting point, the critical point $\hat{a}$ is low due to a relatively low initial line capacity. Secondly are the consequences of congestion which depend upon the magnitude of which the market solution is curbed due to the congestion. This in turn depends upon the size and distribution of the difference $a - \hat{a}$. An implication here is that the overall negative consequences of congestion are higher if the capacity as a starting point is rather low. Hence, the effect of alleviating future congestion by an investment $\Delta\hat{C}$, if the capacity as a starting point is relatively low.

The line investment has different consequences for each group of market participants. For the consumer, as well as for the node 1 producer, we find that their surpluses steadily increase for incremental investments in line 12. For the consumers, the increased surplus follows from increasing levels of consumption, as well as lower prices, since less congestion allows the market to a greater degree to operate on the least-cost production mix. For producer 1 the incremental line investments on one hand increase the quantity to be produced. On the other hand, the occurrence of congestion is reduced, thus reducing the degree to which nodal prices are in force. This implies higher prices for producer 1, since the system price generally is higher than the nodal price in node 1.

For the high-cost producer in node 2, however, investments in line 12 are bad news. Note that producer 2 in times of congestion plays a special role to obtain feasible market solutions; When the network is not congested, the share of total production by each producer is given by the least-cost production mix, of which the high-cost producer 2 has the lower share. When the network is congested, any increments in aggregate production above that level implied by the critical demand level $\hat{a}$, have to be produced by means of a 50/50 production mix by each producer. This is necessary to obtain a feasible network flow according to the flow equations of the network, as an extra production based on the 50/50 production mix is possible as it does not induce any increased flow over line 12. With an investment in line 12, the line capacity increases and the degree of congestion is reduced. For producer 2, firstly the share of time in which the 50/50 production mix is needed is reduced, and secondly the market is increasingly balanced on system prices, which for producer 2 are lower than the nodal prices.

The consequences of grid investments for the net operator are interesting. At a very low capacity, grid surplus is quite low. Though a high degree of congestion, grid revenue is low due to low overall production levels. With modest additional investments, the overall production levels rise, while the degree of congestion is still high, giving a rise in grid revenue due to congestion. For further investments, congestion is reduced, and grid surplus is constantly declining. The figure also illustrates the well-known issue that the socially optimal level of grid investments does not coincide with the level of the maximum grid surplus, also indicating that the grid surplus should not be used as an incentive device for investments.

6 Production Capacity Investments: Investments in Production Technology

Alternatively, increased production in node 2 can also relieve congestion on line 12. This follows as production in node 2 according to Kirchhoff's laws creates a counterflow on the line. An alternative for amending the problems of congestion can thus be to invest in the production capacity of node 2. In our simplified model with no limits on production capacity, this may be illustrated by an investment in a more efficient production technology in node 2, i.e. investments that lower the cost factor c_2. The effect of this investment would be threefold:

- First, with lower costs in node 2 the least-cost mix will be altered and include a greater proportion of node 2 production. As this is the mix of the uncongested market, it implies lower marginal costs, and thus lower prices in times of no congestion.

- Secondly, as the least-cost production mix includes a greater share of node 2 production, this also implies that a greater counterflow over line 12 is induced, with a lower flow over line 12 in the direction from node 1 to 2. Thus, by lowering c_2, the critical level of $\hat{a}$ is raised.

- Thirdly, in times of congestion, the equilbrium resulting production mix above $\hat{a}$ will still be based on the half and half mix for additional quantities above $q_3^d(\hat{a})$. However, due to the lower c_2, the aggregate marginal cost curve in the case of congestion will also be lower, giving lower prices also in times of congestion.

Let us now consider the investment decision and the evaluation of its profitability. In a competitive market production investments are made by the producer, in our case the producer in node 2. The investment's profitability is given by the change in the value of the producer's aggregate surplus, less investment costs. The cost of investment is normally clearly defined, and may be represented by $C(I)$. The main evaluation problem lies in estimating the benefit of the investment. Here we have to calculate the profit in a scenario where the investment has been undertaken, and compare it with the profit in a scenario where the investment has not been undertaken.

The profit of producer 2 for a single period (hour) is $\frac{1}{2}P_2 q_2^s$. The value over the whole lifetime N of the investment, is $\mathbf{PS}_2 = E\left[\int_0^N e^{-rs}(\frac{1}{2}\widetilde{P}_2\widetilde{q}_2^{\,s})ds \middle| a_0\right]$. The effect of the investment is a more efficient production in node 2 with a lower cost factor c_2. In the following we will let c_2 denote the cost factor before the investment, and $\dot{c}_2$ the cost factor after the investment.

As we assume a competitive market, we assume that the producer takes the market price as given, and determines his quantity as a function of the market price, i.e. $q_2^s = P_2 / c_2$. The cost factor c_2 is the relevant cost factor, i.e. c_2 if considering the profit prior to investment, or $\dot{c}_2$ if considering the profit after the investment. We thus see that the supply curve, and thus also the market equilibrium quantity, for a given price is altered if c_2 is reduced. While the producer has full information on his cost factor c_2, the main question, and source of potential pitfalls for the producer in evaluation is related to the estimation and representation of the price process.

Let us now explicitly address the representation of the prices in the evaluation of the investment. In a normal competitive commodity market, the effect of a single investment will normally not affect the price. In that case, the investor may rightfully base his evaluation on the current price process of the market (provided that there are no changes in the stochastics of the underlying fundamentals of the market). In other words, this means that in a normal market a price process estimated on the basis of times series data may provide a fairly accurate representation of the future price process.

In a competitive electricity market, this might well be a reasonable assumption as to the effect of a production capacity investment (e.g. a change in c_2) on the uncongested system price p_s. However, in a congested market, even smaller investments can affect nodal prices. This implies that changes in a cost factor c_2 would be likely to have only minor effect on the system price p_s, but possibly have important effects on the nodal price p_2[139]. To be explicit as to the market situation in which the market prices evolve, we will in the following quote the price as a function of the actual cost factor in the market, i.e. $P_i(c_2)$ for real market prices before the investment, and $P_i(\dot{c}_2)$ for real market prices after the investment.

The investment is profitable for producer 2 if the change in the value of future profits due to the investment is greater than the investment cost, that is $\Pi_{PS_2}(I) = \mathbf{PS}_2\left[\widetilde{P}_2(\dot{c}_2), \dot{c}_2 \middle| a_0\right] - \mathbf{PS}_2\left[\widetilde{P}_2(c_2), c_2 \middle| a_0\right] - C(I) > 0$, if correctly evaluated. The profits in each scenario before and after the investment, are thus represented as a function of two factors. The first factor is the chosen market price process, for example $\widetilde{P}_2(c_2)$ which is the historical nodal price process (i.e. market prices which evolved before the c_2 investment, and given the current state of a_0). The second factor, which is the cost factor c_2 alone, marks the explicit cost factor that the producer uses in the calculation of his supply and resulting profit.

This distinction is important in the following discussion on the representation of the price processes. As future prices are stochastic, the calculation of the investment's profitability, whether using analytical calculations or numerical simulations, requires explicit attention as to how the price process used is defined. Though by first sight seemingly

[139] In our stylized model, it is, however, expected that the investment will affect both.

straightforward, in practice, there are several issues to be considered in making the right assessment of the investment's profitability.

Evaluating the Investment: Finding the Right Price Processes: We will now illustrate two incorrect evaluations of the price process, followed by the correct estimation of the future nodal price distribution.

- *Using the system price process for evaluation:* Most empirical literature has focused on estimating system spot prices processes. In a market where the relative scarcity of energy is important, major attention is normally drawn to the development in the system prices. Following this focus, a natural implication would be to calculate the investment's profitability on the basis of system prices p_s, and their underlying stochastic price process. Moreover, a natural choice would seemingly be the time series based system price process $\tilde{p}_s(c_2)$, as the system price process in a competitive market would normally not change significantly because of a single investment. In this the evaluation based on the time series based system price process may be represented as $\Pi_{PS_2}(I) = \mathbf{PS}_2\left[\tilde{p}_s(c_2), \dot{c}_2 \middle| a_0\right] - \mathbf{PS}_2\left[\tilde{p}_s(c_2), c_2 \middle| a_0\right]$ $-C(I)$[140]. Expectations may be calculated analytically, as in our example, or on the basis of numerical simulations. Assuming a capacity on line 12 of $\hat{C} = 60$, the current starting point of $a_0 = 80$, and an investment that lowers c_2 from a cost factor of 0.8 to 0.6 in our model we then have $\mathbf{PS}_2\left[\tilde{p}_s(0.8), 0.6 \middle| a_0\right] - \mathbf{PS}_2\left[\tilde{p}_s(0.8), 0.8 \middle| a_0\right]$ = 356.3 million − 267.2 million =89.1 million[141]. For the producer in the

[140] We have here assumed that the producer calculates future profits on the basis of the same price process, i.e. that the profit before the investment also is calculated on the basis of system prices. Alternatively, if extrapolating past profits, the before-investment profit could be estimated as $\mathbf{PS}_2\left[\tilde{P}_2(c_2), c_2 \middle| a_0\right]$ which is the true former profit based on historical actual received (nodal prices), extended 10 years into the future. In our illustration we have, however, assumed that the producer uses similar calculations, and thus using the same system price process.

[141] In principle all investments will affect the resulting system price process, giving the new system price process $\tilde{p}_s(\dot{c}_2)$. This implies that the new system price

capacitated network, however, the relevant prices are the nodal prices, which in this area of the network normally are higher than system prices. Thus, the use of system prices here seem to underestimate the value of the investment. The main problem here is thus that the system prices are not the relevant prices faced by the investor.

