KTH Electrical Engineering

On Monte Carlo Simulation and Analysis of Electricity Markets

Mikael Amelin

Doctoral Thesis

Royal Institute of Technology
Department of Electrical Engineering
Stockholm 2004

Cover illustration: Jan Manker
Printed in Sweden
Universitetsservice US-AB
Stockholm 2004
(www.us-ab.com)

TRITA-ETS-2004-05a
ISSN-1650-674x
ISRN KTH/EES/R-0405a-SE
ISBN 91-7283-850-7

ABSTRACT

This dissertation is about how Monte Carlo simulation can be used to analyse electricity markets. There are a wide range of applications for simulation; for example, players in the electricity market can use simulation to decide whether or not an investment can be expected to be profitable, and authorities can by means of simulation find out which consequences a certain market design can be expected to have on electricity prices, environmental impact, etc.

In the first part of the dissertation, the focus is which electricity market models are suitable for Monte Carlo simulation. The starting point is a definition of an ideal electricity market. Such an electricity market is partly practical from a mathematical point of view (it is simple to formulate and does not require too complex calculations) and partly it is a representation of the best possible resource utilisation. The definition of the ideal electricity market is followed by analysis how the reality differs from the ideal model, what consequences the differences have on the rules of the electricity market and the strategies of the players, as well as how non-ideal properties can be included in a mathematical model. Particularly, questions about environmental impact, forecast uncertainty and grid costs are studied.

The second part of the dissertation treats the Monte Carlo technique itself. To reduce the number of samples necessary to obtain accurate results, variance reduction techniques can be used. Here, six different variance reduction techniques are studied and possible applications are pointed out. The conclusions of these studies are turned into a method for efficient simulation of basic electricity markets. The method is applied to some test systems and the results show that the chosen variance reduction techniques can produce equal or better results using 99% fewer samples compared to when the same system is simulated without any variance reduction technique. More complex electricity market models cannot directly be simulated using the same method. However, in the dissertation it is shown that there are parallels and that the results from simulation of basic electricity markets can form a foundation for future simulation methods.

PREFACE

Writing a dissertation is periodically a inconsolable and onerous task. To travesty a former Swedish Minister of Finance:

> *Writing dissertation is damn severe,*
> *but now I prized it as far as here.*

I want to thank all friends and relatives who have had patience with me during the course of the work, and who have given me the necessary support to get finished. I also owe many thanks to my colleagues for good advice, discussions about electricity markets and other issues (as for example toothbrushes), assistance during computer failures and many other things.

This dissertation was originally written in Swedish, and what you are holding now is just a feeble translation made in a hurry by an unprofessional translator (i.e., yours truly). Those who are really interested in my work are recommended to learn Swedish first, and then read the *real* dissertation.

Stockholm
July 2004 *Mikael Amelin*

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1
INTRODUCTION

Starting in the 1980s, but primarily since the end of the 1990s, there has been a global trend to restructure electricity markets in order to promote competition;[1] the intent is of course to utilise the resources more efficiently. Due to the restructuring, the conditions for electricity trading have of course changed radically, which naturally has impacts on the usage of the power system. Besides, a number of new players have appeared in the restructured electricity markets. All in all, this means that the need for analysis methods and planning tools has increased, at the same time as the old models no longer really reflect all aspects of the reality. It can therefore be said without exaggeration that research within the field analysis and simulation of electricity markets is more necessary now than ever before.

Simulation of electricity market is also the theme of this dissertation. I have studied rules which are used or could be used in different electricity market designs and tried to analyse which consequences the rules will have for the electricity trading. Moreover, I have formulated mathematical models which can be used to simulate electricity markets using so-called Monte Carlo methods. A Monte Carlo simulation is simply a sample survey investigating how the electricity market will work in a number of more or less randomly chosen scenarios. From the results of these samples, conclusions may be drawn about the expected behaviour of the electricity market. The advantage of a Monte Carlo simulation is that it is straightforward to include models of different market designs and the strategies used by the players of the electricity market. A disadvantage is that it may require a lot of computations to gain a reliable answer. Therefore, I have also studied and further developed various mathematical tricks, which can increase the efficiency of a Monte Carlo simulation.

1. In everyday speech it is common to use the term “deregulation” when opening a market to competition. However, this word is somewhat misleading, since it gives the impression that the electricity trading should be controlled by less rules than before—actually, it is rather the opposite.

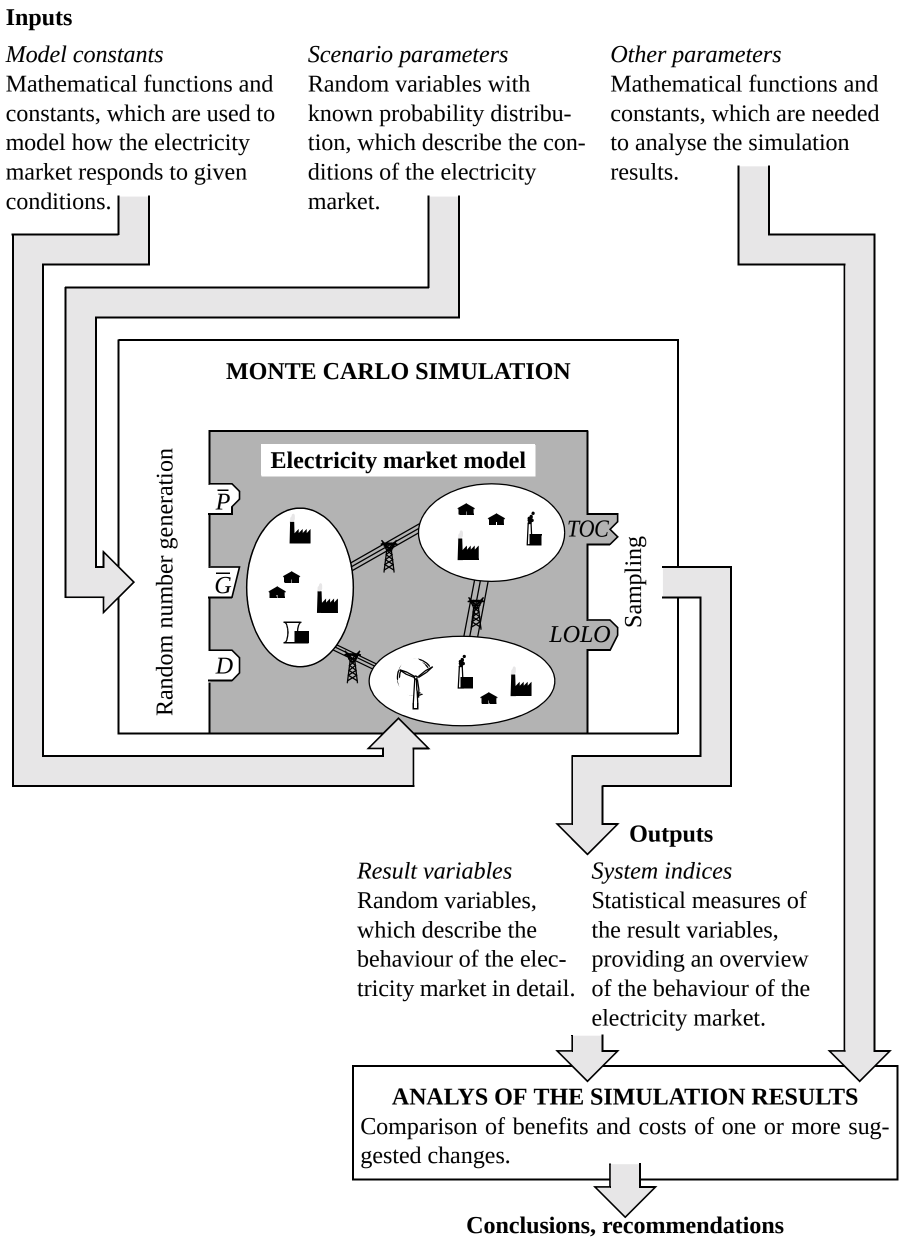


Figure 1.1 *Monte Carlo simulation and analysis of an electricity market. The figure shows the inputs and outputs of a Monte Carlo simulation of an electricity market. The outputs of the simulation describe the consequences of choosing a particular design of the electricity market. These consequences can be studied in a separate analysis (which is not part of the actual simulation) to determine how good or bad certain measures can be expected to be.*

1.1 PROBLEM DESCRIPTION

In order to simulate an electricity market, it is necessary to have a mathematical model, which includes the properties of the power system as well as the behaviour of the players who are involved in the electricity market. (I therefore find the term “power system simulation”, which sometimes is used, to be a little bit to narrow too describe this kind of simulation; hence, I prefer the phrase “electricity market simulation”.) The objective of simulating an electricity market is to study how a change will affect the system, for example concerning electricity prices, reliability or environment. In this section I intend to describe what constitutes an electricity market simulation, and explain some of the terminology, which I will use when I later address the details of modelling and simulation. Finally, I will also comment upon the efficiency requirements the simulation method has to fulfil in order to have any practical value.

Static Electricity Market Simulation

Figure 1.1 provides an overview of how to perform a basic study of an electricity market using Monte Carlo techniques. The two main elements are the simulation itself and the analysis of the simulation results. The simulation shows how the electricity market can be expected to operate under given conditions. This simulation may be performed several times, using different designs of the power system and the rules governing the electricity trading. The results of different simulation runs can then be compared in the following analysis. In this analysis, we must also consider some other parameters which are not directly associated to the simulation itself.

To simulate the expected behaviour of an electricity market it is necessary to determine how it will behave in every possible situation. I have chosen to introduce the designation *scenario* for a particular situation, where both the available resources, the demand and other factors controlling the electricity trading are known. In any electricity market these conditions are varying more or less randomly, which means that there is an infinite number of possible scenarios.

I will here take the opportunity to emphasise that all random events in the electricity market are represented by different scenarios. A scenario cannot in itself include any random factors, but all scenarios are deterministic—if the same scenario should occur twice then the outcome will be exactly the same in both cases.[2]

The conditions of a particular scenario are described by a number of parameters, which I—not very imaginative—designate *scenario parameters*. Each scenario parameter is a random variable, the probability distribution of which is known and part of the input to the simulation. Exactly which scenario parameters there are depends on the chosen model, and I will provide more detailed examples later in this dissertation; in this chapter all the scenario parameters are collected into the random vector Y.

The simulation requires that there is a mathematical model of the electricity market. In this model there are several constants having the same value for all scenarios. These constants I refer to—once again quite unimaginative—as *model constants* and in similarity to the scenario parameters, the model constants constitute inputs to the simulation. We may see the model constants and the structure of the model itself as a mathematical function, g, which shows how the electricity market responds to a certain scenario.[3]

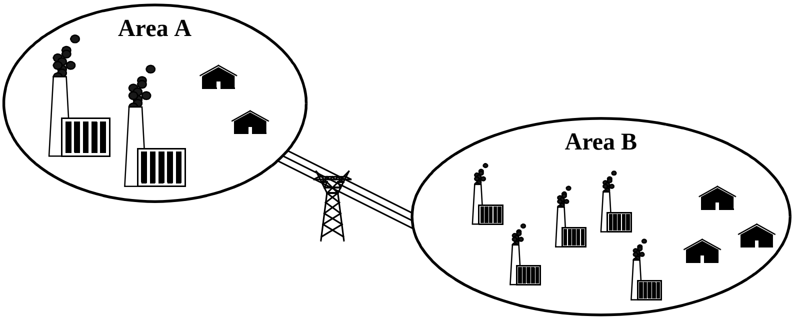
The output of the simulation shows how the electricity market will behave in each possible scenario. Several conceivable characteristics can be studied, as for example resource usage and prices. Depending on which properties we want to study, we may define a number of result variables. For now, I assume that all *result variables* are collected into one random vector, X. As for the scenario parameters, the result variables are random variables, but the major difference is that the probability distribution of the result variables is unknown, before the system has been simulated, as the result variables are a complex function of the scenario parameters. The relation between them is described by the electricity market model:

$$X = g(Y). \tag{1.1}$$

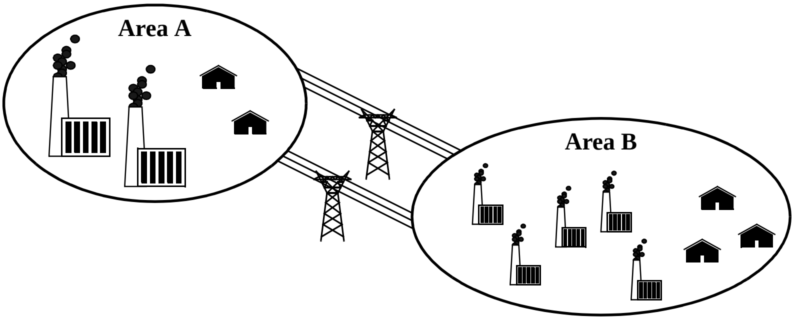
When analysing the results of the electricity market simulation it is generally not practical to consider every possible chain of events (i.e., all possible outcomes of the result variables) in detail. It is preferably to have simple, comprehensible measures, which in a more easy-to-grasp way describe the most important consequences of a particular electricity market model. Most

2. This definition is not self-evident, so let me give a small example to avoid misunderstandings. Assume that there is a power system with a certain load and a certain generation, which covers this load. In other words, everything in the garden is lovely, but then a serious overloading occurs in an important transmission line. If there is at this moment an alert operator in the control room and this operator immediately comes to decision about the appropriate countermeasures, the power system can cope with the disturbance. Conversely, if the operator is on a coffee break or for some reason does not manage to take the right actions in time, then extensive disturbances come up.

   The above situation may seem like an example that the same scenario can result in two different outcomes, but actually my definition means that the two possible events *constitute two different scenarios*! The two scenarios are differentiated by a scenario parameter which either takes the value “alert operator” or “sleeping operator (who probably should have a darn good explanation at hand when the boss comes around and wants to know why half the nation just has gone black)”.
3. The electricity market model g is generally so complex that it can only be defined indirectly from the solution to one or more optimisation problems; cf. section 3.2.2.



a) *A single transmission line with a transmission capability of* 25 MW. *This alternative yields an expected operation cost ETOC* = 2 889 ¤/h *and the risk of power deficit LOLP* = 0.181 %.



b) *Double transmission lines with a total transmission capability of* 50 MW. *This alternative yields an expected operation cost ETOC* = 2 629 ¤/h *and the risk of power deficit LOLP* = 0.166 %.

Figure 1.2 *Example of analysis of an electricity market. Assume that it should be investigated whether it is profitable to increase the interconnection between areas A and B in the system above. Panel a shows the result of simulating the present situation and panel b shows the result if the transmission capability has been increased between the two areas. (The details about the simulation of these two alternatives are further described in section 9.3.) Thanks to the expansion there will be better possibilities to dispatch the power plants of the two areas, which apparently causes the total operation cost (indicated by ETOC) of the system to decrease, at the same time as the reliability (indicated by LOLP) increases.*

The value of the new transmission line is thus determined by the differences between the two alternatives. The operation cost is expected to decrease by about 2.3 M¤ *per year and the time for power deficit is expected to decrease by about one hour and fifteen minutes per year. By comparing these benefits to the investment costs, it can be determined whether or not it is profitable to increase the transmission capability.*

straightforward is to define a number of key values, which I prefer to designate *system indices*, which can be used to compare different alternative electricity markets. The system indices are in practice various statistical measures (generally expectation values) of the result variables. Hence, it can be said that the actual objective of an electricity market simulation is to determine

$$E[X] = E[g(Y)]. \tag{1.2}$$

In figure 1.2 an example is given of how to use system indices—in this case expected operation cost (*ETOC*) and the risk of power deficit (*LOLP*)—to analyse the value of increasing the transmission capability of an interconnection. Further definitions of system indices are given in section 3.2.2.

Dynamic Electricity Market Simulation

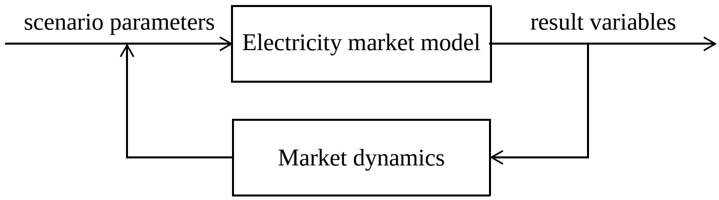
In the static electricity market simulation described above it was presupposed that the probability distribution of the scenario parameters is not just known, but also constant over time. In the short run, this is a fully reasonable assumption, but in the long run the conditions of the electricity trading will change, since the demand of the consumers may increase, new power plants and transmission lines can be built, the legislation may gradually change, etc.

A static electricity market simulation is quite a straight-forward task,[4] but if we want to study how the probability distributions of the scenario parameters vary in the long run, the task becomes a lot more complicated. In some cases the distribution of the scenario parameters change as a result of external conditions; for example, the demand of electric energy is highly depending on the general social development. In other cases—and now things get really tricky—the changes depend on the result variables. The producers in an electricity market can for example be assumed to hold back investments in new power plants until the prices of the electricity market have reached such levels that the investment is considered profitable, which means that the available generation capacity (which is a scenario parameter) will depend on the electricity prices (which are result variables). I refer to this feed-back between the probability distributions of the result variables and the scenario parameters as *market dynamics* (see figure 1.3).

When I theoretically analyse the function of an electricity market, I will sometimes refer to certain rules as having “market dynamic effects”—in those cases I have the above described problem in mind. However, in the models and methods I describe in this dissertation I only study static electric-

4. I dare not use the word “simple” in this context, because— as it hopefully will be clear from the remainder of this dissertation—neither this kind of simulation is a trivial task, but it takes good methods and quite a number of computations to perform a simulation.

a) Static electricity market simulation: the probability distribution of the scenario parameters are independent of the result variables.



b) Dynamic electricity market simulation: the result variables affect the probability distribution of the scenario parameters.

Figure 1.3 *The difference between static and dynamic electricity market simulation.*

ity market simulation. Eventually, the issue of how to appropriately manage market dynamic effects in an electricity market simulation has to be investigated, but that is something which I have had to leave to future research.

Efficiency Requirements

If simulation of electricity markets should be of any practical value, some requirements have to be made on the simulation method. Naturally, the most important requirement is that the simulation is based on mathematical models which include all important aspects of the electricity trading and the physical properties of the power system, but it is also important that a simulation can be performed within reasonable time. Even though the most simple case may require only two simulations to evaluate a certain measure (as the example in figure 1.2), we must in practice be prepared that there are considerably more possible options to be evaluated. Moreover, it is likely that there is some uncertainty in the input data of the simulation; therefore, we may want to perform sensitivity analysis by varying the assumption of the behaviour of the players or the probability distributions of the scenario parameters. All in all, we can assume that a complete survey of a particular electricity market requires dozens or even hundreds of separate simulations. If every simulation took a week, it would be likely that other, less sophisticated, methods of analysis would be chosen. The objective when developing simulation models of electricity markets should therefore be that a simulation preferably should not need more than an hour (and absolutely not more than 24 hours) to be run.

The second requirement we have on the simulation tools is that the precision of the system indices must be sufficient. The difference between two alternatives can be subtle, and it would of course not be acceptable if the uncertainty of the system index calculation is larger than the difference to be determined.

1.2 RELATED RESEARCH

Simulation of electricity market has been studied for some decades and a number of simulation methods have been developed. Below follows a short overview of methods related to my research.

Analytical Methods

The classical method to simulate electricity markets is probabilistic production cost simulation (PPC). This method was first presented in [45] and [48] respectively and have later been further developed by several authors; nowadays, PPC is included in most text books on power system planning (e.g. [11, 31, 32, 34]). In short, the method is based on the usage of an electricity market model having only two scenario parameters: available total generation capacity, $\overline{G}$, and total load, D. To simplify the calculations, these two scenario parameters are combined into one single, by introducing the notion equivalent load, which is defined as the real load plus outages in generation capacity, i.e., $ED = D + \hat{G} - \overline{G}$ (where $\hat{G}$ is the total installed capacity of the system). The duration curve of the equivalent load can be determined using a so-called convolution formula. As there is only one probability distribution now, the system indices can be calculated directly using the definition of expectation value:

$$E[g(Y)] \approx E[g^*(ED)] = \int_0^{\infty} \tilde{F}_{ED}(x)dx, \tag{1.3}$$

where g^* is a simplified model of the system g. The simplified model only takes the equivalent load as input.

To perform a probabilistic production cost simulation it takes a large number of quite simple calculations—in other words, a perfect task for computers. The advantage of the method include that there is no uncertainty about the accuracy of the performed calculations.[5] The disadvantage is that the method presupposes a very simplified model of the electricity market, where only random variations of load and available generation capacity are considered. Attempts have been made to expand the PPC model to include the transmission system as well (see for example [51]), but the calculations become significantly more complicated than in the original method—actually so complicated that as far as I know, this kind of methods has never been used in practice. That somebody should be able to include additional details, as for example the rules of environmental protection, market power and other features of modern electricity markets, seems extremely unlikely. The more complicated electricity market model we wish to use, the more scenario parameters it takes and the harder it becomes to formulate and solve analytical expressions of $E[g(Y)]$.

Another interesting idea to analytically simulate electricity markets has been proposed by Hobbs and his co-workers [53, 54]. The method is based on determining overestimates and underestimates of $E[X]$:

$$\underline{m}_X \le E[X] \le \overline{m}_X. \tag{1.4}$$

Using an iterative process, $\underline{m}_X$ and $\overline{m}_X$ are successively refined until the gap between them is less than the desired accuracy. This is a brilliant idea, but according to what I have heard from Hobbs himself, the method turned out to be difficult to apply to larger systems.

The challenge of an electricity market simulation is that we simultaneously determine several system indices, for example operation costs and reliability. If only reliability is of concern then there are several analytical methods available (see for example [16, 19]). This kind of calculations is however not in focus of this dissertation.

Monte Carlo Simulation

The expectation value of X is according to the definition equal to the mean of a number of samples, when X is distributed *exactly* according to the density function, f_X. The Monte Carlo method is based on the idea that even though a series of samples is not distributed exactly according to the density function, it is likely that the deviation is rather small—at least if the number of samples is large. The mean of an arbitrary number of samples should therefore be approximately equal to the expectation value:

$$E[X] \approx m_X = \frac{1}{n} \sum_{i=1}^{n} x_i, \tag{1.5}$$

5. A reservation is though that generally (1.3) is solved numerically, which of course causes minor errors. If we just choose sufficiently short step size, the numerical errors will be negligible, but it has of course a cost in the form of increased computation time.

where $x_1, \ldots, x_n$ is a number of samples of X (see chapter 8). In an electricity market simulation a sample is obtained by $x_i = g(y_i)$, where y_i is the value of the scenario parameters in scenario i.

The advantages and disadvantages of the Monte Carlo method are more or less the opposite compared to probabilistic production cost simulation: there will be some inevitable uncertainty in the result, because it is based on a random selection, but on the other hand, we can use arbitrarily complicated electricity market models.

A special problem of Monte Carlo methods is that performing a simulation might become very time-consuming. The time to compute the outcome of a scenario, i.e., $x_i = g(y_i)$, will increase as the complexity of the used model increases. Besides, it may take a large number of scenarios to keep the uncertainty of the final results within acceptable levels. To reduce the number of scenarios per simulation it is possible to use so-called variance reduction techniques. An overview of earlier works in this field is given in section 9.2, when I present my own view on which the good sides and the drawbacks of different variance reduction techniques are when simulating electricity markets.

An example of an application of Monte Carlo simulation is the multi-area power scheduling model (EMPS-model), which has been developed by Norwegian SINTEF and is frequently used in the Nordic countries. The primary difference between the EMPS-model and the simulation method I describe in this dissertation is that I assume that the scenarios are randomised according to a given probability distribution, whereas in the EMPS-model scenarios are created using a data base of historical hydrological years [50, 52].[6]

1.3 MAIN CONTRIBUTIONS

A method to simulate an electricity market has two basic components: an electricity market model and an algorithm to perform the probability calculations necessary to describe the expected behaviour of the electricity market. The central theme of this dissertation is about both these problems.

Concerning electricity market models I have defined a basic electricity market model, an **ideal electricity market**, which can be used to simulate simple electricity markets. By studying the conditions which have to be fulfilled on an ideal electricity market, it is possible to identify a number of non-ideal properties of real electricity markets. I have chosen some important non-ideal properties and analysed what is the cause of **the difference between an ideal**

6. The scenarios are defined by hydrological years, because the EMPS-model is mainly focused on modelling the hydro power system in detail; the remainder of the electricity market is more or less represented by model constants.

and a non-ideal electricity market, I have provided an overview of the **market designs which could be introduced to manage non-ideal properties** in an electricity market and **how the strategies of the players are affected**, and I have suggested **mathematical models of the different market designs and strategies**.

Concerning the probability calculations, I have studied how Monte Carlo techniques can be applied to electricity market simulation. In that respect I have primarily studied so-called variance reduction techniques. Much of these studies have revolved about a method called “stratified sampling”. Some of my results on stratified sampling are rather general and could be of interest also to other problems than electricity market simulation; amongst these are **the cardinal error** (which occurs when the samples are not appropriately distributed between different strata) and **the strata tree** (a tool to efficiently and systematically define strata). Moreover, I have developed a **systematic procedure to simulate electricity markets** without energy storage or other important time constants—I refer to this as simulation of short scenarios—and shown that we obtain a **significant efficiency gain compared to not using variance reduction techniques**. Simulation of long scenarios (i.e., when there are energy storage facilities or other important time-depending factors) is more complicated, but I have shown how the theory of short scenarios can be useful also when simulating long scenarios.

1.4 OVERVIEW OF THE THESIS

The research I have performed as a Ph.D. student has more or less revolved around three problems. Two of these—choice of electricity market model and choice of simulation methods—have already been described earlier in this chapter. In addition to that, I have spend time on a solution algorithm for the optimisation problem which arises when simulating certain electricity markets. The result of this work was the NNP algorithm, which is described in detail in my licentiate thesis [7] and which from considerations to space is not further described in this work.[7]

In the first part of the dissertation, I try to provide a general picture of the operation of electricity markets, and how they can be modelled. In chapter

7. That I nevertheless mention the NNP algorithm is because I would like to give a piece of advice to any reader who perchance currently is planning a doctoral project: focus on the central problems and watch out for sidetracks! Although the NNP algorithm was successful in the meaning that it is efficient compared to more general optimisation algorithms [135], I must now afterwards conclude that it would have been more appropriate to use commercial software instead; thus, I would have had more time to study electricity market models and simulation methods.

two I give a general description of the principles for electricity trading, while I at the same time define important terms which will be used in later chapters. The first step of the actual modelling is to consider the simplest possible electricity market (and by that I mean the simplest possible from a simulation point of view), which I have chosen to call an *ideal electricity market*. My definition of the notion ideal electricity market is given in chapter three, where I also present a mathematical model of ideal electricity markets. Using the ideal electricity market as a start point, it is then possible to identify a number of non-ideal properties which may appear in real electricity markets. I have chosen a number of problems which I find important and analysed why the reality is non-ideal, which rules there are (or could be) to manage these problems, and I have suggested how the basic ideal electricity market model can be updated to model different rules and how the strategies of the players are affected. The non-ideal properties that I have chosen to study closer concern environment issues (chapter four), forecast uncertainty (chapter five) and grid costs (chapter six). Besides, there are some more brief analyses of additional market imperfections in chapter seven.

The second part of this dissertation treats the Monte Carlo technique itself, which basically is about how to choose which scenarios should be analysed in the electricity market model. In chapter eight a general description of sampling and variance reduction techniques can be found. In the following two chapters I describe how to practically perform a Monte Carlo simulation of an electricity market in an efficient manner. In chapter nine I consider electricity markets where the time perspective is short. This kind of electricity markets are quite straightforward to simulate, and it is here that I have been most successful to demonstrate considerable efficiency gains by using variance reduction techniques. Simulating electricity markets with a long time perspective is more complicated and I have not had time to investigate this field as thoroughly. In chapter ten I provide an overview of which problems have to be solved and my ideas about how to solve them.

In the last part of the dissertation there is a concluding chapter, where I partly try to summarise my results and partly point out areas where further research would be useful. After this chapter some appendices follow, with short summaries of important mathematical definitions and theorems, and an analytical method to calculate the risk of power deficit in a two-area system. A list of abbreviations and an overview of the notation I use follow after the appendices—those readers who wish to study my electricity market models (i.e., chapters three to seven) in more detail are recommended to first have a look at the general notation. Finally, there are literature references and an index.

Chapter 2

BACKGROUND

Electric energy is quite a difficult good. Firstly, electric energy cannot be stored, but must be generated in the same moment it is consumed.[1] As it is impossible for the producers to know exactly when the consumers are going to use the energy they bought, it is necessary that all electric power systems are equipped with technical systems which automatically balance production and consumption. Secondly, electric energy cannot be distributed in just any way; it takes a grid system connecting producers and consumers. In theory each producer could build an own power distribution grid, but in practice such parallel grids would be unprofitable. All producers and consumers are therefore forced to share a common grid.

An electricity market cannot work unless these problems are managed, which requires that rules are defined to control the responsibilities of each player and to determine the procedures of the electricity trading. However, it is in no way so that there is only one possible design for the rules of the electricity market. In this background chapter I intend to provide a general description of the fundamental features of different trading arrangements, and to introduce important notions which will be used in the following chapters, when different challenges to the electricity market are studied more in detail. However, I must emphasise that the objective of this chapter is not to describe all possible ways of organizing an electricity market. What it is all about is to define a general terminology, which is needed when discussing and comparing different arrangements.

I will also point out that the notions introduced in this chapter are my own. By and large, I try to use generally accepted notions, but in some cases so many designations flourish for similar features, that I have found it easiest to invent my own vocabulary.

1. What we in everyday life consider as storage of electric energy actually stores some other form of energy, which however readily can be transformed into electric energy. For example, a battery stores chemical energy and the reservoir of hydro power plant stores mechanical energy (potential energy).

Goods

The primary good to be traded in an electricity market is naturally electric energy. As technical systems manage the balancing between production and consumption, the trading cannot be performed in real time. It is necessary to introduce *trading periods* instead, the duration of which can be chosen arbitrarily.[2] Selling electricity is equivalent to supplying a certain amount of energy to the grid during a certain trading period and buying electricity is correspondingly the same as extracting a certain amount of energy from the grid during a certain trading period.

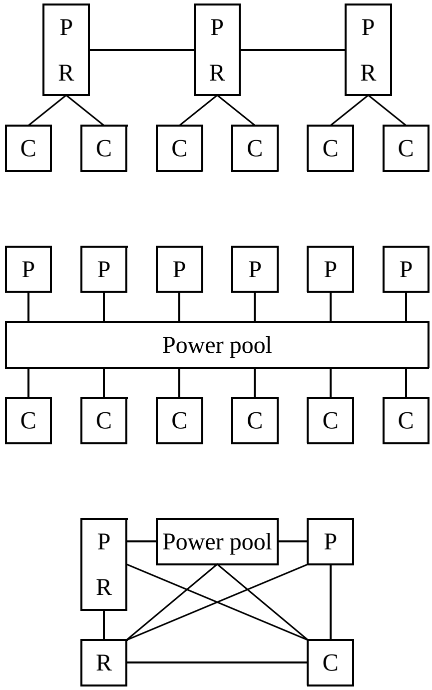
When the players in the electricity market trade in this manner, the average production and consumption during a trading period will be in balance, but they have no responsibility for the instantaneous balance. This responsibility is given to a *system operator* instead. The task of the system operator can vary depending on the rules of the electricity market, but a minimum requirement is that the system operator makes available the technical systems necessary for the function of the power system, and that the system operator assures that the system is operated safely. The system operator can in some cases also be a grid owner or act as *market operator*, i.e., to supply trading places where other players can do business.

Besides trading electric energy, there might be a number of side markets. It may for example be about markets for ancillary services (i.e. when the system operator rather than building necessary technical systems under their own auspices, they buy the services from other players in the electricity market) or markets which are used to lead the players of the electricity market in a desirable direction, for example towards more environmentally benign electricity generation. The side markets which are of interest to electricity market simulation will be more closely described in the following chapters. Below follow further details about the trading of electrical energy.

The Ahead Market

By the ahead market I refer to the trading which occurs before the actual trading period. The ahead market includes among other things trade with price insurances and other financial derivatives, which I however will not further describe in this dissertation.[3] The ahead trading which is of interest is rather something which I—lacking a better designation—call the “physical trad-

2. Most common are half-hour periods (used for example in England-Wales, Australia and New Zealand [14, 27, 29, 30]) and one-hour periods (used for example in the Nordic countries, Spain and the U.S. [14, 20, 26, 43]).
3. Examples of financial derivatives are given in among others [15, 23, 37].



a) Vertically integrated electricity market. The consumers are forced to buy from the local power company. The power companies may trade freely.

b) Centralised electricity market. All producers must sell to the centralised power pool and all consumers must buy from the pool.

c) Bilateral electricity market. All players may trade freely.

P - Producer C - Consumer R - Retailer

Figure 2.1 *Three ways of structuring an electricity market.*

ing”. Actually, this trading is also a purely financial agreement; if, for example, a producer cannot produce in accordance with the sales of the ahead market the physical delivery will be taken care of by the system operator (using the real-time market), while the producer is forced to trade in the post market to fulfil the undertaking from the ahead market. I will return to these procedures later in this chapter.

There are, according to my opinion, three basic ways to structure the ahead market.[4] Figure 2.1 provides the general lay-out of the three main types. The oldest form is the vertically integrated electricity market, where each power company combines the roles of producer, retailer, grid owner and system operator. (The designation vertically integrated actually refers to that the companies manage all steps of the power delivery.) Each company has a franchise for one or more geographical areas, where they have a monopoly of

4. Unfortunately, there is a countless number of different definitions and designations of electricity markets and it is not uncommon to see different authors introducing different designations for what is essentially the same thing. For example, compare how electricity markets are classified in [21] and [24].

retailing; the consumers of a vertically integrated electricity market has no possibility to choose supplier, but must buy from the local power company. To prevent the power companies from taking advantage of their monopoly, the activities are regulated; hence, it is stated which electricity prices they may charge and which other responsibilities they have.

As the power companies of the vertically integrated electricity market do not compete, they can decide themselves how to solve the technical issues of safe system operation in the best manner. To reduce the operation costs they can also get involved in trading with each other in those cases when a power company has unused production capacity which is cheaper than the power plants of the other companies. It is even possible for the power companies to jointly plan the operation of all power plants in the system.[5] However, sometimes there are difficulties also in the trading between vertically integrated companies. An example is when two companies trade, but have no direct electric connection to each other's grids; the trading will take place via the grid of a third company instead, which causes losses (and possibly other inconveniences) for the third company, which therefore wants to be compensated.

The advantage of a vertically integrated electricity market is that simpler technical solutions can be used when one company manages all parts of the power system, and the investment in generation, transmission and distribution may be coordinated. The obvious disadvantage is that there will not be the same pressure to improve the performance as when several companies compete for the favours of the consumers. Hoping to increase the efficiency of electricity markets, a global restructuring process was initiated during the late 20th century. In the restructured electricity markets the vertically integrated companies are divided, so that competitive activities (production and possibly retailing) are separated from monopolies (system operator and grid owners).

Among the restructured electricity markets we can distinguish centralised and bilateral electricity markets. Characteristic of a centralised electricity market is that producers and consumers may not trade directly. The producers have to submit their sales bids to a central power pool, which is managed by the system operator. In some cases the consumers (maybe represented by a retailer) also submit purchase bids to the power pool,[6] whereas in other cases the system operator forecasts the load during the trading period and buys the same amount from the power pool; thus, the system operator serves as retailer for all consumers.[7] For each trading period the pool either determines an elec-

5. Cf. for example the Swedish production optimisation (which existed before the restructuring in 1996) where the major power companies reported their variable production costs and then the total operation cost of the involved companies was minimised [15].
6. An example of this arrangement is the Spanish electricity market [20].

tricity price for the whole market or a number of different electricity prices, which each apply to a part of the market.

In a bilateral electricity market the system operator has a more supervising role. The players do not have to trade through a power pool, but may sell and purchase freely (all transactions must however be reported to the system operator, so that after the trading period it is possible to control that the players have fulfilled their undertakings). There is no official electricity price, but generally there is a power pool also in bilateral electricity markets and the price of the pool serves as a guideline to those players trading bilaterally.

As the players may trade freely in a bilateral electricity market, a business opportunity opens up for pure retailers (which are often referred to as independent traders). Their idea is to buy power directly from the producers or the power pool and sell it on to the consumers. It might appear as if these retailers were just unnecessary, price increasing middlemen—and in the worst case this might actually be true—but they can also supply important functions to the electricity market. Primarily the existence of retailers means larger freedom of choice for the consumers, resulting in increased competition compared to if retailing was run by producers only. The increased competition may not just apply to the electricity price, but it is also possible that enterprising retailers offer better service (for example more employees answering the telephones at the customer service) or special electricity products (such as for example power produced in environmentally benign power plants). The retailers may also take over part of the risks (both towards producers and consumers) by offering stable prices during longer periods than one trading period.

The Real-time Market

The real-time market includes the trading which occurs during a trading period. A real-time market is needed for several reasons. One is that power plants selling to the ahead market may fail and have to be replaced by other generating units. Moreover, when trading in the ahead market it is difficult for many players to predict how much they actually will produce or consume. This is for example the case for wind power plants, where the available generation capacity depends on the wind speed, which is hard to predict even just a few hours ahead [96], or retailers whose customers have so-called take-and-pay contracts, which means that the customer may consume any amount of power up to a specified limit. Finally, it is not certain that the ahead market has taken enough consideration to the limitations of the common grid, which may force the system operator to redispatch production and consumption so

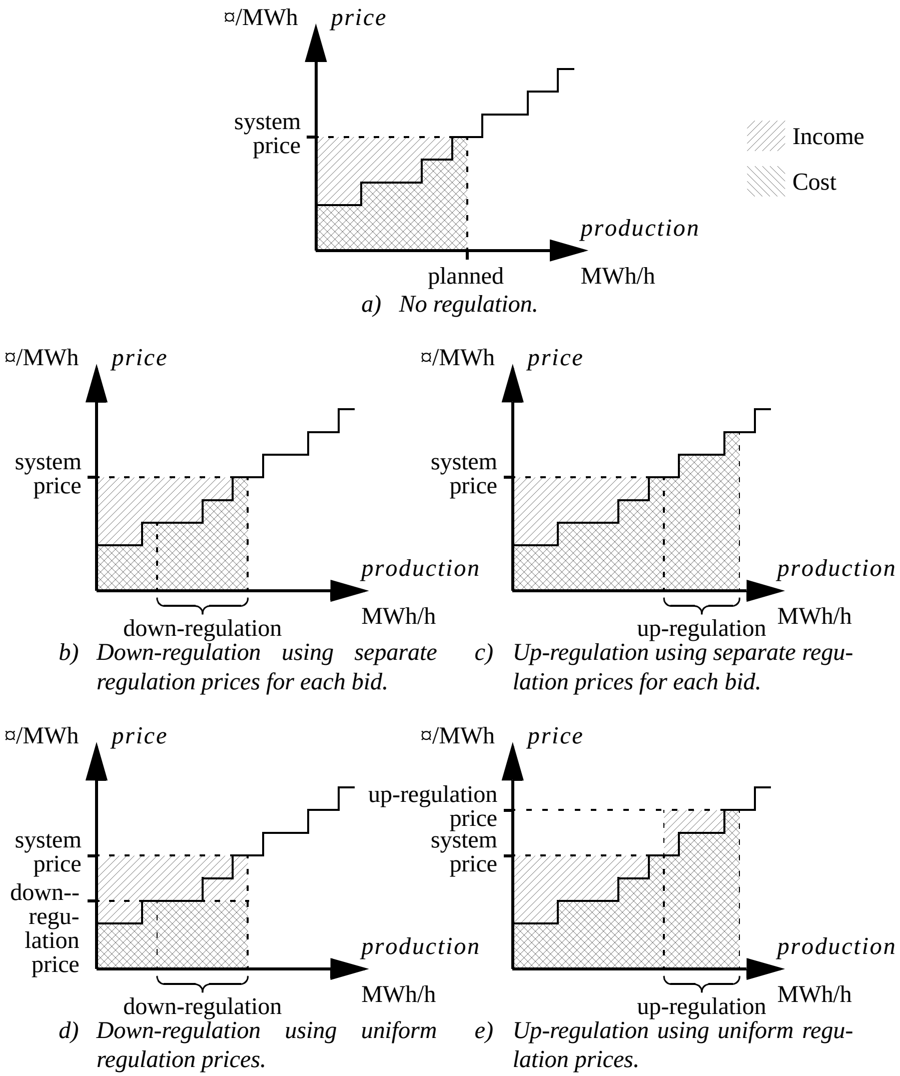
7. This is for example how the electricity market of England-Wales was operated between 1990 and 2001 [30, 69].

that the grid is operated safely.

There are two ways of organizing the real-time trading. The first variant is to establish a regulating market, which means that the players normally decide themselves how much to produce or consume, but if necessary the system operator asks a certain player to change the production or consumption. The second variant is that the system is centrally dispatched by the system operator and the other players are obligated to follow the instructions of the system operator.

Let us start by studying how a regulating market works. During the trading period the system operator will whenever necessary activate bids submitted to the regulating market, so that safe operation is maintained to the least possible cost. Two kinds of bids can be submitted to the regulating market. When down-regulating the system is supplied with less energy than agreed upon in the ahead market (thus, a producer carries out down-regulation by reducing the production, whereas a consumer carries out down-regulation by increasing the consumption). Down-regulation means that the player buys regulating power from the system operator and a down-regulation bid must therefore state how much the player can down-regulate (in MW) and the maximal price (in ¤/MWh) which the player is willing to pay for the regulating power. Similarly, up-regulation means that the player sells regulating power to the system operator, i.e., an up-regulation bid states how much the player can up-regulate and the minimum price for which the player is willing to sell regulating power. Unlike the bids to the ahead market, regulation bids are not just financial but physical undertakings—the system operator measures generation and load and may control that activated regulation bids really have been carried out within time.

The pricing of the regulating market can either be different for each activated bid or there may be uniform prices for up- and down-regulation respectively. Separate pricing means that those players buying regulating power pay exactly the price they stated in their down-regulation bid and the players selling regulating power get paid as much as they stated in their up-regulation bids. Using this pricing scheme the players will—provided that the competition is good—not be able to make any profits from the real-time trading; for example, a producer will buy regulating power at the same price as it would have cost to produce it (cf. figure 2.2b) and will when selling receive just as much to cover the production cost (cf. figure 2.2c). Such arrangements are not very attractive neither for producers nor consumers; it implies that they submits bids out of public duty or that they are simply forced to submit bids whenever possible. To make a regulating market more appealing it is possible to use marginal pricing instead, which means that a down-regulation price is defined, which is equal to the lowest price among the activated down-regulation bids, and an up-regulation price, which is equal to the highest price among the activated up-regulation bids. All activated bids will then obtain



a) No regulation.

b) Down-regulation using separate regulation prices for each bid.

c) Up-regulation using separate regulation prices for each bid.

d) Down-regulation using uniform regulation prices.

e) Up-regulation using uniform regulation prices.

Figure 2.2 *Principles of pricing in a regulating market. In this example there are seven power plants with increasing variable production costs. The planned production according to panel a has been sold in the ahead market, resulting in the indicated system price.*

Each producer now submits bids to the regulating market. The power plants which can down-regulate are willing to pay a price not exceeding their variable production cost (otherwise buying regulating power is a loss). In a similar way, the power plants which can up-regulate are only willing to do so if they get paid at least as much as their variable production costs.

In panels b and c regulating power is traded using separate prices (corresponding to the upper and lower limits respectively of the submitted bids). In panel d and e uniform down- and up-regulating prices are used instead. The first activated regulating bids now receive a more favourable price than the variable production cost.

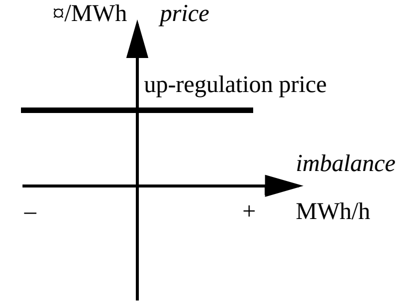
these regulating prices. Thus, a producer may for example buy regulating power to a price which is less than what it would have cost to produce the same amount in an own power plant (cf. figure 2.2d) or may sell regulating power to a price which is higher than the production cost cf. figure 2.2e).[8]

Central dispatch is straightforward to carry out in a centralised electricity market (but it is also possible to consider bilateral transactions; this is for example done in the PJM market in north-eastern USA [26]), because the system operator already has access to the preferences of the players in the form of their bids to the ahead market. To give the players the possibility of correcting bids based on mistaken forecasts, the players may be allowed to adjust their bids before the real-time trading, but the system operator may then require that they can provide a reasonable explanation why they change their bids.[9] In regular intervals, e.g. every fifth minute, the system operator performs an economic dispatch, the result of which is announced to the producers and consumers, who act according to this plan.[10] For each dispatch phase there will be a price (which may differ in different parts of the system); these prices are called real-time prices and can be used in two ways. In for example the national electricity market in Australia all trading use real-time prices and the ahead market is in practice just a basis for the price forecast supplied by the system operator [27]. On the other hand, in several parts of the U.S. the prices are used as a basis for the pricing in the post market [13, 26].

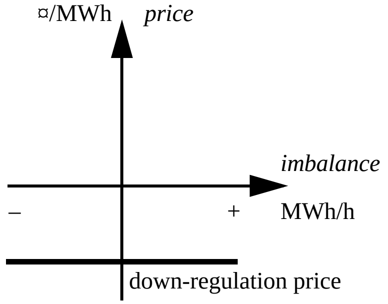
The Post Market

As it is impossible to know already at the time of the ahead trading what will happen during a certain trading period, it is inevitable that smaller or larger deviations will occur between the planned trading and what actually is traded. The post market is necessary to compensate for these deviations and to make sure that somebody pays for all energy supplied to the system during a trading period. However, all players do not need to take responsibility themselves for the differences between ahead trading and what is actually produced and consumed, but it is possible to introduce certain balance responsible players.

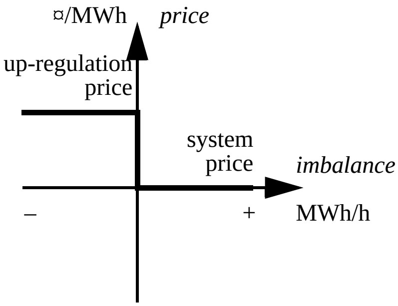
8. It is actually possible to combine the two pricing schemes. An example of this is the Nordic electricity market, where the regulating bids activated to maintain balance between production and consumption are paid uniform up- and down-regulation prices, whereas those bids activated to prevent a part of the grid to be overloaded (so-called counter trading—see section 6.2) receive separate prices.
9. This is the arrangement in for example the national electricity market of Australia; however, an exception is that consumers may decrease their consumption without notifying the system operator [27].
10. It is however not necessary that all consumers are part of the dispatch. Regular residential consumers may for example consume as much as they wish; the system operator will estimate their consumption and include it as a firm load in the dispatch.



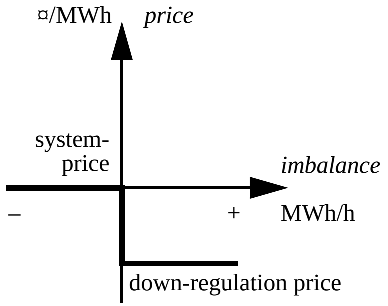
a) Single-price system during an up-regulation period.



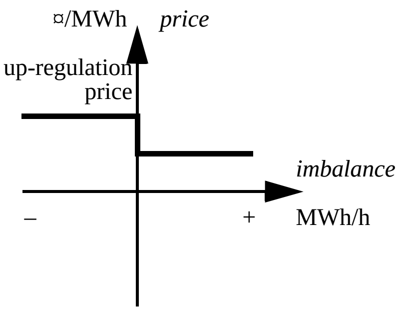
b) Single-price system during a down-regulation period.



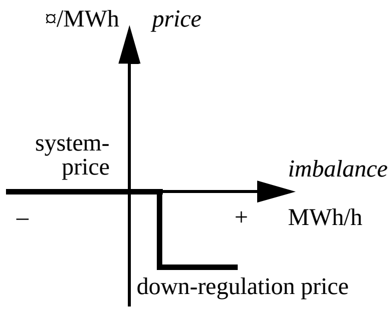
c) Two-price system during an up-regulation period.



d) Two-price system during a down-regulation period.



e) Mixed price system during an up-regulation period. The cost of positive imbalance is the mean value of the system price and the up-regulation price.



f) Two-price system with dead-band during a down-regulation period.

Figure 2.3 *Different pricing schemes for the post market. Players have positive imbalance when supplying more energy during the trading period than they have extracted. The system price refers to the electricity price in the ahead market, whereas up- and down-regulation prices refer to the prices of the regulating market (corresponding to the real-time price if central dispatch is used).*

Being balance responsible is a purely financial undertaking and the balance responsible player does not need to be a producer or consumer, but may act as an agent for others.

When a trading period is ended the system operator can compile how much the balance responsible and his or her clients have actually produced and consumed, as well as how much they have bought or sold in the ahead and real-time markets. This will almost certainly result in a deviation between supplied and extracted energy. The balance responsible players then have to trade at the post market to settle this imbalance. Players having positive imbalances (i.e., they have supplied more energy than they have extracted) sell balance power to the system operator. If there is a negative imbalance instead then the player has to buy balance power from the system operator.

The price of the balance power is generally related to the prices used during the real-time trading and as usual there are several variants. The question is partly whether average pricing is used (as for example in Denmark, England-Wales and Spain [36]) or marginal pricing (as for example in Australia, Finland, Norway and Sweden [27, 36, 39]), and partly whether or not separate prices are used for buying and selling balance power. A one-price system means that all balance power is bought and sold for the same price (this is the case in for example Australia, Norway, Spain and Germany [27, 36]), whereas a two-price system means that there are separate prices for negative and positive imbalances respectively (this is the case in for example Denmark, Finland, Sweden and England-Wales [36, 39]).

An overview of the principles for one-price and two-price systems is given in figure 2.3. If the real-time trading has required both up- and down-regulations then it must first be decided the direction of the net regulation during the trading period; if the system operator has bought more regulating power than they have sold then the trading period counts as up-regulation and vice versa. In a one-price system the up-regulation price is used for all balance power during up-regulation periods (figure 2.3a) and the down-regulation price is used during down-regulation periods (figure 2.3b). If the real-time market uses central dispatch rather than a regulating market, the real-time price is used instead of up- and down-regulation prices.

The two-price system means that a less favourable price is given to those players who are assumed to be the cause of the activation of regulating bids. During an up-regulation period those players who have not supplied enough energy, i.e., which have negative imbalances, must pay the up-regulation price (which is higher than the price in the ahead market), while the players having positive balance receive the same price as in the ahead market (figure 2.3c). During down-regulation on the other hand, those players having positive imbalance are assumed to have caused the need for regulation and therefore are getting paid the down-regulation price (which is less than the ahead market price), while the other players receive the same price as in the ahead market (figure 2.3d). This approach can also be applied to real-time prices originating from a central dispatch; if the real-time price is higher than the ahead market price then there has been an up-regulation period and vice versa.

Using a one-price system a balance responsible having an imbalance in “the right direction” (i.e., positive imbalance when the system has a net up-regulation and negative imbalance when the system has a net down-regulation) may buy or sell balance power to a price more favourable than in the ahead market, which means that there in some cases will be profitable to have an imbalance. Thus, the one-price system introduces a possibility to intentionally obtain an imbalance, but it is hard to see how a player systematically could take advantage of this possibility. However, the one-price system means that the costs decrease for those being balance responsible of unpredictable production or consumption, because the cost of the occasions when the player has an imbalance in “the wrong direction” is to some extent compensated by the income of the occasions when the player has an imbalances in “the right direction”. In a two-price system it is never possible to make any profits from an imbalance, which results in higher costs for the balance responsible players, who accordingly can be assumed to feel more pressure to keep their own balance in every trading period.

Finally, it can be mentioned that there are some other variants of pricing balance power. It might for example be desirable to compromise between a wish to motivate the balance responsible players to keep their balance and a wish to not disadvantage players having difficulties predicting or regulating production and consumption. Such a compromise is to use a two-price system, where those players having imbalance in “the right direction” will not receive a price as favourable as the corresponding regulating price, but still better than the price of the ahead market (for example a mixed price system as in figure 2.3e). It is also possible to refrain from punishing lesser imbalances by an unfavourable price (so-called dead-band; see figure 2.3f).[11]

11. A dead-band has been introduced on trial in Sweden since November 2003 [35].

Chapter 3

THE IDEAL ELECTRICITY MARKET

As we all know, the word ideal has several meanings; hence, the notion *ideal electricity market* can be given several interpretations. An ideal electricity market can be considered a mathematical simplification of real electricity markets, in the same spirit as ideal transistors and ideal resistors are simplified mathematical models of physical components. It is also possible to use the word ideal in the sense “something to strive for”; as the ideal electricity market per definition maximises the benefits to the society, it can be used as a benchmark when evaluating different options to design a real electricity market. When I first used the notion in my licentiate thesis [7] I favoured the former interpretation (mathematical simplification), but nowadays I find the latter interpretation the most interesting.

In the first part of this chapter I provide a definition of ideal electricity markets, and then I analyse which requirements have to be fulfilled in order to consider an electricity market as ideal. In the second part of the chapter there follows a description of a basic, mathematical model of ideal electricity markets. The idea of this model is of course that it should be useful for Monte Carlo simulation.

3.1 PREREQUISITES

A model of an electricity market has two main components; partly an assumption of how the players of the electricity market behave and partly a model of the power system. The term *ideal electricity market* does not refer to any specific design of the electricity market, but I have chosen a general definition which can be applied to restructured as well as vertically integrated electricity markets.

> ***Definition 3.1.*** An electricity market is ideal if it maximises the benefit to the society while considering the physical limitations

of the power system.

The important thing of the definition is the assumption that all players act in such a manner that the benefit to the society is maximised; the power system model (which sets the physical limitations) can on the other hand be chosen quite freely, without making the electricity market non-ideal.