- *Using time-series-based nodal price processes:* Investments in a congested network are thus more correctly evaluated using nodal prices. A seemingly natural way to assess the nodal price process is thus to estimate this underlying stochastic price process by means of a time series analysis of nodal prices. This also is seemingly justified by the normal assumption that the individual participant does not affect market prices in a competitive market. Let us now assume that the producer estimates the nodal price process based on time series data, i.e. on $\widetilde{P}_2(c_2)$. This implies that the profit of the investment is measured as $\Pi_{PS_2}(I) = \mathbf{PS}_2\left[\widetilde{P}_2(c_2), \dot{c}_2 \middle| a_0\right] - \mathbf{PS}_2\left[\widetilde{P}_2(c_2), c_2 \middle| a_0\right] - C(I)$. Assuming a line 12 capacity of $\hat{C} = 60$, we have an increase in producer surplus 2 of $\mathbf{PS}_2\left[\widetilde{P}_2(0.8), 0.6 \middle| a_0\right] - \mathbf{PS}_2\left[\widetilde{P}_2(0.8), 0.8 \middle| a_0\right]$ = 616.8 million − 462.6 million = 154.2 million. This indicates a seemingly better profit, partly due to the use of nodal prices which are higher than system prices, and partly due to the lower cost factor. The major

process will differ from the time series based price process, i.e. $\widetilde{p}_s(c_2) \neq \widetilde{p}_s(\dot{c}_2)$. In real competitive markets, however, the individual investment does normally not significantly affect the overall uncongested market price, and system prices would normally not change to a large extent due to a single investment. A time series estimated system price process, $\widetilde{p}_s(c_2)$, would then be close to the new system price process, $\widetilde{p}_s(\dot{c}_2)$, i.e. $\widetilde{p}_s(c_2) \approx \widetilde{p}_s(\dot{c}_2)$ (given that the fundamental factors in the market do not change). Thus, in the example above we have used the time series based price process. However, in our model with only two producers, any change in c_2 might directly have a significant effect on the system price, and thus on the calculated profit. For example, the profit based on historical system prices is $\mathbf{PS}_2\left[\widetilde{p}_s(0.8), 0.6 \middle| a_0\right]$=356.3 million, while the profit based on new system prices is $\mathbf{PS}_2\left[\widetilde{p}_s(0.6), 0.6 \middle| a_0\right]$=345.3 million.

mistake here is, however, that the nodal prices may be significantly affected by the investment.

- *Using a correct estimated future distribution of nodal prices:* In a capacitated network with loop-flow effects, however, even minor investments may affect nodal prices, if the investment affects the patterns of congestion. To correctly evaluate the profitability of the investment, the future nodal price process has to be estimated. Thus $\Pi_{PS_2}(I) = \mathbf{PS}_2\left[\tilde{P}_2(\dot{c}_2), \dot{c}_2 \middle| a_0\right] - \mathbf{PS}_2\left[\tilde{P}_2(c_2), c_2 \middle| a_0\right] - C(I)$ represents the correct assessment of the profitability of the investment for the producer. To correctly evaluate the investment, the effect on congestion and thus nodal prices has to be evaluated, finding the stochastic distribution of $\tilde{P}_2$. This is not straightforward. For our simple three node model, the distributions may be calculated. For more complex networks and market models, distributions may be simulated. In our model, the correct assessment of the producers change in surplus is given by $\mathbf{PS}_2\left[\tilde{P}_2(0.6), 0.6 \middle| a_0\right] - \mathbf{PS}_2\left[\tilde{P}_2(0.8), 0.8 \middle| a_0\right]$ = 502.7 million − 462.6 million = 40.1 million, and the investment is profitable for producer 2 if $C(I) < 40.1$ million. Also note that the correct after-investment profit of 502.7 million is lower than the profit of 616.8 million based on time series of nodal prices. This is partly because the investment in fact lowers the occurrence of congestion and thus the prices of node 2, which more often are based on system prices. This is not accounted for if the nodal price process is estimated on the basis of historical price data.

Evaluating the Investment - Externalities: The overall profitability of the investment may, however, not be given by the profit allocated to the producer alone. When the investment affects the overall pattern of congestion, the surpluses of other participants are affected as well. Figure 5 shows the correctly evaluated effects of the investment on the value of the surpluses of other market participants, as well as on the aggregate social surplus. A lower production cost c_2 directly raises the profitability of producer 2. It also lowers congestion, thus expanding the production for both producer 1 and 2 in periods that otherwise would have been congested. This causes the expected consumer surplus to rise due to a larger consumption and lower prices. Further, with our model specification, and within relevant intervals of c_2, both producer surplus in node 1 and 2 rise[142].

[142] In our starting point of $c_2 = 0.8$, there is congestion. By lowering c_2, the share of producer 2 in the least-cost production mix will increase. This raises the critical level $\hat{a}$ due to the increased counterflow on line 12 at all production levels. It is also obvious that for $c_2 = c_1$ there will be no congestion, as the two producers will produce equal amounts, thus with a zero flow over line 12. In this case the producer surplus of both producers will be equal, and higher than in our starting point of $c_2 = 0.8$ and $c_1 = 0.2$. In the relevant area of c_2 which is depicted in the figure, both producer surpluses are rising, as congestion is substantially declining in this area. If we were to include further levels of c_2, there are, however, several diverging effects which imply that the surplus curves will not be as smooth as above. Note that whether or not the grid is congested in a specific scenario, in fact depends on the properties of the stochastic *a*, i.e. to the extent to which *a* is likely to exceed $\hat{a}(c_2)$. With our specification of the *a* process, market equilibrium implies that congestion is completely abolished at about $c_2 = 0.35$. If we were to extend the curves of figure 5 to even lower levels of c_2, this would imply a further rise in surplus for producer 2 and a fall for producer 1. However, note that the value of the aggregated surplus virtually is a weighted average of surpluses for different scenarios of *a* over time. And, for a given *a* scenario, we note that the producer 2 surplus as a function of c_2 is a curve which is kinked at the c_2 level where congestion is relieved. The value of the aggregate surplus as a function of c_2, in areas where congestion relief is not the prime motive, does thus not render a smooth curve as a function of c_2.

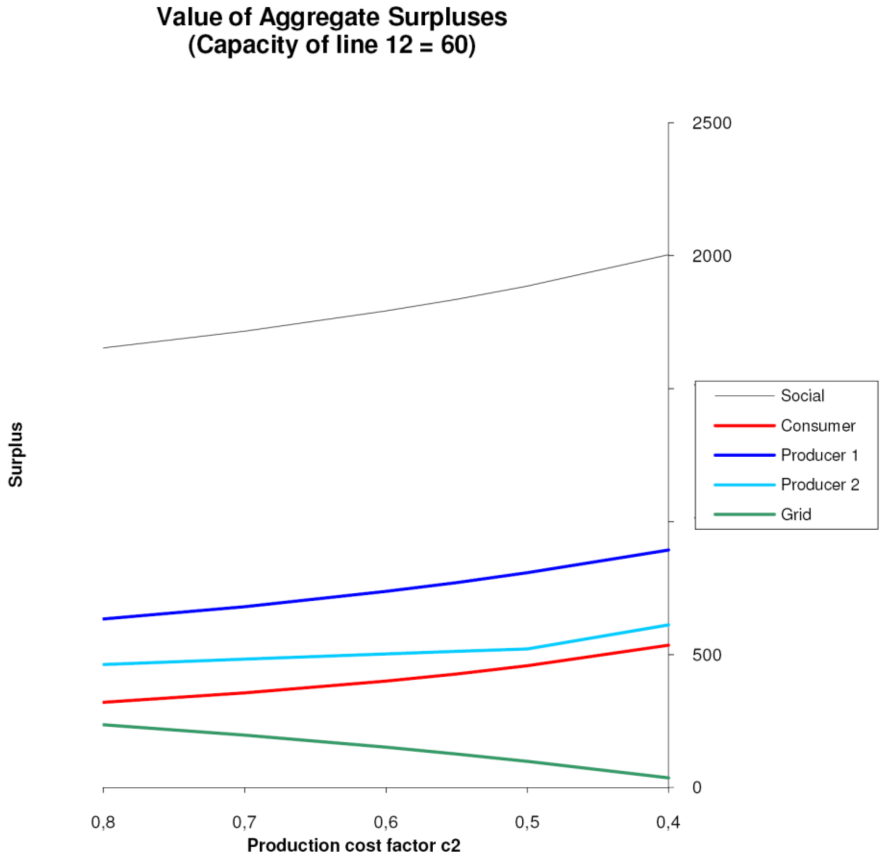


Figure 5: Value of aggregate surpluses as a function of the producer 2 cost function c_2.

The investment is socially profitable if the increase in social surplus exceeds the investment cost. However, while social surplus may be greatly increased, the increase in surplus for producer 2 may be modest. By taking into account investment costs, we might here easily find a scenario in which the investment is unprofitable for the producer, but highly profitable for the society as a whole. For example, considering the investment which lowers the production cost in node 2 from 0.8 to 0.6, we found that the increase in producer surplus in node 2, $\mathbf{PS}_2\left[\tilde{P}_2(0.6),0.6|a_0\right]-\mathbf{PS}_2\left[\tilde{P}_2(0.8),0.8|a_0\right]$, was 40.1 million. The increase in social surplus due to the investment, $\mathbf{SS}\left[\tilde{P}_2(0.6),0.6|a_0\right]-\mathbf{SS}_2\left[\tilde{P}_2(0.8),0.8|a_0\right]$ is, however, 1792.3 million - 1653.1 million = 139.2 million. We thus see that the investment exhibits positive externalities. For any investment costs above 40.1 million, producer 2 will find the investment unprofitable. For investments costs above 40.1 million and below 139.2 million, socially profitable investments will be turned down, as the investment is not profitable for the producer in node 2. This illustrates the issues of network externalities in investments in a network economy, and raises issues of the incentive mechanisms in the market. In this case, by subsidizing the investment costs of producer 2, a higher social surplus may be achieved.