If an electricity market should be ideal, four main conditions must be fulfilled. An ideal electricity market should have *perfect competition, perfect monopolists, perfect information* and *perfect grid tariffs*. In practice, a number of assumptions are concealed behind these headings; some of these assumptions may even fit under more than one heading.[1] Below I will specify the requirements in more detail, but first we should have a closer look at what is meant by "benefit to the society" in definition 3.1.

It should maybe be pointed out that I do not intend to provide a complete mathematical derivation of the requirements of an ideal electricity market; the objective of this presentation is to introduce important notions and to describe fundamental relations. A more strict mathematical analysis of ideal markets in general can be found in textbooks on microeconomics, e.g. [137, 139, 140].

3.1.1 Social Benefit

It would probably be hard to find someone who does not agree with the statement that an ideal electricity market should maximise the benefit to the society, but opinions will most likely differ as soon as the natural follow-up question "What is *really* beneficial to the society?" is asked. In economics it is common to use the notion total surplus as an indicator of the social benefits.[2] The total surplus is the sum of the surplus for every player in the market. Assume that there is a market price λ for a certain good. The surplus of producer p is equal to the income from selling q_p units minus the production cost, i.e.,

$$PS_p = \lambda q_p - C_p(q_p), \tag{3.1}$$

The surplus of consumer c is equal to the benefit of consuming q_c units minus the purchase cost:

$$CS_c = B_c(q_c) - \lambda q_c. \tag{3.2}$$

1. The premier reason why I still continue to use these four main assumption is that they look really striking when the term ideal electricity market should be summarised.
2. This footnote has been added to maintain the footnote numbering in pace with the Swedish edition.

Thus, the total surplus is

$$TS = \sum_c (B_c(q_p) - \lambda q_c) + \sum_p (\lambda q_p - C_p(q_p)). \tag{3.3}$$

This expression can be rewritten as

$$TS = \sum_c B_c(q_c) + \lambda\Big(\sum_p q_p - \sum_c q_c\Big) - \sum_p C_p(q_p). \tag{3.4}$$

To simplify (3.4) we introduce

$$\begin{aligned} q &= \text{total turnover of the market} = \sum_c q_c = \sum_p q_p, \\ B(q) &= \text{total benefit of all consumption} = \sum_c B_c(q_c), \\ C(q) &= \text{total cost of all production} = \sum_p C_p(q_p), \end{aligned}$$

which yields

$$TS = B(q) - C(q). \tag{3.5}$$

If the total surplus should be maximised then it is practical to study the derivative of $B(q)$ and $C(q)$. $MB(q) = dB(q)/dq$ is the marginal benefit function, i.e., it states how much the total value increases if the total consumption is increased by one unit. A more convenient designation for $MB(q)$ is *demand curve*, because the function describes the relation between demanded quantity and the market price. The demand curve corresponds to the consumers' willingness to pay, since it is reasonable to assume that a consumer is not willing to pay more for increased consumption than the consumer's benefit increases. $MC(q) = dC(q)/dq$ is equal to the marginal cost function, which states how much the total production cost will increase when the production is increased by one unit. The function $MC(q)$ is generally referred to as *supply curve*, because it describes the relation between produced quantity and market price. In general, the supply curve corresponds to the variable production costs of the producers.

The total surplus, TS, corresponds to the area between MB and MC, because

$$TS = \int_0^q MB(x)dx - \int_0^q MC(x)dx = \int_0^q (MB(x) - MC(x))dx. \tag{3.6}$$

If we start at the turnover zero and let q increase, this area will increase as long as $MB > MC$, i.e,. until we reach the turnover q^* (cf. figure 3.1). Then the area becomes "negative" and TS will decrease again. Apparently, the total surplus is maximised at the turnover q^*, where marginal value of consumption equals marginal cost of production. A natural market price will be $\lambda^* = MB(q^*) = MC(q^*)$, as both consumers and producers require this price to

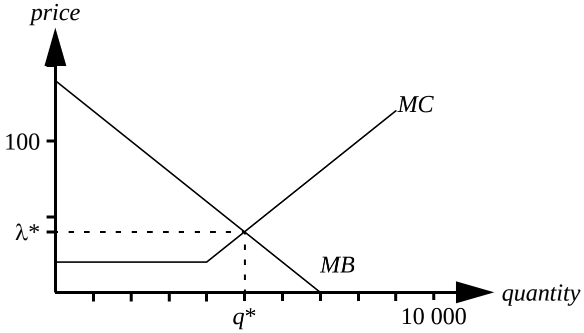


Figure 3.1 *Maximizing the total surplus. In the example we have* $MB = 140 - q/50$ *and* $MC = q/50 - 60$ *for* $q \geq 4\,000$*. The total surplus is maximised at the turnover* $q^* = 5\,000$ *and the market price is then* $\lambda^* = 40$*.*

trade the quantity q^*.

There is an obvious disadvantage of this definition of the social benefits; it is assumed that all benefits and all costs can be measured in money. Both production and consumption may however cause consequences which are hard to put a price tag on, because they depend on our moral values. How much are we for example willing to pay for beautiful nature and biological diversity? What is the value of a human life? Moreover, in some cases it is not just hard to quantify the social benefits, but we must also consider how the surplus is distributed. Although most of us accept that not everybody can have the exact same amount of welfare, there is some kind of limit to which differences we can accept.[3] When studying how *TS* is affected by different measure it is thus necessary to interpret the results with some care.

3. For example, assume that there is a market where a fat, white director in the industrialised country Land gains a surplus of 20 000 ¤, while 1 000 farmers in the developing country Nchi gain a surplus of 1 ¤ each, which yields a total surplus of 21 000 ¤. Suppose that this market was changed so that the distribution became 10 000 ¤ for the director and 10 ¤ each for the farmers. As a consequence of this redistribution the director has to cut down his consumption of golf clubs, liquor and cigars, whereas the farmers now afford to send their children to school, so that they may learn how to read and write. Even though the total surplus in the latter case is just l 20 000 ¤,* I think that most people would be inclined to agree that the benefit to the society nevertheless are larger than in the original distribution.

* All right, the example is somewhat exaggerated, because it could be possible to assign a value to children being able to read and write. If this value would be let us say 1 ¤ and each farmer family has three kids then the total surplus would be 23 000 ¤; hence, the total surplus corresponds to the subjective valuation of the benefit to the society again. But ability of reading and writing is difficult to put a price on and therefore tends to be neglected in the fancy charts of all bean counters.

As the ideal electricity market according to the definition is maximizing the benefit to the society, it is assumed that the social benefits actually can be measured. In an ideal electricity market, there is no indistinctness such as those mentioned above, but *all consequences of the decisions of the players can be assigned a monetary value,* which can be included in the total surplus, *TS*.

3.1.2 Perfect Competition

Perfect competition can be described as a state where the players by themselves act in such a way that the benefit to the society are maximised. A number of constraints have to be fulfilled to achieve this ideal condition, both concerning the good to be traded and how the players should behave.[4] Here follows a short overview of these requirements, without any special order.

We can start by assuming that the players are *rational* (which means that they will always try to maximise their own benefits) and that they are *free to trade* with each other if they want so (i.e., the players decide themselves how much they want to buy or sell and for which price). The free trade also means that it is possible for new players to enter the market if there are unused business opportunities.

The players maximise the benefit to the society, without any central coordination, if the individual players benefits coincide with the social benefits. Consider a player who has such a small share of the market that regardless of how much they choose to produce or consume, the market price is not affected, provided that the rest of the market continues to maximise the benefit to the society. Such a player is referred to as a *price taker*, because the player has to accept the price offered by the market. A price taking producer maximises the individual benefits by increasing the production until the marginal costs correspond to the market price (see figure 3.2).

The situation for consumers is somewhat more complicated and we have to differentiate between *private goods* and *public goods*. Private goods are such goods that are rival. A meat ball is a good example of a private good; if I consume a meat ball (i.e., eat it) then nobody else can consume the same meat ball. If there is no rivalness, the good is public instead. Street lighting is a classical example of a public good; me walking on a lit up street does not prevent other consumers to enjoy the benefits of the same lighting.[5] Whether a good is public or private influence how the individual benefit of a consumer relate to the total benefit of the whole market. The difference is illustrated in figure 3.3. Concerning private goods the situation of the consumers is similar to that of the producers; they have to accept the market price and choose to

4. This footnote has been added to maintain the footnote numbering in pace with the Swedish edition.

consume so much that the marginal benefit is equal to the market price.

As for private goods the benefit to the society is maximised if the marginal benefit of consuming another unit of a public good equals the marginal production cost. The difference is that the consumers in this case do not choose how much they consume—since all consumers receive the same amount of the public good—but the question is how much each consumer is willing to contribute to the procurement of the public good. The problem is that it might be very profitable for a consumer not to contribute at all and let the other consumers pay the costs. Consumers reasoning in this way are referred to as *free riders*. If all consumers try to be free riders, the public good will not be procured, even though it very well could have been optimal for the public welfare.[6] Therefore, free markets are poor at supplying public goods. During the years, several methods to solve this dilemma have been proposed;[7] however, a closer study of these is beyond the scope of this dissertation. Let us just summarise that the most common methods to supply public goods are to appoint a special player (mostly a public authority) to be responsible for production of the public good, or by charity from individual players. *In an ideal electricity market it is assumed that all goods are either private or that there is some player which somehow makes sure that sufficient quantities of the public good is procured to maximise the benefit to the society.*

Let us now return to the assumption that all players are price takers. There are two explanations why the market price does not change very much when the price taker changes his or her turnover. The first is that price takers are so small that they only see a small fraction of the total demand and supply

5. There is a certain ambivalence between private and public goods. One and the same good may sometimes be public, while it under different circumstances becomes private. There is for example a practical limit to how many people who can consume the same street lighting, because it is not possible to squeeze in any number of people in the light of a street light.
6. This can be illustrated by elaborating an example from [139]: Consider ten remotely located real estates, which can only be reached by one way. Assume that it costs 100 ¤ to plough the way when it has been blocked by snow in the winter, and that each of the real estate owners value the benefit of being able to get through on a ploughed way to 20 ¤. It is up to the individual real estate owners to call the snow plough and whoever calls will have to pay the bill. Each owner thus faces the choice of either calling the snow plough and pay 100 ¤ for a benefit of 20 ¤, which clearly is not profitable, or to not call, which does not convey any costs, but in the best case results in a benefit of 20 ¤ (if somebody else calls). If all real estate owners try to individually maximise their benefit nobody will call for the snow plough, even though the total benefit is larger than the cost. A solution which is both beneficial to all real estate owners and fair would be that they each paid 10 ¤ for having the way ploughed; then everybody would make a profit of 10 ¤.
7. General descriptions of different methods are found in both [139] and [140]. Anyone interested in a really in-depth analysis of the problem can tackle [136].

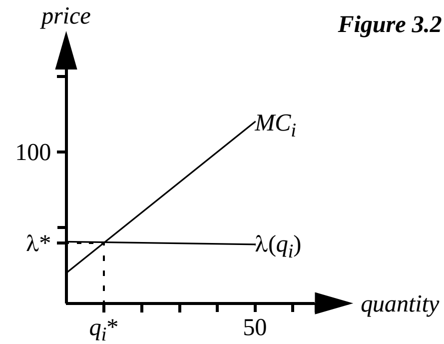
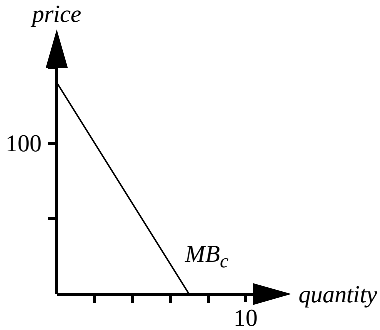
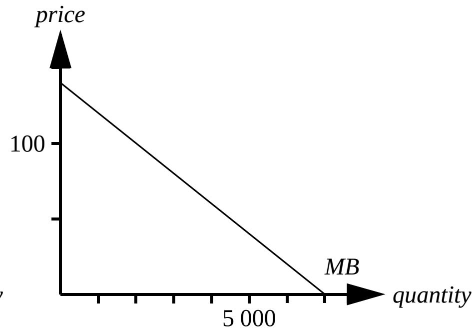


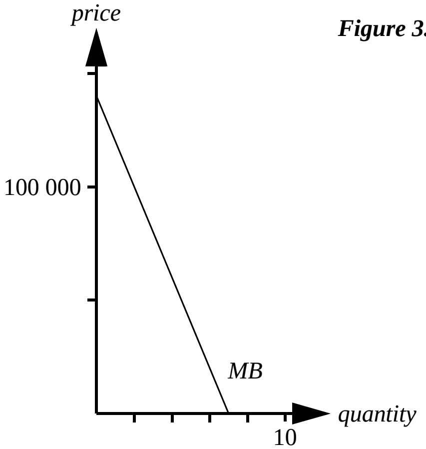
Figure 3.2 *The situation of a price taking producer. The figure shows the marginal cost curve of one of the producers in figure 3.1. The producer's choice of production level, q_i, has clearly just a very small impact on the market price; $\lambda(q_i)$ is approximately equal to the market price λ^* which maximised the total surplus (this presupposes that all other players in the market produce and consume so that the total surplus is maximised). Apparently the surplus of the producer is maximised if the output q_i^* is chosen, in which case the marginal cost of the producer is equal to the market price λ^*. If the producer chooses a larger output then the costs will not be covered and if less is produced, a possible profit is missed.*



a) Individual demand curves of the consumers.

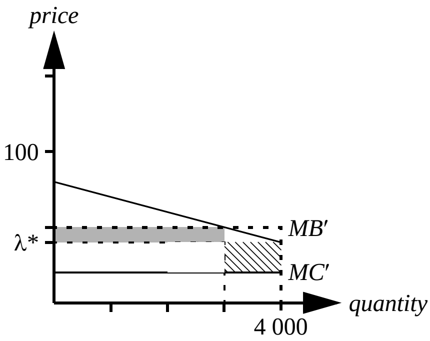


b) Total demand of a private good.

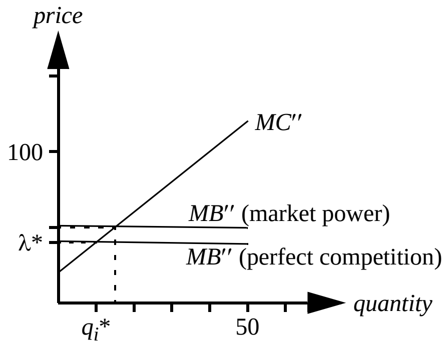


c) Total demand of a public good.

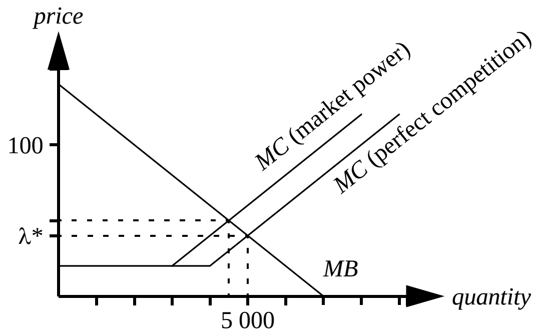
Figure 3.3 *Demand curves for private and public goods respectively. In this example there are 1 000 consumers in the market, each having identical individual demand curves according to panel a. If the good in question is private then all consumers will receive the same market price; the total demand curve will therefore be the "horizontal sum" of the individual demand curves, as shown in panel b. If it is a public good instead, each consumer will receive the same quantity; the total demand curve will then be the "vertical sum" of the individual demand curves, which gives a total demand curve as in panel c.*



a) *The situation of the price setter.*



b) *The situation of a price taker depends on the actions of a price setter.*



c) *Total demand and supply for the whole market.*

Figure 3.4 *The consequences of market power. The demand curve in this example is the same as in figure 3.1, i.e., MB = 140 – q/50. The total supply function is also the same as in figure 3.1, but in this case there is a dominating producer having the supply curve MC' = 20 when $q' \in [0, 4\,000]$ (panel a) and 100 small producers, which each have the supply curve MC'' = 20 + 2q'', $q'' \in [0, 50]$ (panel b). The large producer is a price setter; when this producer decreases the production, the market price increases. The producer maximises the profit by reducing the output from* 4 000 *(which is the production level which would be chosen if market power is not exercised) to* 3 000. *Admittedly the producer will loose income corresponding to the area marked by diagonal lines in area a, but on the other hand there is an extra profit—corresponding to the shaded area—for the remaining production.*

The small producers are price takers, because regardless of how much they produce, it will hardly influence the market price. These producers maximise their profits by choosing the output for which their marginal costs equal the market price. In total, these small producers will produce 1 500 *units if the large producer has an output of* 3 000 *units. The resulting market price is higher than the market price at perfect competition, λ*, as illustrated in panel c.*

respectively. If the player has a major share of the market, the player becomes a *price setter* instead. The actions of a price setter has a clear impact on the market price, even if the rest of the market is perfectly competitive (cf. figure 3.4). A price setter is said to have *market power*. If a player chooses to utilise having market power then that player can increase his or her surplus, but the total surplus of the market will decrease. The one exercising market power will thus gain on behalf of all other players. *In an ideal electricity market it is assumed that there is no market power or that the players who have market power refrain from using it.*[8]

The other explanation that price takers do not influence the market price is that if a price taker tries to change the market price in a favourable direction (from the price taker's point of view) then there will be a business opportunity for a competitor. If for example a price taking producer reduces the production, there will be a competitor whose marginal production cost was slightly above the market price, but who now gets the possibility to produce in place of the supposed price manipulator. This requires of course that the two producers really are competing with each other.[9] If several small players choose to cooperate in a cartel, they will behave as one large player and with that achieve market power. Formation of cartels is—as abuse of market power—controlled by legislation. *In the ideal electricity market it is assumed that there are no cartels.*

A final assumption which has to be fulfilled if the players by their own will should strive to reach the equilibrium (q^*, λ^*) is that there may not be any *externalities*. An externality appears when a transaction between a producer and a consumer does not just involve these two, but is also causing benefits or costs to a third party.[10] If the players only compare the market price to their own benefits and costs respectively, some players may choose a turnover

---

8. Withholding market power may seem as a contradiction to the assumption that all rational players maximise their own benefit. However, in many markets it is illegal to abuse a dominating position; thus, taking advantage of market power includes a risk of unpleasant consequences and then it might be more rational to refrain.
9. It is also required that the difference in marginal production cost does not differentiate too much between the two producers. If the price difference would be large then even a very small producer might be large enough to exercise market power. This prerequisite is therefore a part of the earlier assumption that no players exercise market power.
10. An externality may hence be either something beneficial (external income) or something unfavourable (external cost). The construction of hydro power plants is an example of an activity which cause both external costs and external incomes. The costs are related to the environmental impact and could for example be less income from fishing and tourism. The income is related to the possibility of controlling the water flow downstream the power plant, which may produce a more even water flow; hence, conditions improve for agriculture and irrigation. Besides, the risk of floods is decreased.

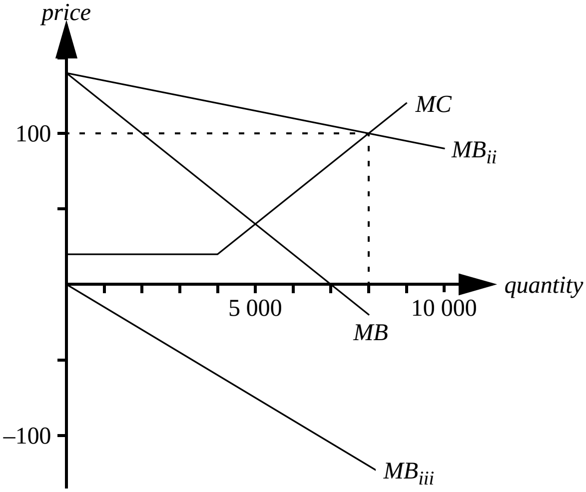


Figure 3.5 *The consequences of external costs. Consider the same market as in the earlier figures, but assume that there are three groups of players: producers (i), consumers (ii) and others (iii). The consumers marginal benefit function is* $MB_{ii} = 140 - q/200$. *Unfortunately their consumption is to the detriment of the other players, who therefore have the marginal benefit function* $MB_{iii} = -3q/200$. *Thus, the total marginal benefit is the same as in figure 3.1. The supply curve is also the same as in the earlier examples.*

If the consumers only consider their own benefits the market price will be $\lambda = 100$ *which yields a total consumption* $q = 8\,000$. *Due to the fact that the consumers disregarded the external costs the total consumption will thus be larger than when the public welfare is maximised. In a similar manner, an external income would have resulted in a too low consumption instead.*

which is too high or too low compared to how they would have behaved under perfect competition (cf. figure 3.5). To cope with this problem the externalities must be made visible to the players of the market, i.e., they must be internalised. *In an ideal electricity market it is assumed that all costs and incomes are internalised*.

How well do the above described assumptions of an ideal electricity market correspond to reality? With suitable legislation it is not particularly hard to give the players the possibility to trade freely. Further, electric energy is a private good, so the electricity trading should not have problems with free riders. In a real electricity market, there are however also some public goods. Among those are for example the grid, which is studied more closely in chapter 6.

Electricity generation is an activity with large economics of scale and it is therefore natural that an electricity market is dominated by a few large companies. Hence, there is in reality a significant risk of market power, which I briefly comment in section 7.1. Electricity generation also causes externali-

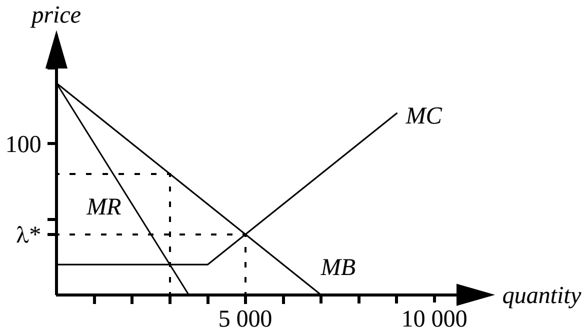


Figure 3.6 *Equilibrium price of a profit maximizing monopolist. Consider the same market as in figure 3.1, but assume that all production is controlled by a monopolist. This producer will not increase the production over the level where the marginal revenue, $MR = dR/dq$, where $R = MB(q)\cdot q$, is equal to the marginal production cost, MC. The result is apparently a market price which is higher than the price when the public welfare is maximised, λ^*.*

tites. Most of them are related to the environment. A more thorough discussion of externalities, electricity markets and environment is provided in chapter 4.

3.1.3 Perfect Monopolists

A monopolist is a player who solely controls all production (or in rare cases all consumption). A monopolist has per definition extreme market power and if the monopolist chooses to use this power, the market price will be higher and the total surplus less than what would have been obtained in a perfectly competitive market (cf. figure 3.6). *In an ideal electricity market it is assumed that any monopolist in the market is not profit maximizing, but altruistically tries to maximise the benefit to the society.* Such monopolists I refer to as *perfect monopolists*.

There are two reasons why I include perfect monopolists in my definition of an ideal electricity market. The first is that the power companies in a vertically integrated electricity market have monopoly, and by introducing the idea of perfect monopolist I want to emphasise that also vertically integrated electricity markets could be ideal. The other reason is that some functions in an electricity market should be operated by a monopolist even in restructured electricity markets.

If we start by investigating the first reason, we may ask ourselves why we should at all bother to restructure electricity markets. Why make things difficult by organizing a competitive market if it works as well with a planned economy, where one or more perfect monopolists run the electricity supply?

In theory, there is certainly no difference between perfect competition and a perfect monopolist, but in practice there are several organisational difficulties with monopolies. A price taker in a perfectly competitive market does not require a large administration to make decisions which maximise the public welfare; all the price taker has to do is basically to compare the market price to the own marginal benefit or marginal cost.

Running a monopoly is a more difficult task. Firstly, the monopolist must as stated above act in such a way that the benefit to the society is maximised, but regardless of how good the intentions of the monopolist are, it is too easy for the organisation to eventually forget about the social benefits and start prioritizing the benefits of the organisation itself—and then in particular the very personal benefits of the leaders of the organisation.[11] There is also a risk that the monopolist has a different view of the social benefits than the rest of the society—the term "benefit to the society" is as mentioned earlier not completely unambiguous. But even if the monopolist actually tries to behave as a perfect monopolist, it is difficult to control centrally a complete market. It is not unusual that the monopolist creates a complicated and inefficient bureaucracy, which cannot control the market in such a way that the benefit to the society is maximised.[12]

There is however no law of nature which yields that a monopolist must be inefficient. An open, democratic organisation with a clearly specified objective of its activities can very well be considered a perfect monopolist. Likewise should a monopoly run by a cooperative, making the customers of the monopoly its owners, be able to fulfil the requirements of a perfect monopolist.

My other reason for including the term perfect monopolists is that in some cases it is preferable to have a monopoly compared to even the most perfect competitive market, since there are so-called natural monopolies, i.e. markets where one single firm can supply all possible levels of output to a cost less than what would be obtained by several competing firms [137]. This situation can arise if there are high fixed costs and low variable costs. An example is

11. The history offers a number of tragic examples of this; maybe the most clear one is the oppression and exploitation of the own population executed by the Soviet elite.
12. Again the Soviet Union is an almost over-explicit example. As pointed out in [142] they had in the 1930's a simple planned economy, which could be controlled by a single ministry of industry. This centrally controlled economy could manufacture heavy industrial products (i.e., the kind of products usable to win world wars), but could hardly produce light industrial products which improve the quality of life for ordinary people. In the 1950's the number of ministries of industry had grown to 40, but still they could not produce food, housing or consumer products in such quantities that the standards of living could reach the same level as the Western World—in spite of a ruthless usage of resources and raw materials, with sometimes devastating consequences for the environment.

the construction of a bridge crossing a river. If the bridge has sufficient capacity to allow all possible traffic to pass then there is no efficiency gain to be made by building a competing bridge next to the first one. In this case the fixed costs (i.e., the cost of building the bridge) are high, whereas the operation costs (the cost of passing the bridge) is practically negligible. The new bridge would just increase the fixed costs; the variable costs—which decide the market price—would not be affected by the competition. In an electricity market, transmission and distribution grids are examples of natural monopolies. The grid is however of such importance to the electricity market, that I will discuss the grid pricing under its own heading (see section 3.1.5).

3.1.4 Perfect Information

The players can only trade in an optimal way if the information they base their rational decisions on is correct. *In the ideal electricity market it is assumed that all players have exact knowledge of all parameters of significance to their decisions.* We say that the players have *perfect information*. We could restrict ourselves to acknowledging that perfect information is a necessary condition for an electricity market to be ideal, but it might be interesting to study a few details about what this really means.

The players need for information depends both on the market structure as well as the technical properties of the good which is traded. If the demand and supply are relatively constant over time, there will be a rather stable market price, and all the individual players need to do is to compare the market price to their own benefits and costs respectively. Consider for example a product like a car; it is not very difficult to get an overview of how much a car of some particular quality should cost. Given this price the producers can decide if they manage to deliver a corresponding car to that price and the consumers can compare the benefits of owning such a car to the cost of purchase. In such a market it is fully possible for the players to collect information which reasonably can be considered as perfect.

For other goods—not the least electric energy—there are considerable variations in both supply and demand, even in the short run, which makes it harder for the players to determine the market price at a given moment. In those cases the market structure will have a larger importance. If the market is designed as some kind of auction (i.e., the players submit bids to some authority, which then decides either to accept or reject bids and sets the prices) then the central authority coordinating the market will assuredly need extensive information, but that information is obtained from the bids of the players. The individual players do not need to know anything more than their own preferences. A centralised electricity market is designed in this way. On the other hand, in a bilateral electricity market, the individual players will require a lot more information, because they have to make their own forecasts about how the market price will vary. A good forecast requires good knowledge of the market and the behaviour of the other players.

If a good can be stored, there will be two consequences; partly the market price becomes more stable, because variations in supply and demand can be evened out using the storage. At the same time, those players who can store the good need good forecasts about the future, as they otherwise cannot optimise the usage of their storage facilities. The possibility to store is thus a factor which creates a demand of forecasts about the future.

Predicting the future is of course an extremely difficult task and it is safe to say that as soon as the players of an electricity market need forecasts, perfect information becomes an impossibility. Studies of how forecast uncertainties affect the electricity market are therefore a pressing problem, which I will treat in more detail in chapter 5.

A special information problem is when some players have asymmetric information, i.e., when some players have access to more extensive or more reliable information than others. For example, the seller of a used car knows more about the condition of the car than the buyer does.[13] The seller thus has better information about his or her marginal cost than the buyer has about the marginal benefit, which may cause them to settle a price which favours the seller. By this means, part of the consumer’s surplus is transferred to the seller, but the total surplus is not affected. However, it could also be possible that the buyer makes such an incorrect assessment that he or she buys a car that would not have been sold if the buyer had known the state of the car—in this case the total surplus has been affected. Hence, asymmetrical information may lead both to redistribution of the total surplus, but in the worst case the total surplus will decrease, too. In an electricity market asymmetrical information could for example occur if different players have access to forecasts of different quality. This kind of problems is however nothing I will study more closely in this dissertation.

3.1.5 Perfect Grid Tariffs

A special property which differs an electricity market from a general market is that all electricity trading has to be performed using a common grid. The consequences of this fact could be included in the earlier described prerequisites of perfect competition, perfect monopolists and perfect information, but I find it quite natural to summarise the analysis under an own heading.

First we may observe that the grid is a natural monopoly. The cost of build-

13. It should be noted that it is not always the seller that has an information advantage towards the buyer. For example, anyone taking out a life assurance probably knows his or her own health better than the insurance company.

ing grids are far higher than the operation costs; therefore it is very unlikely that there would be any benefit to the society if parallel grids were built in order to introduce competition. That the grid is a natural monopoly does however not mean that it has to have one owner; different parts of the grid can be owned by different players, as long as each player has a local monopoly. This ownership structure is for example found in Sweden, where the transmission grid is owned by the state utility Svenska kraftnät, whereas the distribution grids are owned by several different companies.

An electric grid is a complex technical system, where a number of parameters, e.g. frequency, voltage and currents, musts be kept within certain limits in order to keep the system going. That this is done is of course a fundamental necessity to have an electricity market at all. *In the ideal electricity market it is assumed that all players behave so that safe operation of the system is maintained.* Exactly how it is arranged to fulfil this requirement is not specified in detail; safe operation can be maintained spontaneously of all players or because there is a system operator who has the authority to order other players to take appropriate actions when the safety of the system is at hazard.

Naturally, it costs to build and operate a grid and this cost must be included when the total surplus is determined. In an electricity market we therefore have

$$TS = B(q) - C(q) - C_T(q), \tag{3.7}$$

where C_T are the costs of the grid. The turnover maximizing the public welfare, q^*, is thus given by

$$MB(q^*) = MC(q^*) + MCT(q^*), \tag{3.8}$$

which means that the marginal cost of transmitting the quantity q^* over the grid, somehow has to be passed on to the grid users, so that they can compare their marginal benefits to the marginal transmission costs. We may say that the electricity trading causes external costs and the grid tariffs are a method to internalise them. *In the ideal electricity market it is assumed that the grid owners charge perfect grid tariffs, which provides the players of the electricity market correct signals about how the grid should be used to maximise the benefit to the society.*

Safe operation is in this dissertation equivalent to keeping a stable frequency in the system (i.e., to keep balance between production and consumption) and not exceeding the transmission capability of any interconnection. These questions will be discussed in section 5.1. In chapter 6 other operation costs of the grid will be treated; besides, some comments on maintenance and investment costs are given.

3.2 MODELLING

A nice property of ideal electricity markets is that they are easily modelled. Given certain conditions, i.e., what I refer to as a *scenario*, the players of the ideal electricity market will behave so that the benefit to the society is maximised. A scenario in an ideal electricity market can therefore be analysed by solving an optimisation problem in the following form:

maximise	*benefit to the society*	(3.9a)
subject to	*physical limitations of the power system.*	(3.9b)

I call this kind of optimisation problem a *scenario problem*.[14] Notice that the time perspective of the scenario problem is not specified. We may just as well let the scenario problem correspond to the problem of maximizing the benefit to the society in the short run as in the long run. In the short run all resources and the demand are given, whereas if the scenario should maximise the social benefits in the long run, we have to consider market dynamics, as for example that new production and transmission resources may be added to the system and old units can be shut down. In this dissertation I will not treat market dynamics any further, but it is implied that all models consider maximisation of the benefit to the society given the existing resources.

The modelling in itself is limited to finding a proper model to represent the constraint “the physical limitations of the power system”. We may choose any level of details we want. The models described below are intended to be used for Monte Carlo simulation of electricity markets and then it is desirable to keep the number of variables down in each scenario problem, since the computation time of an optimisation problem is not a linear function of the problem size; a twice as large problem will therefore take more than twice as long to solve.[15] Thus, a model having twice as large scenario problems will take considerable longer time to simulate, because thousands of scenarios have to be analysed, and it is therefore important to try to simplify the models as much as possible—of course provided that we do not remove such features which have a major impact on the behaviour of the electricity market.

14. An alternative to formulating a scenario problem is to introduce separate optimisation problems for each producer and consumer; these player problems are then related to each other by one or more balance constraints which apply to the entire market. At the end of the day, we will end up with the same optimality conditions regardless of which alternative we have chosen. When studying non-ideal electricity markets it might be more straightforward to formulate player problems than an overall scenario problem; examples of such cases are found in chapters 4 and 6.
15. The complexity of different optimisation algorithm is almost a science in itself. An insight of the topic is given in [132].

3.2.1 Power System Model

The choice of power system model is primarily depending on the objective of the model. In a really detailed model of a power system, a so-called instantaneous value model, all components are modelled using a mixture of differential and algebraic equations. Such a model requires of course immensely intensive calculations (it may be so that it takes two hours of computations to study what happens in the system during ten seconds— see for example [28]) and if it is possible we prefer to use simpler models. Such a simplification is a fundamental frequency model, where the models of some components are simplified so that part of the differential equations are replaced by algebraic equations [22].

If it is not necessary to study power system dynamics then it is possible to use a load flow model. This model uses rms values for voltage and currents, which enables us to get rid of all differential equations and just use a non-linear system of equations to represent the power system. A further simplification of the load flow model are so-called DC load flow models, where the voltage regulation is neglected and only active power is considered.[16] DC load flow are often used when the power system model is just a part of an economic analysis.[17] In spite of all simplification, the DC flow model still requires a large number of variables to represent the state of all buses of the power system. Consequently, we are forced to solve very large optimisation problems if we want to use this kind of models to study individual scenarios in an electricity market simulation. Therefore, I think that the model which is most appropriate for Monte Carlo simulation, is a so-called multi-area model.

In a multi-area model the real grid is simplified by joining several buses into “equivalent buses” (which we for the sake of simplicity refer to as “areas”). We also develop equivalent models of power plants, transmission lines, load etc. (further details about this will follow below). The final result is a model with a comparatively limited number of variables. The price we have to pay is of course having a less detailed model. The decreased level of details does not have to be restricted to the power system model, but may also affect the economical model. If for example several power plants, which in reality have separate owners, are joined to a single equivalent it is no longer possible to study the surplus of individual producers, but only the sum of the surplus of all producers. This kind of restriction is however fully acceptable in most cases.

Over the years several multi-area models have been proposed, for example [65, 66, 69, 70, 71].[18] The differences between different multi-area models

16. For clarification, load flow models are sometimes referred to as AC load flow as contrast to DC load flows. Both models are described in basic textbooks on power system analysis, e.g. [32, 34].
17. See for example [12, 115, 121].

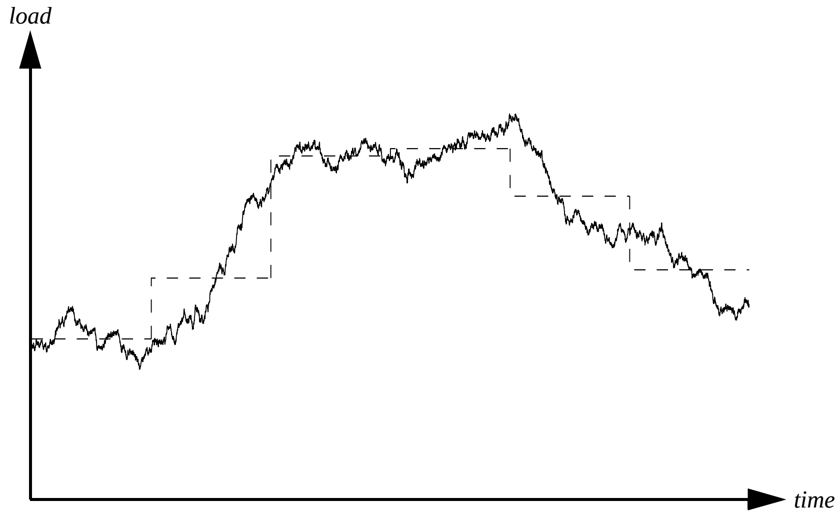


Figure 3.7 *Example of a piecewise constant approximation of continuously varying load.*

are about whether the problem formulation is linear or non-linear, how transmission losses are modelled or if they are neglected, if energy storage facilities (for example hydro reservoirs) are modelled, and if the need for reserves is considered.[19] At the end of the day, the objective of the simulation and the data available decides which model to use. I have chosen to present here a rather general model of an ideal electricity market, which should be appropriate in most contexts and which later can be modified to include non-ideal properties. My model is a further development of the model presented in [69, 70]; the difference is that I use a more general expression for transmission losses and I have added energy limited power plants to the model. Below details follow of how different parts of the power system are modelled.

Time

In a scenario we study an electricity market for a certain period of time. The conditions of the electricity trading varies continuously; the available capac-

18. It should be noted that many authors use the designation “multi-area model” although they use a DC flow model, which might not be completely incorrect, but nevertheless causes some confusion. Personally, I prefer to make a strict difference between these two notions.
19. If the reserves are considered, the multi-area model do not correspond to an ideal electricity market, because there is no need for reserves if perfect information is available.

ity of power plants and the grid may decrease due to failures, the available generation capacity in some power plants is weather dependent, the operation costs may depend on varying fuel prices, consumers increase and decrease their consumption, etc. However, scenarios having continuously varying scenario parameters are impracticable to analyse in digital computers and we are therefore forced to use some kind of approximation. It is common to use piecewise linear function, i.e., each scenario is divided into a number of time periods (which do not have to have the same length) and each scenario parameter is considered constant within each of the time periods (cf. figure 3.7). To emphasise that a piecewise constant approximation is used, we may use the unit MWh/h (average power) rather than MW (instantaneous power).

The piecewise constant approximation causes two kinds of errors. The first is that it is often incorrect to use the period mean of a scenario parameter as input to a function, because in the general case we have

$$h\left(\frac{1}{T}\int_0^T p(t)dt\right) \neq \frac{1}{T}\int_0^T h(p(t))dt. \tag{3.10}$$

However, linear functions will produce the same result regardless of whether the period mean is used or if we take the mean of the instantaneous value. Thus, the impact of this error depends on how we have chosen to model the electricity market. It is fully possible to use completely linear models, but unfortunately quadratic functions are necessary if transmission losses between the areas of a multi-area model should be realistically modelled (cf. [67]).

The second error which can arise is that although the period mean is less than a limit, this does not signify that the instantaneous value is below the limit during the entire period. Thus, there is a certain risk that important events in the system are not detected, as for example when the load exceeds the available generation capacity.

The total duration of a scenario can be chosen more or less arbitrarily. Two main options can be distinguished, and I refer to them as short and long scenarios respectively.[20] In a short scenario we study just a single time period, which means that all scenario parameters are constant during the entire scenario. To reduce the impact of the two errors described above, it is desirable that the time period of the short scenario is as short as possible—when the duration of the time period approaches zero the period mean approaches the instantaneous value. In most cases there is nothing preventing us from letting short scenarios correspond to an infinitesimal time period.[21]

20. In other literature—e.g. [46]—the designations “non-sequential” and “sequential” simulation are found; they have more or less the same meaning.

The disadvantage of short scenarios is that they represent a “snap-shot”, and it is unknown what has happened earlier in the system and which expectations the players have about future events. If there is some sort of energy storage in the system then the players of the electricity market must continuously decide whether they should use the stored energy now or save it and sell it for a higher price at some later occasion. Time constants, for example the start-up time of large thermal power plants or the delay time from the closure of the ahead market until the actual hour of delivery, may cause decisions to seem optimal at one point of time, but at a later time they turn out to have undesirable consequences. Such a time depending course of events is hard to simulate using short scenarios. To overcome this, we are forced to use long scenarios, which includes all scenarios comprising more than one time period.

Errors due to incorrect mean values are inevitable for long scenarios, but we cannot use too short time periods, as the scenario problem then becomes to large. The choice of the duration of each time period is therefore a trade-off between how much computer power we have access to (i.e., how large scenario problems we can manage) and how large errors we find acceptable. The most appropriate compromise is probably varying from system to system, but since at least the load shows large variations within a day it seems reasonable that the time periods should not be less than one hour long. The most natural choice is according to my opinion to let the time periods coincide with the trading period of the simulated electricity market.[22]

In the model, we refer to periods by using the index t. Symbols with too many indices can however be hard to read; therefore, I leave out the period index in general reasoning, where the period division is of no large importance. Likewise, the period index is unnecessary in all models of short scenarios.

Non-dispatchable Power Plants

The available generation capacity of non-dispatchable power is always depending on some factor beyond human control. Among the non-dispatchable power sources are example wind power (where the generation capacity depends on the wind speed), hydro power without reservoirs (where the generation capacity depends on the water flow passing the power plant) and photovoltaics (where the generation capacity depend on the insolation).

The designation “non-dispatchable” might be misleading, because it is

21. The exceptions include simulating so-called market splitting (see section 6.2), because the market during transmission congestion periods will be split during an entire trading period.
22. If a more detailed model of the real-time trading is to be used, it will however be necessary to use time periods shorter than the trading periods (see section 5.1).

actually possible to decrease the generation in these units by spilling power. However, this is never done in reality unless absolutely necessary to maintain safe operation of the power system, because the variable operation cost in non-dispatchable power plants is so low, that it is always more economical to reduce the generation in some other power plant.

Generally, it is assumed that the variable operation cost is zero in non-dispatchable units and their most significant feature is then the available generation capacity in a certain moment. In the general case, the available generation capacity of any area may be summarised into a single equivalent unit, with the available capacity $\overline{W}_n$.

Energy Limited Power Plants

The energy limited power plants have more or less the same properties as non-dispatchable power plants, but with one major exception: these power plants have the possibility to store energy for future use. By that means, there are larger possibilities to control the power plants, as surplus energy can be stored during those times when the available generation capacity is larger than the power output needed, and vice versa. It can be noted that all non-dispatchable power plants can be turned into energy limited power plants by adding an appropriate energy storage; a hydro power plant becomes energy limited if it is provided with a hydro reservoir, wind power and photovoltaics can be connected to batteries, etc.

Several energy limited power plants can be merged into an equivalent power plant. Each equivalent power plant, r, is characterised by having a certain available capacity, $\overline{H}_{r,\,t}$, a certain inflow to the energy storage, $Q_{r,\,t}$, and a certain maximal storage capacity, $\overline{M}_{r,\,t}$. If the inflow is larger than the available generation capacity at the same time as the storage is full, it is then necessary to spill energy from the storage; this option is represented by its own variable, $S_{r,\,t}$. It is also possible to introduce a cost function $C_{Hr,\,t}(H_{r,\,t})$, but since energy limited power plants generally have negligible operation costs, I assume in this dissertation that the operation cost can be omitted.

As a scenario cannot be infinitely long, it is necessary to somehow model how the energy storage facilities of the system have been used before the start of the scenario, which is done by assuming that the initial state, $M_{r,\,0}$, is known for each energy storage. Moreover, it must be considered that the players of the electricity market base their actions on what will happen in the future (seen from the viewpoint of the scenario). They will therefore make sure that they have saved some energy in each energy storage at the end of the last period of the scenario. This can either be modelled by assuming the final contents, $M_{r,\,T}$, to be known in advance, or by defining a benefit function, $B_{Mr}(M_{r,\,T})$, which is included in the objective function. The choice between these two options is arbitrary, since a particular value of stored energy will result in a particular stored energy and vice versa.[23] I have chosen to use predetermined final contents, because the objective function then becomes a little bit shorter.

A very important assumption when modelling energy limited power plants is that there may not be any couplings between the energy storage facilities of two equivalent power plants, as this would make the scenario problem much more complicated to solve. If we for example consider hydro power plants in a river system, electricity generation in one power plant causes the energy storage, i.e., the reservoir contents, to increase in the next power plant downstream. Therefore all hydro power plants in the same river system must be treated as an equivalent power plant.

It is not an easy task to obtain equivalent models of energy limited power plants. There are several factors that make the calculations complicated. For example may the efficiency of a power plant be depending on the contents of the energy storage (for example, the efficiency of a hydro power plant depend on the head, which in its turn depend on how much water there is in the reservoir). The varying efficiency makes it hard to determine exactly how much energy a storage actually contains at a given moment. Besides, couplings between different power plants (as for example the above mentioned hydrological couplings between hydro power plants) make it difficult to identify clear-cut available capacity, maximal energy storage, etc.[24]

Thermal Power Plants

A thermal power plant generates electricity by combustion of some fuel, which is assumed to be available in unlimited amounts.[25] The combustion process can be controlled so that the thermal power plant generates a certain desired output. The operation cost partly consists of costs which depend on how energy is produced—primarily the fuel cost—and partly of the cost to start the power plant. The latter cost varies a lot depending on which technology is used. When starting a diesel generator set it is more or less immediately ready for power generation, whereas a coal-fired condensing power plant may take several hours of heating the boiler—which obviously requires a certain amount of fuel—before it is ready to be operated. Detailed modelling of the start-up costs require the usage of integer variables [90] and that is something we want to avoid in the multi-area model. Optimisation problems

23. But in exceptional cases the relation between stored energy and its value is not entirely unambiguous.
24. Actually, calculation of equivalents for power plants with energy storage facilities is a research project in its own. Some initiatives have been taken at KTH to study equivalents for hydro power.
25. If the access to fuel is limited then thermal power plants are also considered energy limited.

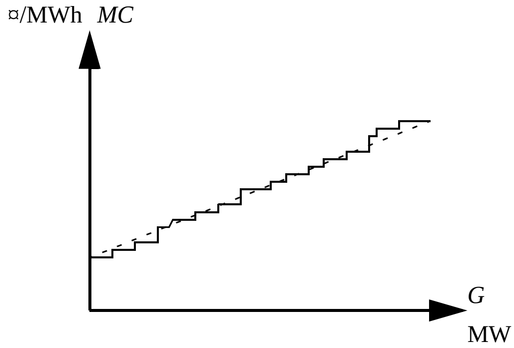


Figure 3.8 *Equivalent unit of power plants with different operation costs. The figure above shows a number of power plants having small capacity compared to the total capacity. The difference in marginal generation cost between any power plant and the next more expensive power plant is also small. In this case the power plant might be represented by a single equivalent unit with a linearly increasing marginal generation cost. The approximation is indicated by a dotted line in the figure.*

including integer variables are generally more complicated to solve than problems with just continuous variables; hence—and this is important for Monte Carlo simulation—they take more time to solve.[26] Therefore, for the time being we neglect the impact of start-up costs (I return to this question in section 5.2).

Several thermal power plants located in the same area can be merged into an equivalent power plant. In this respect it is appropriate that the involved power plants have approximately the same operation cost. It is however also possible to consider a number of smaller power plants as an equivalent unit, as long as there is an equivalent cost function which is a sufficiently good approximation of the price difference between the power plants (cf. figure 3.8). Each equivalent unit, $\underline{g}$, is represented in the multi-area model by an available generation capacity, $\overline{G}_g$, and a cost function, $C_{Gg}(G_g)$.

The Grid

The grid can be divided in a transmission part (high voltage lines for transferring power over long distances) and a distribution part (low voltage lines extending to the consumers). What counts as transmission and what counts as distribution varies from case to case and is primarily depending on the size of

26. Cf. [125], chapter 9.

the considered power system.[27]

The transmission grid is only represented by interconnections between the areas. Between any two areas there may only be one interconnection in each direction and these interconnections are characterised by a certain available transmission capability, $\bar{P}_{n,m}$, and a certain loss function, $L_{n,m}(P_{n,m})$. It is assumed that the players of the electricity market have complete control of the flow between the areas, as long as the transmission capability is not exceeded. This is a fully acceptable approximation for high-voltage DC (HVDC), because the transmission on these lines is controlled by power electronics.AC lines are not controllable to the same extent. If there is an AC grid where power is injected as some buses and extracted from others, the power flows will be distributed according to physical laws. The resulting power flow on each line can be determined using load flow calculations [32, 34], but such calculations are desirable to exclude from the multi-area model. It is preferable to try to find loss functions which approximately resemble the losses which would be obtained if a load flow was performed. In the same manner, we must choose the transmission capability between the areas so that it corresponds to the maximal flows which are possible without having unacceptable currents or voltages in a load flow. Determining appropriate approximations for a multi-area model by comparison to a load flow model is a complicated task and further research is necessary within this area.[28]

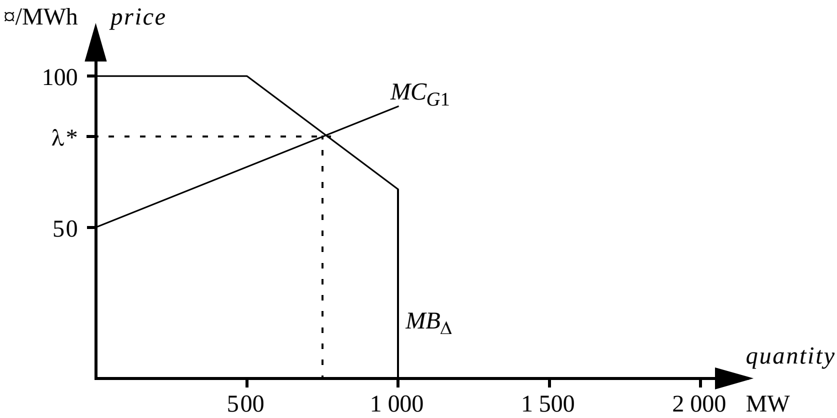
The distribution grid is neglected in the multi-area model, which means that all grid limitations within the areas are disregarded. The distribution losses can either be completely ignored or included in the price insensitive load in each area. In the latter case, it is necessary to obtain an approximate function for the distribution losses; the results of the study in [67] indicate that the internal losses can be approximated as a function of the load in the area.

Load

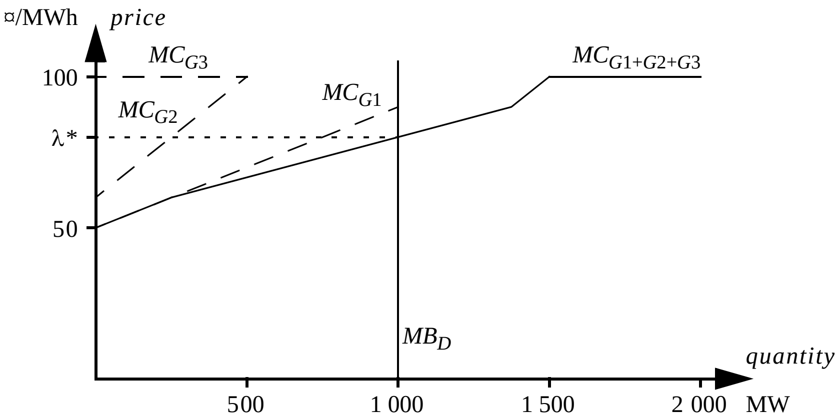
There is generally no reason to distinguish between individual consumers; they can be merged into “equivalent consumers” in a similar manner as when we introduce equivalent power plants. When identifying these demand curves it is appropriate to differ between price insensitive load, D_c, and price sensitive load, Δ_c, because they have to be treated somewhat differently in the model.

Price sensitive load is represented by a benefit function, $B_{\Delta c}(\Delta_c)$, and a

27. In for example Sweden lines having a voltage of 220 kV or more are considered as the transmission grid, whereas in a small local or regional grid as for example Kigoma region [7] the transmission grid consists of 33 kV lines.
28. Cf. for example [67], in which approximation of loss functions was studied, and [68], which discusses calculation of transmission capability.



a) *Supply curve of the power plant G_1 and demand curve of the load Δ. At the electricity price* 80 ¤/MWh *both production and consumption are equal to* 750 MW.



b) *The load has been divided into a price insensitive load, D =* 1 000 MW, *and two fictitious power plants, G_2 and G_3 respectively, which correspond to load reductions. The figure above shows the total supply curve (solid line) of the power plant G_1 and the two fictitious power plants (the supply curve of each power plant is indicated by dashed lines). The demand curve of the price insensitive load is just a vertical line; the load does not change regardless of the price.*

As can be seen, the electricity price is still 80 ¤/MWh, *at which* G_1 = 750 MW, G_2 = 250 MW *and* G_3 = 0 MW. *The price sensitive load is determined by* $\Delta = D - G_2 - G_3$ = 750 MW. *The solution is thus the same as in panel a.*

Figure 3.9 *Example of modelling price insensitive load. In this electricity market there is one power plant, G_1, and one price sensitive load, Δ. In the upper figure the supply and demand curves of the two players are shown. The lower figure shows the supply and demand curves, when the load has been divided in a price insensitive part, D, and two fictitious power plants, G_2 and G_3.*

maximal consumption Δ_c. A load being price insensitive means that the consumers are willing to pay any price for their consumption, which would imply that the benefit of their consumption is infinite. Therefore, it is not possible to define a benefit function $B_{Dc}(D_c)$ for price insensitive load. The solution is to assume that the electricity market always tries to supply the price insensitive load—we could say that price insensitive load is considered as one of the physical limitations of the electricity market. As there is no guarantee that there is always sufficient capacity available to cover the price insensitive load, it is necessary to introduce a possibility to disconnect a part of the price insensitive load. In the multi-area model this possibility is represented by a special variable for disconnected load, U_c, and a corresponding cost function, $C_{Uc}(U_c)$. This cost function does not have to correspond to the social cost of disconnected load (such a cost function could be very hard to identify), but can be chosen arbitrarily as long as it is always preferable to use the most expensive power plant of the system than to disconnect load.

Some simulation methods are based on the assumption that the available resources are compared to given, price insensitive load. This is however not an obstacle to including price sensitive loads too, because a price sensitive load can always be rewritten as a price insensitive load and a fictitious power plant representing load reductions (cf. figure 3.9).

3.2.2 Mathematical Formulation

The benefit, cost and loss functions used in the scenario problem of an ideal electricity market are assumed to be quadratic functions, i.e., functions in the following form:

$$h(x) = \alpha + \beta x + \gamma x^2. \tag{3.11}$$

The resulting optimisation problem can be considered a non-linear network problem and can be solved using the NNP algorithm [7] or some other, more general, solution method for non-linear optimisation problems [135]. To guarantee that there is a unique solution, it is required that the problem is convex; consequently, there will be some conditions on the parameters α, β and γ of the different functions. These conditions are however quite simple to fulfil (cf. the definitions in appendix A) and I will therefore refrain from describing the details about convexity in this presentation. Those interested in the convexity requirements of a multi-area model are referred to [7].

To make the mathematical model of an ideal electricity market more readable I have chosen not to rewrite the problem as a network problem. I have tried to use a notation which hopefully clarifies the relation between the assumptions behind the model and the mathematical expressions. A complete overview of the notation is found on p. 278ff.

Objective Function

The ideal electricity market by definition maximises the total surplus. As it has been shown earlier, this is equivalent to maximizing the value of all consumption minus the cost of all generation and the transmission costs. In the above described multi-area model the transmission costs are only represented by the cost of the electrical losses of the interconnections; hence, there is no reason to differentiate generation intended to be sold to consumers and generation intended to cover transmission losses.

In addition to the value of consumption and the cost of generation we have another term in the objective function of a scenario problem, because we needed to add cost functions for disconnection of price insensitive load. The complete objective function therefore looks like this:

$$\text{maximise} \quad \sum_{t=1}^{T} T_t \left(\sum_{c \in \mathcal{C}_\Delta} B_{\Delta c,t}(\Delta_{n,t}) - \sum_{g \in \mathcal{G}} C_{Gg,t}(G_{g,t}) - \sum_{c \in \mathcal{C}_D} C_{Uc,t}(U_{c,t}) \right) \tag{3.12}$$

Constraints

In an ideal electricity market it is assumed that the system is operated in a safe way, i.e., that frequency, voltages and currents are within certain limits. Voltages and currents in different parts of the grid are not explicitly considered in the multi-area model, but are part of the transmission capability and the loss functions. Neither the frequency is explicitly modelled, but to keep the frequency around the nominal value, balance between production and consumption is required; this balance has to apply to each area of the system. In plain language, a load balance constraint can be formulated as

total generation + import = export + price sensitive load
+ price insensitive load – disconnected load
in each area and time period. (3.13a)

Rewriting this constraint with all optimisation variables in the left hand side yields

$$\sum_{g \in \mathcal{G}_n} G_{g,t} + \sum_{r \in \mathcal{R}_n} H_{r,t} + W_{n,t} + \sum_{m \in \mathcal{P}_{n \leftarrow m}} (P_{m,n,t} - L_{m,n,t}(P_{m,n,t}))$$
$$- \sum_{c \in \mathcal{C}_\Lambda} \Delta_{c,t} - \sum_{m \in \mathcal{P}_{n \rightarrow m}} P_{n,m,t} + \sum_{c \in \mathcal{C}_D} U_{c,t} = \sum_{c \in \mathcal{C}_D} D_{c,t}$$
$$\forall\, n \in \mathcal{N},\ t = 1, \ldots, \mathcal{T}. \tag{3.13b}$$

When there are energy limited power plants in the system, it must be considered that the energy storage has physical limits and cannot be used just anyhow. There is an energy balance, which in words can be expressed as

New energy storage =
old energy storage + inserted energy – extracted energy
in each storage and time period. (3.14a)

Rewriting this constraint with all optimisation variables in the left hand side yields

$$M_{r,t} - M_{r,t-1} + T_t H_{r,t} + S_{r,t} = Q_{r,t}, \quad \forall\ r \in \mathcal{R},\ t = 1, \ldots, \mathcal{T}. \tag{3.14b}$$

Limits

Neither resources nor demand can be infinite; therefore, there must always be an upper limit to each optimisation variable. The upper limit of spillage from an energy storage is however unnecessary to determine; spillage is never worth striving for and will therefore be minimised; hence, in practice it is just the lower limit of the spillage which has any significance. The limits of the optimisation variables are thus the following:

$$0 \le \Delta_{c,t} \le \Delta_{c,t}, \qquad \forall\ c \in \mathcal{C}_{\Delta},\ t = 1, \ldots, \mathcal{T}, \tag{3.15a}$$

$$0 \le G_{g,t} \le \overline{G}_{g,t}, \qquad \forall\ g \in \mathcal{G},\ t = 1, \ldots, \mathcal{T}, \tag{3.15b}$$

$$0 \le H_{r,t} \le \overline{H}_{r,t}, \qquad \forall\ r \in \mathcal{R},\ t = 1, \ldots, \mathcal{T}, \tag{3.15c}$$

$$0 \le M_{r,t} \le \overline{M}_{r,t}, \qquad \forall\ r \in \mathcal{R},\ t = 1, \ldots, \mathcal{T}, \tag{3.15d}$$

$$0 \le P_{n,m,t} \le \overline{P}_{n,m,t}, \qquad \forall\ (n, m) \in \mathcal{P},\ t = 1, \ldots, \mathcal{T}, \tag{3.15e}$$

$$0 \le S_{r,t}, \qquad \forall\ r \in \mathcal{R},\ t = 1, \ldots, \mathcal{T}, \tag{3.15f}$$

$$0 \le U_{c,t} \le D_{c,t} \qquad \forall\ c \in \mathcal{C}_{D},\ t = 1, \ldots, \mathcal{T}, \tag{3.15g}$$

$$0 \le W_{n,t} \le \overline{W}_{n,t}, \qquad \forall\ n \in \mathcal{N},\ t = 1, \ldots, \mathcal{T}. \tag{3.15h}$$

Input and Output of the Scenario Problem

A scenario corresponds to certain given conditions, which mathematically are described by a number of scenario parameters. The designation scenario parameter might be somewhat confusing, since the scenario parameters actually are random variables, although their probability distributions are known (cf. section 1.1). When generating a scenario, an outcome is randomised for each scenario parameter. Then the scenario problem can be formulated according to the general description above, while all constants of the scenario problem are given by the outcome of a particular scenario parameter. The final result is a deterministic optimisation problem.

Even though all constants of the scenario problem thus are scenario parameters, it might be observed that some scenario parameters hardly are random variables, as they gain the same value in all scenarios. I refer to this kind of scenario parameters as model constants. To this category belongs for example the area division. In most cases also the benefit and cost functions are assumed to be the same for all scenarios. The boundary between “real” scenario parameters and model constants is thus something that at the end of the day is decided by the designer of the multi-area model.

The behaviour of the electricity market in a given scenario is described by a number of result variables, which are random variables with unknown probability distributions. When the scenario problem has been solved, the optimal values of the variables in the scenario problem show how the market behaves; hence, the solution to the scenario problem is a set of outcomes of the result variables. All in all, these result variables give a detailed picture of the behaviour of the electricity market, but in many cases we prefer to study more general result variables, which can be introduced by defining new result variables as functions of the old ones. Such important general result variables are for example the following:

> ***Definition 3.2.*** The total operation cost, *TOC*,[29] is the sum of the operation cost in all power plants during the entire scenario, i.e.,
>
> $$TOC = \sum_{t=1}^{\mathcal{T}} T_t \sum_{g \in \mathcal{G}} C_{Gg,t}(G_{g,t}).$$
>
> ***Definition 3.3.*** The result variable *LOLO*[30] indicates if it has been necessary to disconnect any consumers in the system, i.e., whether or not there has been a power deficit. For a single period we get
>
> $$LOLO_t = \begin{cases} 0 & \text{if } \sum_{c \in \mathcal{C}_D} U_{c,t} = 0, \\ 1 & \text{if } \sum_{c \in \mathcal{C}_D} U_{c,t} > 0. \end{cases}$$
>
> Thus, *LOLO* is a binary variable in short scenarios. In long scenarios, *LOLO* becomes a weighted average of the $LOLO_t$ for each time period:

29. This footnote has been added to maintain the footnote numbering in pace with the Swedish edition.
30. Loss Of Load Occasion.

$$LOLO = \frac{\sum_{t=1}^{I} LOLO_t T_t}{\sum_{t=1}^{T} T_t}.$$

Definition 3.4. The energy not served, *ENS*,[31] is the mean of the load which could not be supplied due to power deficit, i.e.,

$$ENS = \frac{\sum_{t=1}^{I} \left(T_t \cdot \sum_{c \in C_D} U_{c,\,t} \right)}{\sum_{t=1}^{T} T_t}.$$

Additional result variables which are relevant to a particular simulation (for example individual loss of load variables in each area) may be defined in a similar manner.

As stated in section 1.1, the objective of the simulation is to determine the value of a number of system indices, which can be used to compare different options. The most common system indices are in a mathematical sense the expectation value of some result variables. Some examples of system indices and the underlying result variables are listed in table 3.1. For the remainder of this dissertation I will focus on the first two system indices, *ETOC* (the

Table 3.1 *Some interesting system indices in a multi-area model.*

Result variable	System index	Unit	Interpretation
TOC	$ETOC = E[TOC]$	¤/h	Expected operation cost.
LOLO	$LOLP = E[LOLO]$	%	Risk of power deficit.
ENS	$EENS = 8\,760 \cdot E[ENS]$	MWh/year	Unserved energy.
E_{tot}[a]	$EE = 8\,760 \cdot E[E_{tot}]$	ton/year	Expected emissions
G_g	$EG_g = E[G_g]$	MWh/h	Expected generation
$P_{n,\,m}$	$EP_{n,\,m} = E[P_{n,\,m}]$	MWh/h	Expected transmission
L_{tot}	$EL = E[L_{tot}]$	MWh/h	Expected losses

a. In each thermal power plant we assume that there is an emission function $E_{Gg}(G_g)$ (cf. [75]); E_{tot} is the sum of the emissions of all power plants. It is possible to define separate emission functions for different hazardous substances—carbon dioxide,sulphur dioxide,nitrogen oxides, etc.—but it may also be preferable to weight all emissions into a single emission index.

31. This footnote has been added to maintain the footnote numbering in pace with the Swedish edition.

expected total operation cost) and *LOLP* (the loss of load probability, i.e., the risk that some consumer involuntarily has to reduce his or her consumption.).

The attentive reader may now ask if the most important system index—the benefit to the society, which is to be maximised by the ideal electricity market—has disappeared from the analysis. This is not the case, but I have taken the liberty of forestalling the difficulty to determine a clear-cut definition of the benefit to the society in a real electricity market. In a completely ideal electricity market, we assumed that all benefits and all costs can be measured and summarised into a total surplus, *TS*. In practice, this works only for small fictitious examples (such as those I use in chapters 4-6), where the numerical values can be chosen rather than determined. For example, in reality the social cost of disconnected load is very difficult to determine and the same is—even more—valid to the costs of environmental damage due to different emissions.

Thus, when simulating real electricity markets we are forced to replace *TS* by several system indices, where those costs which are hard to estimate are separated from directly quantifiable benefits and costs.[32] In this way, the simulation results become more transparent, which should make it easier for the decision makers who ultimately will consider the results.[33]

32. In general I assume that the load is price insensitive, which means that the benefit of consumption is infinite. Maximizing the total benefit minus the total cost is then equivalent to minimizing the cost; the lower the operation cost, the higher the total surplus. It is by the way also possible to use *ETOC* as a measure of the directly quantifiable costs when the load is price sensitive, because a price sensitive load can be divided in a price sensitive part and one or more fictitious power plants—in this case the generation cost of the fictitious power plants should be included in *TOC*.
33. Assume that two options have been simulated, and in the first case the result was *ETOC* = 1 000 M¤/year, *LOLP* = 0.09% and considerable emissions of green house gases, while in the other case the result was *ETOC* = 1 100 M¤/year, *LOLP* = 0.08% and almost zero emissions. In which case is the public welfare the largest? To answer this question, we have to value the risk of power deficit and the environmental impact, and this should clearly be indicated in the data presented to the decision-maker.

Chapter 4
ELECTRICITY MARKETS AND ENVIRONMENT

Today, fossil fuels account for a major share of the world's energy supply.[1] Meanwhile, combustion of fossil fuels causes some of the most serious environmental problems of our time, for example acidification and global warming [72, 83]. If we should achieve the vision of a sustainable development—i.e., "a development that meets the needs of the present without compromising the ability of future generations to meet their own needs" [79]—a changeover to renewable resources is necessary. This changeover will have a huge impact on the electricity markets of the world, because two thirds of all electricity is generated from fossil fuels [88] and this share has to be significantly reduced. Besides, electricity consumption might increase as a consequence of phasing out fossil fuels in other sectors, for example if petrol cars are replaced by electric cars, where the fuel is produced by electrolysis of water.

To create a sustainable development, the players of the electricity market must consider the real costs of electricity generation, i.e., the external costs caused by damage to our common environment must be internalised. In this chapter I will describe different methods to internalise environmental costs and how these methods can be applied to electricity markets. I will focus on such methods which actually are used in real electricity markets. However, I have chosen not to go into details of some methods which might work, but in reality are not used.[2]

1. In the year of 2001 almost 80% of the world total primary energy supply was from fossil fuels [88].
2. An example of such a method which I do not intend to study any further is algorithms for simultaneously minimizing the generation cost and emissions (see for example [74, 77]). In a centralised electricity market it would be possible to require that the production bids not only state a quantity and price, but also the size of the generation will cause. The central power pool could then consider both generation cost and emissions when deciding which bids to accept. Anyway, as far as I know, such a scheme has never been applied in reality.

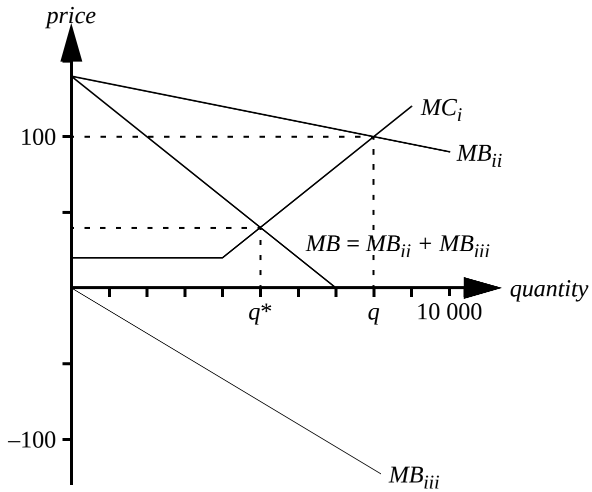


Figure 4.1 *Example of external costs. Consider a market where there are three groups of players: producers (i), consumers (ii) and others (iii), The marginal cost of the producers is MC_i and the marginal benefit function of the consumers is MB_{ii}. Unfortunately, consumption of the good of this market causes damage to the other players, which therefore have the marginal benefit function MB_{iii}. The total marginal benefit is $MB = MB_{ii} + MB_{iii}$. The intersection between MB and MC_i yields the turnover which maximises the total surplus, $q^* = 5\,000$, but if the consumers only consider their own marginal benefit, MB_{ii}, the turnover becomes $q = 8\,000$ instead.*

We can identify two main groups of solutions to problems regarding externalities: private responses (when the involved players by themselves try to manage the situation) or government responses (when the authorities somehow regulate activities causing externalities).[3] The first two sections of this chapter deal with different variants of these two options. The chapter is ended by a discussion of the credibility for different solutions.

4.1 PRIVATE RESPONSE

It is often difficult for a market economy to manage externalities, because the consumers do not get correct price signals of the social cost of their consumption (cf. figure 4.1). Being difficult is however not the same thing as being impossible, and sometimes the market may by its own find a solution which is maximizing the benefit of the society, even when there are externalities present. In this section I will investigate whether private responses can manage environmental problems caused by electricity generation.

3. Cf. [137], section 18.2.

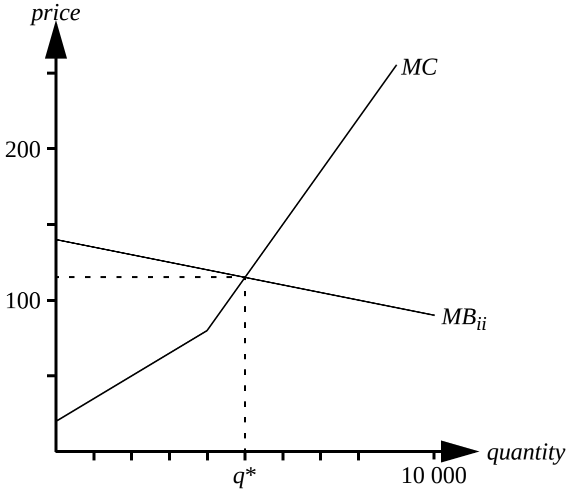


Figure 4.2 *Equilibrium after merging producers and other players. The marginal cost curve of the merged firm is* $MC = MC_i - MB_{iii}$ *and the producers's surplus is about* 381 667*, which can be compared to the surplus* 480 000 *for the producer and* –213 333 *for the others, if they do not merge (cf. figure 4.1). Thus, the sum of the surplus is higher in the merged firm.*

Negotiations and Mergers

The third party who is affected by an external cost does not have to sit silently and accept the situations, but can take actions to increase the own benefit. One way to internalise an external cost is that the party who suffers damage merge with the producers[4] and form a single firm, in which the negative benefit of the third party will then be clearly visible in the accounts. An example of a merger is shown in figure 4.2.

Unfortunately, it is not always easy to merge companies. A wide range of practical difficulties may impede mergers. I will not investigate those any further, but just conclude that a firm internalizing an environmental problem as global warming would have to include more or less all the electricity producers and farmers of the world, as well as a few other industries—in other words, not a very realistic solution.

An alternative to players merging is that they negotiate. A simple example is two persons, A and B, sitting in the same room. A does not like cigarette smoke and values the decreased benefit if B smokes a cigarette to 10 ¤, whereas B values the benefit of smoking a cigarette to 5 ¤. If B takes out a

4. Or take over the company, if the producer is not interested in a merger.

cigarette, it would be possible for A to offer him or her for example 6 ¤ to refrain from lighting the cigarette. The result should be beneficial to both parties compared to if B actually smoked the cigarette, as A pays 6 ¤ to avoid a "cost" of 10 ¤ and B receives 6 ¤ instead of an "income" of 5 ¤.

Although negotiations have the possibility to direct the resource utilisation in such a manner that the benefit to the society is maximised even in presence of externalities, there are plenty of situations where negotiations will not solve the problem or cannot be performed at all.[5] Firstly, there are moral issues to be discussed—is it really righteous that the one subject to a pollution should have to pay to avoid the external cost or is it the polluter who should pay? Secondly, it must be possible to identify the opponent if any negotiations are to take place. A forest owner whose trees have been damaged by acidification cannot possibly identify which of all possible polluters it is who has damaged the trees. If several forest owners try to avoid this problem by consolidating and reaching a bargain so that the total emission of acidifying pollutants diminishes, then there will be a free-rider problem, as there are costs of organizing negotiations of this size. The forest owner who does not contribute to the negotiations will nevertheless obtain the benefits of reduced emissions that are the results of the negotiations. Finally, incomplete information may cause the results of the negotiations to become suboptimal or—in worst case—the negotiations may already in advance be condemned futile and never come off.

Electricity Disclosure

Externalities arise when consumers only compare their own marginal benefits to the direct marginal costs of the producers, without regarding how any third parties are affected. It should however not be taken for granted that neither consumers nor producers behave in this way, because moral values can make us voluntarily internalise external costs. An example of how moral values affect our behaviour is social conventions. Consider the example in [137] how children are taught not to litter, which in economic terms can be described as the individual cost of taking the litter to the closest dustbin is less than the external costs implied of all neighbours if the trash is just dumped in the nature; hence, using the dustbins instead is beneficial to the society.

The environmental debate of the last decades has made many consumers willing to pay a higher price for goods if that avoids or at least decreases the external costs. It is therefore reasonable to assume that if consumers in an electricity market are offered the possibility to choose which kind of electricity they want to buy, then the demand for electricity generated in environmen-

5. See [137], p. 605f.

tally benign power plant will increase.[6]

Let us assume that there in an electricity market are E different electricity products, where one is “grey power” (i.e., electricity from unspecified power plants) and the other correspond to a particular mix of different power sources.[7] There is some player who guarantees that the system is supplied *at least* as much energy from a certain power plant as the consumers of the disclosed electricity products have demanded. The surplus, i.e., the energy which was not specifically demanded by a consumers, is included in the grey power.

The balance between production and consumption of an electricity product is valid for a certain accounting period, which can be chosen arbitrarily. The shorter the accounting period, the harder it will become for non-dispatchable power plants to sell disclosed power. It is common to choose quite a long accounting period, for example a year. If a scenario comprises the periods $1, \ldots, T$ then the scenario is divided in A accounting periods $A_1 = \{1, \ldots, T_1\}$, $A_2 = \{T_1 + 1, \ldots, T_2\}$, …, $A_A = \{T_{A-1} + 1, \ldots, T\}$.[8]

We now introduce the set M_e to denote the equivalent power plants which contribute to electricity product e. The requirement to produce sufficient amounts in these power plants is modelled as an extra constraint to the scenario problem. To make this constraint at least somewhat more readable, we neglect that a particular electricity product might require a certain mix of the power sources in question.[9] We then get the following constraint:

$$\sum_{t \in A_a} T_t \left(\sum_{n \in M_e} W_{n,t} + \sum_{r \in M_e} H_{r,t} + \sum_{g \in M_e} G_{g,t} \right) \geq \sum_{t \in A_a} \sum_{c \in C_e} T_t \Delta_{c,t}, \quad \forall\, e = 1, \ldots, E,\ a = 1, \ldots A, \tag{4.1}$$

where C_e are the equivalent consumers who demand electricity product e.

Eco-labelling

It is difficult for an individual consumer to collect enough information to be able to account for the environmental impact of a product in a purchase decision. A common solution is therefore that a larger organisation (for example an international body like the Nordic Council or the EU, or private associa-

6. Cf. [73].
7. However, some claim that such freedom of choice fools the consumers, because for example electricity from wind power plants cannot be separated from coal power. This kind of objections are treated in section 4.3.
8. The most simple choice is of course $A = 1$ and $A_1 = \{1, \ldots, T\}$, i.e., each scenario comprises exactly one accounting period.
9. It would for example be possible to offer the consumers an electricity product consisting of 80% hydro power and 20% wind power.

tions as the Swedish SNF and KRAV) collects environmental information about different products and conveys this knowledge in form of different kind of eco-labels. Another possibility is of course to have a legislation that states which information a producer must supply about the products.[10] However, from the consumers point of view, it is probably more comfortable to let a few easily identified labels control your purchases, instead of scrutinizing the declarations of contents whenever you go shopping.

Eco-labelling of electricity works in the same manner as disclosure of electricity, but there will be just two electricity products—grey power[11] and eco-labelled power—instead of E products.

4.2 GOVERNMENTAL RESPONSES

If a market economy cannot cope with externalities, public authorities can intervene and establish rules that force the players of the market to utilise the resources in such a manner that the benefit to the society is maximised. Regardless of how the intervention is done, it is required that perfect information is available, if the benefit to the society should be maximised. The responsible authority must know which external costs there are and their origin. In practice, this information can be very hard to collect, because the relation between human activities and environmental damages is complex. It might be hard to value the costs of an environmental damage and in some cases it might also take a long time before the environmental damage even can be confirmed by science—and then it might already be too late to take actions.

With perfect information, different rules will achieve the same resource allocation and consequently the same total surplus (the surplus of individual players may however vary depending on the chosen solution). Without perfect information, it is likely that the consequences of different rules will differ, because the information need varies depending on the chosen solution. Rules which are based on information which can be estimated with reasonable accuracy have better basic conditions to become efficient than rules, which require assumptions about large amounts of uncertain data.

It is not necessary that all parameters controlling the environment rules must be kept constant, but it is possible for the legislator to change the rules if new information about the external costs becomes available. It is also reasonable that the rules are quite simple to fulfil by the way of introduction, and

10. Compare to the legislation in the U.S., where all provisions are provided with detailed nutrition facts.
11. In Sweden some people use the phrase “*ful-el*”, which is a language construction akin to “*ful-öl*” (bad beer). I guess an English counterpart could be “bad power”.

then gradually the requirements are increased until the objective of a sustainable power system (from environmental point of view) is achieved. It is however extraordinary important that there is a clear plan how to achieve the objective and that the authorities really support the system. If someone should be interested in investing in environmental friendly power generation, which is depending on a particular rule, then it must be possible to rely that this rule will last long enough for the investment to pay back.

4.2.1 Restrictions

A simple method to regulate external costs due to environmental impacts is to restrict the activities causing environmental damage. A restriction can be expressed in many ways, for example as a total ban of certain power sources, a limitation of the number of power plants or as a cap on the external cost each power plant may cause.

Prohibition

A total ban is justified if a certain power source causes external costs which are completely unacceptable. In extreme cases it is obvious that it is beneficial to the society if a particular activity is banned,[12] but in many cases it is difficult to determine what is right and what is wrong. This can partly be caused by incomplete information, but sometimes the problem might be that the environmental impact is hard to value, making it difficult to unambiguously define the social benefit (cf. section 3.1.1). An example of such a difficult judgement is nuclear power—some countries allow nuclear power, whereas in others nuclear power plants are forbidden due to the external costs in the shape of environmentally hazardous waste and the risk of severe accidents.

When the external costs are significant, but not entirely unacceptable, a limit to the number of power plants can be imposed instead of a total ban. An example of this is the Swedish ban on exploiting hydro power in the four rivers Vindelälven, Pite älv, Kalix älv and Torne-Muonio älv.

In those cases when valuation issues determine the size of the external cost, it can be argued that if a majority of the population supports a ban then it is reasonable—but far from certain—that the ban is beneficial to the society. For example, the external cost of exploiting the Swedish rivers would primarily be the loss of unspoiled countryside and such a cost is extremely hard to value—at the end of the day it is about the subjective judgement of each indi-

12. Who would for example be willing to allow nuclear waste to be left on the closest refuse dump?

vidual. As there is stable political majority supporting the decision not to exploit the rivers, it seems reasonable to assume that the benefit of not exploiting the rivers as perceived by the Swedish people, is larger than the value of a hydro power expansion. If conditions would change and the value of building new hydro power plants increases, it is possible to make a democratic decision to abolish the prohibition. Thus, it seems reasonable that this decision actually maximises the benefit to the society, both in the short and the long run.

There are however also examples where it is considerably more difficult to determine whether or not a democratic decision will maximise the benefit to the society. Assume for example that a solid majority in a country thinks that electricity generation in nuclear power plants should be allowed, but in each part of the country there is a just as solid majority opposing the storage of nuclear waste in the vicinity. How should the social benefits be maximised in this country, by allowing nuclear power and forcing some part of the country to receive the waste, or by banning nuclear power?

If it is hard to determine what is optimal concerning prohibitions, it is at least simple to include bans in an electricity market simulation. If a certain power source is prohibited then the supply side of the electricity market is affected, but otherwise the market operates as usual. Therefore, the same scenario problem can be used as in an ideal electricity market.

Emission Caps

Prohibitions can be formulated more precisely for such externalities which are easier to measure. It is possible to grant the power plants concessions for certain emissions, for example by allowing a power plant to emit at most a specified amount of pollution per MWh generated. Another variant is to put a limit on the emissions from a particular facility during a year, regardless of how much is produces. The latter alternative makes it more profitable to invest in emission reductions, since it then becomes possible to increase the electricity generation.

This kind of emission caps are quite simple to administrate, but the legislator requires quite a lot of information, which might be hard to obtain, to direct the electricity market towards a resource allocation maximizing the total surplus. It must be known how large the total emissions can be without causing external costs higher than the benefit of the generated electricity. Moreover, it must be known how much electric energy is produced in total in order to set the limit of the emissions per MWh. Another complication is that the cost of investing in reduced emissions might vary from power plant to power plant, i.e., the cost of the same total level of emissions varies depending on in which power plants measures are taken. To minimise the cost, the authorities must basically set separate emission caps in each power plant. Finally, all caps must be updated every time the conditions of the electricity market changes, for example if new power plants are built.

Thus, it can be very hard to maximise the benefit to the society using emission caps. However, this kind of rules should not be totally condemned; if there is for example an immediate environmental problem and quick actions have to be taken, emission caps can be the action which are most easy to implement and which are the fastest to show any results.

The modelling of emission caps is not very complicated. If the emissions are limited per MWh generated then only the supply side of the electricity market is affected, but not the electricity market model. If the emission cap is a cap to the emissions of a particular facility then special emission constraints have to be added to the scenario problem. Assume that period $t, \ldots, \mathcal{T}$ is divided into a number of accounting periods in the same manner as for disclosed electricity (see page 61) and that during each accounting period the permitted emissions are $\bar{E}_{Gg,a}$ for each equivalent power plant. In the equivalent power plant g and accounting period a we get the emission constraint

$$\sum_{t \in \mathcal{A}_a} E_{Gg,t}(G_{g,t}) - \Psi_{g,a} \le \bar{E}_{Gg,a}, \tag{4.2}$$

where $E_{Gg,t}(G_{g,t})$ is an emission function, $\mathcal{A}_a$ is the set of scenario time periods belonging to accounting period a, and $\Psi_{g,a}$ are the excess emissions. Any power plant who exceeds its emission cap is subject to some form of sanction, which is represented by a cost function $C_{\Psi g,a}(\Psi_{g,a})$. The total sanction cost should be included in the objective function:

$$\sum_{a=1}^{A} \sum_{g \in \mathcal{G}} C_{\Psi g,a}(\Psi_{g,a}), \tag{4.3}$$

where A is the number of accounting periods of the scenario.

Tradable Emission Rights

A somewhat more advanced form of limitation is when the legislator only determines the total emissions which can be accepted without causing too much environmental harm. Then, the market itself determines how the permitted emissions should be divided by the generating units. This is done by introducing emission rights, which give the owner permission to let out a certain amount of a certain substance during a certain accounting period, for example one ton carbon dioxide per year. The emission rights are tradable as any other good. By this means, the legislator does not need to determine where the environment investments should produce the largest benefit—these decisions are transferred to the producers themselves. If power plant A wants to increase the electricity generation, but lacks sufficient emission rights, they can either reduce the emissions per MWh in the own facility or buy emissions rights of the owner to power plant B, if it would be less expensive to accomplish the corresponding emission reduction in that power plant instead.

When introducing a system of tradable emission rights, it must somehow be decided how the emission rights initially should be distributed, which can be done in two ways. The most simple is that the government owns all emissions rights at the start, and that they are auctioned to the market. The other method is to distribute the emission rights by "grandfathering",[13] which means that the players are granted emission rights proportional to their emissions during the last years. The total surplus is not affected by the choice between these two methods, but the distribution is more important to individual players. Those who are granted less emission rights than they need are subject to an extra cost for either reducing emissions or for buying more emission rights. Correspondingly, a surplus of emission rights means an extra income. If the emission rights are sold by auction, all players are subject to extra costs, while the government makes a nice profit. Grandfathering has very different consequences for the players—for example, those players who have already performed environment investments are disadvantaged—and it can be hard to agree on what is a fair distribution of the emission rights.

Anyone emitting environmentally hazardous substances without having the corresponding emission rights is subject to a penalty fee. The level of this penalty fee is of course important to make the system work well; if the penalty fee is too low then there will be a risk that too many players choose not to obtain emission rights and just pay their penalties instead. Hence, the penalty fee constitutes a price cap to the emission rights, but this can be avoided by repaying the collected penalty fees to those players who possess the emission rights; the value of an emission right is then equal to the value of not paying the penalty fee plus an expected refund.

Today, there are systems for trading SO_2- and NO_x-emissions in some states of the U.S. According to [85], the system has resulted in considerable emission reductions from the electricity generation and the costs have been significantly less than predicted. The EU is planning to introduce tradable emission rights for greenhouse gases starting 2005 [87].

Let us now consider how to include tradable emission rights in the electricity market model. The largest challenge is the pricing of the emission rights market, because the price is not just depending on the electricity market, but also on the needs of other industries to obtain emission rights.[14] The models described in [76, 78] do not consider this possibility; the trading of emission

13. This footnote has been added to maintain the footnote numbering in pace with the Swedish edition.

14. It is of course conceivable that the tradable emission rights only concerns electricity producers, but that is for example not the case with the EU proposal [87].

rights is restricted to players within the electricity market. Here, I will suggest a basic model, which also includes players outside the electricity market. In the basic model it is assumed that there is perfect competition in the emission rights market and that the players have sufficient information to know how many emission rights they should buy. To simplify the derivation of the model I use an electricity market, which is not divided in areas and each scenario comprises one time period coinciding with the accounting period of the emission rights.[15]

We start by studying a producer problem. The objective function is to maximise the income minus the production cost of the electricity as well as the cost of the emissions. The constraint is that the emission should be equal to the purchased emission rights, Λ_g, plus the excess emission subject to penalty fees, Ψ_g. Thus, the producer problem reads as follows:

$$\text{maximise} \quad \lambda G_g - C_{Gg}(G_g) - \zeta\Lambda_g - \beta_\Psi \Psi_g \tag{4.4}$$

$$\text{subject to} \quad E_{Gg}(G_g) - \Lambda_g - \Psi_g = 0, \tag{4.4a}$$

$$0 \le G_g \le \bar{G}_g, \tag{4.4b}$$

$$0 \le \Lambda_g, \tag{4.4c}$$

$$0 \le \Psi_g. \tag{4.4d}$$

The electricity price λ and the emission right price ζ are given, because the producer is considered a price-taker. The penalty fee of excess emissions, β_Ψ, is defined by the authorities.

The consumers behave exactly as in an ideal electricity market, i.e., they try to maximise their surplus at a given electricity price:

$$\text{maximise} \quad B_{\Delta c}(\Delta_c) - \lambda\Delta_c \tag{4.5}$$

$$\text{subject to} \quad 0 \le \Delta_c \le \bar{\Delta}_c. \tag{4.5a}$$

The surrounding world is treated as a single player, whose objective function is to maximise the benefit of the purchased emission rights. This benefit could be considered a function of both the number of purchased emission rights and the electricity price; when electricity prices are low the production cost of some industries is reduced, which might result in increased output and a larger benefit of obtaining emission rights. Let us neglect this kind of complications in this example, so that the player problem of the surrounding world is simplified to

15. Nothing essential in the following reasoning would change if these simplifications were not done. The difference is that it would be necessary to summarise the emission over each accounting period in the same manner as in (4.2). The formulae would then be harder to read, which I find unnecessary when the model is outlined.

$$\text{maximise} \quad B_{\Theta}(\Theta) - \zeta\Theta \tag{4.6}$$

$$\text{subject to} \quad 0 \le \Theta. \tag{4.6a}$$

For the market as a whole we have that the production and consumption of electric energy must be in balance, and the total number of purchased emission rights may not exceed the emission cap, $\bar{E}$. Together with the optimality conditions of the player problems (4.4)-(4.6) we get a system of equations and inequalities defining how the players will behave:[16]

$$\sum_{c \in C} \Delta_c - \sum_{g \in G} G_g = 0, \tag{4.7a}$$

$$\sum_{g \in G} \Lambda_g + \Theta = \bar{E}, \tag{4.7b}$$

$$MC_{Gg}(G_g) + \iota_g ME_{Gg}(G_g) \ge \lambda, \qquad \text{if } G_g = 0, \tag{4.7c}$$

$$MC_{Gg}(G_g) + \iota_g ME_{Gg}(G_g) = \lambda, \qquad \text{if } 0 < G_g < \bar{G}_g, \tag{4.7d}$$

$$MC_{Gg}(G_g) + \iota_g ME_{Gg}(G_g) \le \lambda, \qquad \text{if } G_g = \bar{G}_g, \tag{4.7e}$$

$$\zeta \ge \iota_g, \qquad \text{if } \Lambda_g = 0, \tag{4.7f}$$

$$\zeta = \iota_g, \qquad \text{if } \Lambda_g > 0, \tag{4.7g}$$

$$\beta_{\Psi} \ge \iota_g, \qquad \text{if } \Psi_g = 0, \tag{4.7h}$$

$$\beta_{\Psi} = \iota_g, \qquad \text{if } \Psi_g > 0, \tag{4.7i}$$

$$MB_{\Delta c}(\Delta_c) \le \lambda, \qquad \text{if } \Delta_c = 0, \tag{4.7j}$$

$$MB_{\Delta c}(\Delta_c) = \lambda, \qquad \text{if } 0 \le \Delta_c \le \bar{\Delta}_c, \tag{4.7k}$$

$$MB_{\Delta c}(\Delta_c) \ge \lambda, \qquad \text{if } \Delta_c = \bar{\Delta}_c. \tag{4.7l}$$

$$MB_{\Delta c}(\Delta_c) \le \zeta, \qquad \text{if } \Theta = 0, \tag{4.7m}$$

$$MB_{\Delta c}(\Delta_c) = \zeta, \qquad \text{if } \Theta > 0. \tag{4.7n}$$

The interpretation of these constraints is that the producers in their marginal cost function must count both the marginal cost of the electricity generation and the emissions.[17] The latter is set by the individual emission right value of the producer, ι_g (dual variable of the constraint in the producer problem), and the marginal emissions. The emission right value is normally equal to the

16. In this system of equations and inequalities we also have the constraints and the variable limits of (4.4)-(4.6), but I have chosen not to repeat them, to at least somewhat shorten the enumeration of the optimality conditions. Besides, it can be noted that the conditions (4.7a)-(4.7n) correspond to the optimality conditions of an aggregated optimisation problem, which in this case is a simple quadratic programming problem.

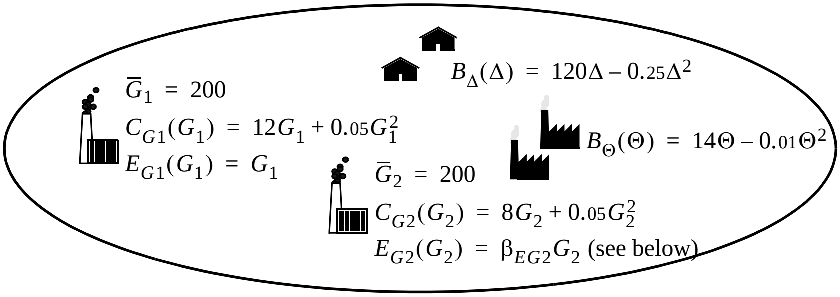
market price of emission rights, which in its turn is determined by the surrounding world. If the demand for emission rights should become large enough to force somebody to pay penalty fees then the market price is equal to the penalty fee, as indicated by (4.7i).

In figure 4.3 I provide some example scenarios showing how an electricity market is affected by tradable emission rights. Power plant 1 has higher marginal production costs than power plant 2, but the emissions per MWh are on the other hand lower. Regardless of the size of the emission cap, producer 2 can increase the surplus by investing in equipment reducing the emissions, but the investment is of course only profitable if the surplus increase is larger than the investment cost. The profitability of the investment depends on the price of emission rights, which in its turn depends on both the total emission cap and the demand of the surrounding world. The emission right trading will thus have dynamic effects on both the electricity market and other markets. Anyone who wants to study the consequences of emission right trading in detail will thus have to collect quite a lot of data.

The total surplus specified in figure 4.3 does not include the external cost. In table 4.1 I have calculated the total surplus for some different values of the true marginal damage caused by the emissions, *MD*. For comparison, I have also calculated the total surplus in an ideal electricity market, where *MD* is internalised in the production costs of the power plants. The ideal electricity market is as expected always more efficient, but the difference is sometimes small. Notice that the efficiency of different emission caps vary depending on the environmental damage caused by the emissions. When *MD* is low, the higher emission cap provides better results, while the lower emission cap apparently limits the emissions too much. For higher values of *MD*, the lower emission cap is preferable, whereas the higher allows too large emissions.

I have no intention to model market dynamics in this dissertation, but let us nevertheless have a look at the willingness to invest in the ideal electricity market and the non-ideal electricity market with tradable emission rights respectively. The ideal electricity market maximises the benefit to the society, i.e., an investment is carried out if it results in an increase of the total surplus which is larger than the investment cost. In table 4.2 it is shown how the total surplus in an ideal electricity market is affected if the emissions of power plant 2 are reduced from 3 ton/MWh to 2 ton/MWh. The larger the marginal damage caused by the emissions, the more profitable the investment becomes. In the non-ideal electricity market the investment is carried out if the cost for producer 2 is less than the producer's increased surplus. In the

17. For comparison we may observe that the cost of the emissions do not have to be the same as the cost of emissions in an ideal electricity market. Here the marginal cost of emissions is equal to $\iota_g ME_{Gg}(G_g)$, whereas in an ideal electricity market it would be $MD(E_{tot}) \cdot ME_{Gg}(G_g)$, where $MD(E_{tot})$ is the marginal damage of the total emissions, i.e., the marginal external cost.



Scenario parameters				
Emission cap, $\bar{E}$ [ton]	1 000	1 000	800	800
Emissions in power plant 2, β_{EG2} [ton/MWh]	3	2	3	2
Electricity price, λ [¤/MWh]	23.29	20.46	28.00	25.08
Emission right price, ζ [¤/ton]	1.81	0.34	4.40	3.72
Consumers				
Consumption, Δ [MWh]	193.41	199.08	184.00	189.85
Value of consumption, $B_\Delta(\Delta)$ [¤]	13 857.38	13 981.32	13 616.00	13 771.15
Purchase cost, $\lambda\cdot\Delta$ [¤]	4 505.35	4 073.41	5 152.00	4 760.75
Surplus, CS [¤]	9 352.03	9 907.91	8 464.00	9 011.39
Producer 1				
Electricity generation, G_1 [MWh]	94.82	81.23	116.00	93.54
Electricity income, $\lambda\cdot G_1$ [¤]	2 208.83	1 662.10	3 248.00	2 345.66
Generation cost, $C_{G1}(G_1)$ [¤]	1 587.46	1 304.69	2 064.80	1 559.93
Emission right cost, $\zeta\cdot E_{G1}(G_1)$ [¤]	171.80	27.50	510.40	348.25
Surplus, PS_1 [¤]	449.57	329.92	672.80	437.47
Producer 2				
Electricity generation, G_2 [MWh]	98.59	117.85	68.00	96.31
Electricity income, $\lambda\cdot G_2$ [¤]	2 296.52	2 411.31	1 904.00	2 415.10
Generation cost, $C_{G2}(G_2)$ [¤]	1 274.69	1 637.16	775.20	1 234.22
Emission right cost, $\zeta\cdot E_{G2}(G_2)$ [¤]	535.87	79.78	897.60	717.13
Surplus, PS_2 [¤]	485.97	694.37	231.20	463.75
Surrounding world				
Emissions, Θ [ton]	609.42	683.08	480.00	513.84
Total emissions of the electricity market, $E_{G1}(G_1) + E_{G2}(G_2)$ [ton]	390.59	310.27	320.00	286.15
Total surplus, $B_\Delta(\Delta) - C_{G1}(G_1) - C_{G2}(G_2)$ [¤]	10 995.24	11 039.48	10 776.00	10 979.63

Figure 4.3 *Example of tradable emission rights. The first producer has a slightly higher generation cost, but lower emissions than the second producer. The ability to compete of the first producer is increased when the emission cap is reduced, because the emission rights become more expensive. The second producer can improve the ability to compete by investing in reduced emissions.*

Notice that the total surplus does not *include the costs caused by the emissions. Thus, it depends on the evaluation of the external costs, which solution is going to maximise the benefit to the society (cf. table 4.1).*

Table 4.1 *Short-term total surplus of the scenarios in figure 4.3.*

Emission cap, $\bar{E}$ [ton]	Emissions from power plant 2, β_{EG2} [ton/MWh]	True marginal damage, *MD* [¤/ton]				
		1	2	3	4	5
Ideal electricity market[a]	3	10 614	10 215	9 843	9 498	9 181
1 000	3	10 605	10 214	9 823	9 433	9 042
800	3	10 456	10 136	9 816	9 496	9 176
Ideal electricity market[a]	2	10 725	10 418	10 121	9 833	9 554
1 000	2	10 723	10 406	10 089	9 772	9 455
800	2	10 691	10 405	10 119	9 832	9 546

a. In the ideal electricity market there is no tradable emission rights, but the true marginal damage of emissions are included in the cost functions of the power plants.

Table 4.2 *Profitability of investing in reduced emissions.*

	True marginal damage, *MD* [¤/ton]				
	1	2	3	4	5
Increased surplus in an ideal electricity market[a]	111	204	278	335	373
Increased surplus of producer 2[b]					
Emission cap $\bar{E}$ = 1 000 ton	208	208	208	208	208
Emission cap $\bar{E}$ = 800 ton	233	233	233	233	233
Is the investment profitable if the cost is 200 ¤?					
Ideal electricity market	No	Yes	Yes	Yes	Yes
Emission cap $\bar{E}$ = 1 000 ton	Yes	Yes	Yes	Yes	Yes
Emission cap $\bar{E}$ = 800 ton	Yes	Yes	Yes	Yes	Yes
Is the investment profitable if the cost is 220 ¤?					
Ideal electricity market	No	No	Yes	Yes	Yes
Emission cap $\bar{E}$ = 1 000 ton	No	No	No	No	No
Emission cap $\bar{E}$ = 800 ton	Yes	Yes	Yes	Yes	Yes
Is the investment profitable if the cost is 240 ¤?					
Ideal electricity market	No	No	Yes	Yes	Yes
Emission cap $\bar{E}$ = 1 000 ton	No	No	No	No	No
Emission cap $\bar{E}$ = 800 ton	No	No	No	No	No

a. Determined by comparing the total surplus according to table 4.1 when the emissions in power plant 2 amount to 3 ton/MWh and 2 ton/MWh respectively.
b. Determined by comparing the total surplus of producer 2 according to figure 4.3 when the emissions in power plant 2 amount to 3 ton/MWh and 2 ton/MWh respectively.

table, it is shown how the surplus of producer 2 is changed. Notice that in this case the profitability is not depending on *MD*, but on the price of the emission rights, which indirectly can be controlled by the choice of emission cap. Therefore, in some cases the investment will be profitable in both the ideal and non-ideal electricity market, whereas in others the results will differ. The example illustrates the importance of choosing the right levels of the parameters controlling the trading (primarily the emission cap, but also the penalty fee of excess emissions), to achieve an efficient resource usage both in the short and the long run.

4.2.2 Fees and Subsidies

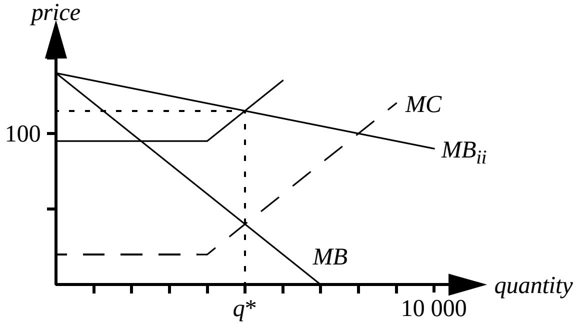
The problem of externalities is, as has already been pointed out, that the players of the market do not get correct economic signals about the consequences of their actions. By introducing fees and subsidies corresponding to the external costs and incomes respectively, it can be assured that the price signals reach the players. To maximise the benefit to the society, the fee must be set so that it corresponds to the external costs at the optimal turnover.[18] To find a correct value of the fee it is required that the supply, demand and externalities are known and can be valued. It should therefore be very hard for an authority to determine the exactly right fee; the best one can hope for is that they get it approximately right. Moreover, whenever a large change occurs in the market, the fees must be updated (cf. figure 4.4).

From a modelling point of view, fees and subsidies are not very difficult, because they directly modify the cost and benefit functions of the players. There are however also some more sophisticated methods to introduce fees and subsidies for externalities. These methods require a somewhat more detailed analysis, which follows below.

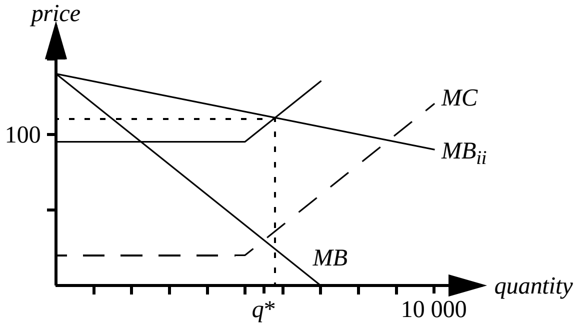
“Feebate” Systems

One way of avoiding the difficulty of determining exact values of the external costs when an environment fee is introduced, is to create a “feebate” system, i.e., a combination of fees and rebates. The fee is proportional to the amount of emissions during a certain accounting period, for example a year. These fees are however not just transferred to the public treasury, but are repaid to the producers according to their share of the total output. The parameters which must be decided by the authorities are then just the fee itself and the accounting period to be used. The system will then create a pressure for the

18. Correspondingly, subsidies should equal the marginal external costs at the optimal turnover.



a) The supply curve when the producers pay a fee of 75 ¤*/unit.*



b) The supply curve when new production capacity has been added, but the fee is still 75 ¤*/unit.*

Figure 4.4 *Consequences of introducing a fee per unit produced. The figures above show the marginal benefit of consumption (MB_{ii}) and the marginal benefits of the consumers and the third party (MB). In panel a the supply curve is the same as in the earlier examples (see figure 4.1), but for each unit produced the producer has to pay a fee of* 75 ¤*/unit. Thanks to the fee, the market will find an equilibrium where the turnover is maximizing the benefit to the society. It can be noted that the total income of the fees is larger than the total external cost.*

In panel b the supply curve has changed, as new production capacity with low marginal cost has been added. The new optimal turnover is 5 500 *units, but if the fee is not increased, the turnover becomes* 5 800 *units. Still, a slightly incorrect fee results is preferable compared to the turnover which would be the result if there was no fee at all.*

producers to reduce their emissions, because those producers who have less than average emissions per generated MWh will receive more than they have paid. They will in other words have a competitive advantage compared to those power plants who cause larger emissions; the higher the fee, the larger the advantage.

As long as the investment costs are low, a producer can make a good earning by reducing the emissions. Feebate systems therefore can have a large impact as they are introduced, but when the most cost efficient investments have been done, it becomes hard to improve any further. This is for example visible in the Swedish environment fee for emission of nitrogen oxides from power plants. The system was introduced in 1992 and during the first year the nitrogen oxide emissions per MWh decreased by 20%. Since then, the decrease rate has gradually declined; in total, the emissions per MWh decreased by 40% between 1992 and 2000 [146]. An advantage of the system is however that the incentive to reduce the emissions never disappears completely; a player who quickly incorporates new, inexpensive emission reducing technology before the competitors will always have an advantage. The risk is of course that this pressure to reduce emissions results in excessive investments compared to what is optimal for the society.

To derive a model of feebate systems I make the same simplifications as when I analysed tradable emission rights, i.e., I consider an electricity market having one area and one time period, which coincides with the accounting period.[19] A producer will try to maximise the income of the electricity market and the rebate minus the production cost and the fee. We then get the following producer problem:

$$\text{maximise} \quad \lambda G_g + \rho G_g - C_{Gg}(G_g) - \beta_E E_{Gg}(G_g), \tag{4.8}$$

$$\text{subject to} \quad 0 \le G_g \le \overline{G}_g, \tag{4.8a}$$

where β_E is the fee and ρ is the rebate per MWh generated.

The consumer problem is the same as in an ideal electricity market, i.e.,

$$\text{maximise} \quad B_{\Delta c}(\Delta_c) - \lambda \Delta_c, \tag{4.9}$$

$$\text{subject to} \quad 0 \le \Delta_c \le \overline{\Delta}_c. \tag{4.9a}$$

For the market as a whole applies that the production and consumption of electric energy must be in balance, and the total rebate should equal the total fees. Together with the optimality conditions of the player problems (4.8) and (4.9) we get a system of equations and inequalities defining how the players will behave:[20]

19. The motivation of these simplifications are—just as before—to make the formulae more readable; there are no technical obstacles of modelling multi-area problems or longer time periods.

$$\sum_{c \in C} \Delta_c - \sum_{g \in G} G_g = 0, \tag{4.10a}$$

$$\rho \sum_{g \in G} G_g = \beta_E \sum_{g \in G_g} E_{Gg}(G_g), \tag{4.10b}$$

$$MC_{Gg}(G_g) + \beta_E ME_{Gg}(G_g) - \rho \geq \lambda, \qquad \text{if } G_g = 0, \tag{4.10c}$$

$$MC_{Gg}(G_g) + \beta_E ME_{Gg}(G_g) - \rho = \lambda, \qquad \text{if } 0 \leq G_g \leq \overline{G}_g, \tag{4.10d}$$

$$MC_{Gg}(G_g) + \beta_E ME_{Gg}(G_g) - \rho \leq \lambda, \qquad \text{if } G_g = \overline{G}_g, \tag{4.10e}$$

$$MB_{\Delta c}(\Delta_c) \leq \lambda, \qquad \text{if } \Delta_c = 0, \tag{4.10f}$$

$$MB_{\Delta c}(\Delta_c) = \lambda, \qquad \text{if } 0 \leq \Delta_c \leq \overline{\Delta}c, \tag{4.10g}$$

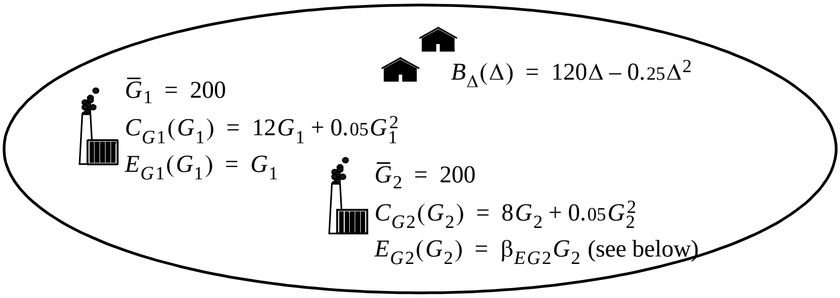
$$MB_{\Delta c}(\Delta_c) \geq \lambda, \qquad \text{if } \Delta_c = \overline{\Delta}c. \tag{4.10h}$$

These conditions say that the marginal cost of the producers is modified depending on how their marginal emissions relates to the competitors. Those who are below the average will have $\beta_E ME_g(G_g) < \rho$; at the end of the day, the consequence of the feebate system is thus a subsidy of these power plants. Those paying the subsidies are the power plants who cause larger marginal emissions. Again, we can notice that the marginal cost of the producers differ in this case compared to an ideal electricity market.[21]

In table 4.5 I provide some example scenarios illustrating the consequences of a feebate system in a small electricity market. Power plant 1 has higher marginal production costs than power plant 2, but the emissions per MWh are on the other hand lower. Regardless of the environment fee, producer 2 can increase the surplus by investing in equipment reducing the emissions, but the investment is of course only profitable if the surplus increase is larger than the investment cost. The higher the fee, the larger the surplus increase when reducing the emissions. Thus we can conclude that the size of the fee has a market dynamic effect, as it influences the willingness to invest.