7 Investments in Production and Grid Capacity

While individual investment alternatives in production or grid capacity may be profitable, the socially optimal investment plan may involve a mix of line investments and production capacity investments. Figure 6 shows the value of aggregate social surplus as a function of the capacity of line 12 for different alternative production costs factors c_2. The figure thus illustrates that different investment mixes may achieve the same gross gain in social surplus. As an example, the figure indicates that a gross social surplus of approximately 1750 million with either an investment in the production technology lowering the cost factor c_2 to near 0.65, or an investment raising the line capacity of line 12 $\hat{C}$ to approximately 108. Alternatively, combinations of line investment and production capacity or technology investments may render the same result. For example, the figure indicates that both the combination of $\hat{C} \approx 73$ and $c_2 \approx 0.7$, as well as the combination of $\hat{C} \approx 86$ and $c_2 \approx 0.75$ will render a gross social surplus of approximately 1750 million. Though the gain is the same, investment costs for the different investment alternatives will differ. For each level of gain illustrated, the least cost alternative, represents the optimal investment mix, and the investment is profitable if the gross gain in social surplus exceeds the investment costs of the least-cost investment combination.

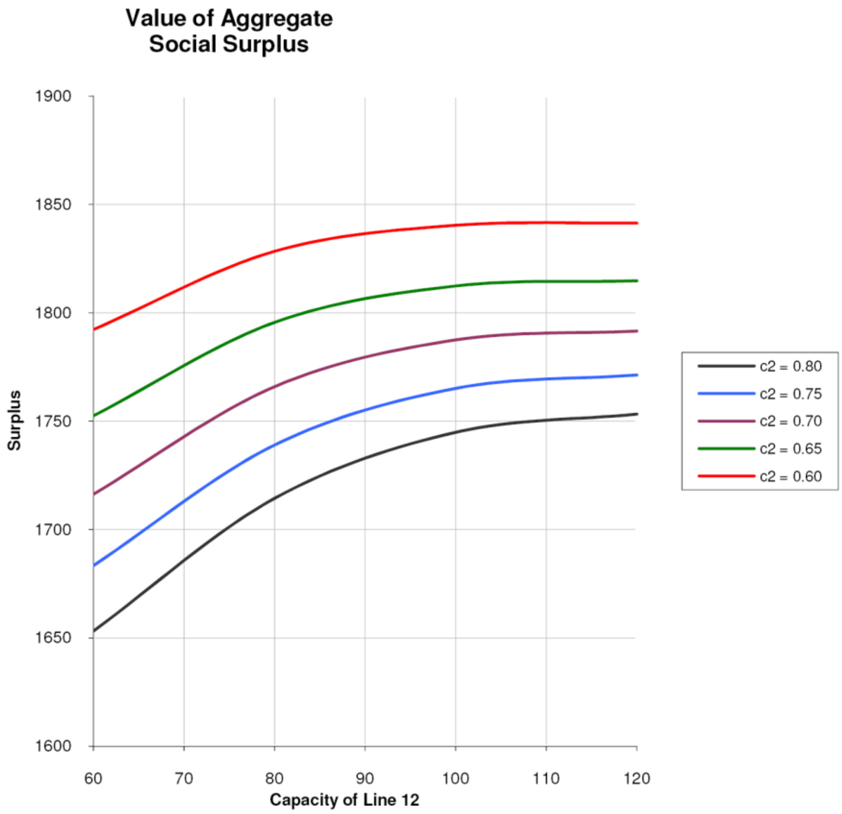


Figure 6: Value of aggregate social surplus as a function of line capacity, and for different alternatives of the producer 2 cost function c_2.

8 Concluding Remarks

The effects of investments in line capacity and production capacity are closely related in a capacitated network with loop-flow characteristics. While the effects are fairly well understood in a one-period deterministic setting, investments in a real world environment are made in a setting of major uncertainty, necessitating explicitly considering uncertainty. In this paper we have illustrated some potential pitfalls of evaluation in using price processes estimated on the basis of time series data. This framework for studying the investment problem gives room for studying further effects of investments in the integrated market, under different assumptions of uncertainty and network configuration, as well as in studying different issues, e.g. incorporating limited nodal production capacities, the effect of the real-world approaches to nodal pricing such as zonal pricing, and using real option theory to value flexible investment strategies.

Appendix

In this section we compute some useful expectations needed to calculate expected the expected surpluses, as prices in the case of congestion differ from the case where there is no congestion. An important function is $I_{\{a_t \le \hat{a}\}}$ which is 1 when $a_t \le \hat{a}$, i.e. when there is no congestion and non-differentiated nodal prices, and 0 otherwise indicating a congested net. Here follows the expected values of $E\left[I_{\{a_t \le \hat{a}\}} | a_s\right]$ and $E\left[I_{\{a_t > \hat{a}\}} | a_s\right]$, $E\left[a_t I_{\{a_t \le \hat{a}\}} | a_s\right]$ and $E\left[a_t I_{\{a_t > \hat{a}\}} | a_s\right]$, and $E\left[a_t^2 I_{\{a_t \le \hat{a}\}} | a_s\right]$ and $E\left[a_t^2 I_{\{a_t > \hat{a}\}} | a_s\right]$.

Assume $t > s$, and that we know a_s. When *a* is modeled by the Ornstein-Uhlenbeck process, we know that it is Gaussian with mean and variance given by equations (12) and (13). Therefore the probability that the net is un-congested at time *t*, given present state a_s, can be written as

$$\begin{aligned}
\overline{F}_t &\equiv E\left[I_{\{a_t \le \hat{a}\}} | a_s\right] \\
&= \int_{-\infty}^{\hat{a}} \left(\frac{1}{\sqrt{2\pi\rho_t}} e^{-\frac{(x-\mu_t)^2}{2\rho_t}} \right) dx \\
&= G^{\mu_t, \rho_t}(\hat{a}) \\
&= G\left(\frac{\hat{a} - \mu_t}{\sqrt{\rho_t}}\right)
\end{aligned}$$

where *G* denotes the cumulative standard normal distribution function. Thus, we also have that the probability that the net is congested at time *t*, given the state a_s, is

$$\begin{aligned}
E\left[I_{\{a_t > \hat{a}\}} | a_s\right] &= 1 - \overline{F}_t \\
&= 1 - G\left(\frac{\hat{a} - \mu_t}{\sqrt{\rho_t}}\right)
\end{aligned}$$

For calculating price expectations etc., we need several calculations. Recall that $\hat{a}$ is defined by (6), and that ρ_t and μ_t depend on s and a_s. We further calculate $E[a_t | a_s, a_t \le \hat{a}] \equiv E[a_t I_{\{a_t \le \hat{a}\}} | a_0]$:

$$\begin{aligned}
\overline{A}_t &\equiv E\left[a_t I_{\{a_t \le \hat{a}\}} | a_s\right] \\
&= \frac{1}{\sqrt{2\pi\rho_t}} \int_{-\infty}^{\hat{a}} x e^{-\frac{(x-\mu_t)^2}{2\rho_t}} dx \\
&= \frac{1}{\sqrt{2\pi\rho_t}} \int_{-\infty}^{\hat{a}} (x-\mu_t) e^{-\frac{(x-\mu_t)^2}{2\rho_t}} dx + \mu_t \int_{-\infty}^{\hat{a}} \frac{1}{\sqrt{2\pi\rho_t}} e^{-\frac{(x-\mu_t)^2}{2\rho_t}} dx \\
&= -\frac{\sqrt{\rho_t}}{\sqrt{2\pi}} e^{-\frac{(\hat{a}-\mu_t)^2}{2\rho_t}} + \mu_t G\left(\frac{\hat{a}-\mu_t}{\sqrt{\rho_t}}\right)
\end{aligned}$$

This clearly implies that

$$\overline{B}_t \equiv E\left[a_t I_{\{a_t > \hat{a}\}} | a_s\right] = E\left[a_t | a_s\right] - \overline{A}_t = \mu_t - \overline{A}_t$$

In addition note that $E\left[a_t^2 | a_s\right] = \rho_t + \mu_t^2$, and that

$$\begin{aligned}
\overline{C}_t &\equiv E\left[a_t^2 I_{\{a_t \le \hat{a}\}} | a_s\right] \\
&= \frac{1}{\sqrt{2\pi\rho_t}} \int_{-\infty}^{\hat{a}} (x-\mu_t)^2 e^{-\frac{(x-\mu_t)^2}{2\rho_t}} dx + \frac{2\mu_t}{\sqrt{2\pi\rho_t}} \int_{-\infty}^{\hat{a}} (x-\mu_t) e^{-\frac{(x-\mu_t)^2}{2\rho_t}} dx \\
&\quad + \mu_t^2 \int_{-\infty}^{\hat{a}} \frac{1}{\sqrt{2\pi\rho_t}} e^{-\frac{(x-\mu_t)^2}{2\rho_t}} dx \\
&= -\frac{\sqrt{\rho_t}}{\sqrt{2\pi}} (\hat{a}-\mu_t) e^{-\frac{(\hat{a}-\mu_t)^2}{2\rho_t}} + \rho_t G\left(\frac{\hat{a}-\mu_t}{\sqrt{\rho_t}}\right) - \frac{2\mu_t\sqrt{\rho_t}}{\sqrt{2\pi}} e^{-\frac{(\hat{a}-\mu_t)^2}{2\rho_t}} + \mu_t^2 G\left(\frac{\hat{a}-\mu_t}{\sqrt{\rho_t}}\right) \\
&= -\frac{\sqrt{\rho_t}}{\sqrt{2\pi}} (\hat{a}+\mu_t) e^{-\frac{(\hat{a}-\mu_t)^2}{2\rho_t}} + (\mu_t^2 + \rho_t) G\left(\frac{\hat{a}-\mu_t}{\sqrt{\rho_t}}\right)
\end{aligned}$$

Finally also observe that

$$\begin{aligned}
\overline{D}_t &\equiv E\left[a_t^2 I_{\{a_t > \hat{a}\}} | a_s\right] \\
&= E\left[a_t^2 | a_s\right] - E\left[a_t^2 I_{\{a_t \le \hat{a}\}} | a_s\right].
\end{aligned}$$

References

Bjørndal, M. (2000): *Topics on Electricity Transmission Pricing*, PhD thesis, Norwegian School of Economics and Business Administration, March 2000.

Johnson, B. and G. Barz (1999): ‘Selecting stochastic processes for modelling electricity prices. In *Energy Modelling and the Management of Uncertainty*, pages 3-21, RISK Books, 1999.

Lucia, J.J. and E.S. Schwartz (2002): ‘Electricity prices and power derivatives: Evidence from the Nordic power exchange. *Review of Derivatives Research*, 5(1):5-50.

Lund, A.-C. and F. Ollmar (2003): ‘Analyzing flexible load contracts in the energy market’,in *An Analysis of Derivatives Prices in the Nordic Power Market*, PhD thesis, Norwegian School of Economics and Business Administration, September 2003.

Lund, A.-C. and L. Rud (2006): ‘Understanding the stochastics of nodal prices: Investments and price processes in a constrained network’, SNF working paper 03/06, or see pp. 199-227 of this thesis.

Weron, R., M. Bierbrauer, and S. Trück (2004): ‘Modeling electricity prices: jump diffusion and regime switching’, *Physica A*, (336):39-48.

ESSAY 5

A Newsboy Model Perspective on the Power Market: The Case of the Wind Power Producer*

Linda Rud

Abstract

Electricity is a continuously supplied and consumed commodity, where aggregate production and consumption momentarily must balance at the time of delivery. With cost and allocational benefits of pre-delivery planning, electricity market systems have in different ways accommodated the necessity of planning. As such, market systems range from integrated markets that optimize planned versus real-time production, to more loosely coupled unbundled markets with separate pre-delivery markets and real-time markets. This paper discusses the decision problem related to a momentary inflexible producer, such as a wind power producer, and how the problem may be interpreted within the classic newsboy model. In a setting of a day-ahead and a real-time market, the results indicate that the optimal sales bid of the wind power producer might diverge from the expected production. This aspect is also found in the setting of market optimization, where the uncertainty of the wind power production has direct implications for the optimal level of planned production by other producers.

* Comments from Petter Bjerksund, Mette Bjørndal, and Kurt Jørnsten are greatly appreciated. Any remaining errors are the responsibility of the author.

1 Introduction

Electricity production and consumption is inherently uncertain. At the same time electricity is also a continuously delivered commodity, and at the time of delivery aggregate production and consumption of electricity must momentarily balance, as electricity cannot be cost-efficiently stored on a large scale. With the cost and allocational benefits of pre-delivery planning, a challenge of any market system is to induce the optimal mix of production and consumption with respect to pre-delivery planned decisions versus momentary real-time delivery decisions. In this paper our focus will be on market participants that prior to delivery face uncertainty with respect to their real-time production or consumption, and who in real-time also are momentarily inflexible meaning that they cannot adjust their level of production or consumption in real-time. Examples here include wind power producers and river generators, where the real-time power production is nature-given and to a large extent uncontrollable. Likewise, also part of the demand for electricity may be momentarily inflexible. With an increasing focus on wind power in many electricity markets, our scope here will in particular be on the wind power producer, and the implications of this production technology for the planned and real-time balance of the electricity market.

Electricity market systems have accommodated the necessity of planning in different ways. As such, market systems range from integrated market systems that optimize the allocation of planned production versus real-time production capacity, to more loosely coupled unbundled markets with separate pre-delivery and real-time markets. In an unbundled market such as the Nordic market, an important question encountered by the wind power producer, is related to the bid to be placed in the day-ahead market. A common approach is to try to accurately forecast the coming electricity load, and submit a bid that reflects the expected power production. Using the insight of the classic newsboy model, we will argue that in a market where real-time load adjustments are more costly than planned electricity load, the optimal bid of the inflexible participant may diverge from the expected load. These results can also be studied in the benchmark model of an optimized electricity plan. In both cases, we find that the optimal action by the inflexible producer will have implications for the optimal planned load of the other market participants. As such the classical newsboy problem sheds light upon the optimal decisions by individual market participants, as well as the optimal benchmark solution in an aggregate market setting.

Section 2 considers the optimal bidding of an inflexible producer, such as a wind power producer, in an unbundled market setting with a day-ahead market and a momentary deviation market. We construct a stylized market setting of a pre-planning period and a momentary delivery period, where the price structure is given, and where the sole uncertainty is related to the level of the real-time power production. The task of the wind power producer is here to find the optimal sales bids in the day-ahead market. This task may be interpreted as a classic newsboy problem, where we find that the optimal bid in the day-ahead market may deviate from expected production.

Turning to the aggregate market, in section 3 we consider the benchmark model of a benevolent optimizer which optimizes the surplus of the entire market. This may also be interpreted as the case of an integrated market which optimizes social surplus, and where participants truthfully have revealed their preferences and costs. Using a stylized representation of the market participants, and with uncertainty only as to production of the inflexible producer, the optimal market plan may also be interpreted in the light of the classic newsboy problem.

In section 4 we consider the full market solution of the unbundled market setting, where the optimal wind power bid of section 2 interacts with the demand bid and the bid of the flexible producer, to form the market equilibrium. Under the given set of assumptions, we show that the market solution equals that of the optimizing model of section 4. We also argue that any restrictions on bidding related to the allowed deviation from the expected power load, will contribute to a lower expected social surplus in the market. Section 5 concludes the paper by discussing challenges for further research.

2 The Momentary Inflexible Participant in an Unbundled Market: The Case of a Wind Power Generator

Consider a generator who has no flexibility to control his electricity load at the time of delivery, which is time period T. An example of the momentarily inflexible generator is that of a wind-power generator, where the generation X in the specific delivery period T is uncertain and nature-given, and where only the distribution of X is known prior to time T. Let us now place this generator in an electricity market, organized as an unbundled market, with a day-ahead market and a real-time market. In addition to the inflexible producer, we assume that the market participants also comprise flexible producers, and demand. Market clearing of the day-ahead market brings forth the spot price, and the corresponding spot commitment to be delivered in time period T. Market clearing at real-time balances real-time production and real-time consumption, and sets the price of settling deviations. Our concern in this section is how the inflexible wind power producer should act in such a market.

In this unbundled market the wind power producer places a bid in the day-ahead market, by stating his supply curve, i.e. which quantity he wishes to sell at different day-ahead prices. The equilibrium market solution then reveals his commitment, which thus represents his planned production. At time T, however, the realized quantity X of the wind power plant will be produced regardless of the day-ahead market price and equilibrium commitment. Any deviations between the realized production and the day-ahead commitment are then settled at the real-time market price. We assume that the producer wishes to maximize his expected revenue of selling X. The problem of the wind power generator is then to find the optimal bid in the day-ahead market.

The problem is analyzed within a stylistic model of the market, with no uncertainty except for the uncertainty related to the amount X of wind power that will be generated at time T. A main feature of our market model is furthermore that the momentary adjustments in the electricity load are more costly than if the changes were made as a planned delivery load. The two markets, the day-ahead market and the real-time market, are modeled as follows:

- *The day-ahead market*, also called the spot market, is a market for planned delivery. The market is cleared at time t on the market price p_t, while delivery takes place during the real-time period of T. The submitted supply and demand bids state the quantity the participant wishes to sell or buy at different prices. The equilibrium market price is found at the intersection of the demand and supply curves submitted by participants in the market[143]. The market equilibrium of the day-ahead market thus constitutes the planned consumption and production of the market participants.

- *The real-time power market* handles all deviations that occur between the planned day-ahead contracted delivery, and the actual delivery at time T. In our model, the real-time price will be either the up-regulation price p_T^u or the down-regulation price p_T^d. We assume that there is no uncertainty as to the real-time prices in the sense that their

[143] We here assume a competitive market and the absence of grid congestion.

structure is known prior to time T, and more specifically given as a function of the resulting day-ahead price. We have defined the real-time up-regulation price to be $p_T^u = \alpha^u p_t$, where $\alpha^u > 1$. The real-time down-regulation price is defined to be $p_T^d = \alpha^d p_t$, where $0 < \alpha^d < 1$. In our setting, both α^u and α^d are known constants. We also assume that the deviation caused by the individual participant is settled by the up-regulation price p_T^u if his deviation causes an *up-regulation* of the load by the flexible participants, and likewise by the down-regulation price p_T^d if his deviation causes a *down-regulation* of the load by the flexible participants.

Note that up-regulation is defined as a real-time ordered increase in production, or equivalently, a real-time ordered decrease in consumption to be carried out by the momentary flexible participants. As such the inflexible generator will cause a need for up-regulation if his actual production is lower than his planned production. Likewise, we define down-regulation as a real-time ordered decrease in production, or equivalently, real-time increase in consumption by the momentary flexible participants. The inflexible generator causes a need for down-regulation if his actual production is higher than his planned production.