The total surplus specified in figure 4.5 does not include the external costs. In table 4.3 I have calculated the total surplus for some different values of the true marginal damage caused by the emissions, *MD*. For comparison, I have also calculated the total surplus in an ideal electricity market, where *MD* is

20. In this system of equations and inequalities we also have the constraints and the variable limits of (4.8) and (4.9), but I have chosen not to repeat them, to at least somewhat shorten the enumeration of the optimality conditions. Besides, it can be mentioned that in this case I have not been able to aggregate the player problems into a single scenario problem, but that has no practical importance. We simply have to develop a special algorithm to find the solutions to (4.10a)-(4.10h). The NNP algorithm [7] may serve as an example of how to design such algorithms.
21. Cf. footnote 17.



Scenario parameters				
Environment fee, β_E [¤/ton]	2	2	4	4
Emissions in power plant 2, β_{EG2} [ton/MWh]	3	2	3	2
Electricity price, λ [¤/MWh]	20.00	19.91	20.73	20.00
Rebate, ρ [¤/MWh]	4.00	3.10	7.19	4.00
Consumers				
Consumption, Δ [MWh]	200.00	200.18	198.54	200.00
Value of consumption, $B_\Delta(\Delta)$ [¤]	14 000.00	14 003.59	13 970.27	14 000.00
Purchase cost, $\lambda\cdot\Delta$ [¤]	4 000.00	3 985.58	4 115.73	4 000.00
Surplus, CS [¤]	10 000.00	10 018.01	9 854.53	10 000.00
Producer 1				
Electricity generation, G_1 [MWh]	100.00	90.09	119.27	100.00
Electricity income, $\lambda\cdot G_1$ [¤]	2 000.00	1 793.69	2 472.47	2 000.00
Total rebate, $\rho\cdot G_1$ [¤]	400.00	279.27	858.04	600.00
Generation cost, $C_{G1}(G_1)$ [¤]	1 700.00	1 486.89	2 142.51	1 700.00
Total fee, $\beta_E\cdot E_{G1}(G_1)$ [¤]	200.00	180.18	477.08	400.00
Surplus, PS_1 [¤]	500.00	405.89	710.92	500.00
Producer 2				
Electricity generation, G_2 [MWh]	100.00	110.09	79.27	100.00
Electricity income, $\lambda\cdot G_2$ [¤]	2 000.00	2 191.89	1 643.27	2 000.00
Total rebate, $\rho\cdot G_2$ [¤]	400.00	341.27	570.28	600.00
Generation cost, $C_{G2}(G_2)$ [¤]	1 300.00	1 486.71	948.35	1 300.00
Total fee, $\beta_E\cdot E_{G2}(G_2)$ [¤]	600.00	440.36	951.24	800.00
Surplus, PS_2 [¤]	500.00	606.09	313.96	500.00
Total emission, $E_{G1}(G_1) + E_{G2}(G_2)$ [ton]	400.00	310.27	357.08	300.00
Total surplus, $B_\Delta(\Delta) - C_{G1}(G_1) - C_{G2}(G_2)$ [¤]	11 000.00	11 029.99	10 879.41	11 000.00

Figure 4.5 *Example of a feebate system. The first producer has a slightly higher generation cost, but lower emissions than the second producer. When the environment fee is increased by* 1 ¤/MWh, *the ability to compete of the first producer increases. The other producer can improve the ability to compete by investing in reduced emissions.*

Notice that the total surplus does not *include the costs caused by the emissions. Thus, it depends on the relation between the external costs and the environment fee,* β_E, *which solution is going to maximise the benefit to the society (cf. table 4.3).*

Table 4.3 *Short-term total surplus of the scenarios in figure 4.5.*

Environment fee, β_E [¤/ton]	Emissions from power plant 2, β_{EG2} [ton/MWh]	True marginal damage, *MD* [¤/ton] 1	2	3	4	5
Ideal electricity market[a]	3	10 614	10 215	9 843	9 498	9 181
2	3	10 600	10 200	9 800	9 400	9 000
4	3	10 522	10 165	9 808	9 451	9 094
Ideal electricity market[a]	2	10 725	10 418	10 121	9 833	9 554
2	2	10 720	10 409	10 099	9 789	9 479
4	2	10 700	10 400	10 100	9 800	9 500

a. In the ideal electricity market there is no tradable emission rights, but the true marginal damage of emissions are included in the cost functions of the power plants.

Table 4.4 *Profitability of investing in reduced emissions.*

	True marginal damage, *MD* [¤/ton] 1	2	3	4	5
Increased surplus in an ideal electricity market[a]	111	204	278	335	373
Increased surplus of producer 2[b]					
Environment fee β_E = 2 ¤/ton	106	106	106	106	106
Environment fee β_E = 4 ¤/ton	186	186	186	186	186
Is the investment profitable if the cost is 50 ¤?					
Ideal electricity market	Yes	Yes	Yes	Yes	Yes
Environment fee β_E = 2 ¤/ton	Yes	Yes	Yes	Yes	Yes
Environment fee β_E = 4 ¤/ton	Yes	Yes	Yes	Yes	Yes
Is the investment profitable if the cost is 150 ¤?					
Ideal electricity market	No	Yes	Yes	Yes	Yes
Environment fee β_E = 2 ¤/ton	No	No	No	No	No
Environment fee β_E = 4 ¤/ton	Yes	Yes	Yes	Yes	Yes
Is the investment profitable if the cost is 250 ¤?					
Ideal electricity market	No	No	Yes	Yes	Yes
Environment fee β_E = 2 ¤/ton	No	No	No	No	No
Environment fee β_E = 4 ¤/ton	No	No	No	No	No

a. Determined by comparing the total surplus according to table 4.3 when the emissions in power plant 2 amount to 3 ton/MWh and 2 ton/MWh respectively.
b. Determined by comparing the total surplus of producer 2 according to figure 4.5 when the emissions in power plant 2 amount to 3 ton/MWh and 2 ton/MWh respectively.

internalised in the production costs of the power plants. The ideal electricity market is as expected always more efficient, but the difference is sometimes small. It can be noted that the larger the difference between the environment fee and *MD* is, the larger the difference between the ideal and non-ideal electricity markets becomes.

I have no intention to model market dynamics in this dissertation, but let us nevertheless have a look at the willingness to invest in the ideal electricity market and the non-ideal electricity market with a feebate system. The ideal electricity market maximises the benefit to the society, i.e., an investment is carried out if it results in an increase of the total surplus which is larger than the investment cost. In table 4.4 it is shown how the total surplus in an ideal electricity market is affected if the emissions of power plant 2 are reduced from 3 ton/MWh to 2 ton/MWh. The larger the marginal damage caused by the emissions is, the more profitable the investment becomes. In the non-ideal electricity market the investment is carried out if the cost for producer 2 is less than the increased producer surplus. In the table, it is shown how the surplus of producer 2 is changed. Notice that in this case the profitability is not depending on *MD*, but on by the size of the fee. Therefore, in some cases the investment will be profitable in both the ideal and non-ideal electricity market, whereas in others the results will differ. The example illustrates the importance of choosing the correct fee to achieve an efficient resource usage both in the short and the long run.

Tradable Green Certificates

To decrease the environmental impact of the electricity market, it is of course necessary to build more environmentally benign power plants. The snag is that environmentally benign electricity generation generally is more expensive than dirty ditto (otherwise the environmentally benign power plants had been built already before people even started to worry about the environment). One method to improve the competitiveness of the environmentally benign power plants is to force the other producers to pay for the environmental damage they cause, for example by introducing tradable emission rights. However, an alternative is to subsidise environmentally benign power plants. To begin with, this might seem objectionable, because the subsidies decrease the electricity price and in that way the consumers get incorrect signals about the real cost of the electricity they consume. However, this can be avoided by letting the consumers pay for the subsidies.

Tradable green certificates[22] are used to subsidise certain kinds of power plants by giving the consumers a certain quota, which states how large share of their consumption which must originate from the chosen “green” power plants. To verify that the consumers fulfil their obligation, green certificates are issued, where each certificate corresponds to 1 MWh energy generated in one of the selected power plants. The certificates can then be traded in a separate certificate market. The result for the producers is that the owners of the certified power plants are subsidised; each MWh is not only sold at the ordinary electricity market, but they also obtain a green certificate to be sold in the certificate market.

The trading of green certificates does not have to be continuous; the quota is defined for a certain accounting period (e.g. one year). After the end of the accounting period, the control authority makes a settlement where the number of green certificates necessary to fulfil the quota are cancelled. Those consumers who do not possess enough certificates at the time of the settlement will have to pay a penalty fee instead.

The definition of which power plants are granted certificates is essential both to the involved players, but also for the possibility to maximise the benefit to the society. The green certificates will for example not guide the development in the right direction if investments in a certain power source are beneficial to the society, but these power plants are neither certified nor capable of competing on their own in the electricity market.

The parameters controlling the green certificate trading are the quota, the penalty fee, the duration of the certificates and the accounting period. The quota is of course very important, because it determines the demand of the certificates. It is possible to—as in the Swedish system—have different quotas for different consumer categories and to increase the quota from year to year. The penalty fee will in practice constitute a price cap, because it can be assumed that the consumers rather pay the penalty fee than buy certificates if the penalty fee is less than the certificate price. However, it is possible to repay the penalty fees to those consumers who fulfilled their quota, which makes it interesting to buy certificates even if the price is higher than the penalty fee—this is done for example in England-Wales.[23] The duration of the certificates is important for the pricing in the certificate market. If the certificates are only valid for one accounting period, there will be a price pressure downwards—those producers who do not manage to sell their certificates will not get any extra income at all. If the certificates can be saved for coming

---

22. The terminology may vary depending on the national legislation—for example in Sweden the term "electricity certificate" is used rather than green certificates (probably due to some unfathomable political reason). I have chosen the term green certificates, because it is the most common in international literature (see for example [80, 81, 82, 84, 144]) and it also emphasises the difference between green certificates, which are used to support environmentally benign power plants, and capacity certificates, which are used to support reserve power plants (i.e., such power plants which are mostly used during peak load periods and which are so rarely dispatched that they have difficulties covering their fixed costs; see for example [18] for further details about capacity certificates).
23. In October 2002 the average price of green certificates in England-Wales was £47.13 although the penalty fee was not larger than £30 [144].

accounting periods, we may expect a price pressure upwards instead, because those producers who do not think that they are paid enough can save the certificates until the prices increase.[24] There is also another possibility for the regulator to control the certificate price, and that is to introduce a redemption price of the certificates. This means that the authorities will buy the certificates to a certain guarantee price, which will constitute a floor to the certificate price in the same way as the penalty fee works as a price cap.

Apparently, there are many aspects to be considered when modelling tradable green certificates. Though the subject is interesting, a complete model would be too extensive and I will limit myself to presenting a basic model, which will demonstrate the general principles. As in the earlier examples of this chapter, I will study an electricity market having only one area and each scenario comprises one time period, which coincides with the accounting period of the certificates.[25] Moreover, I assume that the certificate market is perfectly competitive, that the penalty fee is known in advance, that there is no floor to the certificate price and that both producers and consumers of green certificates have perfect information about supply and demand in the certificate market. Finally, I assume that the electricity market is static, i.e., no new certified power plants will enter the market during an accounting period. The last assumption differs my basic model from the model suggested in [81], which will have consequences for the pricing of the certificates. I will return to this issue below, but let us first have a look at the model I suggest.

We start by studying the producer problem of a certified power plant, the generation of which I designate by G_g^+ (the plus sign indicates that the generation of power plant g is certified). As usual, the objective function is to maximise the income, which partly are due to the electricity market and partly due to the certificate market, minus the generation cost:

$$\text{maximise} \quad \lambda G_g^+ + \xi G_g^+ - C_{Gg}(G_g^+) \tag{4.11}$$

$$\text{subject to} \quad 0 \le G_g^+ \le \overline{G}_g^+, \tag{4.11a}$$

where ξ is the certificate price. Due to the perfect competition, all producers are price takers; hence, ξ is a parameter to the producer in the same way as the electricity price λ.

24. This trend has been seen in the Swedish certificate market. The average price during the period May 1, 2003 (when the certificate trading started) until November 16, 2003, was about 210 SEK/MWh according to statistics from SvK (https://elcertifikat.svk.se/), although the penalty fee of the accounting period May 2003 to April 2004 cannot exceed 175 SEK/MWh. The explanation is probably that during the following accounting period the maximal penalty fee is increased to 240 SEK/MWh [86].

25. And the reason for these simplifications is as before to make the formulae more readable; there would be no difficulties to model multi-area problems and several time periods.

Those producers who are not certified behave as in an ideal electricity market, except that they do not internalise the external costs of their emissions. Thus, their producer problems read as follows:

$$\text{maximise} \quad \lambda G_g - C_{Gg}(G_g) \tag{4.12}$$

$$\text{subject to} \quad 0 \le G_g \le \overline{G}_g. \tag{4.12a}$$

The consumers maximise the benefit of their electricity consumption minus the cost of purchase, which includes both the electricity price, the cost to buy certificates and any penalty fees. As a constraint, the quota must be covered by the number of purchased certificates, Γ_c, plus consumption subject to the penalty fee, Ξ_c. If the quota is designated k_c, we get the following consumer problem:

$$\text{maximise} \quad B_{\Delta c}(\Delta_c) - \lambda\Delta_c - \xi\Gamma_c - \beta_\Xi\Xi_c \tag{4.13}$$

$$\text{subject to} \quad k_c\Delta_c - \Gamma_c - \Xi_c = 0, \tag{4.13a}$$

$$0 \le \Delta_c \le \overline{\Delta}_c, \tag{4.13b}$$

$$0 \le \Gamma_c, \tag{4.13c}$$

$$0 \le \Xi_c. \tag{4.13d}$$

For the market as a whole applies that the production and consumption of electric energy must be in balance. Moreover, the consumption of green certificates may not be larger than the generation. To formulate this condition, we introduce the slack variable, Ω, unused certificates. Together with the optimality conditions of the player problems (4.11)-(4.13) we get a system of equations and inequalities defining how the players will behave:[26]

$$\sum_{c \in C} \Delta_c - \sum_{g \in G} G_g = 0, \tag{4.14a}$$

$$\sum_{c \in C} \Gamma_c - \sum_{g \in G^+} G_g^+ - \Omega = 0, \tag{4.14b}$$

$$MC_{Gg}(G_g^+) - \xi \ge \lambda, \qquad \text{if } G_g^+ = 0, \tag{4.14c}$$

$$MC_{Gg}(G_g^+) - \xi = \lambda \qquad \text{if } 0 \le G_g^+ \le \overline{G}_g^+, \tag{4.14d}$$

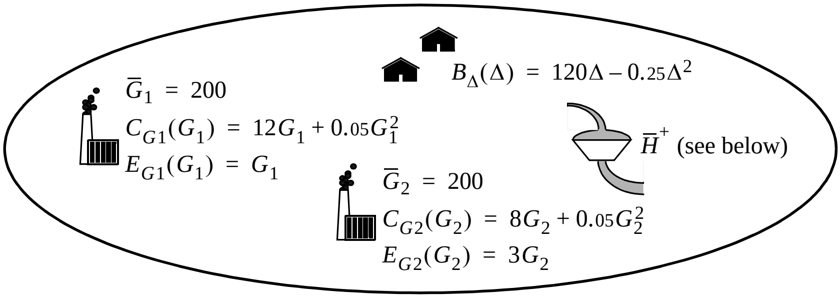
$$MC_{Gg}(G_g^+) - \xi \le \lambda \qquad \text{if } G_g^+ = \overline{G}_g^+, \tag{4.14e}$$

26. In this system of equations and inequalities we also have the constraints and the variable limits of (4.11)-(4.13), but I have chosen not to repeat them, to at least somewhat shorten the enumeration of the optimality conditions. Besides, it can be noted that the conditions (4.14a)-(4.14q) correspond to the optimality conditions of an aggregated optimisation problem, which in this case is a simple quadratic programming problem.

$MC_{Gg}(G_g) \geq \lambda,$	if $G_g = 0,$	(4.14f)
$MC_{Gg}(G_g) = \lambda,$	if $0 < G_g < \overline{G}_g,$	(4.14g)
$MC_{Gg}(G_g) \leq \lambda\ ,$	if $G_g = \overline{G}_g,$	(4.14h)
$MB_{\Delta c}(\Delta_c) \leq \lambda + k_c\kappa_c$	if $\Delta_c = 0,$	(4.14i)
$MB_{\Delta c}(\Delta_c) = \lambda + k_c\kappa_c$	if $0 \leq \Delta_c \leq \overline{\Delta}_c,$	(4.14j)
$MB_{\Delta c}(\Delta_c) \geq \lambda + k_c\kappa_c$	if $\Delta_c = \overline{\Delta}_c,$	(4.14k)
$\xi \geq \kappa_c,$	if $\Gamma_c = 0,$	(4.14l)
$\xi = \kappa_c,$	if $\Gamma_c > 0,$	(4.14m)
$\beta_\Xi \geq \kappa_c,$	if $\Xi_c = 0,$	(4.14n)
$\beta_\Xi = \kappa_c,$	if $\Xi_c > 0,$	(4.14o)
$\xi \geq 0,$	if $\Omega = 0,$	(4.14p)
$\xi = 0,$	if $\Omega > 0.$	(4.14q)

The interpretation of these conditions is that the marginal cost of the certified producers equals the marginal cost of the generation minus the certificate price. The other producers behave as in an ideal electricity market. The consumers must include both the electricity price and their individual certificate value, κ_c (dual variable of the constraint in the consumer problem), when deciding how much they are to consume. The individual certificate value is normally equal to the market price of the green certificates, which in its turn is either equal to the penalty fee (if the certified generation is less than the total quota of the consumers) or equal to zero (if there is a surplus of certificates).

The last observation above is interesting, because it means that the certificate price will toggle between two extreme values if there is perfect competition in the certificate market and the duration of the certificates is limited. This means that the quota must be higher than the certified generation we actually are striving for, since the moment the certified generation exceeds the quota the subsidy disappears. Hence, we are back to square one, i.e., the environmentally benign power plants cannot compete with other power plants. In figure 4.6 I provide some examples of scenarios illustrating this situation. In the example, a number of small hydro power plants are certified, and there are also two larger, non-certified producers. When the quota of the example is 10% it is not profitable to double the capacity of the certified generation, because that would force the certificate price down to zero. However, if the quota is increased to 20% investing in more certified power plants could be profitable, provided that the investment cost is less than the increase of the producers' surplus.



Scenario parameters				
Quota, k [%]	10	10	20	20
Penalty fee, β_X [¤/MWh]	30	30	30	30
Maximal certified generation, $\bar{H}^+$ [MWh]	15	30	15	30
Electricity price, λ [¤/MWh]	19.05	18.64	18.77	18.009
Certificate price, ξ [¤/MWh]	30.00	0.00	30.00	30.00
Consumers				
Consumption, Δ [MWh]	195.91	202.73	190.45	191.82
Purchased certificates, Γ [MWh]	15.00	20.27	15.00	30.00
Penalised consumption, Ξ [MWh]	4.59	0.00	23.09	8.36
Value of consumption, $B_\Delta(\Delta)$ [¤]	13 914.00	14 052.69	13 786.31	13 819.63
Electricity cost, $\lambda\cdot\Delta$ [¤]	3 731.19	3 778.11	3 575.35	3 470.16
Certificate cost, $\xi\cdot\Gamma$ [¤]	450.00	0.00	450.00	900.00
Penalty cost, $\beta_X\cdot\Xi$ [¤]	137.73	0.00	692.73	250.91
Surplus, CS	9 595.08	10 274.58	9 068.24	9 198.56
Certified producers				
Generation, H^+ [MWh]	15.00	30.00	15.00	30.00
Electricity income, $\lambda\cdot H^+$ [¤]	285.68	559.09	281.59	542.73
Certificate income, $\xi\cdot\Gamma$ [¤]	450.00	0.00	450.00	900.00
Surplus, PS_{cert} [¤]	735.68	559.09	731.59	1 442.73
Other producers				
Generation, $G_1 + G_2$ [MWh]	180.91	172.73	175.45	161.82
Electricity income, $\lambda\cdot(G_1 + G_2)$ [¤]	3 445.50	3 219.01	3 293.75	2 947.44
Generation cost, $C_{G1}(G_1) + C_{G2}(G_2)$ [¤]	2 587.30	2 433.14	2 484.16	2 232.82
Surplus, $PS_{övr}$ [¤]	858.22	785.88	809.61	694.63
Total emissions, $E_{G1}(G_1) + E_{G2}(G_2)$ [ton]	401.82	385.45	390.91	363.64
Total surplus, $B_\Delta(\Delta) - C_{G1}(G_1) - C_{G2}(G_2)$ [¤]	11 326.70	11 619.55	11 302.15	11 586.81

Figure 4.6 *Example of tradable green certificates. In this example the hydro power is certified, as it does not cause any emissions. If the quota is increased, the profitability of the certified power plants increase too.*

Notice that the total surplus does not *include the costs caused by the emissions. Thus, it depends on the evaluation of the external costs, which solution is going to maximise the benefit to the society (cf. table 4.5).*

Table 4.5 *Short-term total surplus of the scenarios in figure 4.6.*

Quota, k [%]	Certified generation, $\bar{H}$ [MWh]	True marginal damage, MD [¤/ton]				
		1	2	3	4	5
Ideal electricity market[a]	15	10 936	10 564	10 219	9 902	9 612
10	15	10 925	10 523	10 121	9 719	9 318
20	15	10 911	10 520	10 129	9 739	9 348
Ideal electricity market[a]	30	11 248	10 903	10 586	10 296	10 033
10	30	11 234	10 849	10 463	10 078	9 692
20	30	11 223	10 860	10 496	10 132	9 769

a. In the ideal electricity market there is no tradable green certificates, but the true marginal damage of emissions are included in the cost functions of the power plants.

Table 4.6 *Profitability of investing in increased certified generation.*

	True marginal damage, MD [¤/ton]				
	1	2	3	4	5
Increased surplus in an ideal electricity market[a]	312	339	367	394	421
Increased surplus of certified producers[b]					
Quota k = 10%	–4	–4	–4	–4	–4
Quota k = 20%	884	884	884	884	884
Is the investment profitable if the cost is 200 ¤?					
Ideal electricity market	Yes	Yes	Yes	Yes	Yes
Quota k = 10%	No	No	No	No	No
Quota k = 20%	Yes	Yes	Yes	Yes	Yes
Is the investment profitable if the cost is 400 ¤?					
Ideal electricity market	No	No	No	No	Yes
Quota k = 10%	No	No	No	No	No
Quota k = 20%	Yes	Yes	Yes	Yes	Yes
Is the investment profitable if the cost is 600 ¤?					
Ideal electricity market	No	No	No	No	No
Quota k = 10%	No	No	No	No	No
Quota k = 20%	Yes	Yes	Yes	Yes	Yes

a. Determined by comparing the total surplus according to table 4.5 when the certified generation is 15 MWh and 30 MWh respectively.
b. Determined by comparing the surplus of the certified producers according to figure 4.6 when the certified generation is 15 MWh and 30 MWh respectively.

It is justified to question whether or not the prices in a real certificate market will toggle between extreme values in this manner. According to for example [80, 81, 82] the certificate price can be determined by the long-run production cost (i.e., all costs—investment, operation and maintenance, and a risk reward to the investor—during the life time of the power plant divided by the total output) of the certified power plants which are commenced during the accounting period. I am however not quite convinced that this reasoning is valid, because the power plants which were already in operation at the beginning of the accounting period and those entering later are in the same position; they have already had expenses for the investment and if the price is not high enough to cover the investment costs, they must at least try to minimise the losses. If there is a surplus of certificates, the competition will force the certificate price down to the minimum level. If there is a deficit of green certificates instead, there is no reason, neither for new nor old power plants, to ask for a price much less than the penalty fee. There are however many factors which may influence the pricing, for example incomplete information (the players do not know whether there will be a surplus or deficit of certificates), market power or longer duration of the certificates than one accounting period. More detailed studies of the pricing in certificate markets should be necessary.

If we return to the example in figure 4.6, the external costs are not included in the figure, but in table 4.5 I have calculated the total surplus for some values of the true marginal damage caused by the emissions, *MD*. For comparison, I have also calculated the total surplus in an ideal electricity market, where *MD* is internalised in the production costs of the power plants. The ideal electricity market is as expected always more efficient, but the difference is not particularly large when *MD* is low, whereas the difference is large when *MD* is higher.

I have no intention to model market dynamics in this dissertation, but let us nevertheless have a look at the willingness to invest in the ideal electricity market and the non-ideal electricity market with tradable green certificates. The ideal electricity market maximises the benefit to the society, i.e., an investment is carried out if it results in an increase of the total surplus which is larger than the investment cost. In table 4.6 it is shown how the total surplus in an ideal electricity market is affected when the certified generation increases by 15 MWh. The larger the marginal damage caused by the emissions is, the more profitable the investment becomes. In the non-ideal electricity market the investment is carried out if the cost of the certified producer is less than the increased producer surplus. In the table, it is shown how the surplus of the certified producer is changed. Notice that in this case the profitability is not depending on *MD*, but on the price of the green certificates, which indirectly depends on the quota. Therefore, in some cases the investment will be profitable in both the ideal and non-ideal electricity market, whereas in others the results will differ. The example illustrates the importance of choosing the right levels of the parameters controlling the trading (quota, penalty fee, duration of certificates, etc.) to achieve an efficient resource usage both in the short and the long run.

4.3 CREDIBILITY

Even actions that are necessary to protect the environment usually cause discontent in some places—there is always some party who is subject to increased costs or looses a source of income, and which therefore prefers that everything went on as before. Unfortunately, there are countless arguments for those who want to oppose environmental protection rules; the environmental impact of a particular activity can be denied or played down, it can be questioned whether a certain action will result in the desired improvements of the environment, etc. To rebuff these kinds of arguments it is important that all actions taken to protect the environment are credible.

It goes without saying that regardless of which rules are introduced, they must be applied consequently; if exceptions are made for some producers or consumers credibility is lost—in practice, the result is that environmentally hazardous activities are subsidised. In this section I will therefore focus on two other credibility issues, namely whether or not it is possible to disclose electricity, and what may happen in an electricity market where there is both trading of eco-labelled electricity and green certificates.

The Ownership of a Megawatt-hour

As all electricity trading takes place in a common grid it is in practice impossible to track the generated electric energy from producer to consumer. This fact is frequently used to discredit players who on their own initiative try to consider external costs when choosing electricity supplier. For example, consider this letter to the editor of the Swedish engineering newspaper *Ny Teknik* concerning a petrol station offering hydrogen produced using electricity from wind power:

> "In the article it is stated that this electricity is generated by a wind power plant outside Malmö. I doubt the veracity of this claim—it is as far as I know purely blather, with intent to intimate that the hydrogen should be 'green'. The electrolysis is most likely performed at the petrol station, although this is not clearly stated. It would be madness to have a special electric line between the wind power plant and the petrol station. The truth is of course that the power is taken from the grid…"[27]

There are other debaters who like this writer do not think that a consumer can be said to consume particular electricity generation, but that everybody consumes the same mixture of power sources. Such claims might be physically correct, but are nevertheless misleading.

To explain why, it is most simple to use a metaphor. Consider a simple water market, which is less complicated than an electricity market, but has similar properties. The consumers get their water from a large tank and the producers supply water by pouring it into the common tank. There is also a system operator who makes sure that there is a sufficient amount of water in the tank. Hence, a consumer must not get water exactly at the same time as the producer delivers the water; the trading implies that if the consumer gets a bucket of water then the producer at some time will have to fill up the tank by the same amount of water.

Now assume that there are two consumers, A and B, in this water market. The supplier of consumer A fetches fresh water from a spring in the forest, whereas consumer B has chosen a supplier who fetches the water out of a dirty pool next to an industrial estate. As all water is mixed in the tank, both consumers will drink the same mixture of spring water and filth. Does this mean that they are equally good consumers? This is assuredly a philosophical question, but I can hardly imagine that somebody is prepared to give an affirmative answer—consumer A is paying for a good product being supplied to the system, whereas consumer B is paying for an inferior product. Although different water qualities cannot be physically separated, A should nevertheless be accounted the right of the fresh water. In the same manner I think it is natural that consumers in an electricity market who are paying a producer to supply environmentally benign electricity to the grid, also should be able consider themselves the consumers of that electricity.

A problem which we should look out for, when physical delivery and economical agreements do not follow each other, is *double counting*, i.e., when several players claim the right to the same environmental improving action. If we assume in the water market metaphor, that A and B consume the same amount of water per year, then the water will consist of 50% spring water and 50% filthy water. Assume that B on account of this claims to be consuming 50% spring water, while A claims that he or she consumes 100% spring water. If both were right, the water would consist of 75% spring water, which we know is not the case. Double counting can arise in an electricity market in a similar manner, for example if the share of environmentally benign production is considered in some contexts and the share of environmentally benign consumption in others.

Naturally, any regulation allowing for double counting looses a lot of credibility. Besides, statistics which have been distorted by double counting may

27. The letter is from *Ny Teknik*, no 40, October 1, 2003, p 38. The letter was originally written in Swedish.

give the impression that an environmental problem is smaller than it really is. There are therefore very good reasons to look out for double counting and to make sure that the rules of the electricity market clearly define who should be accounted the right of an action improving the environment.

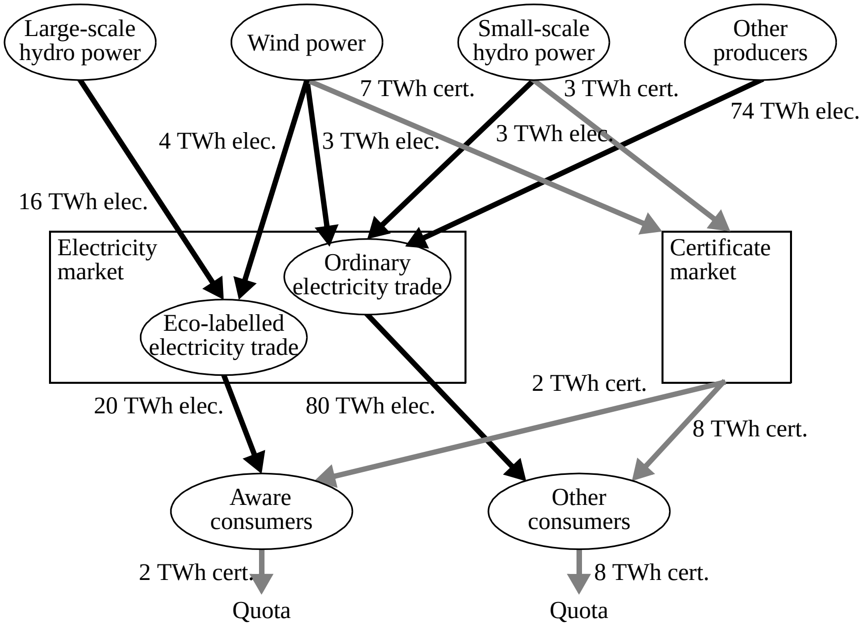
Green Certificates and Eco-labelling

Sometimes green certificates and eco-labelled electricity are presented as two different products. The aim of both systems is however the same; environmentally benign power plants should be paid more, so that they better manage to compete with other power plants. Therefore, my opinion is that it is unfortunate to separate the two systems, because it may lead to double counting.

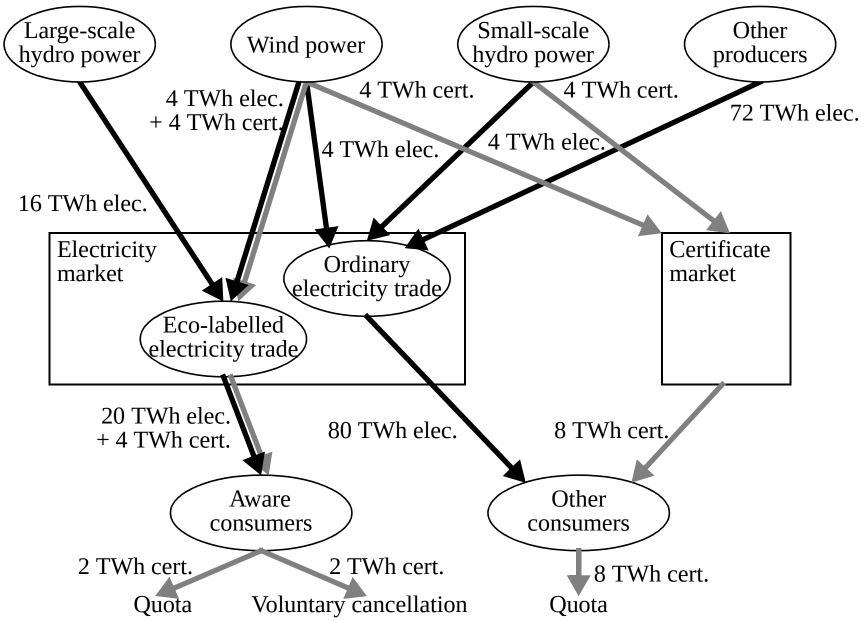
Let us start by observing that certified electricity and eco-labelled electricity does not necessarily have to be the same thing, because the legislator and the eco-labelling organisation may have different views on which power plants to support. This is quite in order—it is even an advantage, because it gives the consumers a larger freedom of choice, if it should be so that one or the other choose an unfortunate definition of what constitutes environmentally benign electricity generation. Let us now consider a simplified electricity market, where there are four power sources representing the four possible combinations of eco-labelling and certification; large-scale hydro power and wind power are eco-labelled, small-scale hydro and wind power are certified, and the other producers are neither eco-labelled nor certified.

The criteria for eco-labelled electricity states that at most 80% may originate from large-scale hydro power and at least 20% must be from wind power. The demand of eco-labelled electricity is 20 TWh/year and as the wind power is more expensive than large-scale hydro power, we assume that this means that those offering eco-labelled electricity will generate 16 TWh hydro energy and 4 TWh wind energy. The quota in the green certificate system is 10%. The production cost of both certified power sources are assumed to be the same, which implies that if the demand for certificates increases by 1 TWh then the wind power will account for 0.5 TWh of that increase and the small-scale hydro for the remaining 0.5 TWh.

Assume that eco-labelled electricity and green certificates are treated as separate products. This means that all consumers, regardless of whether or not they have chosen eco-labelled electricity, must go to the certificate market and purchase certificates corresponding to their quota, as shown in figure 4.7a. The power plants which are both eco-labelled and certified will have extra income both when selling the electricity to a higher price (in the market for eco-labelled electricity) and then another extra income when the certificates are sold. It is hard to see this in any other way than that the producers are paid twice for the same thing. In a similar manner, the environmentally aware consumers are subject to two extra cost increases; first, they



a) Eco-labelled electricity separated from green certificate trading.



b) Eco-labelled electricity including the green certificates assigned to the eco-labelled power plants.

Figure 4.7 *Electricity market with both eco-labelled electricity and green certificates. In this example it is assumed that the eco-labelled electricity consists of 80% large-scale hydro power and 20% wind power, while wind power and small-scale hydro power are certified. The quota is 10%.*

voluntarily pay a little bit extra for eco-labelled electricity and secondly, they have to pay extra for the green certificates.

Assume that the two systems were integrated instead, i.e., when a consumer buys eco-labelled electricity, the deal includes all certificates the producer obtained for the part of the generation which is also certified. If the eco-labelled electricity consists of a lesser share certified generation than the quota, the aware consumers will still have to buy some certificates from the certificate market. But if the share is higher—as it is in our example—then the aware consumers receive more certificates than required to fulfil the quota. This situation is shown in figure 4.7b. If the certificates are valid longer than one accounting period there has to be a possibility to voluntarily cancel the surplus certificates; otherwise the consumer can sell the certificates in the free certificate market and then we are back to a similar situation as in figure 4.7a.

The difference between the two examples is primarily that the wind power in figure 4.7b cannot simultaneously be sold in the market for eco-labelled electricity and the certificate market. The demand of electricity which is both eco-labelled and certified therefore increases when the systems are integrated, which is an obvious advantage. But at the same time, the electricity generation in power plants which are certified but not eco-labelled increases, and it is not self-evident how the organisation behind the eco-label should respond to this. If they consider certified but not eco-labelled electricity to at least be preferable to electricity generation which is neither certified nor eco-labelled then it is better to integrate the systems. On the other hand, if they think that the certification of these power plants is a severe mistake, then it might be better to accept a decreased demand of eco-labelled electricity while decreasing the undesired, certified electricity. In this case separate systems are preferable—if you are prepared to justify the arising double counting.

Finally, I would like to remark that my example is somewhat exaggerated, because I assume that increased demand for certificates is divided equally among eco-labelled generation and generation without the labelling. Hopefully, the environment organisation and the legislator share a more common view of which power plants should be supported, which means that increased demand for certificates mostly results in increased generation in power plants which are favoured by both systems. My personal view is therefore that eco-labelled electricity and tradable green certificates should be integrated to avoid the systems being questioned due to double counting. If for some reason green certificates are issued to directly improper power plants (for example such power plants which are not very environmentally benign or which could manage without support) it is more appropriate to lobby for a change of the definition of green power plants, than to oppose the whole green certificate trading system.

Chapter 5

FORECAST UNCERTAINTY

In an ideal electricity market all players have perfect information, which means that they do no have to make any assumptions about future events, but they know exactly what future awaits them.[1] The players in a real electricity market do not have it as easy, because there is always some uncertainty in forecasts of the future.

We may compile a simple time scale describing the time perspective for different kinds of decisions (see figure 5.1). The time perspective is of importance to which forecasts are necessary, how accurate the forecasts are, and the consequences of incorrect forecasts. In one end we find decisions which are related to the operational security of the system. Here the time perspective is very short; quick decisions have to be made in order to maintain stability of the system; therefore, technology is more important than economy. In the other end the situation is the opposite. Here technology is hardly a limitation, but the question is which investments are profitable. These decisions are characterised by a great deal of forecast uncertainty.

I will in this chapter have a closer look on how to model the consequences of players planning their actions based on partly uncertain information. I will in this respect study three of the time perspectives mentioned in figure 5.1: control, short-term planning and long-term planning. Forecast uncertainties

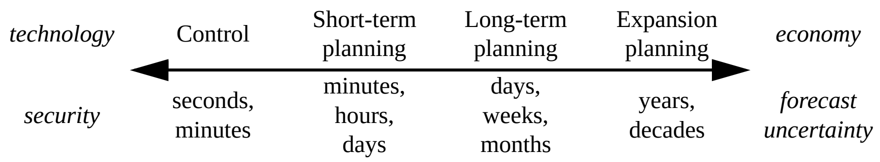


Figure 5.1 *Time perspective in power system planning.*

1. More precisely, if a future event is represented by a random variable, it is not sufficient to know the probability distribution of the random variable to have perfect competition, but we have to know its outcome.

in expansion planning will not be addressed, because the models I work with are intended for that purpose—trying to incorporate into the model how the electricity market is affected by the model itself seems like the beginning of a complicated circle argument.

5.1 CONTROL

In a power system, frequency, voltage and power flows must in every moment be kept within certain safety limits. If we fail to do that, more or less serious disturbances will follow, and in the worst case the system might collapse. It may take days before the system has recovered from such a severe disturbance and the social cost may be enormous.[2]

There are many small changes which immediately affect the operation of a power system and may cause disturbances if they are not compensated by appropriate actions. Examples of such changes are consumers increasing or decreasing their consumption, non-dispatchable power plants changing generation due to the weather and failures in power plants. Obviously, it is impossible to know in advance when a change will occur and it is therefore necessary to install automatic control systems, which rather than forecasting changes, quickly respond to deviations and correct them before they become too large.

Voltage Control

To explicitly model voltage control in an electricity market would require far more detailed models than a multi-area model, which would not be very appropriate for a Monte Carlo simulation (cf. section 3.2). On the other hand, there is no direct need for a detailed model of the voltage control, because the practical consequences of the forecast uncertainty are quite small. As we cannot anticipate when a failure will occur, we need to keep such margins in the system that it can cope with a sudden failure in a line; thus, the available transmission capability between the areas of the system is slightly less than what would had been the case with perfect information. This impact is modelled indirectly when determining the transmission capability between the

2. During 2003 major blackouts occurred in north-eastern USA (including parts of Canada), in southern Sweden (including parts of Denmark) as well as in Italy. It is hard to assess the costs of these blackouts, but in Sweden for example the costs were estimated to about 500 MSEK [38]. This number is probably rather too high—for comparison, it can be mentioned that during the spring 1998 central Auckland was befallen a black-out lasting several weeks, since the four transmission lines supplying power to the city had failed. The social cost was estimated to 60 MNZD per week according to [141].

areas. The voltage control also affects the losses of the system, because the active power losses slightly increase if more reactive power is transferred simultaneously. The increase is however quite marginal and it is justified to assume a constant reactive power when determining the loss functions.

The voltage control causes only small direct operation costs, which are due to internal losses in the components used to control the voltage. There are also indirect costs, since the voltage control influences the losses. Both these costs are however marginal and it is fully reasonable to neglect them. Hence, the costs of voltage control consist more or less only of the investment and maintenance costs of the equipment used for voltage control—these costs can be considered part of the general costs to build grids (cf. section 6.3).

Frequency Control

As soon as there is an imbalance between generation and consumption, the frequency of the system will be affected; the frequency decreases when generation is less than the load and vice versa. To compensate the imbalances which continuously occur due to small changes in load and generation in non-dispatchable power plants, automatic control systems are used to govern the generation of particular power plants. These control systems are divided in primary and secondary control, which roughly correspond to the proportional and integrating parts of a PI controller.[3]

The primary control stabilises the frequency by increasing or decreasing the generation when the system frequency changes. All power plants participating in the primary control have control systems measuring the system frequency and governing the power plant output; the relation between the generation of a power plant, G, and the system frequency, f, can be written as

$$G = G_0 - R(f - f_0), \tag{5.1}$$

where f_0 is the nominal frequency and R is the speed-droop characteristics (expressed in MW/Hz). To allow a power plant to participate in the primary control, the nominal generation, G_0, must be less than installed capacity, so that there is a marginal to increase the generation if the frequency should decrease. It must also be larger than zero, so that there is a marginal to decrease the generation when the frequency increases.

In the ideal electricity market it is the electricity price and the marginal cost of the power plant which determine how much each power plant should produce,[4] but in reality there are some power plants where the generation is

3. Unfortunately, there is some confusion of ideas in this area, as we in the Nordic countries use the term secondary control for manually activated reserves. I have here chosen to use a more internationally practicable terminology, in accordance with [36].

partly controlled by the frequency instead. Hence, there will generally be a difference compared to the operation cost in an ideal electricity market. How large this difference is depends on which power plants participate in the primary control. This is in its turn depending on technical properties of different power plants, i.e., how easy they are to control and which extra operation cost the control will cause.[5] However, we are dealing with relatively small deviations, as the share of the installed capacity which is part of the primary control is quite small. In the Nordel area, where the total installed capacity amounts to about 83 000 MW, the normal primary control reserve is ±600 MW.[6] If we choose to neglect the primary control completely in the electricity market model, the error should not be too large. A more realistic alternative is simply to exclude the primary control reserves from the available generation capacity; in a real tight situation, load will be disconnected before the primary control reserve is the last unused generation capacity of the system.[7]

If we want to model the consequences of primary control more in detail, the following assumptions can be made. The primary control increases the total operation cost, *TOC*, because the power plants participating in the primary control sometimes generate even though the marginal cost is higher than the electricity price, and sometimes are not used although the marginal cost is less than the electricity price. The impact on *TOC* varies from scenario to scenario, but over a longer time period, these variations should even out. Therefore, it seems reasonable to determine an approximate cost function to set aside a certain part of the available generation capacity for primary control. In the electricity market model we may then differ between generation capacity intended for selling energy to the ahead market, $G^F_{g,t}$ (thermal power plants)

---

4. If the marginal cost is higher than the electricity price, the power plant will not generate anything, the power plant will generate its available capacity if the electricity price is higher than the marginal cost of the last MW; otherwise, marginal cost and electricity price should be equal (cf. the examples of producer problems given in chapters 4 and 6).
5. A power plant will generally have one or more maximal efficiency points. Deviating from these levels of generation—which is inevitable when frequency controls generation—means lesser efficiency and hence higher marginal production cost.
6. In total there are 91 000 MW generation capacity in the Nordic countries, but Iceland (1 500 MW) maintain their own frequency and western Denmark (7 000 MW) is part of the central European UCTE area [40, 44].
7. To maintain a stable frequency there has to be a primary control reserve—it is not possible to adjust continuously the balance between production and consumption by disconnecting load. If the available capacity of the Nordic countries is 83 000 MW and the load was allowed to increase to 82 900 MW there would be a really awkward situation if the load increased by another 100 MW or if 100 MW generation capacity was lost. The frequency would then continue to decrease, as there would not be any reserves left, and we would run the risk of a total collapse of the system.

and $H^F_{r,t}$ (energy limited power plants), and generation capacity reserved for the primary control, $G^P_{g,t}$ and $H^P_{r,t}$.[8] The objective function of the scenario problem (3.12) is changed so that it includes both the costs of the planned generation and the costs of the primary control:

$$\text{maximise} \sum_{t=1}^{T} T_t \Bigg(\sum_{c \in C_\Delta} B_{\Delta c,t}(\Delta^F_{n,t}) - \sum_{g \in G} C_{Gg,t}(G^F_{g,t}) - \sum_{g \in G} C^P_{Gg,t}(G^P_{g,t})$$
$$- \sum_{r \in R} C^P_{Hr,t}(H^P_{r,t}) - \sum_{c \in C} C_{Uc,t}(U_{n,t}) \Bigg) \tag{5.2}$$

Notice that the objective function (5.2) contains two different kinds of cost functions: $C_{Gg,t}(G_{g,t})$ refers to the generation cost for physical delivery of a certain average power G_g, while $C^P_{Gg,t}(G^P_{g,t})$ and $C^P_{Hr,t}(H^P_{r,t})$ refer to the cost to set aside a certain capacity for the primary control reserve.

The requirements of maintaining a sufficient primary control reserve, R_t, in the system is represented by a constraint:[9]

$$\sum_{g \in G} G^P_{g,t} + \sum_{r \in R} H^P_{r,t} = R_t, \qquad \forall\, t = 1, \ldots, T. \tag{5.3}$$

Finally, the limits of the power generation (3.15b, 3.15c) are replaced by the following constraints:

$$G^F_{g,t} + G^P_{g,t} \le \bar{G}_{g,t}, \qquad \forall\, g \in G, t = 1, \ldots, T, \tag{5.4}$$

$$H^F_{r,t} + H^P_{r,t} \le \bar{H}_{r,t}, \qquad \forall\, r \in R, t = 1, \ldots, T. \tag{5.5}$$

When using the above modifications of the scenario problem, it will also be necessary to update the definitions of the result variables. For example, the costs of the primary control should be included when calculating *TOC*. The consequences of the other main result variable—loss of load occurrence, *LOLO*—depends on which kind of model that is used. In a short scenario, the scenario problem only include one time period. This time period can be chosen arbitrarily; therefore, a short scenario may be interpreted as observing the

8. The non-dispatchable power plants cannot be used for primary control and it is therefore not necessary to divide their available capacity in this manner. In those cases when some power plants have not been equipped with the control systems necessary to participate in the primary control then this can either be modelled by excluding the optimisation variables $G^P_{g,t}$ and $H^P_{r,t}$ of these power plants, or by setting the corresponding cost functions so high that these power plants will never be used for primary control reserves.
9. It is possible to replace (5.3) by separate constraints for the primary control reserve in each area, if that should be desirable.

electricity market in a particular moment. If instantaneous values of available generation capacity and load are compared, it is possible to determine directly if load shedding will occur or not. In practice, we have to—as mentioned above—consider that load shedding cannot be used for primary control and that due to security reasons, load will be disconnected already when the load exceeds the generation capacity which is not part of the primary control.

In a long scenario, loads (and to some extent generation capacities) are average values for each period (cf. section 3.2.1). Comparing two average values to determine if the generation capacity is sufficient is quite naturally hazardous, but a simple interpretation is that if the average of the marginal cost controlled generation is sufficient to cover the average load then the power plants of the primary control can manage the peaks when the instantaneous load exceeds the average load. This interpretations seems quite reasonable if the average values are calculated using relatively short time periods, for example an hour. If this interpretation is accepted, we may either choose to exclude the primary control reserve from the available generation capacity, or we can use the division between planned generation and reserves according to (5.2)-(5.5).

When a disturbance has been compensated by the primary control, there will still be a frequency error; the primary control restores the balance between production and consumption which prevents the frequency from continuing to decrease or increase, but it does not return the frequency to its nominal value. Therefore, another automatic control system—the secondary control[10]—can be installed. The system collects measurements of system frequency, generation in power plants, power flows in transmission lines, etc., and then transmits new reference values, G_0, to the control systems of the power plants participating in primary control. The objective is to return the frequency to its nominal value, maintain desired transmission between different parts of the power system and minimizing the operation cost by governing the generation of the involved power plant to such levels that the best efficiency is achieved [34]. Unlike primary control, which is an essential function of all power systems, secondary control is not a necessity, but the same function can be achieved by manually activated reserves (see below).

Since the secondary control governs the same generation reserves as the primary control, it has similar consequences for operation cost and other result variables. The possibility to include some economic optimisation in the secondary control may cause the total cost of the automatic frequency control to decrease, which should be considered if cost functions of primary control are included in the model, as described above.

10. Often referred to as *Automatic Generation Control* (AGC).

Manual Reserves

The generation capacity which is set aside for automatic frequency control is not unlimited and must therefore gradually be relieved to allow the system sufficient margins to be able to manage new changes in production or consumption. The relief is performed by the system operator, who manually activates regulation actions using the real-time market (see chapter 2).

Activating reserves may cause an extra cost compared to if perfect information had been available, for example if we are forced to use more expensive, but easily started power plants instead of slower units. Using similar arguments as those used above concerning frequency control, we may claim that it is not unreasonable to approximate the costs of the manual reserves as a fixed cost per scenario (which of course may vary when studying different alternative designs of the electricity market).[11]

If desirable, it is possible to introduce more detailed models of manually activated reserves too. To do this, each scenario has to be divided in several parts. In the first part, which I prefer to call the *market problem*,[12] the trading in the ahead market is simulated. The same kind of models are used in the market problem as in an ordinary scenario problem; the difference is just that some inputs represent the forecasts of the players and not the true outcome. It may also be so that some physical limitations which are normally included in the scenario problem are not included in the market problem, because it is the responsibility of the system operator to manage them.[13]

The actions of the system operator, i.e., the trading in the real-time market, are simulated in the next part of the scenario: the *redispatch problem*.[14] In the redispatch problem all physical limitations are considered (because the system operator has to do that in order to maintain safe operation of the system). Moreover, no forecast values are used as input, because it can be assumed that during a trading period the system operator will have sufficient information to know what is going on in the system. To make this assumption more justified, it is possible to use a shorter period length in the redispatch problem than in the market problem. For example, if the ahead market uses a period length of one hour, we could divide the redispatch problem in twelve five minute periods. The disadvantage of this procedure is that it takes a significantly larger computational effort to analyse each scenario.