In our stylized model, we assume that the inflexible participant is free to set his bid in the day-ahead market, meaning that he is free to choose the degree to which he wishes to use the day-ahead market. In the extreme case the inflexible participant may offer all his production in the real-time market only, i.e. by not submitting any bids in the day-ahead market. However, when he at time T generates a production of X, he will in effect cause an imbalance relative to his planned market schedule, which in the case of no bidding implicitly is a plan of 0 generation. This has consequences real-time for the rest of the market. Note that the equilibrium clearing of the day-ahead market resulted in a demand and production plan for the other participants which in aggregate was in balance. Due to the requirement of an overall real-time momentary balance, the generation of X by the inflexible generator thus necessitates action by the momentary flexible participants. They have to down-regulate by reducing their production or increasing their consumption[144]. Our inflexible producer, having

[144] Flexible producers, having sold an electricity load at the spot price of p_t, will by actively down-regulating, reduce their generation relative to this planned level. Thus they in effect buy electricity back at a price of p_T^d. In this they earn a per unit margin of $p_t - p_T^d$, to cover extra costs and a profit for offering this service. Likewise, flexible consumers who contribute to the real-time ordered down-

generated X, is then remunerated by the real-time down-regulation price p_T^d. With the described price structure, this price is always lower than the day-ahead spot price of p_t. The strategy of solely relying on the real-time market then obviously cannot be the optimal strategy of the inflexible participant. It is more profitable for the inflexible generator, if possible, to sell in the spot market at the day-ahead price.

The wind power generator would thus prefer to sell all his power in the day-ahead spot market. If he were to correctly guess the realization of X, and trade on this basis in the spot market, he would be remunerated by the spot price p_t, and obtain the highest possible selling price. The problem of the inflexible participant is, however, that at time t when the day-ahead market is cleared, the time T realization of X is uncertain. The generator will then cause a real-time imbalance for any realization of X which is to deviate from his day-ahead spot trade Q^w. This imbalance is settled at the real-time price which, relative to the spot price, always is to the disadvantage of the inflexible participant. The question facing the momentarily inflexible participant is thus to find the sales bid in the spot market which will maximize his revenue. In submitting a supply schedule, this is equivalent to stating the supply offer Q^w as a function of p_t. We will in the following analyze this decision within our model, and show that the problem is near equivalent to the classic newsboy problem[145].

By offering the quantity Q^w in the spot market, the producer receives a spot market revenue of $p_t Q^w$. The total revenue of the generator, however, depends upon the realization of X. If the realized quantity is greater than the spot commitment, i.e. $X > Q^w$, the system operator has to order flexible participants to down-regulate their production by the amount of $X - Q^w$. The revenue received by the generator for the extra production $X - Q^w$ is thus $(X - Q^w) p_T^d$. If, however, the realized quantity is less than the spot commitment, i.e. $X < Q^w$, the system operator has to order flexible participants to up-regulate their production by an amount of $(Q^w - X)$ to cover the missing delivery by the participant. In this case the producer has to buy electricity in the real-time market to cover his full obligations of Q^w, paying

regulation, do this by increasing their load. They then in effect buy electricity at a price of p_T^d, and as such earn a compensation by obtaining a rebate relative to the spot price p_t.

[145] See e.g. Khouja (1999) for an overview of literature related to the newsboy (news-vendor) problem.

$(Q^w - X)p_T^u$. Our task is now to find the spot order Q^w at the given spot price p_t which maximizes the expected profit of the inflexible generator.

Let us now denote the down-regulation caused by the inflexible generator as

$$(X - Q^w)^+ = \begin{cases} X - Q^w & \text{if } X \geq Q^w \\ 0 & \text{otherwise} \end{cases}.$$

That is, down-regulation is $X - Q^w$ if generation exceeds the spot trade, and 0 otherwise. Likewise, the up-regulation caused by the inflexible generator is

$$(Q^w - X)^+ = \begin{cases} Q^w - X & \text{if } Q^w \geq X \\ 0 & \text{otherwise} \end{cases}.$$

That is, up-regulation is $Q^w - X$ if generation is less than spot trade, and 0 otherwise.

Assuming that the wind power producer has negligible or zero variable costs, the profit, excluding fixed costs, of the wind power producer for the delivery period T may be represented as:

$$\Pi^w = p_t Q^w + p_T^d (X - Q^w)^+ - p_T^u (Q^w - X)^+ \qquad (1)$$

In the classic newsboy problem, the newsboy has to decide upon the number of papers to buy beforehand, weighing the expected salvage loss of over-stocking against the opportunity costs from under-stocking. For the newsboy, the stock of papers cannot be replenished upon observing the demand. For the case of the inflexible electricity market participant, the problem is slightly different. In this case the inflexible producer beforehand has to decide upon a committed production, but is, depending on the realized production, either compelled to 'replenish' stock to meet the full commitment, or 'salvage' excess production at a lower price. More accurately put, he is compelled to balance his commitment by means of the real-time market if the actual production is to deviate from spot trade. This balancing is made at unfavorable terms compared to the spot market, i.e. by buying at $p_T^u > p_t$ or selling at $p_T^d < p_t$.

On deciding on the pre-planned level of production he has to weigh the expected salvage cost of excess production against the opportunity cost of estimating too low a production. For our inflexible generator, we find that if spot sales Q^w exceed the realized generation X, i.e. $Q^w > X$, the generator suffers a per unit loss of $(p_T^u - p_t)$ for the overestimated quantity of $(Q^w - X)$, i.e. by having to buy each unit at a higher price than the revenue obtained in the spot market. If the realized generation X exceeds spot sales Q^w, i.e. $X > Q^w$, the generator suffers a per unit opportunity cost of $(p_t - p_T^d)$, for the total underestimated quantity $(X - Q^w)$, by having to sell each unit at a lower price than would have been received if offered in the spot market.

Let this motivate the following rearrangement of the profit represented in equation (1): We substitute the term $p_t Q^w$ with $p_t Q^w = p_t(X - (X - Q^w)^+ + (Q^w - X)^+)$, and then rearrange terms[146]. The wind power producer's profit Π^w in (1) may thus be written as $\Pi^w = p_t X - (p_t - p_T^d)(X - Q^w)^+ - (p_T^u - p_t)(Q^w - X)^+$, or in terms of expected revenue as

$$E[\Pi^w] = p_t E[X] - (p_t - p_T^d)E[(X - Q^w)^+] - (p_T^u - p_t)E[(Q^w - X)^+] \qquad (2)$$

The first term $p_t E[X]$ is the spot market revenue that would have been obtained if the expected production were sold on the spot market. In other words, if there were no uncertainty, and the supply was certain to be $E[X]$, this would be the revenue of the generator. The two following terms represent the cost of uncertainty. The first term $(p_t - p_T^d)E[(X - Q^w)^+]$ is the expected cost of bidding too low a quantity in the spot market, where the alternative cost compared with spot sales is $(p_t - p_T^d)$ for each underestimated unit of generation. The second term $(p_T^u - p_t)E[(Q^w - X)^+]$ is the cost of bidding too high a quantity in the spot market, where the generator suffers a loss of $(p_T^u - p_t)$ per unit overestimated, having sold at p_t and having to buy at the higher price of p_T^u.

[146] We have thus inserted for $Q^w = X - (X - Q^w)^+ + (Q^w - X)^+$. Note that if $Q^w > X$, then $Q^w = X - 0 + (Q^w - X) = Q^w$. Likewise, if $Q^w < X$, then $Q^w = X - (X - Q^w) + 0 = Q^w$.

The expected revenue $E[\Pi^w]$ of the generator is maximized by solving the first order condition $\frac{\partial E[\Pi^w]}{\partial Q^w} = 0$, and finding the value of Q^w that solves the equation. First note that[147] $\frac{\partial E[(X-Q^w)^+]}{\partial Q^w} = -\Pr(X > Q^w)$, and $\frac{\partial E[(Q^w-X)^+]}{\partial Q^w} = \Pr(Q^w > X)$ where $\Pr(X > Q^w)$ is the probability that $X > Q^w$, and $\Pr(Q^w > X) = 1 - \Pr(X > Q^w)$ is the probability that $Q^w > X$. The first order condition is then given by

$$\frac{\partial E[\Pi]}{\partial Q^w} = (p_t - p_T^d)\Pr(X > Q^w) - (p_T^u - p_t)(1 - \Pr(X > Q^w)) = 0 \tag{3}$$

By rearranging terms we find that the optimal sales bid is given by the amount Q^w where the probability that X exceeds Q^w is given by the following equation:

$$P(X > Q^w) = \frac{p_T^u - p_t}{p_T^u - p_T^d} \tag{4}$$

With the pricing rules of our model, i.e. $p_T^u = \alpha^u p_t$ ($\alpha^u > 1$), and $p_T^d = \alpha^d p_t$ ($0 < \alpha^d < 1$), this may further be specified as

$$P(X > Q^w) = \frac{\alpha^u - 1}{\alpha^u - \alpha^d} \tag{5}$$

In the following we will assume that X is a uniform random variable on the interval $(0, \hat{X})$, i.e. with the minimum of 0 (no wind) and the maximum of $\hat{X}$ (capacity of the wind power plant). The density function is then

$$f(x) = \begin{cases} \frac{1}{\hat{X}} & if\ 0 < X < \hat{X} \\ 0 & otherwise \end{cases} .$$

Then we have

[147] We have $E[(X - Q^w)^+] = \int_{Q^w}^{\infty} (x - Q^w) f(x)dx$, where $f(x)$ is the density function of X. Then $\frac{\partial E[(X-Q^w)^+]}{\partial Q^w} = \int_{Q^w}^{\infty} - f(x)dx = -\Pr(X > Q^w)$. Also $E[(Q^w - X)^+] = \int_{-\infty}^{Q^w} (Q^w - x) f(x)dx$, and we have $\frac{\partial E[(Q^w-X)^+]}{\partial Q^w} = \int_{-\infty}^{Q^w} f(x)dx = \Pr(X < Q^w)$.

$$P(X > Q^w) = \begin{cases} \frac{\hat{X}-Q^w}{\hat{X}} & for\, Q^w < \hat{X} \\ 0 & for\, Q^w > \hat{X} \end{cases} .$$

Due to the relative price structure of the two markets, it is reasonable to assume that the producer never will choose to place a bid $Q^w > \hat{X}$ which obviously is unprofitable. Thus, we have that $P(X > Q^w) = \frac{\hat{X}-Q^w}{\hat{X}}$, and the first order condition is $\frac{\hat{X}-Q^w}{\hat{X}} = \frac{\alpha^u - 1}{\alpha^u - \alpha^d}$, which implies the optimal bid of

$$Q^{w*} = \hat{X}\,\frac{1-\alpha^d}{\alpha^u - \alpha^d} \tag{6}$$

The optimal bidding strategy thus is dependent upon the price structure of the real-time market, and the relation between the prices of up-regulation and down-regulation. Noting that the expectation of X is $E[X] = \frac{1}{2}\hat{X}$, we see that only in the case that $\frac{1-\alpha^d}{\alpha^u-\alpha^d} = \frac{1}{2}$, i.e. $\alpha^u = 2 - \alpha^d$, will it in our model be profitable for the wind power producer to bid the expected quantity. In all other cases it is profitable to submit a bid which deviates from the expected level of production, as in the original newsboy model. This is contrary to what seems to be a common bidding approach where the wind power producer bids according to expected wind.