Let G_g^F and Δ_c^F denote the resulting trading in the ahead market (i.e., the

11. Exactly how to perform these calculations is a challenge, which I happily leave to other researchers.
12. In [12] the term “market dispatch” was used.
13. An example of such a physical limitation which might be omitted from the market problem is transmission limitations. More about that in section 6.2.
14. In [12] the term “congestion redispatch” was used. I have left out the word congestion, as we here do not simulate just congestion management, but all the real-time trading.

solution to the market problem), while G_g^R and Δ_c^R represent the actual operation during a certain period of the real-time trading. The difference between these (i.e., the performed regulation actions) are denoted by $G_g^{\uparrow}$, $G_g^{\downarrow}$, $\Delta_c^{\uparrow}$ and $\Delta_c^{\downarrow}$. The introduced symbols relate to each other as follows:

$$G_g^{\downarrow} = \begin{cases} G_g^F - G_g^R & \text{if } G_g^F > G_g^R, \\ 0 & \text{if } G_g^F \le G_g^R, \end{cases} \tag{5.6a}$$

$$G_g^{\uparrow} = \begin{cases} G_g^R - G_g^F & \text{if } G_g^F < G_g^R, \\ 0 & \text{if } G_g^F \ge G_g^R, \end{cases} \tag{5.6b}$$

$$\Delta_c^{\downarrow} = \begin{cases} \Delta_c^F - \Delta_c^R & \text{if } \Delta_c^F > \Delta_c^R, \\ 0 & \text{if } \Delta_c^F \le \Delta_c^R, \end{cases} \tag{5.6c}$$

$$\Delta_c^{\uparrow} = \begin{cases} \Delta_c^R - \Delta_c^F & \text{if } \Delta_c^F < \Delta_c^R, \\ 0 & \text{if } \Delta_c^F \ge \Delta_c^R. \end{cases} \tag{5.6d}$$

As described in chapter 2, the real-time trading can either be performed using central dispatch, where the system operator decides how each power plant and controllable load should be operated, or using a regulating market. The modelling of central dispatch is more straightforward, so let us start with that model. To simplify the presentation—a redispatch problem can be very complex if all physical limitations of the power system are considered—we neglect any connections between different points of time (as for example that power plants cannot increase or decrease their generation at any rate). Moreover, assume that the scenario parameters which might change between the ahead market and the real-time market is limited to the price insensitive load and the available generation capacity of the power plants. If a multi-area model is used for the power system, safe operation means that there should be balance between production and consumption in each area, without exceeding any limitations. In the central dispatch, the system operator tries to maximise the benefit of consumption minus the cost of generation, while considering the physical limitations; hence, they solve the following multi-area problem:

$$\text{maximise} \quad \sum_{c \in \mathcal{C}_\Delta} B_{\Delta c}(\Delta_c^R) - \sum_{g \in \mathcal{G}} C_{Gg}(G_g^R) - \sum_{c \in \mathcal{C}_D} C_{Uc,t}(U_{c,t}) \tag{5.7}$$

$$\text{subject to} \quad \sum_{g \in \mathcal{G}_n} G_g^R + W_n^R + \sum_{m \in \mathcal{P}_{n \leftarrow m}} (P_{m,n} - L_{m,n}(P_{m,n})) =$$

$$= \sum_{c \in \mathcal{C}_{D_n}} D_c^R - \sum_{c \in \mathcal{C}_{D_n}} U_c + \sum_{c \in \mathcal{C}_n} \Delta_c^R + \sum_{m \in \mathcal{P}_{n \to m}} P_{n,m}, \quad \forall\, n \in \mathcal{N}, \tag{5.7a}$$

$$0 \le \Delta_c^R \le \overline{\Delta}_c^R, \quad \forall\, c \in \mathcal{C}_\Delta, \tag{5.7b}$$

$$0 \le G_g^R \le \overline{G}_g^R, \quad \forall\, g \in \mathcal{G}, \tag{5.7c}$$

$$0 \le P_{n,m} \le \overline{P}_{n,m}, \quad \forall\, (n, m) \in \mathcal{P}, \tag{5.7d}$$

$$0 \le U_c \le D_c^R, \quad \forall\, c \in \mathcal{C}_D. \tag{5.7e}$$

It could be expected that a model of a regulating market is more complicated than the above optimisation problem, but fortunately central dispatch and a regulating market are in practice equivalent. Assume that all players submit bids to the regulating market, that there is perfect competition and no player has any extra regulating costs.[15] If regulation bids are paid the same price as stated in the bid (see figure 2.2) the cost of the system operator can be written as

$$C_{Gg}^{\uparrow}(G_g^{\uparrow}) = C_{Gg}(G_g^R) - C_{Gg}(G_g^F), \tag{5.8a}$$

$$C_{\Delta c}^{\downarrow}(\Delta_c^{\downarrow}) = B_{\Delta c}(\Delta_c^F) - B_{\Delta c}(\Delta_c^R). \tag{5.8b}$$

Similarly, the benefit of the system operator when selling regulating power is equal to

$$B_{Gg}^{\downarrow}(G_g^{\downarrow}) = C_{Gg}(G_g^F) - C_{Gg}(G_g^R), \tag{5.9a}$$

$$B_{\Delta n}^{\uparrow}(\Delta_n^{\uparrow}) = B_{\Delta c}(\Delta_c^R) - B_{\Delta c}(\Delta_c^F). \tag{5.9b}$$

Hence, the objective function of the system operator should be to

$$\text{maximise} \quad \sum_{g \in \mathcal{G}} \left(B_{Gg}^{\downarrow}(G_g^{\downarrow}) - C_{Gg}^{\uparrow}(G_g^{\uparrow})\right) + \sum_{c \in \mathcal{C}} \left(B_{\Delta c}^{\uparrow}(\Delta_c^{\uparrow}) - C_{\Delta c}^{\downarrow}(\Delta_c^{\downarrow})\right). \tag{5.10}$$

This objective function is also valid if uniform pricing is used in the regulat-

15. Such costs can arise as a consequence of a player being forced to deviate from the optimal operation plan. A hydro power producer who down-regulates might risk to fill the reservoir and become forced to spill water, which means that a future income is lost. Another example is a thermal power plant, which was planned to generate its maximal capacity during four hours. If down-regulated, the start-up cost of the power plant must be divided over a smaller total generation and then there will be a risk that the price of the ahead market will not cover both the variable operation cost and the start-up cost.

ing market; the system operator will still try to use the down-regulation bids which pays the best price for regulating power, etc.

The nice part is that the objective functions of (5.7) and (5.10) are equivalent. The proof is simple and is based on the observation that a single player cannot simultaneously regulate up-wards and down-wards. Thus, the producers can be divided in two groups, $\mathcal{G}_1$ and $\mathcal{G}_2$, where the first group are the producers who down-regulate in the optimal solution, while the producers in the second group up-regulate. The producers who do not regulate at all can be distributed arbitrarily between the groups. With a similar division of the consumers we find that the optimal value of the objective function (5.10) can be written as

$$\begin{aligned}&\sum_{g \in \mathcal{G}}(B^{\downarrow}_{Gg}(G^{\downarrow}_{g}) - C^{\uparrow}_{Gg}(G^{\uparrow}_{g})) + \sum_{c \in \mathcal{C}}(B^{\uparrow}_{\Delta c}(\Delta^{\uparrow}_{c}) - C^{\downarrow}_{\Delta c}(\Delta^{\downarrow}_{c})) = \\ &= \sum_{g \in \mathcal{G}_1} B^{\downarrow}_{Gg}(G^{\downarrow}_{g}) - \sum_{g \in \mathcal{G}_2} C^{\uparrow}_{Gg}(G^{\uparrow}_{g}) + \sum_{c \in \mathcal{C}_1} B^{\uparrow}_{\Delta c}(\Delta^{\uparrow}_{c}) - \sum_{c \in \mathcal{C}_2} C^{\downarrow}_{\Delta c}(\Delta^{\downarrow}_{c}),\end{aligned} \tag{5.11}$$

because the cost and benefit functions (5.8a)-(5.9b) are equal to zero, when the corresponding variable is equal to zero. If we now substitute the definitions of the cost and benefits functions into (5.11), we get

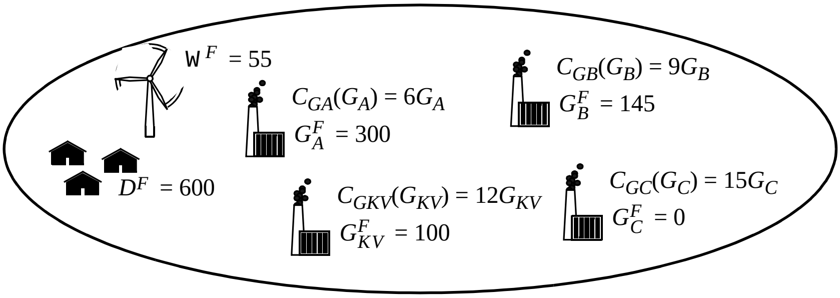
$$\begin{aligned}&\sum_{g \in \mathcal{G}_1}(C_{Gg}(G^{F}_{g}) - C_{Gg}(G^{R}_{g})) - \sum_{g \in \mathcal{G}_2}(C_{Gg}(G^{R}_{g}) - C_{Gg}(G^{F}_{g})) + \\ &+ \sum_{c \in \mathcal{C}_1}(B_{\Delta c}(\Delta^{R}_{c}) - B_{\Delta c}(\Delta^{F}_{c})) - \sum_{c \in \mathcal{C}_2}(B_{\Delta c}(\Delta^{F}_{c}) - B_{\Delta c}(\Delta^{R}_{c})).\end{aligned} \tag{5.12}$$

The results of the ahead market are as mentioned earlier input to the redispatch problem; all terms concerning the ahead market are constants and can be removed from the objective function:

$$\begin{aligned}&\sum_{c \in \mathcal{C}_1} B_{\Delta c}(\Delta^{R}_{c}) + \sum_{c \in \mathcal{C}_2} B_{\Delta c}(\Delta^{R}_{c}) - \sum_{g \in \mathcal{G}_1} C_{Gg}(G^{R}_{g}) - \sum_{g \in \mathcal{G}_2} C_{Gg}(G^{R}_{g}) = \\ &= \{\mathcal{G}_1 \cup \mathcal{G}_2 = \mathcal{G}, \mathcal{C}_1 \cup \mathcal{C}_2 = \mathcal{C},\} = \sum_{c \in \mathcal{C}} B_{\Delta c}(\Delta^{R}_{c}) - \sum_{g \in \mathcal{G}} C_{Gg}(G^{R}_{g}),\end{aligned} \tag{5.13}$$

which is the same expression as in (5.7). Since we have the same constraints (safe operation of the system), the two optimisation problems are equivalent.

Thus, both central dispatch and regulating markets are most simple to simulate by solving an ordinary multi-area problem, but using the real-time values of available generation capacity and price insensitive load etc. If some players cannot or do not want to regulate their generation or load then the corresponding variables in the multi-area problem, G^{R}_{g} and Δ^{R}_{c}, are replaced by constants corresponding to the trading in the ahead market, i.e., G^{F}_{g} and Δ^{F}_{c}



	Redispatch problem					
	1	2	3	4	5	6
Consumers						
Consumption, D [MW]	550	570	580	610	630	610
Wind power producer						
Available capacity, $\overline{W}^R$ [MW]	30	50	70	70	68	60
Generation, W [MW]	30	50	70	70	68	60
Combined heat and power producer						
Available capacity, $\overline{G}_{KV}^R$ [MW]	100	100	100	100	100	100
Generation, G_{KV}^R [MW]	100	100	100	100	100	100
Producer A						
Available capacity, $\overline{G}_A^R$ [MW]	300	300	300	300	300	300
Generation, G_A^R [MW]	300	300	300	300	300	300
Producer B						
Available capacity, $\overline{G}_B^R$ [MW]	150	150	150	150	0	0
Generation, G_B^R [MW]	120	120	110	140	0	0
Activated up-regulation, $G_B^{\uparrow}$ [MW]	0	0	0	30	0	0
Activated down-regulation, $G_B^{\downarrow}$ [MW]	25	0	10	0	0	0
Producer C						
Available capacity, $\overline{G}_C^R$ [MW]	150	150	150	150	150	150
Generation, G_C^R [MW]	0	0	0	0	150	150
Activated up-regulation, $G_C^{\uparrow}$ [MW]	0	0	0	0	150	0
Activated down-regulation, $G_C^{\downarrow}$ [MW]]	0	0	0	0	0	0
Result variables						
Total operation cost, TOC [¤]	680	680	665	710	875	875
Unserved power, ENS [MWh]	0	0	0	0	2	0
Loss of load, $LOLO$	0	0	0	0	1/6	0

Figure 5.2 *Example of simulating manual reserves. The scenario above is divided in a market problem and six ten-minute redispatch problems. The result of the ahead market is shown in the top figure. The generation of the combined heat and power plant is determined by the heat production; therefore no regulation bids have been submitted from this plant. The other dispatchable power plants submits regulation bids and the result of the redispatch problems are displayed above. As can be seen in the table, both load and wind power generation deviate from the forecasted values. Besides, a failure occurs in the power plant of producer B after 40 minutes. The result variables of the entire scenario are* TOC = 4 485 ¤, ENS = 2 MWh *and* $LOLO$ = 1/6.

(cf. the example in figure 5.2). It should also be possible to include extra regulating costs by modifying the cost and benefit functions of the redispatch problem, but I will not consider that problem in this dissertation.

Based on the solutions of the redispatch problems, the values of result variables as *TOC* and *LOLO* can be calculated. If we want to study the surplus of different players, we may also determine real-time prices and imbalances according to the pricing schemes described in chapter 2.

5.2 SHORT-TERM PLANNING

By short-term planning I refer to the process when the players of the electricity market make detailed plans of how much to produce, consume and trade. The time perspective is short—it is about hours and days—and is ultimately decided by technical factors (for example how long time it will take to start up a power plant) and the design of the electricity market (for example how long time that will pass from the closure of the ahead market to the actual hour of delivery).

Planning is in practice the same as solving an optimisation problem; there is a goal which we want to achieve, while we have to consider certain limitations. The multi-area problem, which we use to analyse how the electricity market will behave, corresponds to a simplified short-term planning problem. The difference between the model and the reality is that in real short-term planning, more detailed models are used for power plants and the power system.[16]

If perfect information was available, the players would follow their short-term plans and if the multi-area problem is accepted as a tolerable approximation of the short-term planning problem, we may say that the scenario problem exactly reflects the behaviour of the electricity market. In reality, there will always be small deviations to the short-term planning, due to forecast errors. The question is how to include these deviations in the scenario problem.

Start-up Times

All power plants require some preparation from the moment we decide to start them before they actually are ready to generate power. In many cases the preparation time is negligible; for example, starting a diesel generator set just

16. For example, in the scenario problem a river system with several hydro power plants is represented by one single equivalent power plant, whereas in a short-term planning problem we would normally consider the hydrological couplings between individual power plants (cf. for example [94]).

means pressing a button, the engine starts and the generator is synchronised—a sequence which normally would not take more than a minute. However, large thermal power plants differ in this respect, because they have a boiler which must be heated to its operational temperature before any electricity generation is possible and this process may take up to a couple of hours [31].

When perfect information is available, the start-time will be compensated by starting the power plant sufficiently in advance. Therefore, the availability of a power plant only depends on the technical reliability when simulating an ideal electricity market. The technical reliability is however very hard—if not impossible—to calculate theoretically; hence, we use statistical models based on historical data from the power plant or similar units.

If a power plant due to forecast errors has not been started in time, the consequences will be the same as if the power plant had been subject to a technical error: the generation capacity is not available when needed. The most straightforward method to model the impact of forecast errors in the short-term planning of power plants with significant start-up times is therefore to include both technical errors and planning mistakes in the statistics used to calculate the availability of the power plant.

Forecast errors can also result in starting a power plant, even though it later turns out it was not needed. As some fuel is used for starting the power plant, this kind of planning mistakes cause an extra cost to the owner of the power plant. I hardly think this cost is particularly large and it seems reasonable to use operational logs to estimate an average annual cost of "unnecessary" starts. This cost can then be treated as a fixed operation cost when analysing the results of an electricity market simulation.

Balance Responsibility

Generation in a power system must always be adjusted to the continuous variations of the load. As described in section 5.1, it takes automatic control systems to solve this task, and it would be unreasonable to require that each players maintained his or her own balance in each moment. Therefore, frequency control is considered a part of the infrastructure of the electricity market and it is the duty of the system operator to supply this function. The responsibility of the system operator is however just short-term; eventually anyone who is selling 1 MWh electric energy must either generate or purchase 1 MWh. It is therefore normally required that each player should maintain balance within each trading period, while the system operator is responsible for events within the trading period.

To persuade the players to keep their balance, some players are appointed balance responsible, which means that they have a financial responsibility for any imbalance. A closer description of the notion balance responsibility and how to settle imbalances is found in chapter 2. In this context it is sufficient to establish that forecast errors in the short-term planning will result in the balance responsible players always having larger or smaller imbalances. With a few exceptions, imbalance means increased costs to the balance responsible compared to if perfect information had been available.

The cost of the balance responsibility is however varying between different players. According to [39], the relative costs of balancing power, i.e., the cost of the balance responsibility divided by the total load, are lower for larger balance responsible players than for the smaller players. Among the explanations it is mentioned that large suppliers having many kinds of customers will have relatively smaller load variations as the individual load variations of the consumers will to some extent even out each other; hence, it becomes easier to produce good demand forecasts. Moreover, large players can afford to have 24-hours staff and can therefore continuously revise and correct their demand forecasts. Another difference, which is not mentioned in [39], is that some players also have difficulties in predicting their generation. This is particularly the case for wind power producers, for whom it is not unthinkable that a forecast just a few hours ahead will miss the actual outcome by 100%.[17]

As far as I can judge, the cost of the balance responsibility will however not have any impact on the short-term behaviour of the players; hence, there should not be a need to include this kind of forecast errors in the scenario problem. However, this kind of costs will of course have dynamic effects on the market, because they have an influence on which investments are profitable and how the competition in the electricity market will develop in the long run.

5.3 LONG-TERM PLANNING

Long-term planning is primarily a question for owners of energy limited power plants, because they want to use the stored energy in such a manner that they get paid as much as possible. Other producers may however require long-term forecasts too, for example to plan when to perform major maintenance works.

17. This is partly due to the difficulty of making good wind forecasts, but also because the relation between wind speed and wind power generation is highly non-linear—a small error in the forecasted wind speed may result in a very large error in the forecasted wind power generation. Cf. [96], part 2.

Maintenance Planning

Occasionally, preventive maintenance has to be performed in power plants, which means that the power plant during a longer, continuous time period will not be available for power generation.[18] Unlike corrective maintenance, it is however possible to plan when to perform preventive maintenance and the choice is of course to perform the maintenance during a period when the cost of stopping the plant is as low as possible, which normally means trying to perform preventive maintenance during those periods when electricity prices are low. It is however not guaranteed that the correct period is chosen, as there is an uncertainty in the forecasts of future electricity prices. If a power plant is stopped at an inappropriate time, the total operation cost of the system will increase compared to if perfect information had been available.

Modelling preventive maintenance offers no larger problems. The owners of the power plants stop their power plants when they expect low prices.[19] Thus, there will be a correlation between those scenario parameters which are of large importance to the price (for example load or inflow to the energy storage facilities of the system) and the availability of the power plants. Fortunately, it is fully possible to manage correlated random variables in a Monte Carlo simulation. We may for example use the separation of different time periods described in section 9.2.2.

Seasonal Planning

The owner of an energy limited power plant will of course try to use the available energy at those occasions when they get paid the best prices, while they also must consider that the energy storage facilities have limited capacity. These players need forecasts which extend a shorter or longer time into the future, depending on the relation between the storage capacity and the total inflow. The time span can be months or years—for example, in the Nordic system there are hydro reservoirs which can store about 120 TWh, which is to be compared to the average hydro power generation of about 195 TWh per year [40]. In a smaller system, it might just be necessary to store a few days' or weeks' generation. Regardless of the relevant time perspective, I have chosen to use the term seasonal planning for the problem of determining a long-term operation plan of an energy limited power plant.

18. Transmission lines must also be maintained sometimes, which will have similar consequences as maintenance of power plants—I will therefore not consider grid maintenance under a special heading.
19. Meanwhile, they must of course have make sure that not too many power plants are maintained at the same time.

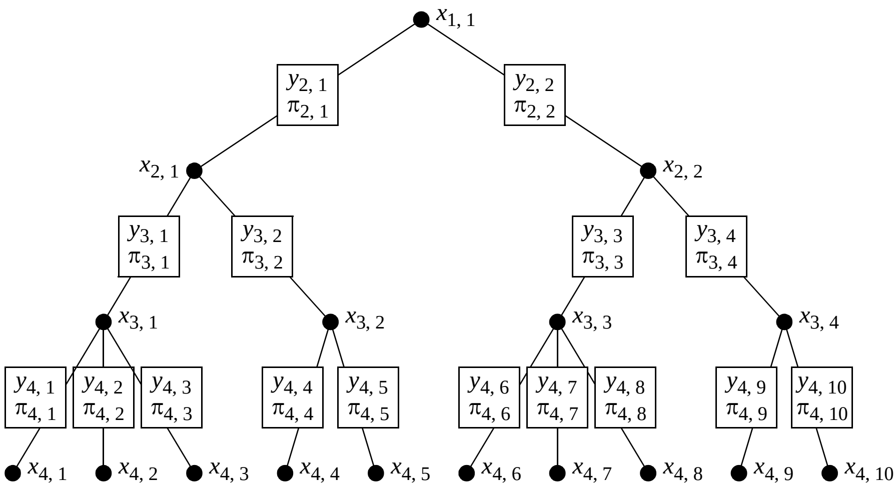


Figure 5.3 *Example of an event tree. The nodes of the tree represent the occasions where decisions are made—the decisions are represented by a vector of decision variables, $x_{t,s}$. All nodes in the same level of the tree belong to the same step, where each step normally represents a certain point of time. Between each step random events occur, the outcome of which ($y_{t,s}$) are represented by new branches starting in the node. Each branch is associated to a particular probability, $\pi_{t,s}$. The difficulty of building an event tree is to choose the discrete outcomes ($y_{t,s}$, $\pi_{t,s}$) in such a manner that a sufficiently good approximation of the underlying stochastic process, $\{Y_t\}$, is obtained. Notice that an event tree does not have to be symmetrical, but it is possible to have different detail levels in different parts of the tree.*

It is obvious that the better the desired approximation, the more branches will be required in the tree. Moreover, the appropriate number of branches depends on the number of elements in the vector Y_t. An event tree can in other words become immensely large when representing fairly accurately the development of multiple random variables during a longer time period—it is common to talk about "the curse of dimensionality".

To perform a seasonal planning, we need a mathematical model representing the forecast. What we have got is a number of random variables, which we for the sake of simplicity collect in one vector, Y. These variables will vary over time; thus, the future is represented by a stochastic process $\{Y_t\}$. When planning, we must somehow consider all possible outcomes of this stochastic process. A common method is to build an *event tree*,[20] which approximately represents how the outcome of different variables varies over time

20. Another term for event tree is "scenario tree". To avoid confusion to a scenario in a Monte Carlo simulation of an electricity market, I prefer the designation event tree, although scenario tree is probably more frequently used in scientific literature.

(cf. figure 5.3). It is not a simple task to build an event tree; cf. [92, 93, 97].

Given an event tree, it is possible to formulate a so-called stochastic programming problem (SP problem), which basically means that the objective is to maximise an expectation value, while considering future developments:

$$\text{maximise} \qquad \textit{expected profit}, \tag{5.14a}$$

$$\text{subject to} \qquad \textit{all possible future developments.} \tag{5.14b}$$

The large challenge about solving SP problems is the size of the problem itself—we have a vector of decision variables for each node in the event tree, and each decision variable corresponds to an optimisation variable in the SP problem. Several methods to solve stochastic programming problems—both general and with specific application to hydro power planning—have been suggested in for example [89, 95, 124], but the fact remains that even when using a brilliant algorithm, solving stochastic programming problems is a time consuming task. To make things worse, the solution of a season planning problem is something perishable. In practice, just the variables of the first step of the event tree are actually used as decision variables, because when the second step is reached, it is likely that improved forecasts are available; making it desirable to modify the event tree accordingly. In practice, a new SP problem is solved before each period. The actual operation plan will be determined by a series of successive season plannings, as illustrated in figure 5.4.

One way of simulating the consequences of forecast errors would be to imitate the actual process of successive season plannings. Such a solution would however not be appropriate for Monte Carlo simulation, because it would mean that for each scenario we have to build a number of event trees and solve the corresponding SP problems; hence, each single scenario would require a tremendous computational effort before any values of *TOC*, *LOLO* and other result variables are obtained

A more convenient solution is to only model the consequences of the season planning of energy limited power plants. Shortly, we may say that the forecast uncertainty sometimes results in players storing more energy than they would have done with perfect information, whereas in other cases they will store less energy. Another way of saying the very same thing is to state that sometimes the energy value will be overestimated and sometimes it will be underestimated. The deviations between forecasted energy value and the energy value if perfect information had been available can be considered random.[21] Such random deviations can be accomplished by adding another term to the objective function of the scenario problem (3.12):

21. Though, possibly they might be correlated to some of the other scenario parameters.

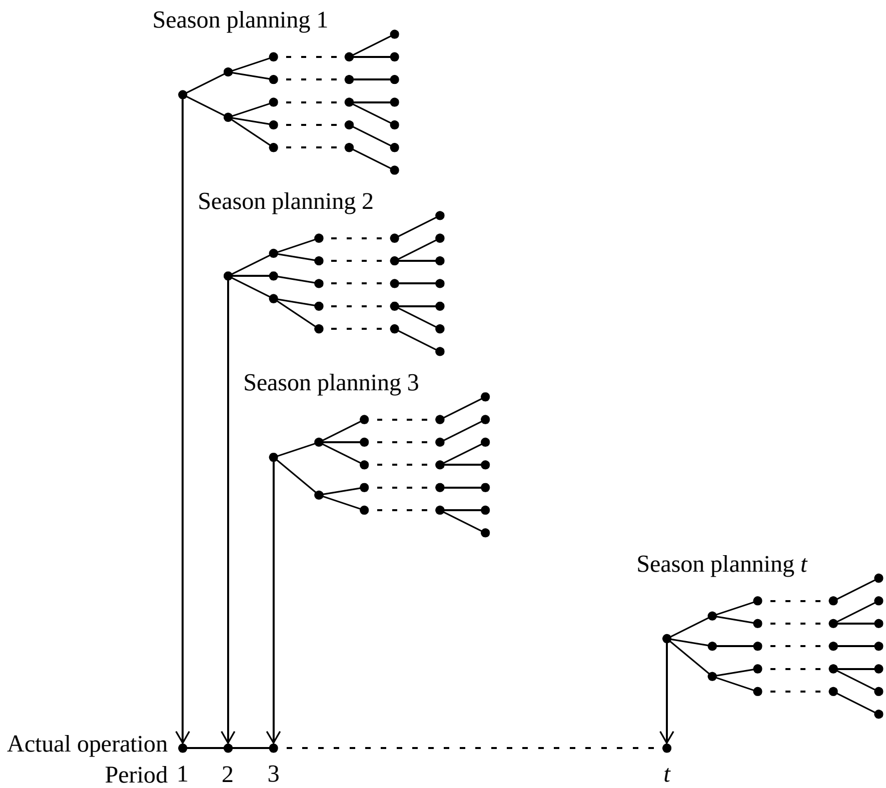


Figure 5.4 *Practical long-term planning. Given the forecast of a player (in the figure represented by an event tree) a seasonal planning is performed. This plan is then used by the player to decide what to do in the first period. However, in the next period, new information has appeared; the future is therefore represented by an up-dated event tree. From this event tree, a new season planning is performed, the result of which is used to determine the actions of period two. This procedure is then repeated for each period.*

$$-\sum_{t=1}^{l} T_t C_{Hr,\,t}(H_{r,\,t}),\tag{5.15}$$

where $C_{Hr,\,t}(H_{r,\,t}) = \beta_{Hr,\,t} H_{r,\,t}$ is a fictitious cost function for the electricity generation in energy limited power plants. That the cost function is fictitious means that it is only meant to be used to modify the behaviour of the electricity market compared to if perfect information had been available; thus, it should *not* be included when calculating *TOC*!

The fictitious cost functions impact on the solution of the scenario problem can be illustrated by studying the optimality conditions. Let $\nu_{r,\,t}$ denote the dual variable of the energy balance constraint (3.14b) of energy storage *r*, period t. The physical interpretation of $\nu_{r,\,t}$ is the *energy value*,[22] i.e., the marginal production cost of the most expensive thermal power plant (or sacrificed marginal benefit in the case of voluntary load reductions) which can be relieved using the stored energy. Moreover, we let $\lambda_{n,\,t}$ denote the electricity price during the period t in the area n, where the energy limited power plant is located; the electricity price is given by the dual variable of the corresponding load balance constraint (3.13b) divided by the period length T_t.

In the case of perfect information, i.e., when the scenario problem is formulated as in chapter 3, we get the following optimality conditions with respect to stored energy, $M_{r,\,t}$:

$$\nu_{r,\,t} \le \nu_{r,\,t-1} \qquad \text{if } M_{r,\,t} = 0, \tag{5.16a}$$

$$\nu_{r,\,t} = \nu_{r,\,t-1} \qquad \text{if } 0 \le M_{r,\,t} \le \overline{M}_{r,\,t}, \tag{5.16b}$$

$$\nu_{r,\,t} \ge \nu_{r,\,t-1} \qquad \text{if } M_{r,\,t} = \overline{M}_{r,\,t}. \tag{5.16c}$$

These conditions state that if the energy storage is empty, the energy value must be less than or equal to the energy value of the proceeding period; otherwise, it would have been better not to empty the storage. Correspondingly, the energy value must be equal or increase if it should be profitable to completely fill up the storage. Normally, when the energy storage is neither completely empty or completely full, the energy value remains the same from one period to another.

Concerning the electricity generation, $H_{r,\,t}$, we get the following optimality conditions:

$$\lambda_{n,\,t} \le \nu_{r,\,t} \qquad \text{if } H_{r,\,t} = 0, \tag{5.17a}$$

$$\lambda_{n,\,t} = \nu_{r,\,t} \qquad \text{if } 0 \le H_{r,\,t} \le \overline{H}_{r,\,t}, \tag{5.17b}$$

$$\lambda_{n,\,t} \ge \nu_{r,\,t} \qquad \text{if } H_{r,\,t} = \overline{H}_{r,\,t}. \tag{5.17c}$$

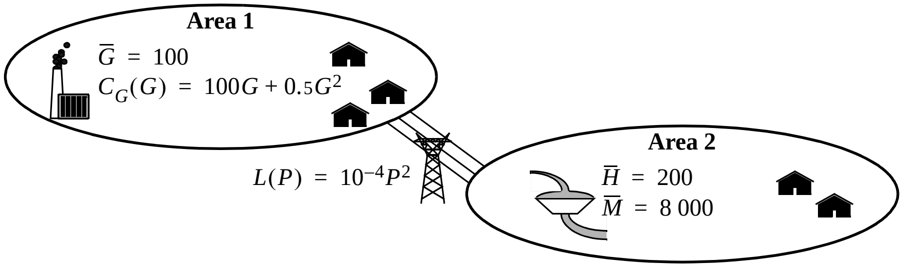
The conclusion of the above conditions is that the energy limited power plant should not be used if the electricity price is less than (or sometimes equal to) the energy value. If on the other hand the electricity price is higher than the energy value then the power plant should be operated at maximal capacity. Finally, the generation can be chosen arbitrarily when electricity price and energy value are equal.

If we include (5.15) in the objective functions, the optimality conditions concerning stored energy remain the same. However, the conditions of the electricity generation, $H_{r,\,t}$, change:

$$\lambda_{n,\,t} \le \nu_{r,\,t} + \beta_{r,\,t} \qquad \text{if } H_{r,\,t} = 0, \tag{5.18a}$$

$$\lambda_{n,\,t} = \nu_{r,\,t} + \beta_{r,\,t} \qquad \text{if } 0 \le H_{r,\,t} \le \overline{H}_{r,\,t}, \tag{5.18b}$$

22. Concerning dispatchable hydro power, the term *water value* is used.



	Period			
	0	1	2	3
Area 1				
Load, D_1 [MWh/h]		100	130	120
Generation, G [MWh/h]				
With perfect information		49.96	50.86	50.56
With forecast errors		47.67	50.53	53.18
Electricity price, λ_1 [¤/MWh]				
With perfect information		149.96	150.86	150.56
With forecast errors		147.67	150.53	153.18
Area 2				
Load, D_2 [MWh/h]		50	60	60
Inflow, Q [MWh]		150	100	120
Storage contents, M [MWh]				
With perfect information	6 000	6 049.71	6 009.93	6 000
With forecast errors	6 000	6 047.39	6 007.28	6 000
Generation, H [MWh/h]				
With perfect information		100.29	139.78	129.93
With forecast errors		102.61	140.12	127.28
Energy value, ν [¤/MWh]				
With perfect information		148.45	148.45	148.45
With forecast errors		148.11	148.11	148.11
Deviation		–2	0	3
Electricity price, λ_2 [¤/MWh]				
With perfect information		148.45	148.45	148.45
With forecast errors		146.11	148.11	151.11
Transmission				
Transmitted power, $P_{2,1}$ [MWh/h]				
With perfect information		50.29	79.78	69.93
With forecast errors		52.61	80.12	67.28
Losses, L [MWh/h]				
With perfect information		0.25	0.64	0.49
With forecast errors		0.28	0.64	0.45

Figure 5.5 *Example of simplified simulation of long-term forecasts errors. The table above shows the solution of the same scenario assuming perfect and incomplete information respectively. The energy storage at the end of period 3 is assumed to be set in advance in both cases.*

$$\lambda_{n,t} \geq \nu_{r,t} + \beta_{r,t} \qquad \text{if } H_{r,t} = \overline{H}_{r,t}. \tag{5.18c}$$

The interpretation of these conditions is similar to the one for (5.17a)-(5.17c), but the difference is that the factitious marginal cost $\beta_{r,t}$ serves as a modifier of the energy value. The fine point is that $\beta_{r,t}$ is not a dual variable, but a scenario parameter. By randomizing a number of values of $\beta_{r,t}$ we will have electricity prices which vary around a certain energy value. An example is shown in figure 5.5. Notice that the variations in the model using random deviations are not centred around the same energy value as obtained with perfect information, which is due to that $\sum\beta_{r,t} \neq 0$ in the example.

Thus, introducing random deviations of the energy values in order to simulate the consequences of errors in the long-term forecasts does not have to make the scenario problem difficult to solve. However, the model presumes that we have a known probability distribution of the deviations and that this probability distribution is chosen so that the consequences are approximately the same as if successive season plannings were performed. Determining an appropriate probability distribution is thus a intricate problem, and further research is required to produce practically applicable methods.

Chapter 6

GRID COSTS

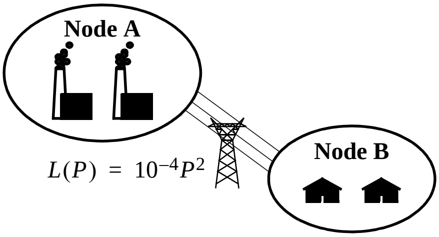
Unlike most other goods, all electricity trading has to pass through an infrastructure, which is common to all players in the market. It is technically difficult to identify and register individual transactions in the common grid. Therefore, the grid can be considered a public good, because all players have access to the services provided by the grid. Moreover, grids are natural monopolies, since it would not be profitable to build parallel, competing grids. As described in section 3.1 both these phenomena result in a risk of inefficient utilisation and decreased benefit to the society. Therefore, it is not self-evident how to design rules, which forces the players of the electricity market to use the grid in a way which is maximizing the benefit to the society.

The costs of the grid can be divided into three parts. Firstly, we have the continuous operation, which means that in every moment the available transmission resources should be utilised in the best possible way. To achieve this, those players who use the grid should pay variable grid tariffs reflecting the losses caused by the players. This issue will be further discussed in section 6.1. On those occasions when the transmission capability is insufficient, the most valuable transactions must be prioritised. Two different methods for so-called congestion management are described in section 6.2.

Secondly, the grid must be maintained, which leads to questions about which level of maintenance is optimal and how the costs should be divided among the users. Thirdly, the grid must be reinforced and expanded in pace with increasing consumption and when new consumers and producers wish to be connected to the grid. Maintenance and grid expansion are briefly treated in section 6.3.

6.1 COST OF LOSSES

All transmission of electric energy causes losses. The costs of these losses are paid by the grid owner, who then must pass them on to the users of the grid;



P_I [MW]	P_{II} [MW]	P [MW]	L [MW]
0	0	0	0
100	0	100	1
0	200	200	4
100	200	300	9

Figure 6.1 *The problem of loss allocation. There are two players, I and II respectively in the example above, and both want to inject power at node A and extract it at node B. If just one of the players used the interconnection, they would cause losses of* 1 *and* 4 MW *respectively. The sum of the individual losses is however not equal to the losses when both players use the interconnection simultaneously, due to the non-linear nature of the losses. The question is how large losses each player has caused in the latter case. There is no natural answer.*

otherwise, losses become an external cost, which causes decreased total surplus (cf. section 3.1.2). It is however not obvious how the grid users should pay the costs of the losses, because there is no unambiguous method to decide how large losses a particular player has caused. It is a technical challenge to measure continuously how the players use the grid (i.e., how much and where they inject energy and how much and where they extract energy). Besides, the physical laws are such that even if it was known how much each player transferred through the grid, it would in most cases not be possible to associate a certain loss to a specific transaction. Generally, there are several ways from a point of injection to the point of extraction, and it is extremely complicated to determine in a fair way which transactions takes which way through the grid. But even if we overcame this problem, there would still be difficulties, as the losses are not a linear function of the transmitted power. This means that the losses when power is transferred between two nodes of the grid will vary depending on what other players are doing (cf. figure 6.1).

In spite of all these problems, is it possible to identify a perfect loss price? The ideal electricity market maximises the benefit to the society, so let us study under which circumstances that goal is achieved. For the sake of simplicity, we only consider one line and we assume that the rest of the grid is already used in the best possible way. The question is thus which transmission from a certain node *n* to another node *m* will maximise the benefit to the society including the cost of the losses caused by the transmission. The total surplus when transmitting $P_{n,\,m}$ MW from node n to node m is

$$TS_{nm} = B_m(P_{n,\,m}) - C_n(P_{n,\,m}) - C_m(L(P_{n,\,m})). \tag{6.1}$$

In this expression $B_m(P_{n,\,m})$ is the value of consuming $P_{n,\,m}$ MW in node m, $C_n(P_{n,\,m})$ is the cost of injecting $P_{n,\,m}$ MW in node n and $C_m(L(P_{n,\,m}))$ is the cost of supplying node m enough power to cover the transmission losses. The surplus of the interconnection between the nodes is maximised when

$$MB_m(P_{n,\,m}) - MC_n(P_{n,\,m}) - MC_m(L(P_{n,\,m}))\cdot ML(P_{n,\,m}) = 0. \tag{6.2}$$

If we assume that all players are price takers, the marginal costs and the marginal benefits should equal the electricity price. In this case we must apparently have different electricity prices in the two nodes:

$$\lambda_n = MC_n(P_{n,\,m}), \tag{6.3a}$$

$$\lambda_m = MC_m(L(P_{n,\,m})) = MB_m(P_{n,\,m}). \tag{6.3b}$$

The relation between the two node prices can be written as

$$\lambda_m - \lambda_n = \lambda_m \cdot ML(P_{n,\,m}). \tag{6.4}$$

Since electricity prices as well as marginal losses are larger than or equal to zero, the electricity price in the importing node m must be higher than the electricity price of the exporting node n, which seems natural. It is interesting that the price difference, which corresponds to the perfect loss price, depends both on the size of the marginal losses and the price paid to compensate losses. Both these quantities are varying continuously; thus, in reality it is impossible to determine the perfect loss prices in advance.

Depending on which method we choose to manage the cost of losses, we will have different deviations between the perfect loss prices and the real, non-ideal prices. To evaluate these deviations and determine their impact on the total surplus, the same system can be simulated both using perfect loss prices and pricing according to some other method. As far as I known, no such studies have been made on real systems, but the principles have been demonstrated for a small fictitious system in [103]. In the following sections I will describe three methods to cover the cost of losses, and simple simulation models of the different methods.

Internalised Cost of Losses

In a centralised electricity market it is the system operator who decides how the power system should be operated and which prices should be applied. The other players reveal their preferences by submitting bids, which are used by the system operator to perform an economic dispatch. If the system operator then includes the losses, for example by solving an optimal power flow problem or a multi-area problem, the economic dispatch will result in price differences between different parts of the system. We may consider the loss price to be internalised in the node or area prices which are obtained in this kind of electricity market.

Although internalised cost of losses have all the qualities to achieve an optimal usage of the resources, far from all centralised electricity markets have utilised this possibility. It rather seems like it is most common to use post allocation of transmission losses, which I will study in more detail later in this chapter. But there are examples of electricity markets where the losses are internalised in the electricity price, for example the national electricity market of Australia and in New Zealand, although they do use simplified loss models [27, 29].

Simulating internalised cost of losses is straightforward; the same economic dispatch tool is used as the one used by the system operator. It can however be difficult to evaluate how well a certain economic dispatch algorithm corresponds to the perfect loss prices. By and large, the price differences between node or area prices will correspond to the loss prices according to (6.4). There are however two factors which may cause minor deviations: one is that the dispatch might be based on a somewhat simplified power system model and the other is that the dispatch is not updated in real-time. The first problem can be studied by comparing the results of a particular dispatch algorithm with those of another algorithm, which is using a more detailed power system model. The impact of the length of the trading period is preferably studied by comparing how the same algorithm behaves if the length of the trading period is reduced by for example 50%. However, if the comparison should be fair, the frequency control should be included in the simulation (see section 5.1), because shorter trading periods will not only result in reduced costs of losses, but also a decreased need for frequency control.

Feed-in Tariffs

In a bilateral electricity market, loss pricing is more complicated, because the electricity price is determined by the producers and consumers themselves, which prevents the system operator from directly internalising the cost of losses in the same way as is possible in a centralised electricity market. The cost of losses rather has to be internalised indirectly, which the system operator can do by charging the grid users for the losses they cause.[1] One method

1. For the sake of simplicity, I assume that the system operator is also the grid owner. Nothing essential would change in the reasoning if these functions are divided between several players.

to cover the cost of losses is to introduce a feed-in tariff. In each connection point, a feed-in tariff is specified, which states how much the grid users will have to pay for each MWh inserted at the node. The method is straightforward to apply and works both on bilateral electricity markets and centralised electricity market, where for some reason the costs of losses are not internalised in the economic dispatch.

The feed-in tariffs of the different nodes can be either positive or negative. Positive feed-in tariffs mean that the grid users have to pay for injection, whereas extraction is rewarded by the system operator. Negative tariffs work the other way around; the system operator pays for injection and requires payment for extraction.[2] Some grid users will in other words get paid for using the grids, which at a cursory glance might seem odd, but there is a simple technical explanation: if energy is injected to a net importing node then the import will decrease; hence, the losses will also decrease. The payment received by the grid user is thus a reward for decreasing the losses. In a net exporting node it is instead extraction which results in decreased losses.

It is possible to choose the feed-in tariffs so that they provide the same results as perfect loss prices. Let τ_n denote the feed-in tariff in node n. A price taking producer connected to this node will then solve the following player problem:

$$\text{maximise} \quad \lambda G_g - C_{Gg}(G_g) - \tau_n G_g, \tag{6.5}$$

$$\text{subject to} \quad 0 \le G_g \le \bar{G}_g. \tag{6.5a}$$

The optimality conditions of this problem are

$$MC_{Gg}(G_g) \ge \lambda - \tau_n, \qquad \text{if } G_g = 0, \tag{6.6a}$$

$$MC_{Gg}(G_g) = \lambda - \tau_n, \qquad \text{if } 0 < G_g < \bar{G}_g, \tag{6.6b}$$

$$MC_{Gg}(G_g) \le \lambda - \tau_n, \qquad \text{if } G_g = \bar{G}_g. \tag{6.6c}$$

A positive feed-in tariff will as already mentioned result in an income for a consumer, which gives us the following player problem:

$$\text{maximise} \quad B_{\Delta c}(\Delta_c) + \tau_n \Delta_c - \lambda \Delta_c, \tag{6.7}$$

$$\text{subject to} \quad 0 \le \Delta_c \le \Delta c, \tag{6.7a}$$

and the optimality conditions

$$MB_{\Delta c}(\Delta_c) \le \lambda - \tau_n \qquad \text{if } \Delta_c = 0, \tag{6.8a}$$

$$MB_{\Delta c}(\Delta_c) = \lambda - \tau_n \qquad \text{if } 0 \le \Delta_c \le \Delta c, \tag{6.8b}$$

2. If desirable, it is of course possible to define extraction tariffs in a similar way; the extraction tariffs simply get the opposite sign of the feed-in tariffs. Within the Nordel area, feed-in tariffs have been chosen, although Svenska kraftnät uses the more neutral “energy tariff” in their price list [112].

$$MB_{\Delta c}(\Delta_c) \geq \lambda - \tau_n \qquad \text{if } \Delta_c = \Delta c. \tag{6.8c}$$

We see that the feed-in tariff acts as a modifier of the system price (i.e., the electricity price of the whole market). As the feed-in tariff varies from node to node, we will in practice have separate node prices for each part of the grid. If the node prices are to be equivalent to the perfect loss prices then we must choose feed-in tariffs according to (6.4), i.e.,

$$(\lambda - \tau_n) = (\lambda - \tau_m) - (\lambda - \tau_m) \cdot ML_{n,\,m}(P_{n,\,m}), \tag{6.9}$$

which can be rewritten as

$$\tau_m = \lambda - \frac{\lambda - \tau_n}{1 - ML_{n,\,m}(P_{n,\,m})}. \tag{6.10}$$

Apparently there is a certain degree of freedom when choosing the feed-in tariffs, because the perfect loss prices only determine the difference between the tariffs of two nodes. Some node in the system has to be appointed reference node; then all feed-in tariffs of the other nodes can be calculated in relation to the reference node. Assume that node *n* is the reference node and that the feed-in tariff of the reference node has been chosen to $\tau_n = 0$. It then follows from (6.10) that

$$\tau_m = \lambda - \frac{\lambda}{1 - ML_{n,\,m}(P_{n,\,m})} = \lambda \cdot \frac{-ML_{n,\,m}(P_{n,\,m})}{1 - ML_{n,\,m}(P_{n,\,m})}. \tag{6.11}$$

The feed-in tariff can thus be written as the electricity price multiplied by a sensitivity coefficient. As node *m* is importing from the reference node when $P_{n,\,m} > 0$, the sensitivity coefficient must be negative (injection decreases the losses, extraction increases them).

Feed-in tariffs are used in the Nordic electricity market.[3] Sensitivity coefficients have been calculated for each point of connection to the Nordic main grid and for four different periods (peak load workdays, peak load other days, low load workdays, low load other days). The sensitivity coefficients show how much injection in a certain point affect the total Nordic losses, which corresponds to the term $-ML_{n,\,,m}(P_{n,\,,m})/(1 - ML_{n,\,,m}(P_{n,\,,m}))$ in (6.11).[4] Each period is also associated to an electricity price, which is based on the price which the Nordic system operators pay for loss energy; these prices correspond to λ in (6.11). Sensitivity coefficients and electricity prices are updated regularly,[5] but will of course pretty often deviate more or less from the perfect loss prices.

When simulating feed-in tariffs, we must include the tariffs in the player problems of producers and consumers, as shown above. Moreover, a player

3. See for example [112].
4. The calculation of the sensitivity coefficients are described in more detail in [104].

problem has to be formulated for the system operator. As the system operator generally is not allowed to own production facilities, they have to buy the loss energy from other players. These purchases can be managed in several ways; among others, the system operator could sign a take-and-pay contract with a retailer[6] or perform the purchase in the real-time market. A detailed model of the different options for loss energy purchases would be a little bit too extensive for this presentation, so I will restrict myself to assuming that the system operator has perfect information and therefore can buy all losses in the ahead market, paying the same electricity price, λ, as all other players. The system operator will of course try to minimise the cost of the losses, while buying as much loss energy as required to maintain the balance between production and consumption in each part of the system:

$$\text{minimise} \quad \lambda \sum_{(n,m) \in \mathcal{P}} L_{n,m}(P_{n,m}) \tag{6.12}$$

$$\text{subject to} \quad \sum_{g \in \mathcal{G}_n} G_g(\lambda) + \sum_{m \in \mathcal{P}_{n \leftarrow m}} (P_{m,n} - L_{m,n}(P_{m,n}))$$

$$- \sum_{c \in \mathcal{C}_n} \Delta_c(\lambda) - \sum_{m \in \mathcal{P}_{n \to m}} P_{n,m} = 0, \qquad \forall\, n \in \mathcal{N}, \tag{6.12a}$$

$$0 \le P_{n,m} \le \bar{P}_{n,m}, \qquad \forall\, (n,m) \in \mathcal{P}. \tag{6.12b}$$

The optimality conditions of (6.12) looks as follows:

$$\lambda \cdot ML_{n,m}(P_{n,m}) - \mu_m(1 - ML_{n,m}(P_{n,m})) + \mu_n \ge 0 \qquad \text{if } P_{n,m} = 0, \tag{6.13a}$$

$$\lambda \cdot ML_{n,m}(P_{n,m}) - \mu_m(1 - ML_{n,m}(P_{n,m})) + \mu_n \ge 0 \qquad \text{if } 0 \le P_{n,m} \le \bar{P}_{n,m}, \tag{6.13b}$$

$$\lambda \cdot ML_{n,m}(P_{n,m}) - \mu_m(1 - ML_{n,m}(P_{n,m})) + \mu_n \ge 0 \qquad \text{if } P_{n,m} = \bar{P}_{n,m}. \tag{6.13c}$$

The market price, λ, is given by the link between the player problems of the

5. However, the Nordic system operators have somewhat different desires concerning how often the feed-in tariffs should be up-dated. Norway prefers to update the tariffs at least once a year, whereas Finland and Sweden prefer stable feed-in tariffs, which are updated about every third year [111].
6. A take-and-pay contract means that the consumer may consume any amount of power up to a certain limit. The consumption does not have to be constant, but can vary during the duration of the contract, as long as the maximal power is not exceeded. Cf. [25], p. 34f.

system operator, producers and consumers, i.e., that the total production should be balanced by the total consumption (including the loss purchases of the system operator):

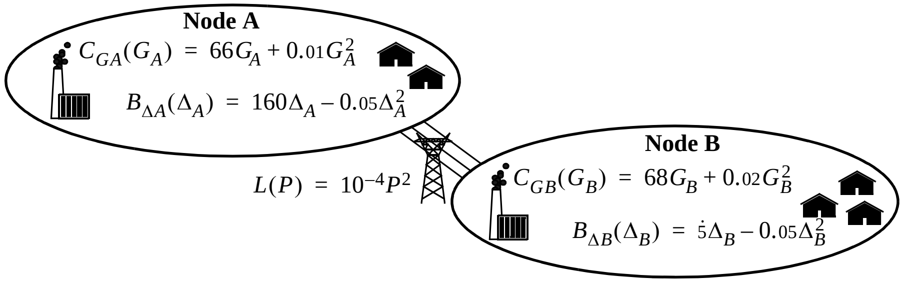
$$\sum_{g \in G} G_g - \sum_{c \in C} \Delta_c - \sum_{(n, m) \in P} L_{n, m}(P_{n, m}) = 0. \tag{6.14}$$

The market balance constraint, the optimality conditions of the producer, consumer and system operator problems, (6.6a)-(6.6c), (6.8a)-(6.8c) and (6.13a)-(6.13c), as well as the variable limits and constraints of these problems form a system of equations and inequalities, the solution of which describes how the system will be used.[7] A detail in this context is that the area balances (6.12a) and the market balance are linearly dependent. We therefore have to fix the dual variable μ_n in some area.[8] On the other hand, the dual variables have no practical usage and we may quite as well disregard (6.13a)-(6.14) and not bother to compute μ_n.

A simple example of how feed-in tariffs influence an electricity market is given in figure 6.2. The example shows that it does not matter which node is selected as reference node; as long as the difference between the feed-in tariffs corresponds to the perfect loss price, the market will behave in the same way as an ideal electricity market. The only difference between different choices of reference node is the system price, which is adjusted up- or downwards so that the practical price (system price minus feed-in tariff) remains the same. However, if the difference between the feed-in tariffs deviate from the perfect loss price, the practical prices change and with that the production and consumption. In the figure, a scenario is shown where the difference between the feed-in tariffs is too small, which results in an increased transmission between the areas. This favours the producers in the exporting node and the consumers of the importing node. In total there will be a small decrease of the total surplus.

---

7. The same solution can also be obtained using the iterative method described in [103]. In brief, the method starts from a market problem, where producers and consumers can trade freely. From this trade, the system operator chooses the power flows $P_{n, m}$ to minimise the losses. Now a new market problem can be solved, where the system operator enters as a consumer, buying energy corresponding to $\Sigma L_{n, m}$. As a result, the demand increases; hence, the market price increases too. The new trading results in changes in the power flows, which causes new losses. We may continue to iterate in this manner until the values of λ and $P_{n, m}$ has converged.
8. This is a recurrent phenomenon of so-called transport problem. Cf. for example [132], p. 253.



Scenario parameters	Ideal			
Feed-in tariff A, τ_A [¤/MWh]	electricity	0	10	9
Feed-in tariff B, τ_B [¤/MWh]	market	–10	0	0
Electricity prices				
System price, λ [¤/MWh]		90.00	100.00	99.39
Node A, λ_A [¤/MWh]	90.00			
Node B, λ_B [¤/MWh]	100.00			
Consumers in node A				
Consumption, Δ_A [MWh/h]	700.00	700.00	700.00	696.06
Marginal benefit, $MB_{\Delta A}(\Delta_A)$ [¤/MWh]	90.00	90.00	90.00	90.39
Surplus,* $B_{\Delta A}(\Delta_A) + \tau_A\Delta_A - \lambda\Delta_A$ [¤/h]	24 500.00	24 500.00	24 500.00	24 225.06
Producers in node A				
Generation, G_A [MWh/h]	1 200.00	1 200.00	1 200.00	1 219.69
Marginal cost, $MC_{GA}(G_A)$ [¤/MWh]	90.00	90.00	90.00	90.39
Surplus,* $\lambda G_A - C_{GA}(G_A) - \tau_A G_A$ [¤/h]	14 400.00	14 400.00	14 400.00	14 876.54
Consumers in node B				
Consumption, Δ_B [MWh/h]	1 275.00	1 275.00	1 275.00	1 281.06
Marginal benefit, $MB_{\Delta B}(\Delta_B)$ [¤/MWh]	100.00	100.00	100.00	99.39
Surplus,* $B_{\Delta B}(\Delta_B) + \tau_B\Delta_B - \lambda\Delta_B$ [¤/h]	81 281.25	81 281.25	81 281.25	82 055.88
Producers in node B				
Generation, G_B [MWh/h]	800.00	800.00	800.00	784.85
Marginal cost, $MC_{GB}(G_B)$ [¤/MWh]	100.00	100.00	100.00	99.39
Surplus,* $\lambda G_B - C_{GB}(G_B) - \tau_B G_B$ [¤/h]	12 800.00	12 800.00	12 800.00	12 319.70
System operator				
Transmission, $P_{A,B}$ [MWh/h	500.00	500.00	500.00	523.63
Losses, $L_{A,B}(P_{A,B})$ [MWh/h]	25.00	25.00	25.00	27.42
Surplus,** $\tau_A(G_A - \Delta_A) + \tau_B(G_B - \Delta_B) - \lambda L_{A,B}(P_{A,B})$ [¤/h]	2 500.00	2 500.00	2 500.00	1 987.40
Total surplus, $B(\Delta_A) + B(\Delta_B) - C(G_A) - C(G_B)$ [¤/h]	135 481.25	135 481.25	135 481.25	135 464.38

* In the ideal electricity market the node prices (λ_A and λ_B) are used instead of the system price. No feed-in tariffs are used; hence, $\tau_A = \tau_B = 0$.

** In the ideal electricity market the system operator buys $P_{A,B}$ MWh/h in node A and sells $P_{A,B} - L_{A,B}(P_{A,B})$ MWh/h in node B, which yields the surplus $\lambda_B(P_{A,B} - L_{A,B}(P_{A,B})) - \lambda_A P_{A,B}$.

Figure 6.2 *Example of an electricity market using feed-in tariffs.*

Post Allocation of Transmission Losses

When using feed-in tariffs, the losses are charged in advance (in the sense that the loss prices are announced before each trading period—the actual payment can be done at a later occasion). The opposite, i.e., when prices are determined after the trading period, is also practicable in both bilateral and centralised markets. The principle is simple; after each trading period, the system losses are known and each player is made responsible for a certain share of the losses. In practice, the losses are thus treated as consumption in the balance of each player and are paid to market price or the price of balancing power, which in its turn is based on the market price.[9]

The problem of post allocation is that there is no self-evident method to determine which player caused a certain loss (cf. figure 6.1). There are a number of alternative methods of calculation to be used (see for example [99, 101, 102, 106, 107]) and it is just to choose the one, which is considered to give the most desirable signals to the grid users. Please notice that “desirable signals” is a very subjective judgement in this context, because the different methods may have widely differing impacts on different players (cf. [101]). Although post allocation of transmission losses thus always can be criticised for disadvantaging some player, the method is used in many electricity markets. Two examples are Spain and England-Wales [100, 105].

The choice of loss allocation algorithm does not affect the actual operation of the system (unless the players include forecasts of their share of the losses in their player problems; more about that later in this section), but only the financial transactions between different players; therefore, I find it unnecessary to summarise all methods here and I restrict myself to demonstrating the principle of how to simulate them. As example I have chosen proportional allocation, because it is the most simple method.[10] In proportional allocation half of the losses are allocated to the producers and the other half to the consumers. Within the groups, each player receives a share corresponding to their share of the total production and consumption respectively, i.e.,

$$L_{Gg} = \frac{1}{2} L_{tot} \frac{G_g}{\sum\limits_{k \in \mathcal{G}} G_k}, \tag{6.15a}$$

9. Cf. chapter 2.
10. Which does not mean necessarily mean that it is the best method—rather the opposite according to [101].

$$L_{\Delta c} = \frac{1}{2} L_{tot} \frac{\Delta_c}{\sum_{k \in C} \Delta_k}. \tag{6.15b}$$

The trading in the ahead market is simulated in the same way as for feed-in tariffs, if it is assumed that the players' behaviour in the ahead market is not affected by the cost of losses they will receive later. Thus, we may use the same model as for feed-in tariffs, but with the difference that all τ_n are set to zero.

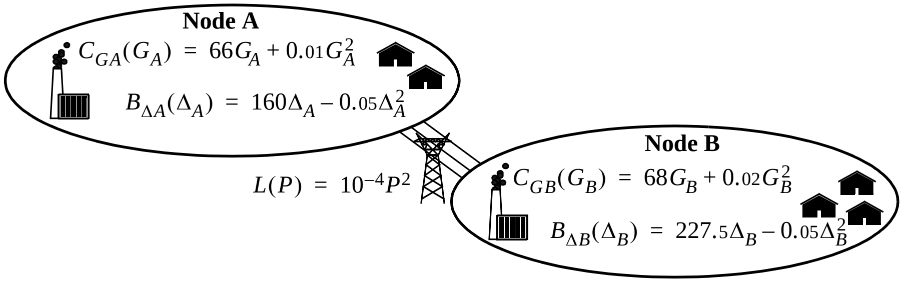
An example of post allocation of transmission losses is given in figure 6.3. The input is the same as in the example of feed-in tariffs, which enables us to make direct comparisons. Apparently, post allocation produces a lower total surplus than the feed-in tariffs, even if the feed-in tariffs deviate somewhat from the perfect loss price—the difference would however obviously be smaller if a worse choice of feed-in tariffs had been made. Notice that post allocation produces the same total surplus regardless of which loss allocation method we have chosen; if we change the loss allocation, there will just be a redistribution of the cost of losses between different players, but the sum will remain the same.

With feed-in tariffs we have already seen that there is a theoretical possibility to achieve the same public welfare as an ideal electricity market, and we may wonder if this also applies to post allocation. Assume for example that the players should include an expected cost of losses in their player problems, so that for example the producer problem would be formulated as

$$\text{maximise} \quad \lambda G_g - \lambda \varphi_{Gg} G_g - C_g(G_g), \tag{6.16}$$

$$\text{subject to} \quad 0 \le G_g \le \bar{G}_g, \tag{6.16a}$$

where the parameter φ_{Gg} is the producer's expected share of the losses. If the producer should expect $\lambda \cdot \varphi_{Gg}$ to be equal to the feed-in tariff, τ_n, in figure 6.2 then we would get the same result for feed-in tariffs and post allocation. However, it is impossible to fulfil this condition for all players, as long as they have at least approximately correct forecast about the loss allocation; the total cost of the grid users is less in post allocation compared to using feed-in tariffs. The difference is that in post allocation the grid users pay an average price corresponding to the system operator's cost of losses, whereas feed-in tariffs are marginal cost based pricing, which generates a surplus to the system operator (cf. figures 6.2 and 6.3). Post allocation of losses therefore lacks the possibility to be as efficient as an ideal electricity market. The method may however be more efficient than poorly chosen feed-in tariffs.



	Ideal electricity market	Proportional loss allocation
Electricity prices		
System price, λ [¤/MWh]		94.00
Node A, λ_A [¤/MWh]	90.00	
Node B, λ_B [¤/MWh]	100.00	
Consumers in node A		
Consumption, Δ_A [MWh/h]	700.00	660.03
Allocated losses, $L_{\Delta A}$ [¤/MWh]	0.00	9.05
Surplus,* $B_{\Delta A}(\Delta_A) - \lambda(\Delta_A + L_{\Delta A})$ [¤/h]	24 500.00	20 930.79
Producers in node A		
Generation, G_A [MWh/h]	1 200.00	1 399.86
Allocated losses, L_{GA} [¤/MWh]	0.00	18.69
Surplus,* $\lambda(G_A - L_{GA}) - C_{GA}(G_A)$ [¤/h]	14 400.00	17 839.28
Consumers in node B		
Consumption, Δ_B [MWh/h]	1 275.00	1 335.03
Allocated losses, $L_{\Delta B}$ [¤/MWh]	0.00	18.31
Surplus,* $B_{\Delta B}(\Delta_B) - \lambda(\Delta_B + L_{\Delta B})$ [¤/h]	81 281.25	87 393.57
Producers in node B		
Generation, G_B [MWh/h]	800.00	649.93
Allocated losses, L_{GB} [¤/MWh]	0.00	8.68
Surplus,* $\lambda(G_B - L_{GB}) - C_{GB}(G_B)$ [¤/h]	12 800.00	7 632.53
System operator		
Transmission, $P_{A,B}$ [MWh/h	500.00	739.83
Losses, $L_{A,B}(P_{A,B})$ [MWh/h]	25.00	54.73
Surplus** [¤/h]	2 500.00	0.00
Total surplus, $B_{\Delta A}(\Delta_A) + B_{\Delta B}(\Delta_B) - C_{GA}(G_A) - C_{GB}(G_B)$ [¤/h]	135 481.25	133 796.17

* In the ideal electricity market the node prices (λ_A and λ_B) are used instead of the system pride.

** In the ideal electricity market the system operator buys $P_{A,B}$ MWh/h in node A and sells $P_{A,B} - L_{A,B}(P_{A,B})$ MWh/h in node B, which yields the surplus $\lambda_B(P_{A,B} - L_{A,B}(P_{A,B})) - \lambda_A P_{A,B}$. In the case of post allocation of the losses, the system operator receives exactly as much as they have paid for the losses.

Figure 6.3 *Example of an electricity market using post allocation of transmission losses.*

6.2 CONGESTION MANAGEMENT

A well designed grid should normally have sufficient transmission capability to accommodate all desired transactions. Thus, during normal operation the grid is a public good—one player using the grid is not blocking the possibility of the other players to use the grid. Under certain conditions, for example when the grid is heavily loaded or if there are disturbances in important transmission lines, some parts of the grid will reach their capacity limit. The available transmission capability through these cuts must then somehow be rationed.

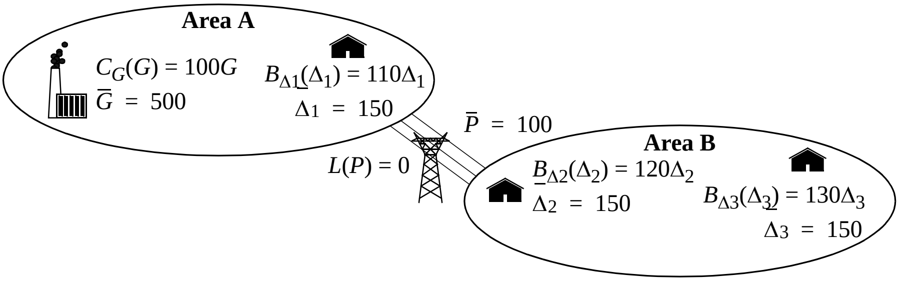
There are several methods to solve this so-called congestion management problem; figure 6.4 provides some examples. The different methods can be divided in market based and other solutions [108]. The core of the market-based solutions is that the transmission capability through a congested cut is a private good, i.e., if one player is using a part of the available capacity then the available capability for other players is decreased.[11] In other words, the transmission capability can be efficiently distributed by somehow allowing the players to buy and sell transmission capability to each other. The two basics principles of a market based solution are counter trading and market splitting. These two methods are described in more detail below.

In the other methods, it is not the players' valuation of the transmission capability which decides which transmission should be given priority, but some other criterion. It might of course be possible to find advantages with such methods, but there is also an obvious risk that the total surplus is decreased (this is for example the case in figure 6.4). As there seems to be a general agreement that market-based congestion management is preferable,[12] I have chosen not to model the other methods.

Market Splitting

One of the basic conditions of a free market is that the players can trade freely, This is obviously not the case when there are congestion problems and therefore it seems natural to consider the areas on both sides of the congested cut as two separate markets, where each market has its own electricity price.

11. The transmission capability of a grid is thus an example of a good which is partly public and partly private. Cf. chapter 3, footnote 5.
12. The EU has for example stated that "Network congestion problems shall be addressed with non-discriminatory market based solutions which give efficient economic signals to the market participants and transmission system operators involved." [109]



Consumer	Contract signed	Period of validity
1	1/6 10:00	1/6 10:00 - 1/6 11:00
2	1/1 9:00	1/1 0:00 - 31/12 24:00
3	1/6 11:00	1/6 10:00 - 1/6 11:00

Method	Actions	Generation [MWh/h]	Consumption [MWh/h]		
		G	Δ_1	Δ_2	Δ_3
Market splitting	Consider the two areas as separate markets.*	250	150	0	100
Counter trading	The system operator pays the consumers in area B to decrease their consumption until the transmission capability is not exceeded.*	250	150	0	100
Pro rata	All consumers in area B have to decrees their consumption by the same share.	250	150	50	50
First come, first serve	The consumers who submit their purchase bids first are prioritised.	250	150	100	0
Type of contract	Long-term contracts are prioritised before short-term contracts.	250	150	100	0

* More elaborate examples of market splitting and counter trading are found in figures 6.5-6.7.

Figure 6.4 *Examples of congestion management. If the transmission capability was not limited, the electricity price would be* 100 ¤/MWh *and all three consumers should consume their maximal load, i.e.,* $\Delta_1 = \Delta_2 = \Delta_3 = 150$ MWh/h. *Due to the limitation, the consumption in area B must be decreased to* 100 MWh/h. *The table above briefly describes some methods to achieve this. The solution which maximised the benefit to the society is to prioritise the consumption which has the highest marginal benefit, i.e.,* Δ_3. *The two market based solutions, counter trading and market splitting, result in this solution, whereas the other methods produce a lower total surplus.*

This basic idea can be implemented in some different ways.[13] Sometimes congestion actually occurs on the border between two markets (this is for obvious reasons most common concerning international interconnections) having different system operators, different trading periods, different trading arrangements, etc. On each side of the congested cut, there will be a separate electricity price, λ_n and λ_m, according to the rules valid for each market. If we neglect the losses and assume that the player who owns the interconnection is a price taker in both markets then the owner will try to maximise the income of buying power at one market and selling it at the other:

$$\text{maximise} \quad (\lambda_n - \lambda_m)P_{n,m} + (\lambda_m - \lambda_n)P_{m,n} \tag{6.17}$$

$$\text{subject to} \quad 0 \le P_{n,m} \le \overline{P}_{n,m}, \tag{6.17a}$$

$$0 \le P_{m,n} \le \overline{P}_{m,n}. \tag{6.17b}$$

The above player problem is not at all depending on who the owner is of the right to use the interconnection—it does not have to be the grid owner who trades via the interconnection, but the transmission right might be distributed using various kinds of auctions [108]. However, the surplus of individual players will be affected by how the transmission rights are made available.

The producers and consumers behave as usual; they try to maximise their profits given the price of the area where they are located:

$$\text{maximise} \quad \lambda_n G_g - C_{Gg}(G_g), \tag{6.18}$$

$$\text{subject to} \quad 0 \le G_g \le \overline{G}_g, \tag{6.18a}$$

and

$$\text{maximise} \quad B_{\Delta c}(\Delta_c) - \lambda_n \Delta_c, \tag{6.19}$$

$$\text{subject to} \quad 0 \le \Delta_c \le \Delta_c. \tag{6.19a}$$

If the congestion occurs in the middle of an electricity market, a division in separate market prices can be enforce by introducing price areas. A condition is that all trading between the price areas is performed at a power pool (trading within a price area may still be bilateral). The price areas must be defined before the trading is started in the power pool, which however is not a practical problem, as there is no need to change the price area division unless the grid is expanded or some other major change occurs. Each bid submitted to the power pool should state in which price area injection or extraction will occur. The power pool accepts the bids maximizing the total surplus considering the transmission limitations between the price areas. In other words, the power pool solves a multi-area problem, where each price area corresponds to

13. Sometimes different forms of market splitting are considered as separate methods—see for example [108, 110]—but I prefer to see them as variants of the same method.

an area of the multi-area problem. In the following analysis it is assumed that the power pool does not consider the transmission losses between the price areas. If all producers and consumers are price takers, they will bid their real cost and benefit functions, which yields the following optimisation problem:

$$\text{maximise} \quad \sum_{c \in C} B_{\Delta c}(\Delta_c) - \sum_{g \in G} C_{Gg}(G_g) \tag{6.20}$$

$$\text{subject to} \quad \sum_{g \in G_n} G_g + \sum_{m \in P_{n \leftarrow m}} P_{m,n} = \sum_{c \in C_n} \Delta_c + \sum_{m \in P_{n \leftarrow m}} P_{n,m}, \qquad \forall\, n \in N, \tag{6.20a}$$

$$0 \le \Delta_c \le \Delta c, \qquad \forall\, c \in C, \tag{6.20b}$$

$$0 \le G_g \le \overline{G}_g, \qquad \forall\, g \in G, \tag{6.20c}$$

$$0 \le P_{n,m} \le \overline{P}_{n,m}, \qquad \forall\, (n, m) \in P. \tag{6.20d}$$

It can be noted that regardless of whether there are two or more completely separate markets, as in (6.17)-(6.19), or if there is one market divided in several price areas, as in (6.20), the optimality conditions will result in the same system of equations and inequalities, showing what the trading of the ahead market will become:[14]

$$\sum_{g \in G_n} G_g + \sum_{m \in P_{n \leftarrow m}} P_{m,n} = \sum_{c \in C_n} \Delta_c + \sum_{m \in P_{n \leftarrow m}} P_{n,m}, \qquad \forall\, n \in N, \tag{6.21a}$$

$$MC_{Gg}(G_g) \ge \lambda_n, \qquad \text{if } G_g = 0, \tag{6.21b}$$

$$MC_{Gg}(G_g) = \lambda_n, \qquad \text{if } 0 < G_g < \overline{G}_g, \tag{6.21c}$$

$$MC_{Gg}(G_g) \le \lambda_n, \qquad \text{if } G_g = \overline{G}_g, \tag{6.21d}$$

$$MB_{\Delta c}(\Delta_c) \le \lambda_n \qquad \text{if } \Delta_c = 0, \tag{6.21e}$$

$$MB_{\Delta c}(\Delta_c) = \lambda_n \qquad \text{if } 0 \le \Delta_c \le \Delta c, \tag{6.21f}$$

$$MB_{\Delta c}(\Delta_c) \ge \lambda_n \qquad \text{if } \Delta_c = \Delta c, \tag{6.21g}$$

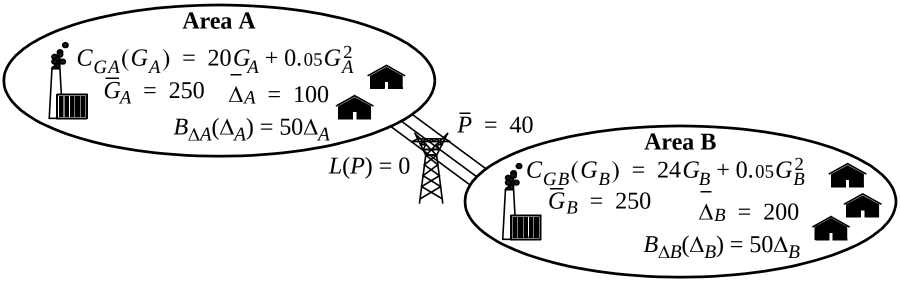
$$\lambda_n \ge \lambda_m, \qquad \text{if } P_{n,m} = 0, \tag{6.21h}$$

$$\lambda_n = \lambda_m, \qquad \text{if } 0 < P_{n,m} < \overline{P}_{n,m}, \tag{6.21i}$$

$$\lambda_n \le \lambda_m, \qquad \text{if } P_{n,m} = \overline{P}_{n,m}. \tag{6.21j}$$

The interpretation of (6.21a)-(6.21j) is quite trivial. Each price area will

14. The constraints and variable limits of (6.17)-(6.19) and (6.20) respectively is also included in the optimality conditions, but has been left out to somewhat shorten the enumeration of the optimality conditions.



	Area A	Area B
Electricity price, λ [¤/MWh]	34	40
Generation, G [MWh/h]	140	160
Consumption, D [MWh/h]	100	200
Transmission, $P_{A,B}$ [MWh/h]	40	

Figure 6.5 *Example of market splitting. The economic results of the different players are shown in figure 6.7.*

have its own area price. If an interconnection is not fully utilised then the price will be the same in both areas; if desirable, the two areas may be seen as one. In those cases when the power flow between two areas reaches the maximal value, there will be a lower area price in the exporting area and a higher price in the importing area.

Market splitting cannot solve all congestion problems, because the method only affects the interchange between the areas; internal limitations within a price area have to be solved with another method (which in practice means counter trading if market based solutions are to be used). This problem can be avoided by using nodal pricing, i.e., by letting each connection point to the grid become a separate price area.

There is also an information problem, because the area prices are determined in the ahead market. Here there are some practical differences depending on how the market splitting is accomplished. If there are separate markets, the player who owns the transmission right must be able to predict the electricity prices on both sides of the interconnection to be able to trade in such a manner that the interconnection is used in the best possible way. We do not have the same problem in a central power pool, because the decision of how to use the transmission lines is made simultaneously as the prices are determined in each area. Common for both methods is however that it is not possible to account for changes which occur after the closure of the ahead market. There is in other words a certain risk that the market is split although it later turns out that there was no congestion and vice versa. To consider this uncertainty in the simulation, the real-time market has to be simulated, as described in section 5.1.

An example of market splitting is provided in figure 6.5. Each area corresponds to a price area. In area A there is a surplus of inexpensive generation capacity, but due to the transmission limitation only 40 MWh/h can be transmitted to area B. The result of the ahead market is therefore a lower electricity price in area A and a higher price in area B.

Counter Trading

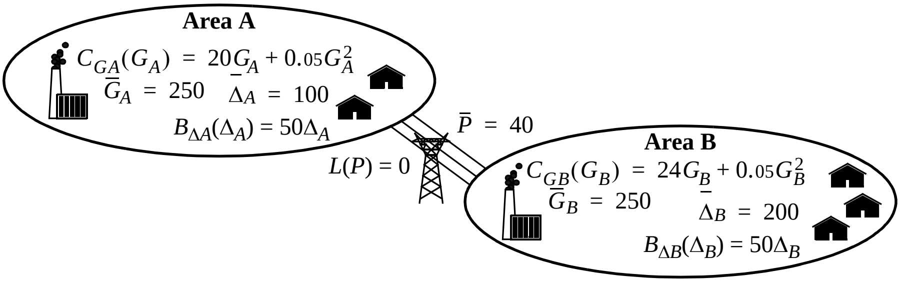
If counter trading is used, the system operator chooses which transactions should be prioritised when there are congestion problems. Those players who have the least benefit from using the congested interconnection are persuaded to change their production or consumption. At the exporting side, a power plant can decrease the generation or a consumer can increase the consumption; in both cases the net export is reduced. At the importing side, the net import should be reduced, which can be done by increasing the generation in a power plant or by decreasing the load of a consumer.

To modify the production and consumption in different parts of the system, the system operator uses the real-time market.[15] In theory each up regulation on the importing side should correspond to an equally large down regulation on the exporting side, but in practice it is common to combine counter trading and frequency control. If for example the frequency is too low, it is sufficient to perform an up regulation on the importing side. By that means, we both solve the congestion problem, while supplying more generation, which makes the frequency increase.

Obviously, counter trading can be performed in real-time, which of course increases the possibilities to utilise the grid in the best possible way; we are not only eliminating the risk that forecast errors influence the congestion management, but we also gain the possibility to redistribute production and consumption during a part of a trading period, if the congestion problem should only appear during a shorter time. Moreover, counter trading can be applied to congestion problems everywhere in the system and the method can be used both in bilateral and centralised electricity markets.

To simulate counter trading, we use the same model as for manually activated frequency control reserves (see section 5.1). Simulating the real-time trading in detail will however in many cases be unnecessarily detailed. The model can then be simplified by assuming that the players in the ahead market have perfect information about what will happen during a trading period, and that generation and consumption do not deviate more from the ahead

15. Sometimes, counter trading is differentiated between the case when the system operator may order regulation actions and when they have to use a regulating market. This is for example the case in [108, 110], but I prefer to consider them as two variants of the same method.



	Area A	Area B
Market problem		
Generation, G^F [MWh/h]	170	130
Consumption, D^F [MWh/h]	100	200
Transmission, $P_{A,\,B}$ [MWh/h]	70	
Electricity price, λ [¤/MWh]	37	
Redispatch problem		
Producers		
Generation, G^R [MWh/h]	140	160
Activated up-regulation, $G^{\uparrow}$ [MWh/h]	0	30
Activated down-regulation, $G^{\downarrow}$ [MWh/h]	30	0
Consumption, D^R [MWh/h]	100	200
Transmission, $P_{A,\,B}$ [MWh/h]	40	
Up-regulation price, $\lambda^{\uparrow}$ [¤/MWh]	37	40
Down-regulation price, $\lambda^{\downarrow}$ [¤/MWh]	34	37

Figure 6.6 *Example of counter trading. The result of the trading in the ahead market is a transmission of* 70 MWh/h *between the two areas, but the transmission capability is just* 40 MWh/h*. The system operator has to activate a down-regulation of* 30 MWh/h *in area A and an equally large up-regulation in area B. The player who is willing to pay the highest price when purchasing regulating power is activated in area A, i.e., the producer. In area B, it is the player who will require the least payment when selling regulating power, who is activated; here too, regulating the generation is preferable.*

The economic results of the different players are shown in figure 6.7.

trading than what can be managed by the automatic frequency control. Hence, we will only need one redispatch problem per trading period and the same input is used in the redispatch problem as in the market problem—the difference is that in the market problem we disregard the transmission limitations, whereas we include them in the redispatch problem. By comparing the solutions of the two problems, we can determine which regulation actions have been performed and which regulating power prices should be applied. An example of this simulation method is shown in figure 6.6. In the market problem, the players trade freely, without considering the transmission limitation between the two areas. The result is a transmission which exceeds the available capability by 30 MWh/h. As the consumers in the example are practically price insensitive, the only solution is to decrease generation in area A by 30 MWh/h and increase it by the same amount in area B.

Impact on Individual Players

Generally, I focus on the benefit to the society in this dissertation, when I analyse different market designs. If we assume perfect information and compare market splitting and counter trading, we find that they produce the same total surplus in the short run. The explanation is that the same producers and consumers, whose trading over congested interconnections is not accepted when the market is split, will be performing up- and down-regulation in the counter trading (presuming that all players active in the ahead market also participate in the real-time trading and that they do not change their bids). In each scenario, we will end up with exactly the same production and consumption, regardless of which method has been used for congestion management; hence, the total surplus will of course be the same. I will not prove this claim, but the reader has to be content with the verification provided by the example in figure 6.7. I would however like to point out some interesting observations about how the choice between market splitting and counter trading affects individual players.

Let us study an electricity market, which due to congestion problems has been divided in N price areas. We assume that congestion only appears between price areas, i.e., there are no internal congestion problems within any price area. Also assume that the losses are negligible and that all players are price takers. In each area there is a cost function for the supply, denoted $C_{Gn}(G_n)$. There is also a benefit function for the demand, but to simplify the reasoning, we assume that the demand is constant and equal to D_n. The benefit of consumption is given by $B_{Dn}(D_n) = \beta_{Dn} D_n$, where the parameter β_{Dn} is chosen so that the marginal benefit of consumption is always larger than the most expensive power plant of the system. Basically, this assumption is equivalent to assuming that the load is not price sensitive—the difference is that if the load would be really price insensitive then $B_{Dn}(D_n)$ would be infinite and when the benefit of consumption is not well-defined, the calculation of the total surplus would be misleading. The validity of the reasoning is not affected by the assumption, because price sensitive load can always be modelled as price insensitive load combined by a fictitious power plant, corresponding to load reductions (see section 3.2.1).

Between the different price areas there is a physical transmission capability, $\hat{P}_{n,m}$. If the transmission is exceeding this limit, counter trading will be applied up to a certain limit, $B_{n,m}$, and then the market will be split. The ahead market results in an area price, λ_n, in each price area; the exact area prices depend on $\hat{P}_{n,\,m}$ and $B_{n,\,m}$, but are of no importance in this reasoning.

The counter trading is assumed to be performed in a regulating market and all trading in the ahead market is passed on to this regulation market, without any additional regulation costs.[16] Moreover, there is an up-regulation price $\lambda_n^{\uparrow}$, which is paid to all activated up-regulation bids in the area, and a down-regulation price, $\lambda_n^{\downarrow}$, which is paid to all activated down-regulation prices in the area. The exact regulating prices do not have any further importance either.

Let us now study the surplus of each player. To simplify the notation, I use the symbol

$$\sum_{m_i}$$

to denote summation over all nodes m which are importing from another node n. In a similar manner, the index m_e is used to denote the nodes which are exporting to node n.

The consumers in this example are in practice not price sensitive and will therefore not submit any bids to the regulating market. The consumers' surplus in each area is therefore the value of consumption minus the purchase cost, i.e.,

$$CS_n = B_{Dn}(D_n) - \lambda_n D_n = \beta_{Dn} D_n - \lambda_n D_n. \tag{6.22}$$

The incomes and costs of the producers originate both from the trading in the ahead market and the trading of regulating power. Let us study these terms separately. The power sold in each price area is equal to the consumption within the area plus the total export from the area minus total import to the area, where the total export and import correspond to the physical transmission capability plus counter trade. All trading in the ahead market is using the area price and the income can thus be written as

$$\lambda_n G_n^F = \lambda_n \left(D_n + \sum_{m_i} (\hat{P}_{n,\,m_i} + B_{n,\,m_i}) - \sum_{m_e} (\hat{P}_{m_e,\,n} + B_{m_e,\,n}) \right). \tag{6.23a}$$

The producers also gain income from selling regulating power to the system operator. The counter trading means that the system operator needs to buy regulating power in the importing price area. The price of this regulating power is equal to the up-regulation price; hence, the income of the producers is

16. By this is meant that if a producer for example has a power plant with the generation capacity 100 MW, where 60 MW is sold in the ahead market, the producer will submit an up-regulation bid of 40 MW and a down-regulation bid of 60 MW to the regulating market.

Figure 6.7 *Comparison of counter trading and market splitting. If no counter trading is used ($B_{A,B} = 0$) then the solution of figure 6.5 is obtained. Unlimited counter trading yields the same results as in figure 6.6.*

Scenario parameters			
Maximal counter trading, $B_{A,B}$ [MWh/h]	0	20	∞
The market			
System operator			
Income of sales to the ahead market, $\lambda_B(\hat{P}_{A,B} + B_{A,B})$ [¤/h]	1 600	2 280	0
Income of sold regulating power, $\lambda_A^{\downarrow} \cdot \Delta G_A^{\downarrow}$ [¤/h]	0	680	1 020
Cost of purchase from the ahead market, $\lambda_A(\hat{P}_{A,B} + B_{A,B})$ [¤/h]	1 360	2 160	0
Cost of buying regulating power, $\lambda_B^{\uparrow} \cdot \Delta G_B^{\uparrow}$ [¤/h]	0	800	1 200
Surplus, MS [¤/h]	240	0	–180
Total surplus, TS [¤/h]	12 100	12 100	12 100
Area A			
Area price, λ_A [¤/MWh]	34	36	37
Down-regulation price, $\lambda_A^{\downarrow}$ [¤/MWh]	—	34	34
Consumers			
Consumption, D_A [MWh/h]	100	100	100
Value of consumption, $50D_A$ [¤/h]	5 000	5 000	5 000
Purchase cost, $\lambda_A \cdot D_A$ [¤/h]	3 400	3 600	3 700
Surplus, CS_A [¤/h]	1 600	1 400	1 300
Producers			
Sales in the ahead market, G_A^F [MWh/h]	140	160	170
Generation, G_A^R [MWh/h]	140	140	140
Down-regulation, $G_A^{\downarrow}$ [MWh/h]	0	20	30
Income from sales in the ahead market, $\lambda_A G_A^F$ [¤/h]	4 760	5 760	6 290
Generation cost, $C_{GA}(G_A^R)$ [¤/h]	3 780	3 780	3 780
Cost of buying regulating power, $\lambda_A^{\downarrow} \cdot G_A^{\downarrow}$ [¤/h]	0	680	1 020
Surplus, PS_A [¤/h]	980	1 300	1 490

(continues next page)

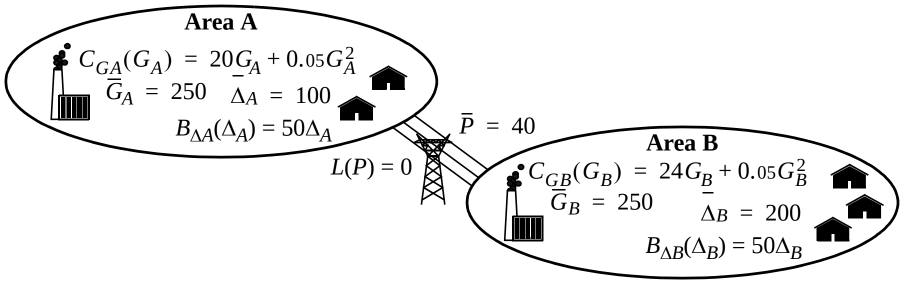
$$\lambda_n^{\uparrow} G_n^{\uparrow} = \lambda_n^{\uparrow} \sum_{m_e} B_{m_e, n}. \tag{6.23b}$$

In the exporting price area, the system operator is selling regulating power to down-regulation price. The producers cost of buying regulating power is

$$\lambda_n^{\downarrow} G_n^{\downarrow} = \lambda_n^{\downarrow} \sum_{m_i} B_{n, m_i}. \tag{6.23c}$$

Finally, we have the generation cost of the producers, which is

Figur 6.7 (cont.)



Scenario parameters Maximal counter trading, $B_{A,\,B}$ [MWh/h]	0	20	∞
Area B			
Area price, λ_B [¤/MWh]	40	38	37
Up-regulation price, $\lambda_B^{\uparrow}$ [¤/MWh]	—	40	40
Consumers			
Consumption, D_B [MWh/h]	200	200	200
Value of consumption, $50D_B$ [¤/h]	10 000	10 000	10 000
Purchase cost, $\lambda_B{\cdot}D_B$ [¤/h]	3 400	3 600	3 700
Surplus, CS_B [¤/h]	6 600	6 400	6 300
Producers			
Sales in the ahead market, G_B^F [MWh/h]	160	140	130
Generation, G_B^R [MWh/h]	160	160	160
Up-regulation, $G_B^{\uparrow}$ [MWh/h]	0	20	30
Income of sales in the ahead market, $\lambda_B G_B^F$ [¤/h]	6 400	5 320	4 810
Income of sold regulating power, $\lambda_B^{\uparrow} \cdot G_B^{\uparrow}$ [¤/h]	0	800	1 200
Generation cost, $C_{GB}(G_B^R)$ [¤/h]	5 120	5 120	5 120
Surplus, PS_B [¤/h]	1 280	1 000	890

$$C_{Gn}(G_n^F + G_n^{\uparrow} - G_n^{\downarrow}) = C_{Gn}\Big(D_n + \sum_{m_i} \hat{P}_{n,\,m_i} - \sum_{m_e} \hat{P}_{m_e,\,n}\Big), \tag{6.23d}$$

because the actually generation in area n consists of the consumption in the price area plus physical export minus physical import. If we denote the actual generation by G_n^R and summarise the income minus the costs according to (6.23a)-(6.23d), we get the producers’ surplus

$$PS_n = \lambda_n D_n - C_{Gn}(G_n^R) + \sum_{m_e}(\lambda_n^{\uparrow} - \lambda_n)B_{m_e,\,n} + \sum_{m_i}(\lambda_n - \lambda_n^{\downarrow})B_{n,\,m_i}. \tag{6.24}$$

Market splitting means that the system operator sells a certain amount of electricity in the deficit areas to the corresponding area price, which produces an income:

$$\sum_{n} \lambda_n \sum_{m_e} (\hat{P}_{m_e, n} + B_{m_e, n}). \tag{6.25a}$$

The corresponding amount has been bought in the surplus areas to the corresponding area prices, which causes the following cost:

$$\sum_{n} \lambda_n \sum_{m_i} (\hat{P}_{n, m_i} + B_{n, m_i}). \tag{6.25b}$$

Counter trading provides income to the system operator when selling regulating power:

$$\sum_{n} \lambda_n^{\downarrow} \sum_{m_i} B_{n, m_i}. \tag{6.25c}$$

In those cases when the system operator is buying regulating power, there will be a cost instead:

$$\sum_{n} \lambda_n^{\uparrow} \sum_{m_e} B_{m_e, n}. \tag{6.25d}$$

The surplus of the system operator can thus be written

$$\begin{aligned} MS \; &= \sum_{(n, m)} (\lambda_m - \lambda_n) \hat{P}_{n, m} \\ &- \sum_{n} \left(\sum_{m_e} (\lambda_n^{\uparrow} - \lambda_n) B_{m_e, n} + \sum_{m_i} (\lambda_n - \lambda_n^{\downarrow}) B_{n, m_i} \right). \end{aligned} \tag{6.26}$$

Summarizing (6.22), (6.24) and (6.26) yields the total surplus:

$$TS = \sum_{n} (CS_n + PS_n) + MS \; = \sum_{n} (B_n(D_n) - C_n(G_n^R)). \tag{6.27}$$

As expected, all terms depending on the prices and the amount of counter trade disappear; hence, the total surplus is only depending on the supply, demand and physical transmission capability between the price areas.

How are then the individual players affected? The larger the maximal counter trading, the more rare will market splitting become; thus, there will be a trend towards uniform area prices in the whole system. This is of course beneficial to those consumers who are located in those price areas which regularly imports, because those consumers would have to pay a higher area price if the market was split. Uniform prices also favour producers in price areas which are likely to export, because they would sell for a lower area price when the market is split. We may in other words conclude that counter trading favours consumers in import areas and producers in export areas, whereas consumers in export areas and producers in import areas are disadvantaged (cf. figure 6.7). Here we see a disadvantage of counter trading—which will only appear in the long run—since counter trading tends to direct new generation to export areas and new consumption to import areas, which eventually increases the need for transmission capability across the congested interconnection. However, this disadvantage can be compensated by fixed fees, which I will return to in section 6.3.

Counter trading is also profitable to the players, whose regulating bids are activated, regardless of where in the system they are located. This is seen in (6.24), where the two last terms always are positive, because the up-regulation price always is at least as large as the area price $(\lambda_n^{\uparrow} \geq \lambda_n)$ and the down-regulation price is always at most as high as the area price $(\lambda_n \leq \lambda_n^{\downarrow})$.

The incomes which the counter trading is generating for the players performing up- and down-regulations are paid by the system operator, who therefore will have higher costs. However, when the market is split, the system operator receives an income, because they always buy for a low area price and sell to a higher price; thus, the first term in (6.26) is positive. If only market splitting is used, the system operator will have a positive surplus, which can be used to cover other costs (see section 6.3 below). Conversely, if only counter trading is used, the system operator will have negative surplus, which must be compensated somehow.

If there would be a player with a dominating position in a certain price area (but not in the whole market), counter trading reduces this player's influence in the ahead market. On the other hand, the player remains dominant in the real-time market (if pricing is free in that market), but there market power has a smaller impact on the other players than it has in the ahead market, because the regulating prices only affect the players involved in regulation actions or who have an imbalance in the post market (cf. chapter 2).

6.3 OTHER GRID COSTS

In addition to the direct operation costs related to losses and congestion, the grid also requires maintenance and investments. An ideal electricity market is not only efficient in the short run, but also in the long run. This means that in an ideal electricity market, exactly so much maintenance is performed as needed and the grid is expanded in correct pace and in the right way. In reality it is unfortunately difficult to determine what is sufficient maintenance and a proper pace of investments, and it can safely be assumed that the ideal model and the reality will differ. However, these deviations are of no importance to the kind of static electricity market simulation, which I am dealing with in this dissertation; maintenance and investment costs are mostly fixed costs in the short run and we have to include market dynamics in the simulation model to study those costs. Although I therefore do not intend to present any models of maintenance and investments, I would like to comment briefly upon a few details in this context.

When discussing maintenance, we differentiate between corrective maintenance and preventive maintenance. Corrective maintenance refers to reparations which are performed after the failure of a component. The costs of corrective maintenance is thus a variable cost, which could be included in the total operation cost, *TOC*. In practice, such a solution would cause unnecessary difficulties in a simulation. To begin with, we would have to define at which moment the reparation cost should be counted; when the component fails, when it is repaired or some other time? We also have to include time in the scenarios we study; in a short scenario (i.e., a snap-shot of the electricity market) we have no information about how long time an unavailable component has been out of order. If we include the time, which according to my terminology means that we simulate long scenarios, the actual simulation will become more difficult to perform (see chapter 10). A better method to include the costs of corrective maintenance is to introduce a separate system index for expected maintenance costs and determine or estimate this index outside the electricity market simulation.[17]

The objective of the preventive maintenance is to reduce the risk of disturbances, for example by "exercising" components which are rarely used or by replacing ageing (or otherwise damaged) components before they completely cease to function. Assuredly, the material costs of preventive maintenance may vary somewhat, but by and large, the preventive maintenance can be considered a fixed cost per year; thus, this cost does not have to be included inn the simulation part. If anything, the result of the preventive maintenance affects the input of the electricity market simulation; the more maintenance is done, the higher the availability of transmission lines and power plants should become. Determining this relation and identifying which preventive maintenance is the most important for the reliability of the system is a—to say the least—complicated problem, which I will not address any further.[18] Investments cause fixed costs in a similar way as preventive maintenance. (The boundary between the two is not very clear—should for example a major overhaul of an over-aged distribution grid be seen as maintenance or a new investment?)

Apparently, there are quite a lot of fixed costs for the grid. How should the

17. Cf. for example the case study in my licentiate theses [7], where I quite simply assumed that the maintenance costs of grid and some power plants were an annual cost corresponding to 5% of the original investment.
18. Those who are interested in these issued are referred to [98].

grid users be charged to cover these costs? Depending on how the cost of losses are covered and which congesting management method is chosen, the running operation may either generate a surplus to the system operator (for example when using marginal pricing of losses or market splitting to solve congestion problems) or a deficit (for example as a consequence of counter trading). But even if the final result of the system operator is a gain, this gain is normally not sufficient to finance the maintenance and investment costs of the grid [111].[19] In other words, additional fees are required to provide full cost coverage for the system operator. These fees (which in [111] are referred to as *residual tariffs*) can be designed more or less arbitrarily, but the most common is probably to have a fixed, annual fee, which is proportional to the maximal injection or extraction of the grid user, or—in the case of an investment cost—as a fixed single payment.

The residual tariffs do not have to be uniform for all grid users, but can be differentiated both between different parts of the grid and between different categories of grid users. One reason to have different residual tariffs in different parts of the grid is that the cost of connecting a certain maximal power varies; the cost per connected grid user is obviously higher the fewer connections there are per km of line. Moreover, it was shown in the previous section that counter trading tends to favour generation in export areas and consumption in import areas, which causes an increased need of transmission resources and hence higher fixed grid costs. If this kind of geographical variations is not reflected in the grid tariffs then there will be a risk that new power plants and new load centres are placed in the wrong part of the system, resulting in a decrease of the benefit to the society, due to the increased costs of maintenance and investments in the grid.

On the other hand, access to electricity supply is of such fundamental importance to our modern society, that there might be good reasons to subsidise the grid costs in sparsely populated areas, if the grid costs otherwise would become so high that it would affect the possibility to finance basic public functions. It is however possible to keep the fixed grid costs quite uniform for most grid users, without giving up the possibility to provide economic signals about where new investments should be located. The solution is to differentiate the residual tariffs for different groups of grid users.

The cost of grid connection is probably of less importance for most smaller load centres; a new residential area is build where there is a housing shortage, not where access to the existing grid is as close as possible, and this is probably beneficial to the society. Large consumers, for example energy intense industries, and—above all—large producers may nevertheless have approximately the same benefits and costs regardless of where new facilities are

19. I here assume, as in the previous sections, that the system operator also is the grid owner.

built, and then the residual tariff may actually have an impact on the investment decision. It is more important to persuade the larger grid users to locate their facilities in the right part of the grid—it is self-evident that a new nuclear power plant of 1 600 MW has a considerably larger impact on the need for grid investments than a weekend cottage with a 16 A main fuse. Thus, it can be justified to differentiate the residual tariffs, so that small grid users pay uniform tariffs, while the larger gird users have a tariff system which to a larger extent reflects geographical differences in the grid costs. However, when small and large grid users are separated in this way, it is important that the share of residual tariffs which is paid by the small grid users is not too large, because the variations of the tariffs of the large grid users might then become so small that it does not affect their investment decisions (cf. [143]).

Chapter 7

OTHER MARKET IMPERFECTIONS

As I have shown in chapter 3, a number of conditions must be fulfilled if an electricity market should be considered ideal. In the previous chapters, I have studied how external costs, uncertain forecasts and grid tariffs differ between ideal and real electricity markets. If all the electricity markets in the world were examined then countless other differences between reality and the ideal model would be revealed. Such a complete survey is for natural reasons too extensive for a dissertation. However, there are a few additional imperfections which I would like to comment upon briefly, but without describing or modelling them in detail.

7.1 MARKET POWER

Most electricity markets are dominated by a single or a few players. This is partly a heritage from the days of the vertically integrated monopolies and partly a consequence of the economics of scale in power generation. Undoubtedly, possibilities to exercise market power arise when the production resources are concentrated like this, and it is not surprising that this issue has attracted a lot of attention the last few years; [113-121] is just a small selection of all the scientific papers about market power in electricity markets which have been published. Here, I will restrict myself to a few general comments on the subject of market power and the most important models, which can be used.

As already known, market power arises when a player has such a large share of the market that he or she becomes a price setter, i.e., a player who can influence the market price. The price setter has not total control over the price, but must still consider the reactions of the other players, but the control is sufficient to allow the price setter to increase his or her own benefit (compared to an ideal electricity market) on behalf of the total surplus. An example of how this was given in figure 3.4.

In this section I will assume that market power is exercised by producers. This is not an attempt to discredit the industry, but just a way to avoid lengthy descriptions covering all possible cases. That I choose producers as example and not consumers is because most electricity markets have many small and medium consumers and quite a few large producers, which makes it more realistic that it is the producers who may possess market power.

Legislation

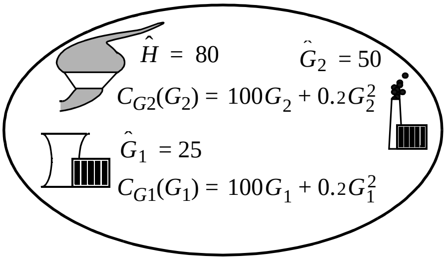
Any in-depth discussion of the legislation concerning market power should not be necessary—exercising market power reduces the total surplus and is therefore forbidden.[1] Generally, the legislation is designed so that it is not prohibited to be a dominant player in a market; what is prohibited is using the dominating position in such a manner that other players are damaged.

It should however be noted that the regulatory authorities face a major challenge when surveilling the electricity market. Admittedly, afterwards it is easy to point out a certain action as suspected price manipulation, but then it remains to prove that the action deliberately was performed to influence the price. It is however difficult to exclude other possible explanations to the behaviour of the player. (Anyone who wants to hide attempted price manipulations can probably every now and then blame forecast errors; cf. figure 7.1.) To convict someone for exercising market power, it is likely that written evidence of irregularities have to be found in a raid at the suspected player. This happened for example in the Enron bankruptcy when U.S. authorities confiscated documents showing that Enron developed and also practised various strategies to exploit weaknesses of the Californian electricity market [122].

Although it thus can be feared that there are plentiful opportunities—for those who want—to fairly unnoticed exercise market power, the situation may yet not be that bad, because the risks of the dominating companies also have to be accounted for. After all, if they are caught, the punishment can be severe and they will hardly gain any sympathies from the general public. This applies especially to those power companies which are still entirely or partly state-owned. I find it hard to believe that any politician in a working democracy would neither want to nor dare to sanction a state-owned power company fleecing the electricity consumers of the nation and it seems just as unlikely that the board of the company would initiate such a risky venture.

I could imagine that in most electricity markets there are players who possess market power, but they choose to use it only in a small scale. Proving or rejecting this thesis however requires more detailed studies of this issue.

1. Of course, this does not only apply to trading in the electricity market, but all other goods, too.



Out-come, s	Proba-bility, π_s [%]	Available hydro generation, $\bar{H}$ [TWh]	
		Forecast 1	Forecast 2
1	25	55	56
2	50	60	61
3	25	65	66

Figure 7.1 *Example of the difficulty of proving price manipulation. Consider an electricity market where the electricity is generated by dispatchable hydro power, nuclear power and other thermal power plants. Assume that there are neither generation capacity limitations, transmission limitations nor reservoir capacity limitations, resulting in a constant electricity price during the whole year. The load of the system is price insensitive and the demand is always* 100 TWh/year. *In this market the electricity price is determined by the inflow to the hydro reservoirs.*

Assume that one single company owns all nuclear power plants. The company musts before the beginning of the year decide how much fuel they should load their nuclear power plants with. If the nuclear power plants are fully loaded, it is possible to use the entire technical potential, $\hat{G}_1$, *but it is also possible to set the available capacity,* $\bar{G}_1$, *to a lower value. The problem is that any unused fuel causes a cost due to lost income from interest. Assume that the storage cost is* $\beta_K = 5$ ¤/MWh. *In the other thermal power plants, the fuel storage is filled as time goes; hence, the entire technical potential is available each year, without extra storage costs.*

If there is perfect competition in this market, the producers would minimise the expected operation cost. Assume there are $\mathcal{S}$ *discrete outcomes of the inflow and that each inflow is associated with a certain probability,* π_s, *and a certain available hydro power generation,* $\bar{H}_s$. *The market can then be simulated by solving the following problem:*

$$
\begin{array}{lll}
\text{minimise} & \sum_{s \in \mathcal{S}} \pi_s (C_{G1}(G_{1,s}) + C_{G2}(G_{2,s}) + \beta_K K_s) & \\
\text{subject to} & \bar{G}_1 - G_{1,s} - K_s = 0, & \forall s \in \mathcal{S}, \\
 & H_s + G_{1,s} + G_{2,s} = D, & \forall s \in \mathcal{S}, \\
 & 0 \le H_s \le \bar{H}_s, & \forall s \in \mathcal{S}, \\
 & 0 \le G_{g,s} & \forall g \in \mathcal{G}, s \in \mathcal{S}, \\
 & 0 \le K_s & \forall s \in \mathcal{S}.
\end{array}
$$

If this problem is solved using data according to the first forecast above, we find that the nuclear power plants should be loaded with 21,88 TWh *fuel. Assume that the company just loads* 21,04 TWh, *which is the optimal solution if the second forecast is used instead. Does this mean that the company is trying to increase the market price by withholding some nuclear capacity? In our example, we may define the first forecast as the correct one, and conclude that the company deliberately overestimated the hydro power generation when using the second forecast, but in reality it would be very hard to show which forecast corresponds best to the real probability distribution.*

Modelling

In an ideal electricity market the actions of the producers are only depending on the market price and the marginal costs of the producers (cf. the optimality conditions of the producer problems in chapter 4-6). Price setters have the possibility to deviate from this behaviour and in that way increase their own surplus on behalf of the other players. A price setter may choose between many different strategies. To model strategic behaviour, we end up in a branch of mathematics called game theory. A market can be considered as a game, where the players choose their moves (i.e., the bids they submit to the market) depending on what they believe the other players will do. The objective of the game is to gain the largest possible surplus.

There are many ways to design a game model of a market and the choice of model may have a significant impact on the final result. This is not just about the game model itself (i.e., the strategies of the players and how the game is played[2]) but also the extent to which the model includes the special conditions in an electricity market (for example that a comparatively small player can have market power, because the competitors are blocked by transmission congestion problems). Below follows a brief summary of the general models found in economic literature, as well as some comments about their application to the electricity market.

- **Bertrand model.** A Bertrand producer chooses a price and then sells as much as is demanded at this price [137, 140]. In many cases a Bertrand model produces the same result as a perfectly competitive market,[3] but if there is a large demand, which is not very price sensitive, the producer can increase their surplus on behalf of the consumers (see figure 7.2). This kind of behaviour is said to have been observed in the centralised electricity market in England-Wales [117, 119].
- **Cournot model.** The Cournot producer chooses a quantity and then adjusts the price until the market demands the chosen quantity [137, 140]. The result of a Cournot model is somewhere between the monopoly case and perfect competition [137, 140]. The Cournot model is frequently used when simulating electricity markets; see for example [114, 120, 121].
- **Stackelberg model.** Both the Bertrand and Cournot models as-

2. It is for example a considerable difference between a game which is played in a single round compared to if there are multiple rounds, allowing the players to study each others moves and adept to the strategies of the other players. Cf. [140], section 15.6.
3. Cf. the example in [137], p. 504ff.

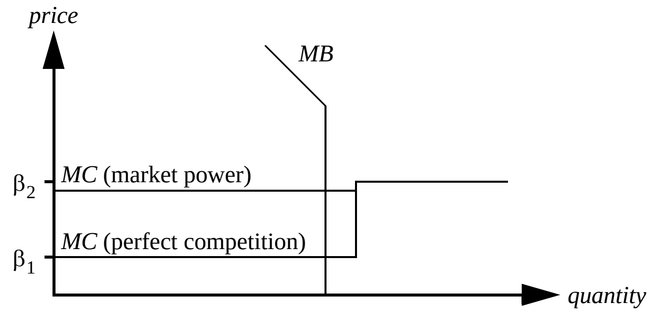


Figure 7.2 *Example of a Bertrand model. In the market above there are two producers. The first producer knows that they can sell a quantity corresponding to the entire demand, as long as they do not request a higher price than the offered by the other producer. Thus, they can set a market price which is just slightly less than the marginal cost of the other producer, β_2. If the same supply had been available on a perfectly competitive market, the market price would have been β_1 instead.*

sume that all players submit their bids without knowing the bids of the other players. The players in the Stackelberg model make their decisions in a certain order. The simplest case is to consider a large, leading firm and a smaller company, but other constellations are also possible. Given the quantity produced by the leading company, the smaller firm chooses the production which maximises their profits. However, the leading firm is assumed to be able to predict the reactions of the smaller company, which hence can be included in their own profit maximisation problem. This advantage makes the Stackelberg model more profitable to the leading firm than the Cournot model [118, 139, 140].[4] An example of how to apply the Stackelberg model to an electricity market is found in [118].[5]

- **Supply function equilibrium.**[6] Rather than choosing between selling for a certain price or selling a fixed quantity, the produc-

4. As can be seen, the Stackelberg model resembles the Cournot model, as they are both a game in quantities. There is also a corresponding market leader model where the producer chooses which price to offer (in similarity to the Bertrand model), but in this case it is not profitable to be market leader [140]; hence, such a model is of no practical interest.
5. Besides, I might mention that in figure 3.4 I used a Stackelberg model.
6. This footnote has been added to maintain the footnote numbering in pace with the Swedish edition.

ers bid could consist of an arbitrary supply function, i.e., it is stated which quantity is offered for a certain price [138]. The Bertrand and Cournot models are in a way two extreme cases of supply functions (in the Bertrand model the producer chooses a horizontal supply function and in the Cournot model they choose a vertical ditto). The supply function has been applied to study market power in electricity markets in for example [113, 116].

- **Conjectured supply functions.**[7] This model partly resembles the previous one, but here the players also consider how they expect the other players to change their production due to price changes [115]. This is a very general model, which with different choices of parameters can be made to correspond to any of the above mentioned models. The description in [115] is applied both to bilateral and centralised electricity markets.
- **Cartels.** The models above assume that each player makes his or her decision independently from the others. It is also possible to create models where some players cooperate by forming a cartel. The choice whether to participate in the cartel or not becomes a game in itself, because under some conditions it can be profitable to break the cartel agreement [137].

Apparently, there is a big selection of useful models to study market power, but this does not mean that more research in this field is not required. It may for example take some fine-tuning of the above mentioned models to apply them in combination with a multi-area model of the electricity market. New models may also have to be developed to manage for example cross subsidiaries.[8] Other interesting research issues would be to study how the possibilities of exercising market power is affected by other rules in the electricity market (tradable green certificate, choice of method for transmission congestion management, etc.).