3 The Market Optimizer

The wind power generator typically operates in a market with demand and other producers, and where at least part of the other participants are able to flexibly adjust their load in real time. Before turning to the setting of a day-ahead and a real-time market in the next section, let us start by considering the benchmark case of a benevolent optimizer that has the task of finding the maximum social surplus of the market, and study the optimal market plan in this setting. To focus on the essence of the planning problem related to the inflexible and stochastic producer, we will still operate within a highly stylized model with the sole source of uncertainty connected to the production of wind power. Further, an important feature of the model is the assumption that the costs of real-time load adjustment by flexible producers are higher than their costs in producing a planned load. We also, more specifically, assume that the market consists of three representative participants:

- *The wind power producer:* The wind power producer produces a quantity X at time T. No variable costs occur during production. At time $t < T$, X is a stochastic variable, however, with a known distribution. For simplicity we assume that X is a random uniform variable with the minimum of 0 and the maximum of $\hat{X}$.

- *The flexible producer:* The flexible producer generates power at time T. The production may be planned on beforehand at time $t < T$, or decided upon in real-time at time T. Planned production is in general cheaper than any production following from production decisions made at time T. We assume there are no capacity limits, nor any uncertainty with regard to production costs. More specifically, we assume that the producer has constant marginal costs. The marginal cost of planned production is c_t. At time T the producer may up-regulate his production at a per unit marginal cost of $c_T^u = \alpha^u c_t$, where $\alpha^u > 1$, i.e. at a higher cost than planned production. At time T the producer may likewise reduce his production relative to the planned production at time t. Though the producer by the one unit down-regulated is spared the production of the unit, the full cost of c_t is not recovered, as the per unit cost recovery only is $c_T^d = \alpha^d c_t$, where $0 < \alpha^d < 1$. Thus planned production is more expensive than unplanned production[148].

- *Demand:* In our stylistic model we also assume that there is no uncertainty in demand. Demand is characterized by a falling marginal benefit, where the total benefit of a demand D is given as $\Pi_{benefit}^D = aD - \frac{1}{2}bD^2$, which would correspond to a linear demand curve.

[148] To see this, let the required quantity be the certain amount of D, and let us for a moment assume that there are no other producers such as the wind power producer. Let Q^f be the planned quantity of the flexible producer. If $D = Q^f$, the total cost of production by the flexible producer is $c_t D$. If $D > Q^f$, a real-time up-regulation of the amount $D - Q^f$ is needed, at the cost of $c_T^u = \alpha^u c_t$ per unit. The total cost of producing D given the planned quantity of $\mathrm{Q}^\mathrm{f} < \mathrm{D}$ is then $c_t Q^f + \alpha^u c_t (D - Q^f) = c_t D + (\alpha^u - 1) c_t (D - Q^f)$, which is a higher cost. If $D < Q^f$, a down-regulation of the amount $Q^f - D$ is required. Since planned costs are $c_t Q^f$, the real-time down-regulation means a cost reduction of $\alpha^d c_t (Q^f - D)$, with the total costs of $c_t Q^f - \alpha^d c_t (Q^f - D) = c_t D + (1 - \alpha^d) c_t (Q^f - D)$ which thus also is a higher cost compared to having the total demand covered by planned production.

The task of the optimizer is to find the level of demand and planned production that maximizes the expected social surplus, which is defined as the benefit of total demand less expected costs. Without the wind power producer, this is straight forward within our model. In this case there is no uncertainty. We established above that it is not profitable, nor necessary, in this scenario to rely on any unplanned production, as this only would increase production costs. In this case the optimal production plan is given by $Q^f = D$, and the social surplus is $\Pi^{SS} = aD - \frac{1}{2}bD^2 - c_t D$. Optimal demand is $D = \frac{a-c_t}{b}$, found by the equation $\frac{\partial \Pi^{SS}}{\partial D} = a - bD - c_t = 0$, and with a total social surplus of $\Pi^{SS} = \frac{1}{2b}(a-c_t)^2$.

What is then the contribution of the wind power producer? The main benefit of the wind power production in this model is related to the generation of power at low marginal costs, which here are assumed to be 0. This benefit is obvious in the case of certainty: Let us assume that the capacity of the wind power producer always is lower than any relevant total demand level, i.e. $X < D$. Consider a generation of X by the wind power producer that is known with certainty. The amount to be generated by the flexible producer is then $D - X$. In this case of certainty, it is not profitable to rely on unplanned production, so that the planned production by the flexible producer is exactly $Q^f = D - X$. In the case of certainty, the social surplus when including the wind power producer is $\Pi^{SS} = aD - \frac{1}{2}bD^2 - c_t(D-X)$. The optimal demand level is as before $D = \frac{a-c_t}{b}$, as the first order condition is unchanged, i.e. $\frac{\partial \Pi^{SS}}{\partial D} = a - bD - c_t = 0$. The total optimal social surplus is, however, now $\Pi^{SS} = \frac{1}{2b}(a-c_t)^2 + c_t X$, where the addition of $c_t X$ represents the benefit of including the wind power generator[149].

In addition to offering low marginal costs, a main effect of wind power is, however, related to the uncertainty of generation, i.e. the level of X is uncertain prior to time T . This uncertainty implicitly has implications for the optimal running of the remaining power system, in this case for the optimal production plans of the flexible producers. For a given level of demand, wind power will cover part of the demand, and the rest will be generated by the flexible producer. As the amount of wind power is uncertain, the exact amount to be produced by the flexible producer cannot be planned with certainty, and real-time adjustments in generation will be

[149] Note that this is the benefit of including the production of a wind power plant that already is built. As such, this is by no means a total analysis of the profitability of wind power investment, which must include an analysis of investment costs, future prices, potential generation, and on an aggregate economic level also network effects and implications for the cost of system operation.

needed. The costs of real-time adjustment are higher than for the equivalent planned production, thus implying that the marginal cost saving of using wind power is lower than illustrated in the case of certainty.

Let us study this issue in terms of the decision problem of the benevolent optimizer in the case of uncertain wind power production. We assume that the task of the optimizer is to find the course of action that gives the highest expected social surplus. The optimizer has to decide upon two questions, i.e. what is the optimal level D^* of demand, and what level of production Q^{f*} should be planned for the flexible participants. Note that there are no decisions to be made regarding the wind power producer, as the generation by this participant is nature-given.

We will solve the decision problem of the optimizer in two stages. We first find the optimal level of planned production for each possible demand level. The objective is here to establish the production strategy, i.e. the level of planned versus unplanned production by the flexible producers, that will render the lowest expected costs of meeting a given demand level. Then, having established the relevant cost function, we find the optimal level of demand.

i) Optimal Q^f for a given demand level D

Our task is now to find the production strategy that covers a given demand level of D at the lowest possible expected cost. The production of wind power is given by X which is revealed at time T. Given a demand of D, the flexible producer will have to generate an amount of $D - X$, which thus is uncertain at the time of planning. This may be done with a combination of planned production Q^f, and unplanned production which is the residual $D - X - Q^f$. Planned production Q^f is generated at a per unit cost of c_t. If the planned production turns out to be lower than the required level at time T, i.e. $Q^f < (D - X)$, up-regulation is required, and an unplanned quantity of $(D - X - Q^f)^+$ has to be produced at a per unit cost of c_T^u. If planned production turns out to be higher than the required level at time T, i.e. $Q^f > (D - X)$, down-regulation is required, and the producer has to lower his production by $(Q^f - D + X)^+$, lowering the total costs by this quantity multiplied by the per unit amount of c_T^d. The total cost Π_{cost}^{SS} to cover the demand level D is thus $\Pi_{cost}^{SS} = c_t Q^f + c_T^u (D - X - Q^f)^+ - c_T^d (Q^f - D + X)^+$, or alternatively expressed[150] as

[150]We have here substituted for $c_t Q^f = c_t ((D - X) - (D - X - Q^f)^+ + (Q^f - D + X)^+)$.