7. This footnote has been added to maintain the footnote numbering in pace with the Swedish edition.
8. By cross subsidiaries I refer to a situation where companies participating in the competitive part of the electricity market are owned by a player who is also controlling companies in the regulated monopoly side of the electricity market; by transferring money from the monopoly company, it is possible to subsidise the company which is subject to competition. For example, a grid company could buy power to cover the losses at a price higher than the market price. By that means, the generation company can attract new customers by lowering their price in the regular electricity market. The price reduction are paid by the customers of the grid company, which have no possibility to choose another supplier.

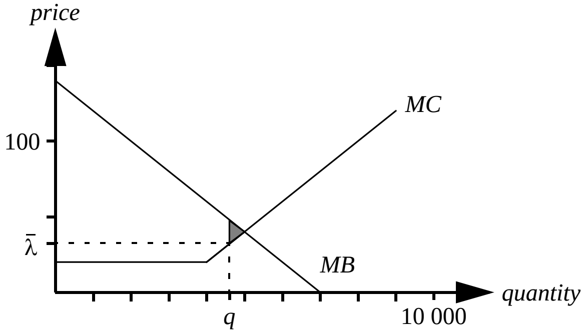
7.2 IRRATIONAL PLAYERS

Every now and when players can be observed, who seem to act irrationally, i.e., they are not maximizing their surplus. For example, I have been told that owners of hydro power plants sometimes generate electricity, although it would be more profitable to save the water and use it at a later occasion when the electricity prices are expected to be higher. The explanation for this behaviour is that if the electricity prices are lower than expected when the budget was decided, the company compensates by selling more instead. Another, more obvious example, is consumers who do not switch supplier, although they would receive a lower cost with another supplier.

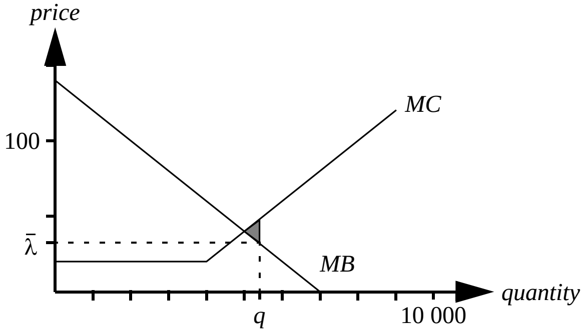
That players seem to behave irrationally is however not the same as that they really are irrational, but there might be other explanations for the behaviour. A player may for example need to consider some factor outside the electricity market; the hydro power producer who can fulfil the budget might gain some advantage from banks or share-holders. It may also be as discussed in section 3.1.1, i.e., that the benefits perceived by a player cannot only be measured in money, but other values may also play a role. A small consumer may not save more than a trifling amount of money every year when changing supplier, and may therefore find that the inconvenience of contacting possible alternative suppliers and comparing prices is more costly than to pay slightly more expensive electricity bills.[9] Another possible explanation is incomplete information; consumers who during all these years have been obliged to buy electricity from the local power company may not be aware that after a restructuring they are free to choose any supplier they like in a bilateral electricity market.

The deviations from an ideal electricity market mentioned above could be modelled using similar methods as has been suggested to simulate the impact of forecast uncertainty in the seasonal planning, i.e., by introducing a random disturbance on cost or benefit functions. However, the impact of differences between rational behaviour in an ideal electricity market compared to the real case is most likely so small that it is reasonable to neglect it. For as long as it is pure ignorance which make players behave irrationally, it is better to counter this by information campaigns from relevant authorities.

9. At least I myself, happily pay a couple of hundreds crowns to avoid having to call the customer service of a power company, just to be disconnected after half an hour of waiting in a telephone queue.



a) If a price cap just is introduced, no producer will be willing to supply production when the marginal cost is higher than the price cap. The consequence is that the turnover in the market is lower than what would have been the case in a perfectly competitive market; hence, the total surplus is reduced by an amount corresponding to the shaded area in the figure.



b) Here, a price cap is introduced while the producers are ordered to cover the entire demand, whenever physically possible. The consequence is that the turnover in the market is higher than for perfect competition; hence, the total surplus is reduced by an amount corresponding to the shaded area in the figure.

Figure 7.3 *Impact of price caps.*

7.3 LIMITATIONS IN ELECTRICITY TRADING

All players are limited by the legislation and other rules controlling the electricity market. Some of these limitations are justified comparing to the requirements on an ideal electricity market; it is for example nothing wrong with punishing emission of environmentally hazardous agents or requiring balance responsibility, etc. Other limitations are however such that they in a obvious way contradict the conditions of an ideal electricity market and therefore can be questioned.

An example of an unnecessary limitation is bilateral electricity markets where not all consumers have the possibility to choose supplier (which clearly is a departure from the requirement on free trade). If the objective is to simplify the transition from a vertically integrated electricity market to a restructured ditto, this kind of limitation might be justified, but in the long run all consumers should have access to the competitive electricity market.

Another example is when attempts are made to regulate the electricity prices, for example by introducing a price cap. Such actions are sometimes claimed to be necessary to counter the consequences of market power, but it is an action which risks to fail. If the price cap is set lower than the market price in a perfectly competitive market, the market will be prevented from achieving the equilibrium where the total surplus is maximised, as shown in figure 7.3. An example of how a price cap can cause problems is the Californian energy crisis from the summer 2000 to the spring 2001, during which several disturbances in the power supply stroke the consumers, while the large power suppliers were brought to the brink of ruin [147]. A single, crucial factor responsible for the crisis cannot be identified, but so much is at least sure that it was not beneficial to the society to prohibit the power suppliers to raise the prices paid by the final consumers, even though the suppliers paid more when buying power than they received from the customers. To prevent the power suppliers from going bankrupt the state of California had to support them with huge amounts. Thus, the artificial low electricity prices resulted in overconsumption—if the real electricity price had been visible, the most price sensitive consumers would probably had turned off their air conditioning when the power shortage was at its worst—and the consumers finally had to pay for the whole show via taxes instead.

Chapter 8

MONTE CARLO TECHNIQUES

Monte Carlo techniques can be considered a general designation for methods where a mathematical problem is solved by studying randomly chosen samples.[1] Today, Monte Carlo techniques are used within almost all areas of science, from opinion polls to computer simulation of complicated technical systems.

In this chapter I will give a summary of the notions and methods which are relevant for performing Monte Carlo simulation of electricity markets. Most of the chapter is thus a summary of mathematical theory, which has been known for decades. Some parts of the chapter, more exactly some thoughts about duogeneous populations and the idea of the strata tree, are however results of my own research.

8.1 SIMPLE SAMPLING

In its moste simple form, Monte Carlo simulation means collecting a number of completely random observations of a population. This is referred to as *simple sampling.* In this section, the most important formulae of simple sampling are summarised. Please note that I am only considering simple sampling with replacement, which means that the same sample may appear any number of times during the sampling procedure. Some formulae take a somewhat different expression if sampling without replacement is used.[2]

I have tried to use the same symbols as in most standard works on Monte Carlo techniques—in particular [127]—but I have made some changes to avoid confusion with other symbols appearing in this dissertation. I will define the symbols used as they are introduced, but it can be worthwhile to already from the start clarify the general principles which I have used in my

1. In some literature the notion “Monte Carlo technique” is used as a synonym for variance reduction technique.

choice of symbols:

- Random variables are generally designated by Latin capital letters, e.g. Y.
- Observations of a random variable are designated with the corresponding Latin lower-case letter. Generally, an index is used to distinguish different observations, e.g. y_i.
- The exact values of parameters in probability distributions are designated by Greek lower-case letters. The random variable to which the parameter is attributed is indicated by an index, e.g. μ_Y.
- Estimates of parameters in probability distributions are designated by the corresponding Latin lower-case letter. Here too, the random variable is indicated using an index, e.g. m_Y.

Estimation of Expectation Value and Variance

If there is a number of samples from a random variable and these samples are distributed completely according to the density function, the mean of the samples will be equal to the expectation value of the variable—this is simply the meaning of the definition of expectation value (cf. definition B.6). In practice, an arbitrary number of samples will only be distributed approximately according to the density function; as a result the mean of the samples will only approximately be equal to the expectation value. This can be expressed in a theorem:

Theorem 8.1. If there are n independent samples, $y_1, \ldots, y_n$, of the random variable Y then the mean of these samples,

$$m_Y = \frac{1}{n}\sum_{i=1}^{n} y_i,$$

2. The difference between sampling with and without replacement can be illustrated by the following example: Assume that there is an urn full of green and white marbles. When sampling without replacement, a marble is taken out of the urn, its colour is noted and it is then put aside. If sampling with replacement is used then the marble would be returned to the urn after its colour had been noted.

   It might seem strange to replace samples, because it inevitably leads to an increase of the uncertainty of the result. (With replacement there is a risk that for example the same marble is selected in each trial, whereas without replacement it is certain that eventually the whole population will have been studied and thus we will have *exact* knowledge of its composition.) However, this increase in uncertainty is negligible for large populations; moreover, it is often time consuming to exclude those samples that already have appeared. For simulation of electricity markets, sampling with replacement is definitely preferable.

is an estimate of $E[Y]$.[3]

The idea of the Monte Carlo technique is to use this possibility to determine expectation values by random samples. If there is a random variable, which has an expectation value that is unknown and too complicated to calculate analytically, then the expectation value is estimated instead. It is also possible to rewrite a complicated, but essentially deterministic, problem in such a manner that the solution is expressed as an expectation value, which then is estimated using Monte Carlo techniques. An example of this is Buffon's needle, which I mentioned in [7] and which also is described in many textbooks or in popular science literature.

It is important to notice that the estimate m_Y is a also a random variable with its own probability distribution. It can be shown that the expectation value of the estimate is equal to the expectation value of Y, i.e., $E[m_Y] = E[Y]$; otherwise there would be a systematic error in the estimate. The variance of the estimate, which is a measure of how accurate the result is expected to be, is given by the following theorem:

> ***Theorem 8.2.*** The variance of the estimated expectation value in simple sampling is given by
>
> $$Var[m_Y] = \frac{Var[Y]}{n}.$$ [4]

The theorem simply states that the more samples that are collected, the lower the variance of the estimate. In plain language this means that the deviation between the estimate and the real value is expected to decrease as the number of samples increase. However, as $Var[m_Y]$ never becomes equal to zero, there will always be a certain random error in the estimate m_Y. It should also be noted that even though the error is expected to decrease as more samples are collected, this does not mean that it always decreases, but in some cases the estimate may deteriorate if some more samples are collected.[5]

In many cases it is not sufficient to know the expectation value of a random variable, but it is also necessary to know the variance. If for example the value of an investment is investigated then it is probably not sufficient to conclude that the investment is expected to be profitable (i.e., the expectation

3. According to my opinion, the nicest proof of this theorem is given in [127], p. 28f.
4. A proof can be found in [127], p. 23f.
5. This situation can be illustrated by considering a coin which is tossed twice. If it is an ordinary coin (no hanky panky—the probability is 50% that the outcome is heads and 50% that it is tails) and we had tails in the first trial and heads in the second then we would estimate the probability of heads as 50%, which is the correct answer. If another trial is made, the estimate will either be 33% or 67%—in both cases the estimate is deteriorated. If the result had been either just heads or just tails in both trials then the third trial would either result in an improvement of the estimate or that the estimate remained the same.

value of the gain is positive), but it would also be interesting to know how large the risk is (i.e., how large is the variance of the gain). The estimate of the variance may also be required for internal calculations of the simulation, as we will see later in this section.

Theorem 8.3. If there are n independent samples, $y_1, \ldots, y_n$, of the random variable Y then

$$s_Y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - m_Y)^2$$

is an estimate of $Var[Y]$.[6]

An interesting practical detail in this context is that the expression of s_Y^2 can be reformulated as

$$\begin{aligned} s_Y^2 &= \frac{1}{n}\sum_{i=1}^{n}(y_i^2 - 2m_Y y_i + m_Y^2) = \frac{1}{n}\left(\sum_{i=1}^{n} y_i^2 - 2m_Y\sum_{i=1}^{n} y_i + n \cdot m_Y^2\right) = \\ &= \left\{\sum_{i=1}^{n} y_i = n \cdot m_Y\right\} = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - m_Y^2. \end{aligned} \tag{8.1}$$

In other words, in order to calculate the estimates m_Y and s_Y^2, it is sufficient that after each new sample collected the sums Σy_i^2 and Σy_i are updated Thus, it is not necessary to store the whole set of samples, $y_1, \ldots, y_n$, to perform simple sampling, unless it should be necessary to go back and study certain samples more closely after the simulation is completed.

Estimation of Probability Distributions

In many contexts the expectation value and the variance provide sufficient information of the behaviour of a random variable, but sometimes it is desirable to have more detailed information about the variable. It is then possible to estimate the density function of a random variable using Monte Carlo technique.

The most straightforward method to estimate the density function is to simply store all samples y_i and at the end of the simulation estimate $f_Y(x)$ as the number of samples $y_i = x$ divided by the total number of samples, n. For continuous random variables, the probability is extremely low that two identical samples are collected, which makes the estimated density function will

6. A proof of the theorem is for example given in [127]. In this proof the denominator $n - 1$ is used instead of just n; the reason is to simplify some formulae. If the number of samples is large then the difference is negligible.

appear somewhat strange. In this case it is probably preferable to present the result as a distribution function or duration curve instead.

Storing all samples might be impractical if there are many result variables and a lot of samples are required. To reduce the need for storage capacity the sample space can be divided in a number of segments and estimate the share of samples belonging to each segment. In practice this means that a continuous random variable is approximated by a discrete variable, but the result will still be rather good if the number of segment is sufficiently large.[7]

Confidence Intervals

Theorem 8.2 shows that simple sampling always leads to some amount of uncertainty of how well the estimate corresponds to the true expectation value. Thus, it is appropriate to take this uncertainty into consideration before any conclusions are drawn from the estimate. It is common to state a *confidence interval* for the estimate. A confidence interval is an interval which has a certain probability, the *confidence level*, to include the true value. The confidence level to be used may be chosen arbitrarily, but usually it is 95%, 99% or 99.9%.

A simple method to calculate the confidence interval of m_Y is to assume that the estimate is normally distributed around the true value μ_Y, i.e., $m_Y \in N(\mu_Y, \sqrt{Var[m_Y]})$.[8] The confidence level is the probability that μ_Y is within the interval $m_Y \pm \delta$, i.e.,

$$P(\mu_Y \in (m_Y - \delta, m_Y + \delta)) = P(\mu_Y \le m_Y + \delta) - P(\mu_Y \le m_Y - \delta). \tag{8.2}$$

Using theorem 8.2 and 8.3 we obtain the following expression for the confidence level in (8.2):

$$\Phi\left(\frac{\delta\sqrt{n}}{s_Y}\right) - \Phi\left(-\frac{\delta\sqrt{n}}{s_Y}\right) = 2\Phi\left(\frac{\delta\sqrt{n}}{s_Y}\right) - 1. \tag{8.3}$$

Each desired value of the confidence level is thus corresponding to a specific value of $t = (\delta\sqrt{n})/s_Y$. Table 8.1 shows t for some confidence levels.[9] Given a desired confidence level the confidence interval is calculated by

7. In this manner I presented the results of the case study in [7].
8. Whether or not this is a justified assumptions has been investigated by several mathematicians, cf. [127], p. 39ff. Briefly, it can be said that if the number of samples is large then m_Y tends to be normally distributed. In [62] it is claimed that the system index *LOLP* should be gamma-distributed, but no direct proof of this statement is given. Cf. chapter 9, footnote 36.
9. Appropriate values of t for other confidence intervals can be found in most mathematical handbooks, e.g. [134].

Table 8.1 *Calculation of confidence intervals.*

Confidence level	t
95%	1.9600
99%	2.5758
99.9%	3.2905

$$\delta = \frac{t \cdot s_Y}{\sqrt{n}}. \tag{8.4}$$

Heterogeneous, Conformist and Diverging Units

The estimate of the standard deviation, s_Y, is important for several reasons, for example because it is used to calculate confidence intervals, stopping criteria and—as we will se in section 8.2.5—for calculation of sample allocation. It is therefore worthwhile to reason a little about the possible outcome of s_Y for different kinds of random variables. Above all, it is important to study the possibility that the estimate $s_Y = 0$ is obtained, because such an estimate is equivalent to estimating $Var[m_Y]$ equal to zero. This in its turn would imply that an exact result had been obtained, but such are reserved for analytical methods. When using Monte Carlo technique there is *always* a certain amount of uncertainty in the final result—otherwise one should suspect that something has gone awry.

The sample space of a general random variable can be seen as a population consisting of a number of units (or individuals), which of course may be infinite. Each unit has a certain value y_i and all units are equally likely; if there is an outcome ψ_1 which is more probable then another one, ψ_2, then there must be relatively more units for which $y_i = \psi_1$ than there are units for which $y_i = \psi_2$. Collecting samples of the random variable is equivalent to randomly choosing a unit and observing its value.

In some cases the units of a population are *heterogeneous*, which means that it is hard to find two units having the same value, i.e., it is unlikely that $y_i = y_j$ for two randomly chosen units i and j. Consequently, it is sufficient to collect just a few samples to obtain $s_Y > 0$.

In other cases the population consists of a large number of *conformist units* and a small group of *diverging units*. Let us refer to this as a *duogeneous population*.[10] If the diverging units are homogenous (all diverging units have the same value) or heterogeneous (the diverging units may assume different

10. The terms duogeneous population as well as conformist and diverging units are my own inventions; I have not seen similar notions in the textbooks I have studied.

values) is of less importance; the important thing is that they are few compared to the conformist units. If a small number of units is randomly chosen, the probability is large that only conformist units are chosen, which yields the estimate $s_Y = 0$. Thus, it is obvious that the estimate m_Y does not take the diverging units into consideration; therefore, it is more or less erroneous. The size of this error is of course depending on the proportion of diverging units in the population and how much they differ from the conformist units.[11]

Table 8.2 *Probability of just choosing conformist units.*

Number of samples, *n*	Proportion of diverging units in the population						
	50%	25%	10%	5%	1%	0.1%	0.01%
8	0.0039	0.1001	0.4305	0.6634	0.9227	0.9920	0.9992
32	0.0000	0.0001	0.0343	0.1937	0.7250	0.9685	0.9968
64	0.0000	0.0000	0.0012	0.0375	0.5256	0.9380	0.9936
256	0.0000	0.0000	0.0000	0.0000	0.0763	0.7740	0.9747
1 024	0.0000	0.0000	0.0000	0.0000	0.0000	0.3590	0.9027

It can be interesting to study the probability of getting the estimate $s_Y = 0$. Table 8.2 shows some examples of the probability for just obtaining conformist units from a duogeneous population. The table does not say anything about the number of samples necessary to get a good estimate of m_Y, but is intended to provide an indication of how many samples are needed to differentiate a duogeneous population from a completely homogenous population, which in its turn is a minimum requirement in order to obtain a good estimate of m_Y. The trend is clear; the smaller the proportion of diverging units, the more samples are necessary.[12]

11. In particular the relative error becomes large when the conformist units correspond to the value zero. In a population consisting of 99% conformist units with $y_i = 0$ and 1% diverging units where $y_i = 1$, it is likely that a simulation based on a few samples will produce the estimate $m_Y = 0$, even though $\mu_Y = 0.01$. A relative error of 100% in other words, although the error is only 0.01 in absolute terms.
12. Another way of understanding the difficulty is to imagine a situation where 100 samples have been collected and the result was 99 conformist units and one diverging unit. This result does not provide enough information to differentiate between for example the following two interpretations:

    One unit out of ten is diverging, but unfortunately the samples are such that the number of diverging units became unusually low.

    One unit out of a thousand is diverging, but the samples are such that a diverging unit appeared already among the first 100 trials.

    In order to be fairly certain about the proportion of diverging units, it would be necessary to have at least 10 diverging units among the samples. The smaller the proportion of diverging units, the longer it will take before a sufficient number of diverging units appear in the chosen samples.

Stopping Rules

As already mentioned, it is never certain that the result of simple sampling is the true expectation value, but if a sufficient number of samples is collected the probability is large that the estimate obtained is accurate enough. To determine how many samples are needed it is thus necessary to define what is meant by "accurate", as well as deciding which probability of an accurate result that we desire. Given these parameters it is possible to determine a suitable number of samples using rough calculations.[13]

An alternative to choosing the number of samples in advance is to use some kind of test—a *stopping rule*—in order to decide to during the course of the simulation whether or not the estimate seems to be accurate enough. It is not necessary to check the stopping rule after each collected sample. I usually divide the simulation into a number of batches, where each batch includes a certain number of samples. The stopping rule is checked after each batch and if it is not fulfilled then another batch is run. This procedure is most practical when using stratified sampling, because then there will be another reason for dividing the simulation in batches (see section 8.2.5).

It is possible to design a number of stopping rules. A common method is to study the so-called coefficient of variation, which according to [47] is defined as

$$a_Y = \frac{\sqrt{Var[m_Y]}}{m_Y} \approx \{\text{use theorem 8.2 and 8.3}\} \approx \frac{s_Y}{m_Y \cdot \sqrt{n}}. \tag{8.5}$$

If a_Y is less than some relative tolerance ρ then the result is considered as sufficiently good and the simulation can be terminated. The lesser the chosen value of ρ, the more accurate results can be expected.

For a heterogeneous population it might be sufficient to study the coefficient of variation, but for a duogeneous population I think that another requirement should be added to the stopping rule, more exactly that $s_Y > 0$. As described earlier, the estimate $s_Y = 0$ means that no diverging units are among the collected samples, which in many cases means that the estimate m_Y is completely wrong. Therefore, the sampling should continue until diverging units have been encountered (i.e., until $s_Y > 0$) and the estimate of there relative frequency is reliable (which it will be when $a_Y < \rho$).

8.2 VARIANCE REDUCTION TECHNIQUES

We resort to Monte Carlo techniques when a problem is too complicated to

13. Cf. for example [127], p. 72f, [47], p. 35ff, or [11], example 4.35.

be solved analytically. However, there is often some knowledge about the solution; in some cases it might even be possible to solve a part of the problem analytically. By using this knowledge it can be possible to increase the accuracy of a Monte Carlo simulation. There are several methods to utilise what is known about the solution of a given problem; a general designation for these methods is *variance reduction techniques*. The designation refers to attempting to reduce the variance of the estimate, $Var[m_Y]$, which means that high accuracy can be achieved even with fewer samples.

Short descriptions of the variance reduction techniques I have acquainted myself with follow below, as well as a short analysis of the benefits of each method. In the latter analysis I assume that the problem to be solved is to calculate $E[X] = E[g(Y)]$, where Y is a vector of random variables with known probability distribution, f_Y, and g is a model of the system to be studied. As described in section 1.1, static electricity market simulations fit into this general format.

8.2.1 Complementary Random Numbers

The idea behind complementary random numbers is to reduce the influence of the random fluctuations, which always appear in sampling, by creating a negative correlation between the samples. In plain English, this means that random numbers are generated in such a manner that the probability of a good spread over the whole population increases.

Principle

Assume that m_{Y1} and m_{Y2} are two different estimates of the expectation value μ_Y of a certain random variable, i.e., $E[m_{Y1}] = E[m_{Y2}] = \mu_Y$. If we have a look at the average of the estimates then we find that it is

$$E\left[\frac{m_{Y1} + m_{Y2}}{2}\right] = \frac{1}{2}E[m_{Y1}] + \frac{1}{2}E[m_{Y2}] = \mu_Y. \tag{8.6}$$

The average of m_{Y1} and m_{Y2} is thus also an estimate of μ_Y, which maybe is not so surprising. The interesting thing is that the variance of the average is

$$Var\left[\frac{m_{Y1} + m_{Y2}}{2}\right] =$$
$$= \frac{1}{4}Var[m_{Y1}] + \frac{1}{4}Var[m_{Y2}] + \frac{1}{2}Cov[m_{Y1}, m_{Y2}]. \tag{8.7}$$

The variance of the average is apparently less than the variance of m_{Y1} and m_{Y2} respectively. Above all, we can utilise the fact that the covariance may become negative. A straightforward method to generate two estimates having a negative covariance is to use complementary random numbers.

If U is $U(0, 1)$-distributed then the complementary random number of U is given by $U^* = 1 – U$. Obviously, U^* is also a $U(0, 1)$-distributed random variable and, using definitions B.11 and B.12, the correlation coefficient can be calculated as –1, i.e., the strongest possible negative correlation. Random numbers of an arbitrary distribution are obtained by transforming $U(0, 1)$-distributed random numbers.[14] Assume that Y is a random variable with an arbitrary probability distribution and that Y has been obtained by transforming U, while Y^* has been obtained from U^* using the same transform. The standard transforms maintain at least a part of the negative correlation between U and U^*; hence, Y and Y^* will also be negatively correlated.

A practical aspect concerning the usage of complementary random numbers is that it is not necessary to have two separate estimates of m_Y based on the original samples and the complementary random numbers respectively. Rather than managing two series of n observations each, it is thus possible to consider $y_1, \ldots, y_n$ and $y_1^*, \ldots, y_n^*,$ as one single series of $2n$ observations. The formulae stated for simple sampling can the be used right away, even though the samples are not really independent.

Benefits of Complementary Random Numbers

Complementary random numbers can create a negative correlation between the input, Y, but in order to achieve a variance reduction a negative correlation is required between the observations of the variable which is actually sampled, i.e., the output, $X = g(Y)$. Hence, the strong negative correlation between U and U^* is attenuated twice; first, when U is transformed to Y and U^* to Y^*, and then when $X = g(Y)$ and $X^* = g(Y^*)$ are calculated. If complementary random numbers are to be of any use, it is required that the total attenuation is not too large.

Most transforms are such that the negative correlation is maintained, but some probability distributions—above all extremely asymmetrical distributions—may substantially weaken the correlation. An example of such an asymmetrical distribution is a binary variable having a low probability that $y_i = 1$. If this probability is denoted p and is less than 50% then it holds that

$$Cov[Y, Y^*] = E[Y \cdot Y^*] – E[Y] \cdot E[Y^*] = 0 – p^2. \tag{8.8}$$

As

$$Var[Y] = Var[Y^*] = E[Y^2] – (E[Y])^2 = p – p^2, \tag{8.9}$$

we get the correlation coefficient

14. See appendix C for details.

$$\rho(Y, Y^*) = \frac{Cov[Y, Y^*]}{\sqrt{Var[Y]Var[Y^*]}} = \frac{-p^2}{p - p^2} = -\frac{p}{1 - p}. \qquad (8.10)$$

When p approaches zero, the expression (8.10) will also approach zero, i.e., the correlation is almost disappearing, and $Var[m_Y]$ using complementary random numbers is more and more equal to $Var[m_Y]$ for simple sampling.

Whether g is such a function that $g(Y)$ and $g(Y^*)$ remains negatively correlated depends entirely on the system at hand.

8.2.2 Dagger Sampling

Dagger sampling[15] and complementary random numbers are based on similar ideas. However, dagger sampling is especially appropriate for random variables with only two possible outcomes and the probability is low for one of the outcomes—as seen in section 8.2.1 complementary random numbers is not particularly efficient for this kind of variables. The original description of dagger sampling is found in [130].

Principle

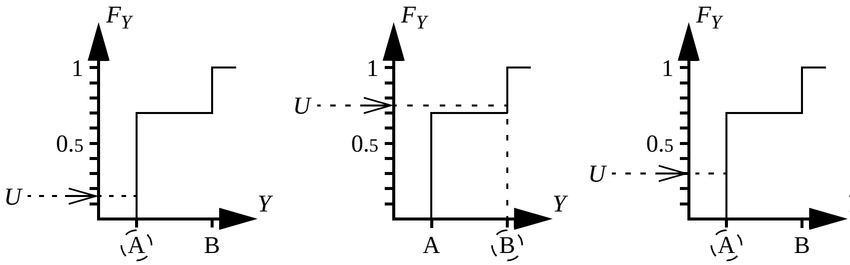
Consider a random variable Y, which has the density function

$$f_Y(x) = \begin{cases} 1 - p & x = \text{A}, \\ p & x = \text{B}, \\ 0 & \text{all other } x, \end{cases} \qquad (8.11)$$

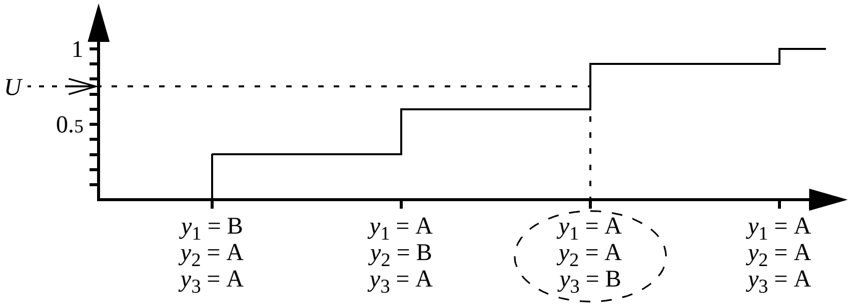
where $p < 0.5$. Usually, samples are created by using the inverse transform method. This method is based upon transforming a $U(0, 1)$-distributed random number to the desired distribution using the inverse of the distribution function of the variable (see appendix C). This can be viewed graphically by starting at a randomly chosen point of the F_Y-axis (which is graded between 0 and 1) and follow a line parallel to the Y-axis, until the curve of the distribution function is reached. Then a vertical line is drawn to the horizontal axis, where the transformed random value can be read.

In dagger sampling a scale between 0 and 1 is instead divided in subintervals with the width p. The number of subintervals, which is referred to as the dagger cycle length,[16] is thus equal to the largest integer S which is less than or equal to $1/p$. In addition to a number of subintervals, there may also be a

15. This footnote has been added to maintain the footnote numbering in pace with the Swedish edition.



a) Three samples of Y created by the inverse transform method.



b) Three samples of Y created by dagger sampling.

Figure 8.1 *The principle of the dagger sampling technique.*

remainder part. The finesse of dagger sampling is that one single random number generates S samples. This is done by creating a $U(0, 1)$-distributed random number and if this random number falls into subinterval k then the outcome becomes y_k = B, while the remaining y_i, $i = 1, \ldots, S$, $i \neq k$, are equal to A. If the random number falls into the remainder part then the outcome is A in all samples. The difference between this procedure and using the inverse transform method is illustrated in figure 8.1. With a generous portion of imagination it can be perceived as if one single random number “cuts” through several samples—hence the designation “dagger sampling”.

Let us now study the expectation value and variance of an estimate based on samples generated by dagger sampling. In order to simplify the calculations, I will assume that the outcome A corresponds to the numerical value 0 and B correspond to 1, which means that $E[Y] = p$. Considering the expectation value of S observations it is obvious that the probability is $S \cdot p$ that there will be exactly one outcome where $y_i = 1$ and the probability is $1 - S \cdot p$ that

16. This footnote has been added to maintain the footnote numbering in pace with the Swedish edition.

only outcomes where $y_i = 0$ are obtained. The expectation value of the estimate can therefore be written as

$$E[m_Y] = E\left[\frac{1}{S}\sum_{i=1}^{S} y_i\right] = \frac{1}{S}(Sp \cdot 1 + (1 - Sp) \cdot 0) = p. \tag{8.12}$$

The variance of the estimate is

$$Var[m_Y] = Var\left[\frac{1}{S}\sum_{i=1}^{S} y_i\right] =$$
$$\frac{1}{S^2}\left(\sum_{i=1}^{S} Var[y_i] + 2\sum\sum_{i<j} Cov[y_i, y_j]\right). \tag{8.13}$$

The variance of a single observation y_i is of course equal to $Var[Y]$. If there was no correlation between the observations $y_1, \ldots, y_S$ then (8.13) would be simplified to

$$Var[m_Y] = \frac{1}{S^2} S \cdot Var[Y] = \frac{Var[Y]}{S}, \tag{8.14}$$

which corresponds to the variance of the estimate for simple sampling (cf. theorem 8.2). But during dagger sampling there is clearly a correlation between the observations, because it is known that if $y_j = 1$ then the remaining $y_i = 0$. The product $y_i \cdot y_j$ is therefore always equal to zero; the definition of covariance then yields

$$Cov[y_i, y_j] = E[y_i \cdot y_j] - E[y_i]E[y_j] = 0 - p \cdot p = -p^2. \tag{8.15}$$

Since the covariance terms are negative, the variance of the estimate according to (8.13) must be less than for simple sampling.

Benefits of Dagger Sampling

As for complementary random numbers it is required that the studied model, g, is a fairly nice function, so that the negative correlation between the inputs, Y, will be preserved in the output, X. Also the practical effect of dagger sampling is similar to complementary random numbers, i.e., dagger sampling improves the spread of the observations. If S samples are produced using simple sampling then the probability is $(1 - p)^S$ that only observations where $y_i = 0$ are obtained. If p is small then this probability is very high. However, in dagger sampling the probability of generating S samples where $y_i = 0$ is equal to the size of the remaining part, i.e., $(1 - S \cdot p)$. As $S \approx 1/p$, this probability is small, regardless of the value of p.

Dagger sampling is very well suited for reliability analysis [47]. When considering the state of a system, this state is depending on the state of a whole set of components, i.e., $X = g(Y_1, \ldots, Y_k)$. In a general case these components will have different failure rates, which means that the dagger cycle length is different for different components. The question which dagger cycle length that should be used is treated more closely in [59].

8.2.3 Control Variates

The method of control variates enables the usage of simple, analytical models to improve the result of a Monte Carlo estimate. The idea is to sample the difference between the problem at hand and a simplified model, which can be treated analytically.

Principle

Assume that X is a random variable with the expectation value μ_X. Moreover, assume that there is another random variable, a *control variate, Z*, which expectation value $E[Z] = \mu_Z$ is known (from analytical calculations or earlier investigations). Rather than estimating $E[X]$ we choose to estimate the expected difference between X and Z. An estimate of $E[X]$ is then calculated by

$$m_X = m_{(X-Z)} + \mu_Z, \tag{8.16}$$

because

$$E[m_{(X-Z)} + \mu_Z] = E[X-Z] + \mu_Z = E[X] - \mu_Z + \mu_Z = \mu_X. \tag{8.17}$$

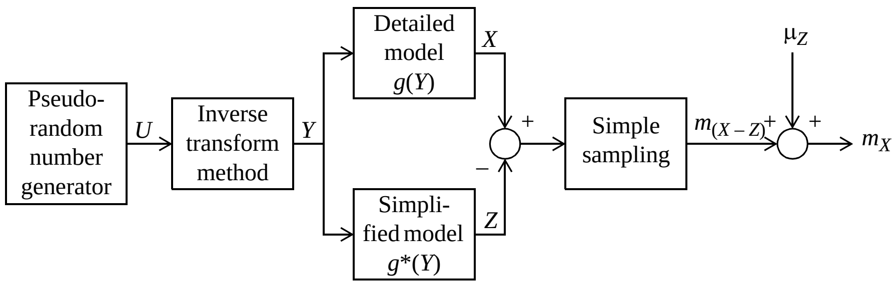
The variance of the difference $X - Z$ is

$$Var[X-Z] = Var[X] + Var[Z] - 2Cov[X, Z]. \tag{8.18}$$

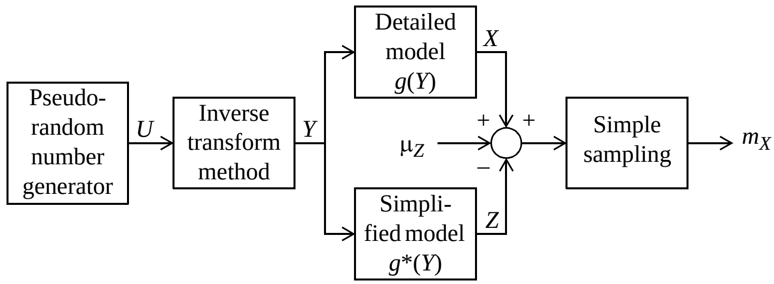
If a control variate Z can be found which is strongly positively correlated to X it is then possible that $2Cov[X, Z] > Var[Z]$; this results in $X - Z$ having a smaller variance than X. Simple sampling of $X - Z$ is then according to theorem 8.2 resulting in less variance of the estimate than if X is sampled directly.

Benefits of Control Variates

The prerequisite of making the control variate method efficient is of course that a suitable control variate can be found. As earlier mentioned, we assume



a) The expectation value of the control variate is added after sampling.



b) The expectation value of the control variate is included in every sample.

Figure 8.2 *The principles of sampling using a control variate.*

a random variable Y with a known distribution. The random variable to be sampled is a function of Y, i.e., $X = g(Y)$. The desired mathematical model, $g(Y)$, is generally a fairly complicated model, which is difficult to treat analytically—otherwise Monte Carlo methods would not be very interesting. However, it is often possible to create a simplified mathematical model of the same system, $g^*(Y)$. The simplified model can then be used as control variate, as shown in figure 8.2. The simplified model should behave approximately the same way as the complex model and thus it will be sufficiently correlated to X in order to produce a variance reduction. A condition is of course that the expectation value of the simplified model, $\mu_Z = E[g^*(Y)]$, can be calculated analytically

In figure 8.2a simple sampling is applied to the difference between X and the control variate Z; the expectation value of the control variate, μ_Z, is not added until the difference has converged. An alternative approach is to add the expectation value of the control variate before sampling, as shown in figure 8.2b. The difference between the two alternatives is that they affect stopping rules based on the coefficient of variation. In the first alternative, the estimated variance is

$$s_{(X-Z)} = \frac{1}{n}\sum_{i=1}^{n}(x_i - z_i - m_{(X-Z)})^2. \tag{8.19}$$

The estimated variance in the other alternative is

$$s_{(X-Z+\mu_Z)} = \frac{1}{n}\sum_{i=1}^{n}(x_i - z_i + \mu_z - m_{(X-Z+\mu_Z)})^2. \tag{8.20}$$

However, (8.19) and (8.20) are equally large, because

$$m_{(X-Z+\mu_Z)} = \frac{1}{n}\sum_{i=1}^{n}(x_i - z_i + \mu_Z) = \mu_Z + \frac{1}{n}\sum_{i=1}^{n}(x_i - z_i) =$$
$$= \mu_Z + m_{(X-Z)}, \tag{8.21}$$

which means that

$$x_i - z_i + \mu_Z - m_{(X-Z+\mu_Z)} = x_i - z_i - m_{(X-Z)}. \tag{8.22}$$

The resulting coefficients of variation are thus

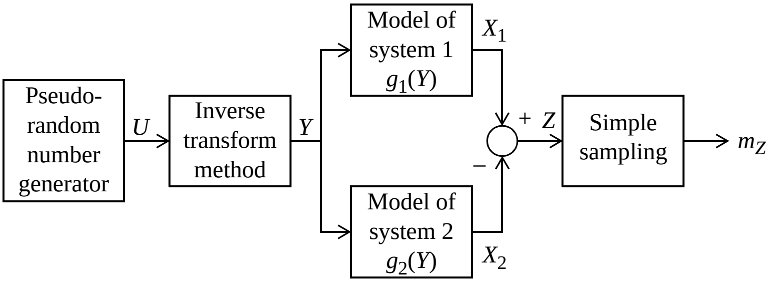
$$a_{(X-Z)} = \frac{s_{(X-Z)}}{m_{(X-Z)} \cdot \sqrt{n}}, \tag{8.23a}$$

$$a_{(X-Z+\mu_Z)} = \frac{s_{(X-Z+\mu_Z)}}{m_{(X-Z+\mu_Z)} \cdot \sqrt{n}} = \frac{s_{(X-Z)}}{m_X \cdot \sqrt{n}}. \tag{8.23b}$$

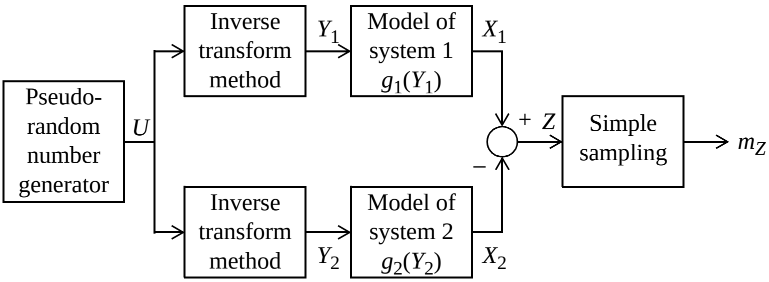
The coefficients of variation differ only depending on how m_X is related to $m_{(X-Z)}$; exactly the same series of samples will yield a smaller coefficient of variation in the first alternative if $m_{(X-Z)} > m_X$. This must be considered when deciding a suitable relative tolerance in the stopping rule.

8.2.4 Correlated Sampling

Correlated sampling is based on similar principles as the control variate method. The control variate method studies the difference between a detailed and a simplified model of the same systems, whereas in correlated sampling the difference is studied between two separate systems which have similar properties. A typical application is to study what happens if an existing system is modified.



a) The two systems use inputs with the same probability distribution.



b) The two systems use inputs with different probability distributions.

Figure 8.3 *The principles of correlated sampling.*

Principle

Assume that the difference between the expectation values of two random variables X_1 and X_2 should be studied. Rather than estimating $E[X_1]$ and $E[X_2]$ in separate simulations, we choose to study $Z = X_1 - X_2$. Apparently, $E[Z]$ is equal to the difference we are looking for, because

$$E[Z] = E[X_1 - X_2] = E[X_1] - E[X_2]. \tag{8.24}$$

The variance of Z is

$$Var[Z] = Var[X_1] + Var[X_2] - 2Cov[X_1, X_2]. \tag{8.25}$$

If the two random variables are strongly positively correlated to each other then it is possible that simple sampling of Z is more efficient than sampling X_1 and X_2 separately.

Benefits of Correlated Sampling

The benefits of correlated sampling is utterly depending on whether or not the two studied variables X_1 and X_2 are sufficiently positively correlated to produce a variance reduction according to (8.25). As earlier mentioned, we start by input with known probability distributions. To apply correlated sampling, we should have one system g_1 with output X_1 and another system g_2 with output X_2. The two systems can either use the same input (figure 8.3a) or inputs with different probability distributions (figure 8.3b). Apparently it is more likely that X_1 and X_2 are positively correlated in the first case, because the correlation then only depends on how similar the two systems g_1 and g_2 are. In the other case, the correlation is attenuated, since different transforms are used to create Y_1 and Y_2 respectively.

A disadvantage with correlated sampling is that even though the estimate of $E[X_1 - X_2]$ is very good, this does not mean that very accurate estimates of $E[X_1]$ and $E[X_2]$ are obtained.

8.2.5 Stratified Sampling

If there is some knowledge about the characteristics of a population then we can use it to concentrate the samples to those parts of the population which are of most importance. Stratified sampling achieves this effect by dividing the population in smaller parts, which are then studied separately.

Principle

Assume that there is a random variable Y with the sample space $\mathcal{Y}$. In stratified sampling the sample space is divided in L strata, where stratum h comprises the outcomes $\mathcal{Y}_h$, which is a subset of $\mathcal{Y}$. Two strata may not overlap; in other words, each outcome should belong to exactly one stratum:

$$\mathcal{Y}_h \cap \mathcal{Y}_j = \emptyset \ \forall \ h \neq j, \tag{8.26a}$$

$$\bigcup_h \mathcal{Y}_h = \mathcal{Y}. \tag{8.26b}$$

Each stratum is given a weight according to how large the part of the population is that belongs to the stratum:[17]

17. In the literature, e.g. [127], the symbol W is used for stratum weight. As this symbol is frequently used in my electricity market model, I have chosen to use ω for stratum weight.

$$\omega_h = \frac{N_h}{N} = P(Y \in \mathcal{Y}_h), \tag{8.27}$$

where

ω_h = stratum weight of stratum h,
N_h = number of units (i.e., possible outcomes) in stratum h,
N = number of units in the whole population.

As seen above, the stratum weight can also be expressed as the probability that an observation of Y will fall into a certain stratum h.

It is now possible to consider L separate random variables Y_h, $h = 1, \ldots, L$, each with the sample space $\mathcal{Y}_h$. The expectation value of each stratum is calculated separately. In the best case, $E[Y_h]$ can be calculated analytically; otherwise an estimate is obtained by simple sampling:

$$m_{Yh} = \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h}, \tag{8.28}$$

where

m_{Yh} = estimate of the expectation value for stratum h,
y_{hi} = value of the i:th sample of stratum h,
n_h = number of samples from stratum h.

The expectation value of the whole population, $E[Y] = \mu_Y$, is then estimated by weighting the results of each stratum:

$$m_Y = \frac{\sum_{h=1}^{L} N_h m_{Yh}}{N} = \sum_{h=1}^{L} \omega_h m_{Yh}.\text{[18]} \tag{8.29}$$

If it has been possible to analytically calculate $E[Y_h] = \mu_{Yh}$ then $m_{Yh} = \mu_{Yh}$ should of course be used in (8.28).

The variance of the estimate m_Y in stratified sampling is

$$Var[m_Y] = \sum_{h=1}^{L} \omega_h^2 \frac{Var[Y_h]}{n_h}.\text{[19]} \tag{8.30}$$

If strata have been defined properly then it is possible that (8.30) is less than the variance for simple sampling (cf. theorem 8.2). Notice that the opposite is

18. A proof that $E[m_Y] = \mu_Y$ is given in [127], p. 91.
19. A proof is given in [127], p. 92.

also possible; a poor stratification may result in a higher variance for the estimate of the expectation value $E[Y]$!

If $E[Y]$ is to be estimated using in total n samples, how should then these samples be divided between the different strata? It can be shown that (8.30) is minimised if the samples are distributed according to the so-called Neyman allocation:[20]

$$n_h = n\frac{\omega_h \sigma_{Y_h}}{\sum_{k=1}^{L} \omega_k \sigma_{Yh}}, \tag{8.31}$$

where

σ_{Yh} = standard deviation of stratum h, i.e., $\sqrt{Var[Y_h]}$.

If $Var[Y_h]$ could be calculated analytically then it would also be possible to calculate $E[Y_h]$ analytically and then there would be no need at all to apply any Monte Carlo methods. Thus, it must be expected that all the σ_{Yh} are unknown. Estimates of σ_{Yh} have to be used instead when (8.31) is used:

$$s_{Yh} = \sqrt{\frac{1}{n_h}\sum_{i=1}^{n_h}(y_{hi} - m_{Y_h})^2}, \tag{8.32}$$

where

s_{Yh} = estimate of σ_{Yh}.

In other words, it is necessary to have a few samples from each stratum before the Neyman allocation can be applied. This is solved by starting the sampling with a small pilot study, where a predetermined number of samples is taken out of each stratum. The number of samples can either be the same in each stratum, so-called proportional sampling, or be adjusted to the expected properties of the individual stratum.

Better and better estimates of σ_{Yh} can be expected as more samples are collected. It may therefore be worthwhile to stop every now and then and use the improved estimates to determine a new sample allocation. A practical solution is to divide the simulation into batches; after each batch the stopping rule is checked and if it is not fulfilled then a new sample allocation is calculated.

Benefits of Stratified Sampling

In [127] stratified sampling is used for heterogeneous populations.[21] I per-

20. A proof is given in [127], p. 98f.

sonally think that stratified sampling is most useful when studying a duogeneous population (see page 156), because it is then possible to use the stratification to concentrate samples to those parts of the population where the diverging units are found. Ideally, all conformist units are put in some strata and the diverging units in others. Those strata which only have conformist units do not need any further studies, because μ_{Yh} is obvious. At the same time, it is much quicker to estimate the expectation value of the strata containing only diverging units; indeed, if things are nice, there is only one kind of diverging units, so that μ_{Yh} is obvious in these strata, too.

If it is impossible to achieve the ideal division then one should attempt to design strata having only conformist units or a mixture of conformist and diverging units. In the latter, the proportion of diverging units can become larger than it is for the whole population; hence, these proportions will be easier to estimate (cf. the reasoning around table 8.2).

However, this strategy has a potential weakness, derived from the fact that we are forced to use a pilot study before we can determine how to divide the samples between the strata in the remainder of the study. If the result of the pilot study is too erroneous, it might have catastrophic consequences for the final result. Let me show an example of how this can happen.

Figure 8.4 shows an irregular, black shape, the area of which should be determined. This can be done by a Monte Carlo method. Each sample consists of a randomly chosen point within the surrounding, rectangular frame. Thus, we have a population composed of a majority of grey points, which constitute the conformist units. Besides, there is a small proportion of black points, which consequently are the diverging units. The black surface constitutes the same share of the surrounding rectangle as the proportion of diverging units in the population. Since the area of the rectangle is simple to calculate, it is possible to estimate the black surface on the basis of an estimate of the proportion of diverging units.

Assume that three strata have been defined as in the figure. In the pilot study, 8 samples are taken per stratum. In stratum 1 all samples will be conformist units, resulting in the estimate $s_{Y1} = 0$. According to the Neyman allocation (8.31) zero samples should be taken from stratum 1. The samples that already have been collected in the pilot study cannot be undone; the essential is that no more samples will be allocated to stratum 1. This is correct and efficient, as stratum 1 is a completely grey surface; thus, no new information would be gained from further sampling of stratum 1.

If we consider stratum 2 instead, the probability that a certain sample is constituted by a black point is about 50%. The probability that 8 samples would comprise only black or only grey points is less than one per cent. Most

21. Most examples in the book considers issues of population statistics, as number of inhabitants in cities, number of acres per farm, etc.

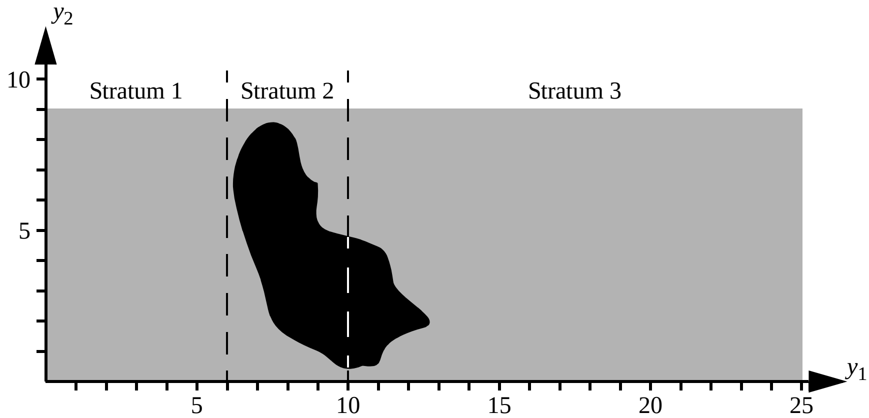


Figure 8.4 *Area determination problem. The share of the grey rectangle which is covered by the black figure can be estimated by randomly selecting a number of points (y_1, y_2). By that means it is possible to estimate the area of the figure, because the area of the rectangle is easy to calculate.*

likely, the estimate $s_{Y2} > 0$ is obtained, which is a prerequisite that more samples will be allocated to this stratum. Exactly how many more samples there will be depends on the stratum weight ω_2 and how s_{Yh} relates to the estimated standard deviation of the other strata.

So far we have not had any trouble, as problems arise first when we incorrectly estimate s_{Yh} to zero, because then we risk to exclude important contributions to the final result. This situation can arise in if we study stratum 3 in the example. The black part is here only 5% of the total area of the stratum, which means that the probability is about 2/3 that we solely collect 8 grey samples from stratum 3; then we get the estimate $s_{Y3} = 0$ and thus no more samples will be allocated to stratum 3. The black part of stratum 3 will therefore never be detected in the remainder of the sampling procedure, *even though it comprises about a third of the total black surface*!

The error described above is far more serious than the random errors which are inevitable in Monte Carlo simulation, because this error does not approach zero as the number of samples increase.[22] Therefore, I have introduced the designation *cardinal error* for those cases there the pilot study has yielded the result $s_{Yh} = 0$ although $\sigma_{Yh} > 0$.

Another trouble caused by cardinal error is that a stopping rule including the requirement that $s_Y > 0$ may never be fulfilled. This problem can be managed in several ways. One is to continue the pilot study until $s_Y > 0$. A more

22. The simulation will converge towards an incorrect estimate instead, as illustrated in [7], figure 4.5.

simple alternative is to just introduce an upper limit to the number of samples in a simulation; if this limit is exceeded, the simulation is terminated regardless of whether the stopping rule has been fulfilled or not.

Generally, stratified sampling can be very efficient for duogeneous populations, as long as cardinal errors can be avoided. One should however notice that the presence of cardinal error in a simulation does not always have a catastrophic impact on the final result. If just a small fraction of the black surface had been in stratum 3 in the example above, then the consequences of neglecting this part had been rather insignificant.

When estimating $E[X] = E[g(Y)]$ the strata has to be defined for the input Y, but it is the properties of $X = g(Y)$ which should be kept as similar as possible. For heterogeneous populations it is possible to deduce stratification strategies which at least approximately minimises $Var[m_X]$.[23] These methods assume that there is a fairly strong correlation between Y and X, and if strata are defined so that Y is quite similar in each strata then the corresponding values of X will also be similar. I have not made any experiments with such methods, because I think that the stratification in the first place must aim at separating conformist and diverging units according to the principles discussed above. It is possible to achieve this by using a simple tree structure, which I call a *strata tree*.

A strata tree is used to systematically divide a population which is characterised by a number of inputs, $Y_1, \ldots, Y_J$, in such strata that the output $X_1, \ldots, X_K$ gets similar properties in each stratum. The tree structure which is used should have the following properties:

- Each node specifies a subset of the sample space for a certain input Y_j, $j \in 1, \ldots, J$. The only exception is the root, which does not contain any information at all.
- Each node has a *node weight*. The node weight is equal to the probability that Y_j belongs to the specified subset. The root always has a node weight equal to 1.
- Each branch of the tree should specify one subset of all Y_j, $j = 1, \ldots, J$. This means that each branch specifies a subset of the population; thus, it may be considered as a stratum. The stratum weight is simply obtained by multiplying the node weights along the branch, assuming that all Y_j along a branch are independent of each other.
- All nodes must not specify subsets of the inputs Y_j. As long as it is possible to calculate a node weight, it is possible to introduce different categories of help variables.