$\Pi_{cost}^{SS} = c_t(D-X) + (c_T^u - c_t)(D-X-Q^f)^+ + (c_t - c_T^d)(Q^f - D + X)^+$. The expected cost for meeting a demand of D is thus:

$$E[\Pi_{cost}^{SS}] = c_t E[D-X] + (c_T^u - c_t)E[(D-X-Q^f)^+] + (c_t - c_T^d)E[(Q^f - D + X)^+] \qquad (7)$$

The first part shows the costs if the entire necessary quantity by other producers of $\mathrm{D-X}$ were to be planned. The second and third parts show the expected extra costs by respectively planning too low a quantity, and too high a quantity. The minimum expected cost for meeting a demand of D is found by solving the first order condition of $\frac{\partial E[\Pi_{cost}^{SS}]}{\partial Q^f} = 0$, and finding the value of Q^f that solves the equation. To derive the first order condition, first note[151] that $\frac{\partial E[(D-X-Q^f)^+]}{\partial Q^f} = -\Pr(X < (D-Q^f))$, and $\frac{\partial E[(Q^f-D+X)^+]}{\partial Q^f} = \Pr(X > (D-Q^f))$. Also note that $\Pr(X > (D-Q^f)) = 1 - \Pr(X < (D-Q^f))$. We then have

$$\frac{\partial E[\Pi_{cost}^{SS}]}{\partial Q^f} = -(c_T^u - c_t)\Pr(X < (D-Q^f)) + (c_t - c_T^d)(1 - \Pr(X < (D-Q^f))) = 0 \qquad (8)$$

which implies that the optimal production plan is given by the Q^f that satisfies the equation of

$$\Pr(X < (D-Q^f)) = \frac{c_t - c_T^d}{c_T^u - c_T^d} = \frac{1-\alpha^d}{\alpha^u - \alpha^d} \qquad (9)$$

To find the optimal Q^f for meeting the demand level D, we thus have to insert the appropriate representation of $\Pr(X < (D-Q^f))$. In our model, X is a random variable on the

[151]To see this, note that $E[(D-X-Q^f)^+] = \int_{-\infty}^{D-Q^f} (D-x-Q^f) f(x)dx$. Then $\frac{\partial E[(D-X-Q^f)^+]}{\partial Q^f} = \int_{-\infty}^{D-Q^f} -f(x)dx = -\Pr(X < (D-Q^f))$. Also, $E[(Q^f - D + X)^+] = \int_{D-Q^f}^{\infty} (Q^f - D + x) f(x)dx$, and we have $\frac{\partial E[(Q^f-D+X)^+]}{\partial Q^f} = \int_{D-Q^f}^{\infty} f(x)dx = \Pr(X > (D-Q^f))$.

interval $(0,\hat{X})$, which implies[152] that $\Pr(X<(D-Q^f))=\frac{D-Q^f}{\hat{X}}$. By inserting this in (9), and solving for D, we thus have the optimal planned quantity of

$$Q^f = D-\hat{X}\,\frac{1-\alpha^d}{\alpha^u-\alpha^d} = D-\phi \tag{10}$$

where $\phi=\hat{X}\,\frac{1-\alpha^d}{\alpha^u-\alpha^d}$. We see that the planned quantity equals the entire demand level, reduced by a quantity ϕ due to wind power generation. Note that the term ϕ equals that of the optimal sales bid by the wind power producer found in equation (6). We thus see that the planned production Q^f, for any $\alpha^u \neq 2-\alpha^d$, deviates from total demand, not by the expected wind power production, but by the term $\hat{X}\,\frac{1-\alpha^d}{\alpha^u-\alpha^d}$.

ii) Optimal demand level

The optimal demand level may now be found by maximizing the expected social surplus. Social surplus is given by the total benefit less total costs, ie. $\Pi^{SS}=aD-\frac{1}{2}bD^2-c_t(D-X)-(c_T^u-c_t)(D-X-Q^f)^+-(c_t-c_T^d)(Q^f-D+X)^+$, which, by substituting for $Q^f=D-\phi$, may be represented by the expression $\Pi^{SS}=aD-\frac{1}{2}bD^2-c_t(D-X)-(c_T^u-c_t)(\phi-X)^+-(c_t-c_T^d)(X-\phi)^+$. Due to the certainty of demand, we see that the up-regulation and down-regulation costs are in fact not affected by the choice of demand. The expected social surplus is:

$$E[\Pi^{SS}]=aD-\tfrac{1}{2}bD^2-c_tE[D-X]-(c_T^u-c_t)E[(\phi-X)^+]-(c_t-c_T^d)E[(X-\phi)^+] \tag{11}$$

The optimal demand level is thus found by the first order condition $\frac{\partial E(\Pi^{SS})}{\partial D}=a-bD-c_t=0$, where the demand level thus is given as in the case of certainty, i.e.

$$D^*=\frac{a-c_t}{b} \tag{12}$$

[152] This is under the assumption that we always have $(D-\hat{X})<Q^f$, which is a reasonable assumption as a lower choice of Q^f implies that we with certainty plan an amount to be produced at the real-time cost, a strategy which argued above is not cost efficient.

By inserting this into expression (11) for the expected social surplus, inserting for $c_T^u = \alpha^u c_t$ and $c_T^d = \alpha^d c_t$, and noting that with our uniform distribution of X we have the expectations[153] $E[X] = \frac{1}{2}\hat{X}$, $E[(\phi - X)^+] = \frac{1}{2}\hat{X}(\frac{1-\alpha^d}{\alpha^u - \alpha^d})^2$, and $E[(X - \phi)^+] = \frac{1}{2}\hat{X}(\frac{\alpha^u - 1}{\alpha^u - \alpha^d})^2$, we then find that the expected surplus may be expressed as

$$E(\Pi^{SS}) = \frac{(a - c_t)^2}{2b} + \frac{1}{2} c_t \hat{X} \frac{1 + \alpha^d(\alpha^u - 2)}{\alpha^u - \alpha^d} \tag{13}$$

The first part $\frac{(a-c_t)^2}{2b}$ is the social surplus in the case of certainty without the wind power producer. The second part is the expected reduced production costs due to production by the wind power producer with zero marginal costs. This part is a product of two factors, $\frac{1}{2}c_t\hat{X}$ and $\frac{1+\alpha^d(\alpha^u-2)}{\alpha^u-\alpha^d}$. The first factor $\frac{1}{2}c_t\hat{X}$ represents the reduction in planned production costs if the expected production $\frac{1}{2}\hat{X}$ of wind power were to be produced with certainty. However, this is not the case. The contribution to the social surplus of the wind power producer is less than this, hence multiplication by the second term $\frac{1+\alpha^d(\alpha^u-2)}{\alpha^u-\alpha^d}$, where $\frac{1+\alpha^d(\alpha^u-2)}{\alpha^u-\alpha^d} < 1$ under our assumptions of α^u and α^d. This factor reflects the extra costs of unplanned production by the flexible producers. In the case that there are low extra costs of unplanned production, i.e. for α^d and α^u drawing close to 1, the factor also draws closer to 1, indicating a total cost saving due to wind production of nearly $\frac{1}{2}c_t\hat{X}$. However, in the scenario of higher unplanned costs, i.e. for higher levels of α^u and lower levels of α^d, this factor rises, and consequently, the expected benefit of wind power production is lowered.

The social surplus represented in (13) is thus the highest possible expected surplus to be achieved. Here it is derived by the benevolent optimizer. The next section comments upon the optimal solution derived by market mechanisms in a setting of a day-ahead and a real-time market.

[153] Noting that the density function is $f(x) = \frac{1}{\hat{X}}$ for the uniform distribution of X, we have $E[X] = \int_0^{\hat{X}} x f(x)dx$, $E[(\phi - X)^+] = \int_0^{\phi} (\phi - x) f(x)dx$, and $E[(X - \phi)^+] = \int_{\phi}^{\hat{X}} (x - \phi) f(x)dx$.

4 Market Solution

In this section we now turn to the scenario of a market setting where production and consumption decisions are made by the individual market participants. In principle the market solves the above optimization problem by making all participants adjust to common equilibrium market prices that balance the market. Let us now assume that the market is organized with a day-ahead market that is cleared at a price p_t, and a real-time market which is cleared at a price p_T. We will assume that there is no market power. Further assumptions of each representative participant and their derived bids are as follows:

The flexible producer operates actively in two markets, and thus has to submit several supply schedules: In addition to the supply bid to the day-ahead market, he operates actively in the real-time market, meaning than he has to submit a supply bid for both up-regulation and down-regulation in this market. The total profit of the flexible producer is given by $\Pi^f = (p_t Q^f - c_t Q^f) + (p_T^u q^u - c_T^u q^u) + (c_T^d q^d - p_T^d q^d)$ where q^u and q^d are the resulting quantities to be up- or down-regulated by the flexible producer. We have in the model assumed that the flexible producer has no capacity limits, and that he faces certain and constant marginal costs. This means that he can operate unlimited in the real-time market and regardless of the level of planned production. Furthermore, we assume that the flexible producer is able to deliver his planned production Q^f without any deviation[154]. The flexible producer may also earn a profit by actively contributing to the real-time market. To find his optimal bid, note that the real-time market is settled after the day-ahead market. We also assume that bidding in this market is after the settlement of the day-ahead market. These issues imply that the bidding decision for the day-ahead market, for up-regulation, and for down-regulation, can be separated. First order conditions for these three bidding decisions, where we respectively have maximized with respect to Q^f, q^u and q^d, thus imply the following three supply bids of (14), implying that the flexible producer offers an unlimited amount at these three prices.

[154] Note that if he were to passively contribute to an imbalance in the real-time market by deviating from his planned production Q^f, this deviation would be settled by deviation prices, with lower sales prices, and higher buying prices. This is, however, not a profitable strategy for the producer.

$$p_t = c_t$$
$$p_T^u = c_T^u = \alpha^u c_t \qquad (14)$$
$$p_T^d = c_T^d = \alpha^d c_t$$

Demand is as above represented without any uncertainty. As argued above, it is then not profitable for consumers to knowingly use the real-time market to cover demand, as any imbalance is settled to the disfavor of the passive participant in the real-time market. Further, demand is characterized with a falling marginal utility of consumption, indicating a price-sensitive demand function. The profit of our representative demand is given by $\Pi^D = aD - \frac{1}{2}bD^2 - p_t D$, where the first order condition thus gives the demand function of

$$p_t = a - bD \qquad (15)$$

The case of *the wind power producer* was the topic of section 1, where the optimal bid was implied by equation (4) where $P(X > Q^w) = \frac{p_T^u - p_t}{p_T^u - p_T^d}$, which in our case with a uniform distribution of X, thus implies $\frac{\hat{X} - Q^w}{\hat{X}} = \frac{p_T^u - p_t}{p_T^u - p_T^d}$, i.e. a supply bid of

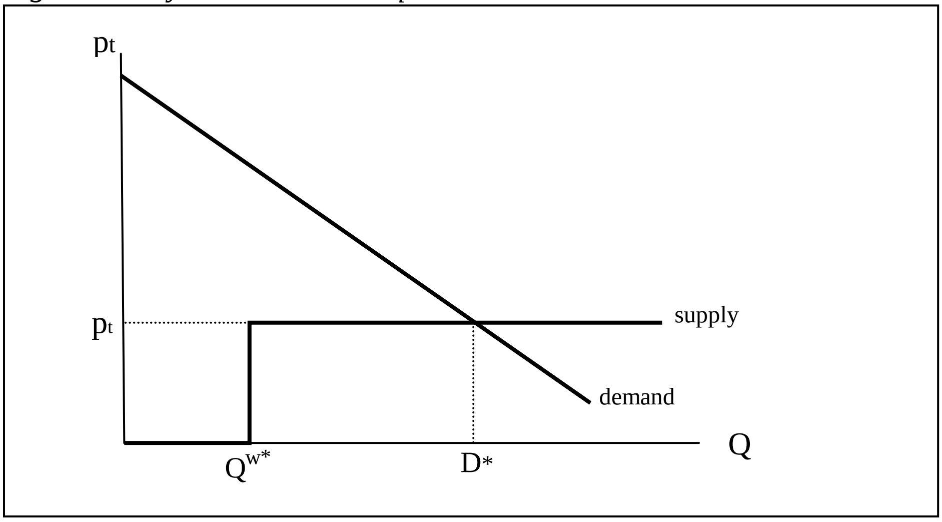
$$p_t = p_T^d + \frac{p_T^u - p_T^d}{\hat{X}} Q^w \qquad (16)$$

The bid of the wind power producer is thus dependent upon his assumptions and expectations of the up-regulation price and the down-regulation price. In our model there is no uncertainty as to the production costs of the flexible producer. In the scenario of the optimizer in section 2, we assumed the optimizer held this information. To parallel this scenario, we will here assume that this information is readily available for the market participants, i.e. we assume that all participants know that $p_T^u > p_t$ and $p_T^d < p_t$. In our simple model without any uncertainty other than for X, we will also assume that the structure of prices at time T is known, i.e. that $p_T^u = \alpha^u p_t$, and $p_T^d = \alpha^d p_t$ where we will assume that α^u and α^d are known constants. This knowledge might e.g. have been gained from analyzing former real-time prices, which thus reflect the underlying cost structure of the flexible producers. Then $\frac{p_T^u - p_t}{p_T^u - p_T^d} = \frac{\alpha^u p_t - p_t}{\alpha^u p_t - \alpha^d p_t} = \frac{\alpha^u - 1}{\alpha^u - \alpha^d}$. Inserting this information, equation (16) translates into a bid by the inflexible producer, where the wind power producer, regardless of the day-ahead price p_t, bids

$$Q^w = \hat{X}\,\frac{1-\alpha^d}{\alpha^u - \alpha^d} \tag{17}$$

With the supply and demand schedules of these three groups of participants, we can now illustrate the market equilibrium of the day-ahead market. The supply curve of the market is given as a kinked supply curve, where a quantity Q^w is offered at a zero price, while quantities above Q^w are offered at a price that equals the marginal cost c_t of planned production by the flexible producers, i.e. at $p_t = c_t$. Demand is a linear downward sloping demand-curve. The market is then cleared at the intersection of the demand curve and the supply curve. The market solution for the day-ahead market is illustrated in figure 1.

Figure 1. Day-ahead market equilibrium



We see that total demand is found for $p_t = a - bD = c_t$, i.e. a total demand of $D = \frac{a-c_t}{b}$. Of this quantity, the commitment of the inflexible producer is $Q^w = \hat{X}\,\frac{1-\alpha^d}{\alpha^u - \alpha^d}$, while the remaining part of D is to be covered by the planned production of the flexible producer, i.e. $Q^f = \frac{a-c_t}{b} - \hat{X}\,\frac{1-\alpha^d}{\alpha^u - \alpha^d}$. Under these assumptions of full information, we see that the market solution, not surprisingly, equals that of the optimizer. This is in accordance with general economic theory, where in the absence of market imperfections and under similar assumptions, the equilibrium of the market maximizes the expected social surplus, i.e. is equal to the maximum attainable surplus as in the benchmark model of the benevolent optimizer.

Finally, let us shortly comment on practices and rules as to bidding in the day-ahead market relative to the real-time market. An employed rule in some markets is that the participants are required to place bids in the day-ahead market that balance their market transactions and underlying commitments. The rule might be motivated as a measure for minimizing real-time transactions, and thus also minimizing the need for reserve capacity. In our model, and in the case of the inflexible producer, this may be represented by the requirement that the sales bid in the day-ahead market has to reflect the expected production, i.e. $Q^w = E(X) = \frac{1}{2}\hat{X}$. The optimal solution in our model was $Q^w = \hat{X}\frac{1-\alpha^d}{\alpha^u-\alpha^d}$. By bidding this optimal bid, the wind power producer will, however, not obey the rule, and thus faces any incorporated penalty in the system. By bidding in accordance with the rule, i.e. bidding the expected production $\frac{1}{2}\hat{X}$, the wind power producer will not obtain his maximum profit. From the above model we also saw that this restriction thus results in a non optimal solution for the market as a whole[155].

5 Concluding Remarks

Our focus in this paper has been on a wind power producer, where real-time production is stochastic and non-controllable. We have studied the optimal bidding of such a participant in the setting of a day-ahead spot market, and a real-time deviation market, as well as within the benchmark case of directly optimizing social surplus. Our model has been simplified to focus on the aspect that real-time adjustments are likely to be more costly than planned adjustments in production and consumption. By isolating this aspect, and focusing on the attribute that wind power production is largely non-controllable, our main conclusions show that the optimal bid of the wind power producer is not equal to the expected wind power generation. This was mirrored in the benchmark case of the optimizer, where we found that neither did the resulting optimal planned production by other producers mirror the actual expected wind power production.

Having focused on these main properties, a more detailed representation of the real-time market and of participants, may give further insight as to the optimal bidding strategies, as well as the effect of wind power plants and other stochastic and inflexible generation or consumption on market equilibrium. We will briefly indicate some areas of extension:

[155] The solution is not optimal, except for the case where the underlying cost structure is given by $\alpha^d = 2-\alpha^u$, since, for $\alpha^d = 2-\alpha^u$, we have $\frac{1-\alpha^d}{\alpha^u-\alpha^d} = \frac{1-(2-\alpha^u)}{\alpha^u-(2-\alpha^u)} = \frac{1}{2}$, in which the two bids will be equal.

- *Representation of real-time price formation:* While it is reasonable to expect that $p_T^u > p_t$ and $p_T^d < p_t$, it is, however, not reasonable to believe that the functional relationships between p_T^u and p_t, and between p_T^d and p_t are certain or fixed. Using our terminology, it is likely that the factors of α^u and α^d are stochastic, and that the relationship might also not be multiplicative as in our model. Up- and down-regulation prices will in principle depend on the aggregate imbalance in the market, and the cost functions of the specific active participants in the real-time market.

- *Price settlement of passive imbalance:* In our model, the producer imbalance is settled by the up-regulation price if he causes an up-regulation, and the down-regulation price if he causes a down-regulation. In practice, this is not necessarily so. For example, until recently, the settlement price of the Norwegian real-time market has been the up- (down-) regulation price if the market as a whole needs to be up- (down-) regulated. In this pricing regime, the participant has profited if his imbalance is in contrast to the total market imbalance, e.g. if the market imbalance calls for up-regulation, while the inflexible producer has produced too large a quantity. Another pricing regime example is the proposed two-price rule: If imbalances by passive participants enhance the total market imbalance, they are settled by the corresponding dis-favorable up- or down regulation price. If passive participant imbalances have relieved the total market imbalance, they are settled by the day-ahead spot price for the period in question.

- *Representation of market participants:* The participants of the market may be more realistically modeled, e.g. by modeling the uncertainty of demand or uncertainty of the flexible producers. Furthermore, we have in our model assumed that there are no capacity limits for the inflexible producers. By defining a restricted capacity, the bidding decision of the flexible producer becomes much more complicated. The producer has to decide upon which market to allocate his production capacity. As such this represents a two-stage stochastic problem for the flexible producer.

- *Security of supply considerations:* With restricted capacity, a major issue with respect to the immediate security of supply is to allocate sufficient real-time adjustment production/consumption capacity to system operation to ensure the security of supply. This is also a question of the relative incentives provided by the day-ahead market and the real-time market, and with the addition of other markets, as for example different types of reserve capacity options.

Several of these issues represent large challenges in modeling the market interaction. They also have important implications for which effect the inclusion of a considerably larger degree of inflexible participants in the market may have, as most of the issues mentioned will elaborate on the complexity in optimally handling the increased uncertainty of these participants for the real-time balance of the market.

References

Khouja, M. (1999): 'The single-period (news-vendor) problem: literature review and suggestions for future research', *Omega, The International Journal of Management Science*, 27, 537-553.