A stratification requires that all possible inputs belong to a stratum and that there are no overlaps (8.26a), (8.26b). To guarantee that the strata tree really

23. See [127], section 5A.7.

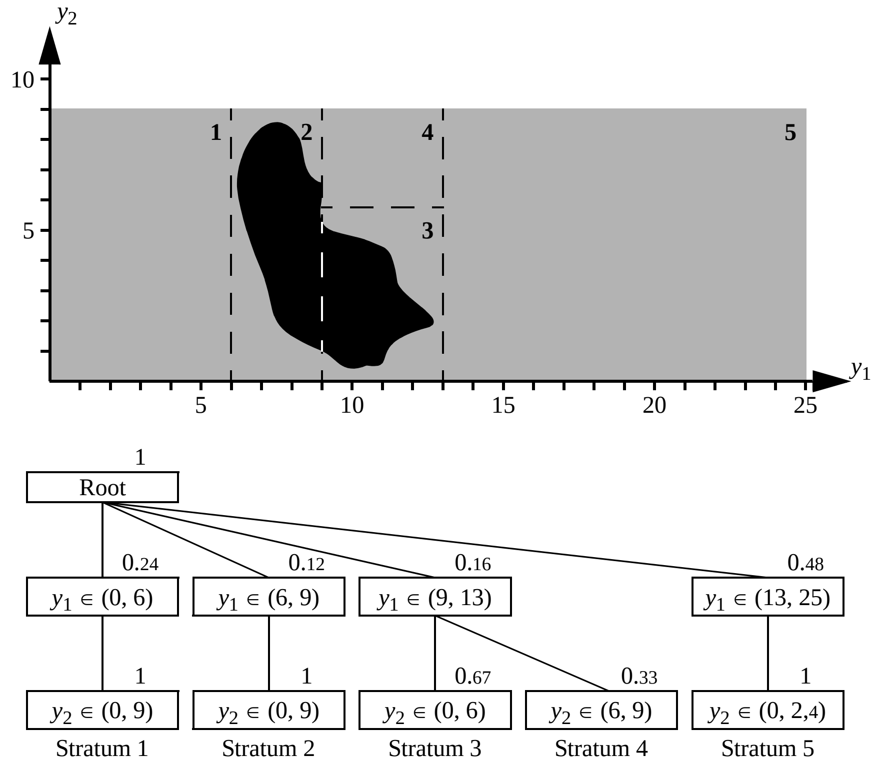


Figure 8.5 *Stratification for the area determination problem in figure 8.4. The proportion of black points compared to the grey points vary a lot along the y_1-axis; therefore, the y_1-coordinate is placed in the first level below the root of the strata tree. Given which interval of the y_1-axis a point belongs to it is possible to differentiate the y_2-coordinates where the black points can be found.*

covers all inputs it is possible to apply the following two simple rules:

- The children of a certain node must define subsets of the same input Y_j.
- The sum of the node weights of the children should be exactly 1.

The principle of how to utilise the strata tree is illustrated in figure 8.5.

8.2.6 Importance Sampling

Importance sampling improves the estimate in a similar manner as stratified sampling, i.e., by concentrating samples to the most significant parts of the population. In importance sampling this effect is achieved by modifying the random number generation so that another probability distribution is used than the real. Those parts of the population which as a result of this are over-represented are given a lesser weight when the expectation value is estimated.

In the literature it is common to derive the equations of importance sampling from the problem of calculating the integral $\int h(x)dx$.[24] Here I will present an alternative derivation, which is based on a general problem where we want to determine $E[X] = E[g(Y)]$.

Principle

Let Y be a random variable with a known density function f_Y defined on the sample space $\mathcal{Y}$. We wish to determine $E[X]$, where $X = g(Y)$, but rather than sampling X we introduce another variable, Z. The density function of Z, f_Z, is called the *importance sampling function*[25] and should be such that $f_Z(\psi) > 0$ $\forall\ \psi \in \mathcal{Y}$. For each outcome $Y = \psi$ we have

$$X = g(\psi), \tag{8.33a}$$

$$Z = g(\psi)\frac{f_Y(\psi)}{f_Z(\psi)}. \tag{8.33b}$$

It is simple to verify that $E[Z]$ equals the expectation value we are looking for, i.e., $E[X]$. According to definition B.6 we get

$$E[X] = \int_{\psi \in \mathcal{Y}} g(\psi) f_Y(\psi) d\psi \tag{8.34}$$

and

$$E[Z] = \int_{\psi \in \mathcal{Y}} g(\psi)\frac{f_Y(\psi)}{f_Z(\psi)} f_Z(\psi) d\psi = \int_{\psi \in \mathcal{Y}} g(\psi) f_Y(\psi) d\psi \tag{8.35}$$

According to definition B.7 we obtain

$$Var[Z] = E[Z^2] - E[Z] = \int_{\psi \in \mathcal{Y}} g^2(\psi)\frac{f_Y^2(\psi)}{f_Z(\psi)} d\psi - \mu_X^2. \tag{8.36}$$

As can be seen $Var[Z]$ depends on the importance sampling function $f_Z(x)$, which may be chosen arbitrarily. Obviously $Var[Z]$ is minimised if we choose

$$f_Z(\psi) = \frac{g(\psi) \cdot f_Y(\psi)}{\mu_X}. \tag{8.37}$$

24. See for example [47, 59, 133].
25. This footnote has been added to maintain the footnote numbering in pace with the Swedish edition.

However, this choice presupposes that μ_X is known and if that was the case, there would be no reason to practice Monte Carlo simulation at all. The point is that although the variance cannot be reduced to zero, a clever choice of importance sampling function may result in $Var[Z] < Var[X]$. According to theorem 8.2, this means that simple sampling of Z is more efficient than simple sampling of X. However, it must be kept in mind, that the opposite is also possible, i.e., a poorly chosen importance sampling function may result in $Var[Z] > Var[X]$.

Benefits of Importance Sampling

Importance sampling works in a similar manner as stratified sampling. Both methods choose samples in such a manner that those units, which contribute the most to the final results are more likely to appear among the samples than during simple sampling. This overrepresentation is compensated by giving the samples a lesser weight when estimating the expectation value; in importance sampling this weight is given by the quota $f_Y(x)/f_Z(x)$ and in stratified sampling it is given by the stratum weight, ω_h. The difficulty of both methods is to choose the parts of the population where the samples should be concentrated. In stratified sampling, this choice is done in two steps: partly by designing strata and partly by allocating samples between different strata. The two steps combined correspond to the choice of importance sampling function.

If there is a sharp limit which can be identified between different parts of a population then I think that stratified sampling is the more straightforward method to use; the task of defining strata can be rather conveniently managed by a strata tree. If the limit however is less well-defined or if its exact location cannot be determined for practical reasons, importance sampling becomes a more appealing alternative. An incorrectly chosen division between different strata will convey a significantly increased risk of cardinal error, whereas a weight sampling function can be designed to concentrate the samples to the area where the limit is assumed to be, but without excluding other parts of the population.

Chapter 9

SHORT SCENARIOS

Among all the models that can be used for simulation of an electricity market, there are two main types that can be distinguished, namely those where time is an explicit part of the model and those which constitute a “snap shot” of the electricity market. I refer to the latter as *short scenarios*. (The opposite is long scenarios, which are treated in the following chapter.) The time elapsed during a short scenario can be chosen arbitrarily, but must be so short that none of the conditions of the electricity market has time to change, i.e., all scenario parameters are constant during the whole scenario. It is possible to imagine a short scenario which lasts as long as a trading period in the electricity market, or the short scenario may correspond to an infinitesimal time period. In the former case it is of course an approximation to assume that all scenario parameters are constant.

In a short scenario the result variables depend only on the state of the electricity market in one particular moment; it does not matter what has happened earlier in the system or which expectations the players have on future events. In reality there is of course no electricity market which works in this manner. If there is some kind of energy storage in the system then the players of the electricity market must continuously make a choice between using the stored energy now or saving it and selling it for a higher price at a later occasion. In such systems earlier states and expectations of the future play an important role; therefore, short scenarios are absolutely unreasonable. Time constants, for example long start-up times in certain power plants, have similar consequences.

Although short scenarios thus always are a simplification of reality, studies of simulation methods for short scenarios are relevant in the highest degree. Firstly, there are electricity markets where energy storage and time constants only have a negligible impact on the behaviour of the players and in those cases it is a completely reasonable approximation to simulate short scenarios. Secondly, it is easier to study long scenarios once one has mastered simulation of short scenarios.

9.1 SCENARIO PARAMETERS

A scenario was defined in section 1.1 as a situation with given conditions for the electricity market. A scenario is represented mathematically as a fixed outcome for each of the scenario parameters (random variables with known distribution) that appear in the electricity market model. Exactly which scenario parameters are necessary to define a short scenario depends of course on which model has been chosen. Let us for the sake of simplicity limit ourselves to the most basic scenario parameters, namely those that are part of a multi-area model of an ideal electricity market.

- Available generation capacity in thermal power plants: $\overline{G}_g$ (one value for each equivalent unit).
- Available generation capacity in non-dispatchable power plants: $\overline{W}_n$ (one value for each equivalent unit, which in practice means one value per area).
- Available transmission capability: $\overline{P}_{n, m}$ (one value for each interconnection)
- Load: D_n (one value for each area).[1]

In order to apply variance reduction techniques, it is necessary to have some knowledge about the behaviour of the electricity market already before the simulation has started. It is however not easy to predict operation cost or loss of load occasions from just a glance at the above mentioned scenario parameters. To gain a better overview it is a good idea to introduce special scenario parameters which represent the total system resources and system demand respectively. I call these scenario parameters *primary scenario parameters* and the remaining—which provide more details about how the primary parameters are distributed in the system—I call *secondary scenario parameters*.

Available Generation Capacity

Concerning the available generation capacity it as actually sufficient to use one single primary scenario parameter which defines the state of all power plants in the system (both thermal and non-dispatchable)—no secondary scenario parameters are needed. The principle is straightforward from a mathematical point of view; all possible combinations of available and unavailable power plants are listed and the probability is calculated for each combination, as well as the total available capacity in non-dispatchable power plants, $\overline{W}_{tot}$, and thermal power plants, $\overline{G}_{tot}$, respectively. In practice this enumera-

1. We assume that there are only price insensitive loads in the system. This does not restrict the simulation model, because price sensitive load can be simulated using price insensitive load and fictitious power plants (see section 3.2.1).

tion of possible combination may be quite time consuming, as the number of combination grows exponentially with the number of power plants in the system. I will return to this issue in section 9.2.2. For the moment we assume that all combinations have been identified and then we arrange a distribution function, $F_{\overline{W}+\overline{G}}$, where each combination corresponds to one state. It is important that the states are sorted according to $\overline{W}_{tot} + \overline{G}_{tot}$; otherwise the inverse transform method will not provide a correlation between the original sample, U, and $\overline{W}_{tot} + \overline{G}_{tot}$. Complementary random numbers will not work without such a correlation (see section 8.2.1).

An important prerequisite to create a common probability distribution of $\overline{W}_{tot} + \overline{G}_{tot}$ is that the available capacity in individual power plants only can assume discrete states. In reality, the available capacity of non-dispatchable power plants is a continuous random variable; thus, a discrete approximation must be used for the capacity in these power plants. It is normally not difficult to design such approximations.[2] Neither does it result in any greater loss of accuracy.

It may be worthwhile noticing that correlations between available capacity in different power plants can be managed within the concept of a common probability distribution $F_{\overline{W}+\overline{G}}$. Dependent variables only affect the calculating of the probability for a certain state, but $F_{\overline{W}+\overline{G}}$ is used in the same manner as when all power plants are independent of each other.

Available Transmission Capability

In the same way as for the available generation capacity, it is possible to enumerate and determine the probability of every possible combination of available and unavailable transmission lines. The result is a single distribution function, $F_{\overline{P}}$. However, unlike the generation capacity it is not possible to define any meaningful "total transmission capability" of each state—just adding up available capacity of each interconnection does not tell us anything about the behaviour of the system. I have not performed any detailed study whether or not there is some way of arranging the states which is better than others; this is a question which remains to be answered.

Load

It is simple to define primary scenario parameter of the total load, D_{tot}, in an electricity market. As the load is an continuous random variable it is however not possible to include the geographical distribution of the load in the primary

2. Concerning wind power, I have described how to determine discrete approximations in my licentiate thesis [7], appendix C. The same description can also be found in [11], p. 81ff.

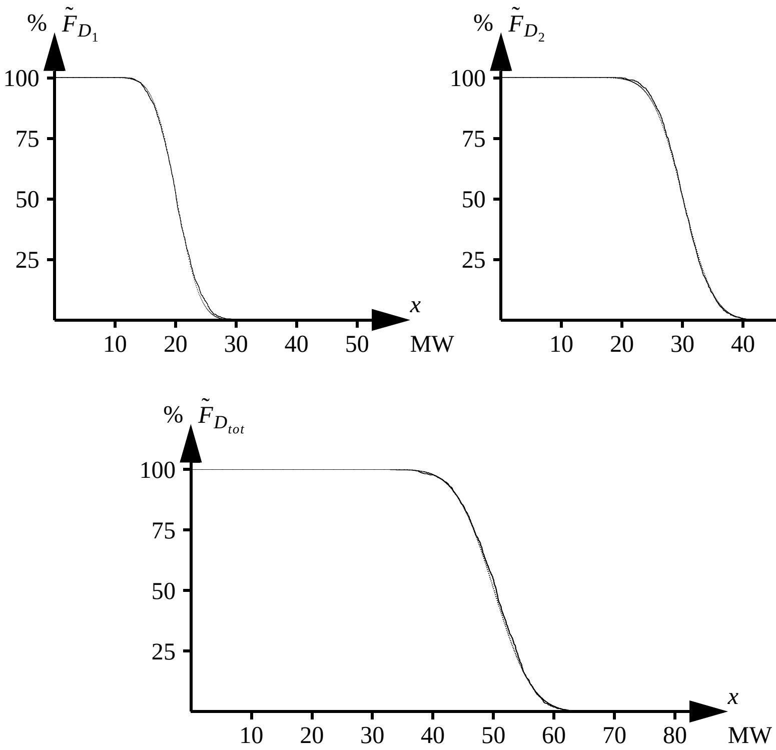


Figure 9.1 *Disturbance when using the scaling method. Here 1 000 samples have been generated using the scaling method. The load in the two areas are independent of each other; in area 1 it is* $N(20, 3)$*-distributed and in area 2 the load is* $N(30, 4)$*-distributed. The sampled duration curves are indicated by solid lines and the theoretical curves are dotted. The disturbance is barely visible to the naked eye.*

scenario parameter, unless a discrete approximation is used. A discrete approximation of course means lost accuracy and therefore it is better and even easier to introduce secondary scenario parameters, which show how large the share of the total load is located in each of the areas of the system is.

Given the probability distribution of the load in each area it is possible to calculate the probability distribution of D_{tot}.[3] When generating the scenario parameters of the load, the first thing to do is to randomise a value of D_{tot} according to this distribution. Then preliminary load values, $D_n{}^*$, are randomised in each area according to their given probability distributions. If there are correlations between the load in different areas then this is considered when the preliminary loads are randomised.[4] However, generally the sum of the area loads will not correspond to the randomised total load (unless there has been an almost uncanny strike of luck). This is solved by simply scaling the preliminary loads so that they match the total load:

$$D_n = \frac{D_{tot}}{\sum_{m \in N} D_m^*} D_n^* \ \forall\, n \in N. \tag{9.1}$$

It should be noted that this scaling method causes a slight disturbance of $f_D(x_1, \ldots, x_N)$. I have not made any general mathematical analysis of the possible size of this disturbance, but I have been content with establishing that it is not large enough to have any practical importance; the errors caused by the scaling method are rather small compared to the random errors of a Monte Carlo simulation (cf. figure 9.1).

9.2 POSSIBILITIES FOR VARIANCE REDUCTION

Monte Carlo techniques are frequently used in scientific papers concerning power systems and electricity markets, but comparatively little have been written about the application of variance reduction techniques, which maybe can be taken as an indication that most authors have been satisfied with using simple sampling. Concerning calculation of reliability indices in power systems there are at least two thorough reviews of the possibilities and applications of variance reduction techniques [47, 59-60]. There are also some papers where one or more variance reduction techniques are applied to reliability calculations in power systems, for example [46, 56, 58, 61]. The most common variance reduction techniques in these contexts seem to be complementary random numbers and control variates, but stratified sampling is used in [56], and [61] uses an idea that resembles stratified sampling and importance sampling very much. Concerning calculation of expected operation costs most authors seem to focus on the three variance reduction techniques complementary random numbers, control variates and stratified sampling

3. In this context it is a major advantage if it can be assumed that the load is normally distributed and that the load in each area is independent of the load in the other areas. In many cases this requirement can be fulfilled by using different probability distributions of the load during different time periods. Assume for example that the load during daytime is $N(100, 15)$-distributed in a certain area and $N(50, 10)$-disturbed in another, and that these loads are independent of each other. By night the loads are $N(60, 10)$ and $N(30, 5)$-distributed instead. In this system there is a clear correlation between the load of the two areas, because the load decreases in both areas during the night, but within each time period (day and night) the loads can still be considered independent.
4. There is an example of a method to generate correlated random numbers from a general distribution in appendix C.

[49, 55, 57, 63, 64]. However, nobody seems to have studied variance reduction techniques for electricity market simulations, i.e., when one and the same simulation is used to estimate both reliability indices, as for example *LOLP*, and production cost indices, as for example *ETOC*.

In this section I will describe how the different variance reduction techniques described in chapter 8 could be used for simulation of short scenarios. I pay most attention to stratified sampling, as this is the method which is most complicated to apply. To gain maximal efficiency of stratified sampling it is of extraordinary importance to design the strata with care; yet surprisingly few studies have been made about strata design. In for example [49, 57, 64] strata are chosen more or less based on engineer's intuition. In [56] so-called poststratification is used,[5] which results in a certain loss of efficiency as the Neyman allocation can<not be applied. In [55] Huang describes a stratification strategy based on the "cum $\sqrt{f(y)}$-rule". The derivation of this rule is based on the assumption of a heterogeneous population,[6] but my view is that stratified sampling produces the largest gain when considering a duogeneous population, where a number of diverging units have a particular importance to the expectation value (loss of load scenarios are typical examples of diverging units in electricity market simulations; the remaining scenarios are conformist) and then another stratification strategy than the one used by Huang is necessary.

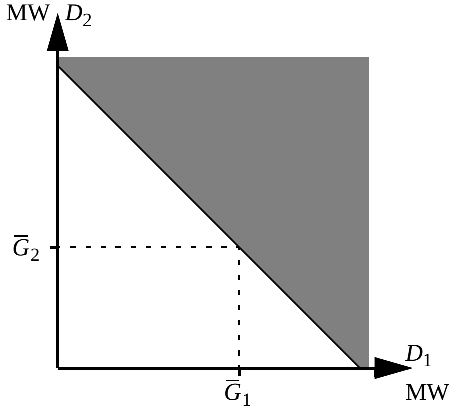
Below follows a description of the principles of my stratification strategy, which we may call *the strata tree method*. The first part of the description is theoretical; I start by identifying which scenarios have common properties (section 9.2.1) and then I show what a strata tree should look like in order to place scenarios having similar properties in the same stratum (section 9.2.2). Then a description of the practical implementation of stratified sampling follows (section 9.2.3). Finally, there is an analysis of the usage of the other variance reduction techniques in simulation of electricity markets (section 9.2.4).

9.2.1 The Properties of the Scenario Population

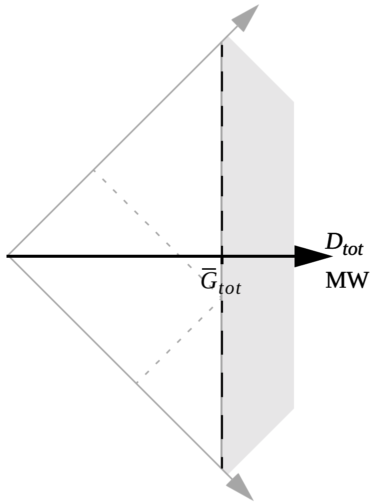
Stratified sampling produces a variance reduction if it can be managed to design homogeneous strata, i.e., when all units belonging to a certain stratum have similar properties [127]. In other words, the stratification demands knowledge about the properties of the population. Let us now study which properties short scenarios can be expected to have and how to separate different kinds of scenarios from each other in advance.

Assume that there is a power system with a given total available generation

5. See [127], section 5A.9.
6. See for example [127].



a) The set of scenarios. The shaded area corresponds to the loss of load scenarios.



b) Projection of the D_1D_2-plane on a D_{tot}-axis.

Figure 9.2 *Loss of load scenarios in a loss-free two-area system without transmission limitations.*

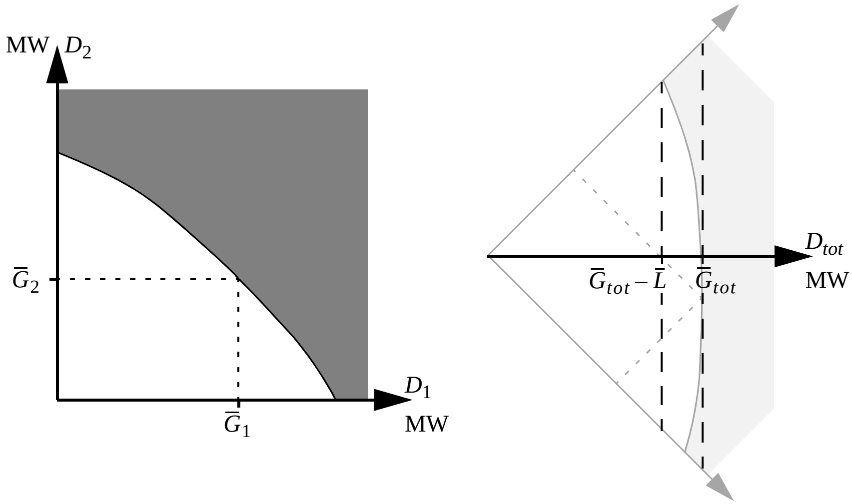
capacity, $\overline{G}_{tot}$, divided in N areas, so that the available generation capacity in each area is $\overline{G}_n$.[7] Moreover, assume that there is a certain available transmission capability, represented by the matrix $\overline{P}_{n,\,m}$. As the generation capacities and transmission capabilities are assumed to be fixed, the set of possible scenarios is equal to the set of possible load levels in the different areas. Let us now consider in which scenarios there will be a loss of load.

We could define a subset $D_N \in \boldsymbol{R}^N$, which includes all loss of load scenarios. However, it is more convenient to project the N-dimensional space of possible scenarios to a single dimension D_{tot}-axis, where D_{tot} is the total load in a scenario. All points on the hyperplane defined by

$$\sum_{n \in N} D_n = c, \tag{9.2}$$

where c is an arbitrary constant, will be represented by a single point on the D_{tot}-axis. Thus, the D_{tot}-axis can be seen as an axis orthogonal to the $(N-1)$-

7. Normally I use the symbol $\overline{G}_{tot}$ for available *thermal* generation capacity. To get away from the trouble of repeatedly writing $\overline{G}_{tot} + \overline{W}_{tot}$ I will sometimes use $\overline{G}_{tot}$ to designate *total* generation capacity in the following discussion. It should be clear from the context what is meant.



a) The set of scenarios. The shaded area corresponds to the loss of load scenarios.

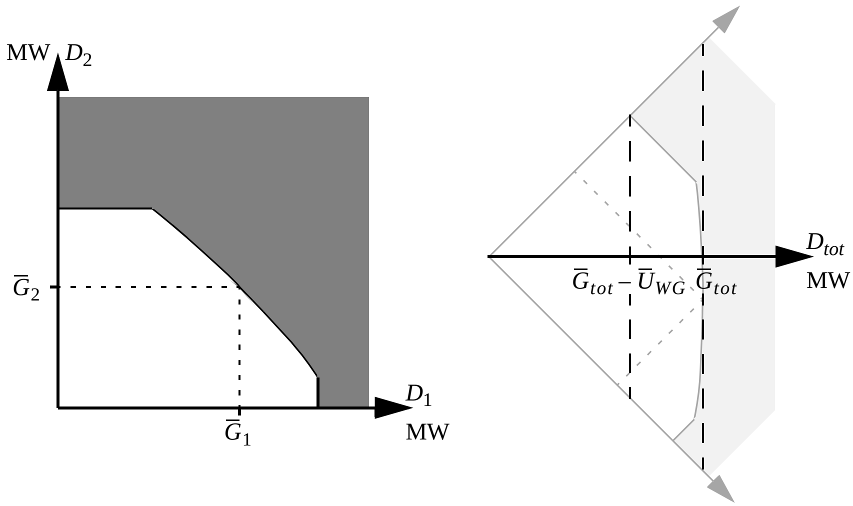
b) Projection of the D_1D_2-plane on a D_{tot}-axis.

Figure 9.3 *Loss of load scenarios in a two-area system with transmission losses but no transmission limitations. It may be noted that the losses in this system have been chosen excessively large to make the figure more clear.*

dimensional hyperplanes defined by (9.2). The idea can be illustrated for a two-area system, where the hyperplanes correspond to lines where $D_1 + D_2 = c$; an axis perpendicular to these lines is a bisector between the D_1- and D_2-axes, as shown in figure 9.2b.

In the most simple case we have loss-free transmission interconnections with unlimited capacity. Loss of load (i.e., scenarios where $LOLO = 1$) will then occur whenever the total load exceeds the total generation capacity. We may therefore divide the D_{tot}-axis in two partitions, leaving all loss of load scenarios in one partition, while the remaining scenarios are found in the other. The limit between these two partitions is $\overline{G}_{tot}$. Figure 9.2a illustrates this situation for a two-area system; the shaded area corresponds to the loss of load scenarios. In figure 9.2b it is shown how all scenarios in the shaded area are projected on the same part of the D_{tot}-axis.

The conclusion above is essentially quite obvious. It gets more interesting when transmission losses are added to the system. Now it is not sufficient that $D_{tot} \leq \overline{G}_{tot}$ to avoid loss of load, because the generation capacity must also cover the losses in the system. However, the losses depend on how generation resources and the load are distributed between the areas. If $D_n = \overline{G}_n$ in all



a) The set of scenarios. The shaded area corresponds to the loss of load scenarios.

b) Projection of the D_1D_2-plane on a D_{tot}-axis.

Figure 9.4 *Loss of load scenarios in a two-area system with transmission losses as well as transmission limitations.*

areas no power will be transmitted on any line and the losses will be zero. The larger the difference is between the geographical distribution of the generation capacity compared to the load, the larger the transmission losses will become (cf. figure 9.3a, which shows the scenarios which will be subject to loss of load in a two-area system with transmission losses).

In order to distinguish scenarios with different properties we now need to introduce the stratification parameter $\overline{L}$ as the largest possible losses in this system. Such a parameter must exist, because the losses cannot become infinite in a power system; however, in practise it might be difficult to calculate $\overline{L}$ analytically—but let us put that problem aside for the moment.[8] Given $\overline{L}$ there are three intervals of a D_{tot}-axis where the scenarios have different properties:

- None of the scenarios where $D_{tot} \leq \overline{G}_{tot} - \overline{L}$ is subject to loss of load.
- If $D_{tot} \in (\overline{G}_{tot} - \overline{L}, \overline{G}_{tot})$ then it is not possible to determine whether or not there will be any loss of load without studying the exact relation between $\overline{G}_n$ and D_n.

8. The practical consequences when $\overline{L}$ cannot be calculated exactly are treated in section 9.2.3.

- If $D_{tot} > \overline{G}_{tot}$ then loss of load is inevitable.

Finally we add transmission limitations to the system. With that there is a possibility that loss of load occurs even when $D_{tot} \leq \overline{G}_{tot} - \overline{L},$ because there may be situations where a part of the available generation capacity cannot be utilised due to transmission congestion. These situations mostly occur when the load has a significantly different geographical spread compared to the available generation capacity (cf. the example in figure 9.4).

We now introduce another stratification parameter, $\overline{U}_{WG},$ which is the maximal unused generation capacity. As with $\overline{L}$ it is generally not possible to determine the exact value of $\overline{U}_{WG},$ but we definitely know that there is a maximal unused generation capacity. If it was known, we would find four interesting intervals on the D_{tot}-axis:

- If $D_{tot} \leq \overline{G}_{tot} - \overline{U}_{WG}$ then loss of load is impossible.
- When $D_{tot} \in (\overline{G}_{tot} - \overline{U}_{WG}, \overline{G}_{tot} - \overline{L})$ loss of load may occur due to transmission congestion.
- When $D_{tot} \in (\overline{G}_{tot} - \overline{L}, \overline{G}_{tot})$ both transmission congestion and losses may cause loss of load.
- If $D_{tot} > \overline{G}_{tot}$ then loss of load is inevitable.

The same reasoning which above was applied to *LOLO* can also be applied to study how the size and distribution of the load will affect the operation cost, *TOC*, in a system where there is a significant possibility that the generation capacity having negligible operation costs, $\overline{W}_{tot},$ is sufficient to supply the total load, D_{tot}. The shaded areas in figures 9.2-9.4 will in this case correspond to scenarios where $TOC > 0$, whereas the other scenarios are such that $TOC = 0$. The stratification parameters necessary to separate the different partitions of the D_{tot}-axis are the maximal losses, $\overline{L}$ (the same parameter as when considering *LOLO*) and the maximal unused generation capacity having negligible operation costs, $\overline{U}_{W}$.

The analysis above can be used to compile seven main types of scenarios, the properties of which concerning *LOLO* and *TOC* are shown in table 9.1. Depending on how the available generation capacity relates to the stratification parameters, there may also appear some special types of scenarios. Most of these special types are so rare that they do not need to be given special treatment; if they are neglected due to cardinal error then it will only have an insignificant impact on the final result. The only cases I have encountered where scenarios of special types have had great importance, are systems where there is a significant risk[9] that there is no thermal generation capacity available, while there is a certain generation capacity with negligible costs. Such systems are not common, but they may appear when studying for example small, isolated systems supplied by non-dispatchable units and one or a

9. Which means at least a couple of per cent. Exactly how large the risk must be depends on which accuracy is desired.

Table 9.1 *The main types of scenarios.*

Type	*TOC*	*LOLO*	Part of the D_{tot}-axis
I	0	0	$D_{tot} \leq \overline{W}_{tot} - \overline{U}_W$
II	≥ 0*	0	$\overline{W}_{tot} - \overline{U}_W < D_{tot} \leq \overline{W}_{tot} - \overline{L}$
III	≥ 0**	0	$\overline{W}_{tot} - \overline{L} < D_{tot} \leq \overline{W}_{tot}$
IV	> 0	0	$\overline{W}_{tot} < D_{tot} \leq \overline{W}_{tot} + \overline{G}_{tot} - \overline{U}_{WG}$
V	> 0	0 or 1*	$\overline{W}_{tot} + \overline{G}_{tot} - \overline{U}_{WG} < D_{tot} \leq \overline{W}_{tot} + \overline{G}_{tot} - \overline{L}$
VI	> 0	0 or 1**	$\overline{W}_{tot} + \overline{G}_{tot} - \overline{L} < D_{tot} \leq \overline{W}_{tot} + \overline{G}_{tot}$
VII	> 0	1	$\overline{W}_{tot} + \overline{G}_{tot} < D_{tot}$

* Depending on the impact of transmission limitations.
** Depending on the impact of transmission losses.

Table 9.2 *Some special types of scenarios.*

Type	Properties	Part of the D_{tot}-axis	Condition
VIII	*TOC* = 0 *LOLO* = 0 or 1*	$\overline{W}_{tot} - \overline{U}_W < D_{tot} \leq \overline{W}_{tot} - \overline{L}$	$\overline{G}_{tot} = 0$
IX	*TOC* = 0 *LOLO* = 0 or 1**	$\overline{W}_{tot} - \overline{L} < D_{tot} \leq \overline{W}_{tot}$	$\overline{G}_{tot} = 0$

* Depending on the impact of transmission limitations.
** Depending on the impact of transmission losses.

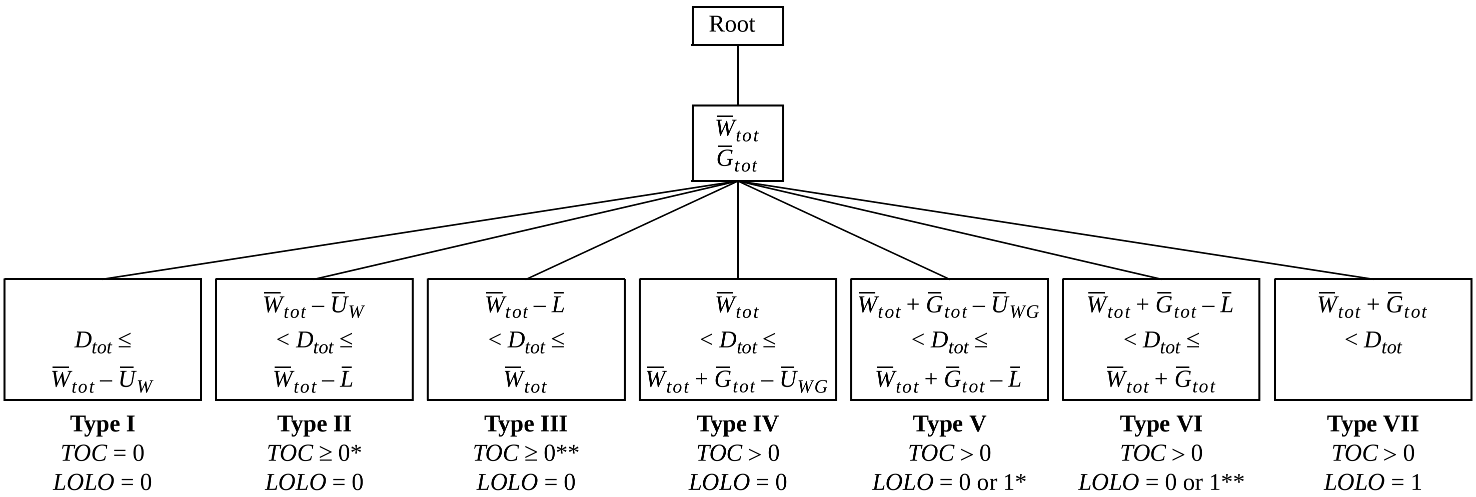
few diesel generator sets. When simulating those systems, it is recommended to consider scenarios of type VIII and IX according to table 9.2.

9.2.2 Strata Trees

By using a strata tree we may divide the set of possible scenarios into subpopulations, where all scenarios in a subpopulation are of a certain type (cf. table 9.1). The basic idea is to compare the total available resources with the total demand, which can be done in the strata tree by placing the scenario parameters describing the resources above the scenario parameters describing the demand. In this section I will describe how to design strata trees according to this basic idea and how the strata tree then is used to define strata. Finally, I will give an example of a poor method to design strata trees, so that there is no hesitation on how *not* to do.

Basic Strata Trees

Let us for the sake of simplicity start by studying stratified sampling in a multi-area system having just two scenario parameters: available generation



* Depending on the impact of transmission limitations
** Depending on the impact of transmission losses

Figure 9.5 *Part of a strata tree. Each state of the available generation capacity is located in a node just below the root of the strata tree (the complete strata tree therefore comprises further nodes holding the other possible states of the available generation capacity). For each such node we can create a number of child nodes, corresponding to the partitions of the D_{tot}-axis which was identified in the theoretical analysis. As can be seen in the figure, the properties of the scenarios differ between the different branches, but are relatively homogeneous within each branch. It can be noted that branches of type V and VI to begin with seem to have the same properties, but in general it is a significant difference between the* share *of scenarios for which LOLO* = 1. *The same goes for TOC and branch II and III respectively.*

capacity and load. The proper way to stratify such a system is to have a strata tree with two levels below the root. In the level just beneath the root we create a separate node for each state of the available generation capacity. As each node thus specifies a well-defined generation capacity, we have achieved the situation analysed in section 9.2.1, i.e., the properties of the scenarios only depend on the load. In the lowest level we can now add nodes which specify intervals of the total load. These intervals are chosen to correspond to the partitions from table 9.1.[10] Each branch of the strata tree will only contain scenarios of a certain type, as shown in figure 9.5.

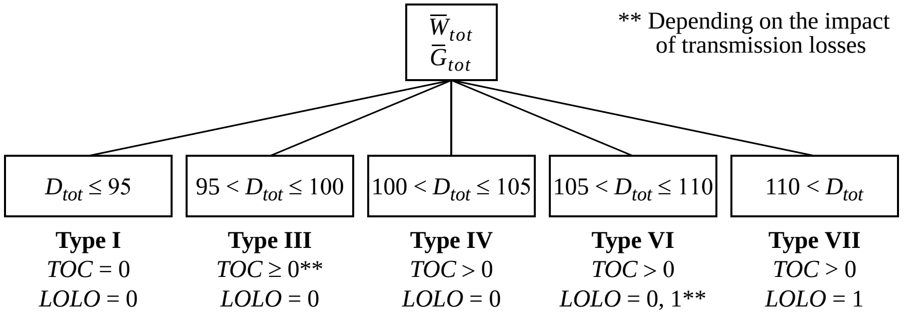
All types of scenarios do not have to appear in each part of the strata tree. If for example $\overline{W}_{tot} = 0$ there will be no scenarios of type I-III. Besides, there may also be other types of scenarios than the main types I-VII. When building the strata tree one node may be added for each special type (cf. figure 9.6b) or the special cases may be neglected and sorted under a similar main type (cf. figure 9.6c). As concluded in the end of the previous section, the special types shown in table 9.2 might be important for the final result, but in many cases it is possible to not separate the special cases, which simplifies the life of lazy programmers.[11]

When building the strata tree it may for numerical reasons be necessary to neglect nodes having a very small node weight.[12] For example, if there is a load which is normally distributed and the mean is 1 000 MW and the standard deviation is 100 MW, while the total available generation capacity amounts to 2 000 MW, then the probability that the load exceeds the available generation capacity is in practice zero; hence, there is no point in including a branch of type VII in this case. The scenarios which would have belonged to the cut off branch are transferred to the preceding D_{tot}-node instead.

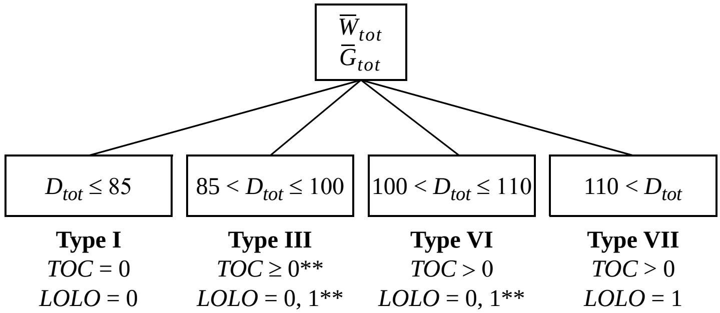
Additional Scenario Parameters

Electricity market models having more scenario parameters than available generation capacity and load can be managed by adding more levels to the strata tree. If it is considered that not just power plants but also transmission lines are subject to failures then the scenario parameter available transmission capability must be added to the strata. Since the derivation of the different

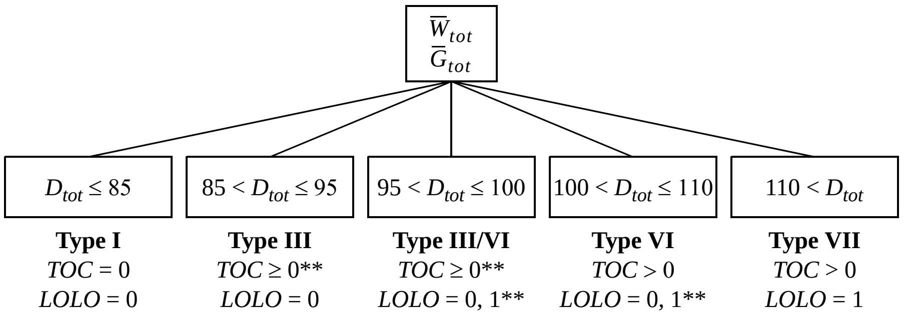
10. To choose these intervals correctly, we need to know the stratification parameters $\bar{U}_W$, $\bar{U}_{WG}$ and $\bar{L}$. How to calculate these is a question which is treated more closely in section 9.2.3; for the moment, we concentrate on how an appropriate strata tree would look like.
11. Which of course refers to myself.
12. I cannot specify an exact limit on how large the node weight must be in order to include a node in the strata tree, because it depends on how large errors that can be accepted.



a) $\overline{W}_{tot} = 100$, $\overline{G}_{tot} = 10$, $\overline{L} = 5$.



b) $\overline{W}_{tot} = 100$, $\overline{G}_{tot} = 10$, $\overline{L} = 15$.



c) $\overline{W}_{tot} = 100$, $\overline{G}_{tot} = 10$, $\overline{L} = 15$.

Figure 9.6 *Example of managing special types of scenarios. The maximal losses in figure a are less than the available thermal generation capacity and the load nodes follow the standard pattern. However, as there are no transmission limitations partition I and V are omitted. In figures b and c the losses are larger than the thermal generation capacity, which causes partition III and VI to partly overlap each other. In figure b this has been solved by ignoring partition IV. Moreover, the overlapping part is put in partition III, which means that there is a certain risk of load shedding in partition III (which normally should have sufficient generation capacity)—by that means a small risk of cardinal error is introduced. In figure c the overlapping part is forming its own partition instead, and its properties is a combination of partition III and VI.*

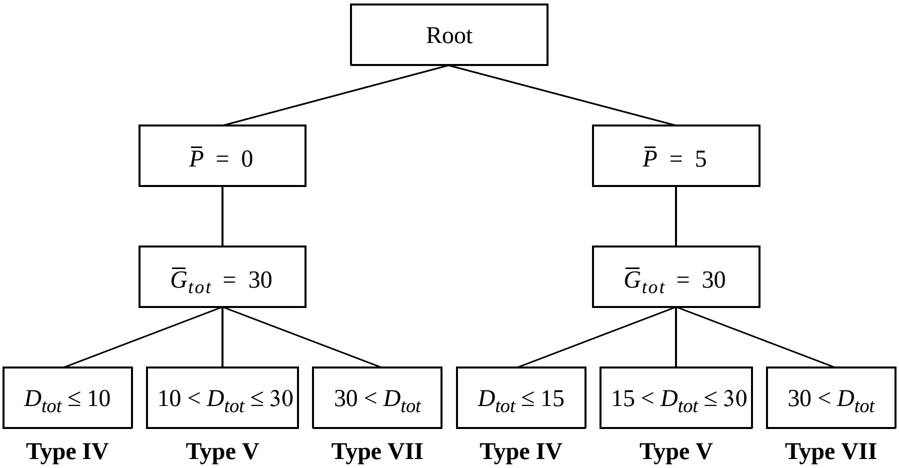


Figure 9.7 *Strata tree including failures in transmission lines. The figure shows the strata tree of a simple two-area system. In the first area there is a power plant with the capacity* 10 MW *and in the other there is a* 20 MW *power plant in the other area. Both power plants are assumed to be completely reliable. The areas are interconnected by a loss-free transmission line, which can transmit* 5 MW *when available. In this case the largest unused generation capacity,* $\bar{U}_{WG}$*, equals* 15 MW. *If the transmission line is unavailable* $\bar{U}_{WG}$ *increases to* 20 MW, *which result in somewhat different limits between the partitions in the left part of the tree compared to the right part. It should however be noted that the extra level in the strata tree does not introduce any new types of branches than those defined in i figure 9.5.*

partitions concerning the level of the total load in section 9.2.1 was based on studying the power system for a specified generation capacity and a specified transmission capability we need to arrange the strata tree so that each branch corresponds to a fixed value for each of those two scenario parameters.[13] If we want to, we may simply merge these parameters to a single scenario parameter (available resources) and locate each state of the resource parameter in the top level below the root in the strata tree. The alternative I most frequently use—mainly due to its perspicuousness—is to have available transmission capability and available generation capacity on separate levels in the strata tree. The principle is illustrated in figure 9.7.

Correlations

Another situation when it might be desirable to expand the strata tree is to

13. Hence, a strata tree can become huge—see page 193 for additional comments around this problem.

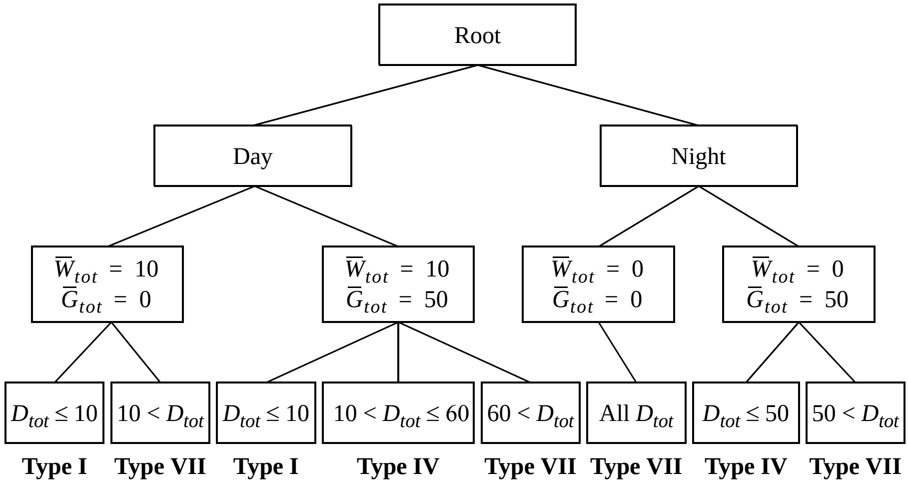


Figure 9.8 *Strata tree with "point of time level". The figure shows a strata tree for a small system with photovoltaics and a diesel generator set. The photovoltaics are assumed to generate* 10 kW *constantly during the bright hours of the day and nothing at all during the night. The diesel generator set can generate* 50 kW *when it is available. The transmission losses are negligible and there are no transmission limitations.*

Notice that the extra level in the strata tree does not introduce any new types of branches than those defined in i figure 9.5.

manage some correlations between scenario parameters. According to the definition of strata tree (see section 8.2.5) the scenario parameters along a branch must be independent of each other. (Notice that this does not mean that all scenario parameters must be independent; it is for example no problems to have correlations between the load in different areas of the system, because it is only the total load which is represented in the strata tree.) In most cases it is also reasonable to assume that this requirement is fulfilled, but there are some obvious exceptions. The most clear example could be electricity generation in photovoltaic modules. During the day, when the load is comparatively high, the photovoltaic generation is maximal, and during the night, when the load decreases to lower levels, the photovoltaic modules generates nothing at all; apparently there is a certain positive correlation between the photovoltaic generation and the load. Similar correlations, both positive and negative ones, can also be found between other non-dispatchable power plants and the load.

Correlations can in many cases by managed by the observation that if we limit ourselves to certain time periods—part of the day, season or whatever it may be—then it is once again reasonable to consider the scenario parameters as independent of each other. It is for example hard to find any correlation between the photovoltaic generation during daytime and the load; a cloud temporarily blocking the sunshine and decreasing the power generation should not influence the load in any particular direction. Thus, the correlation may be managed by locating different points of time in different parts of the strata tree on the top level just below the root. This idea is illustrated infigure 9.8. By separating different points of time it is also possible to indirectly manage correlations between the load in different areas.[14]

From Strata Tree to Strata

The strata tree is as mentioned earlier a tool for sorting the scenario population. Starting from this sorting it is then possible to finally define strata. The most simple way of defining strata from a strata tree is to let each branch represent a strata.[15] I have chosen to call this strategy *"complete stratification"*. As each branch only includes scenarios of one type, we have achieved the goal of keeping each stratum as homogeneous as possible. The disadvantage is however that if the number of states for the available generation capacity is large then the number of strata might become annoyingly large. When the simulation is started, a large number of strata is a problem, because the simulation requires that a few samples are taken from each stratum. In other words, if the number of strata is large then the least possible number of scenario swill be large. To reduce the number of scenarios, it is therefore recommended to minimise the number of strata.

An alternative to the complete stratification is to let each stratum comprise several branches of the strata tree; all branches of type I have similar properties concerning *TOC* and *LOLO*, and may therefore be put in a common stratum, etc. With that, the number of strata is limited to the number of scenario types (see table 9.1 and 9.2). This strategy I call a *"reduced stratification*".

Large Strata Trees

In a really large power system with hundred of power plants, the number of possible states of the available generation capacity might become very large; consequently, the strata tree is even bigger. Things are even worse if there in addition is a large number of possible states for the transmission capability between the areas; when we identify the different types of scenarios each possible combination of available generation capacity and transmission capability must be treated separately.

The question must be raised whether the gain of using stratified sampling is

14. See footnote 3.
15. This is how the technique was originally presented [10].

worthwhile the effort to calculate $F_{\overline{W} + \overline{G}}$ and building the strata tree. This is partly a question about the algorithm which is used—the difficulty is the problem size and not that the computations are particularly complicated in their own. I have not performed any closer studies of which algorithms might be useful and how many states it is reasonable to manage. However, quite a good guess would be that there is a pain threshold, where the number of states of the generation capacity simply is too large.[16]

This does not mean that stratified sampling should be impossible to use for really large systems. Probably there will be a limited number of large power plants, and a larger number of smaller power plants. By just including the larger power plants in the common distribution function $F_{\overline{W} + \overline{G}}$, we can limit the size of the strata tree. The available generation capacity in the smaller power plants is randomised independently of each other (or maybe using dagger sampling) and is thus not part of the stratification. When comparing the total load and the total available capacity in the larger units, it has to be considered that there is also an undetermined generation capacity in the smaller power plants. The consequences of this unknown parameter corresponds to the consequences of not knowing in advance the size of the losses in a particular scenario, as long as the total generation capacity in the power plants regarded as "smaller" is small compared to the total generation capacity of the other power plants.

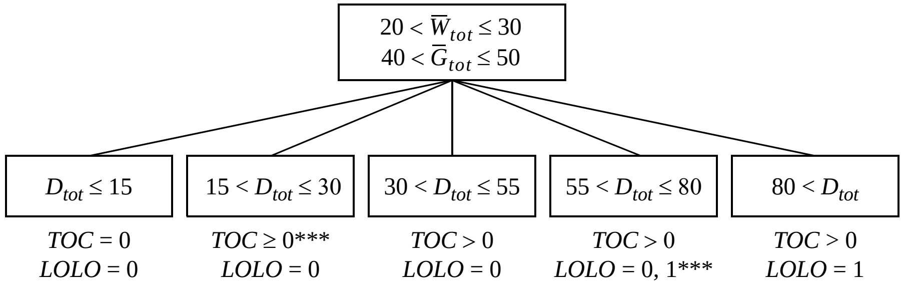
Similar reasoning should also be applicable when there is a large number of states for the available transmission capability. The problems of large strata trees does however require further studies, in order to find out how to best manage them best.

Unsuitable Strata Trees

Above I have described how a strata tree should be designed and how the tree should be used to define strata. I have at some occasions[17] used another method to build strata trees; the difference is that the nodes specifying the available generation capacity comprises an interval rather than just a single state. This is unfortunately not a very good solution. Below I will briefly explain why, so that there is no confusion about how *not* to do when defining strata.

The idea behind the method—which we may refer to as "*multiple state nodes*" was the same as for the reduced stratification, i.e., to limit the number of strata. The problem is that when comparing an interval of available resources to the total load, the partitions where the properties of the scenarios cannot be predicted, but where *TOC* and *LOLO* have to be calculated, are

16. Cf. [61], figure 2.
17. For example [7, 10].



*** Depending on available generation capacity as well as transmission losses

Figure 9.9 *Strata tree with multiple state nodes. In this example it is assumed that* $\bar{L} = 5$.

comparatively large, as illustrated in figure 9.9. Hence, $Var[m_{TOC}]$ and $Var[m_{LOLO}]$ will increase and consequently a larger number of samples will be required to achieve a certain level of accuracy.

It can also be noted that the risk of cardinal error increases with multiple state nodes—especially if there are states with very different probabilities in the same node.[18] In order to avoid cardinal error, one could try to develop a lot of special rules which prohibit certain states to be located in the same node, but it is difficult to simultaneously keep the rules simple and general enough to be possible to implement.

Finally, it can be concluded that although the number of strata decreases compared to the complete stratification, a stratification using multiple state nodes will result in a lot more strata than usage of the reduced stratification.

9.2.3 Stratified Sampling in Practice

In the previous section I have identified scenarios with different properties and shown how the scenarios can be sorted in a strata tree. Using this knowledge, we may formulate a general method to apply stratified sampling in practice.

18. For example, assume that there is a multiple state node where we have 99% probability that the available generation capacity is 120 MW and 1% probability that only 100 MW is available. In the interval $100 < D_{tot} \leq 120$ it is then very likely that the generation capacity is sufficient (if we disregard the losses), but there is also a small possibility that the capacity is insufficient. This small risk is easily overlooked if only a few samples are collected from the interval.

Presimulation

To define strata we need to know the stratification parameters $\bar{U}_W$, $\bar{U}_{WG}$ and $\bar{L}$. Unfortunately, it is in practice impossible to calculate these analytically in systems with more than two areas. Without an analytical method of calculation we may—of course—use Monte Carlo techniques; we simply perform a presimulation, the primary objective of which is to estimate the stratification parameters necessary to build the strata tree. However, just studying the stratification parameters would be uncalled-for; when analysing scenarios it takes only a negligible extra work to also sample the usual result variables (*TOC* and *LOLO*) too. The results of all scenarios included in the presimulation should therefore be stored, together with the values of the scenario parameters in each studied scenario. (The latter are needed in order to sort the presimulation scenarios in the right stratum after the stratification.)

In the presimulation we cannot use stratified sampling, but the other variance reduction techniques may be applied as usual. It is desirable to use as few scenarios as possible in the presimulation, as stratified sampling accounts for a large part of the variance reduction. The question is therefore how many scenarios are needed to get estimates of the scenario parameters, which are so accurate that the resulting stratification is efficient.

The maximal losses actually depend on the available generation capacity in each area. However, in practice it does not matter if the maximal losses are somewhat overestimated, and it is therefore possible to use the same $\bar{L}$ in the entire strata tree. By those means, it is sufficient to create a number of scenarios, consider the total losses in each scenario, L_i, and then estimate $\bar{L}$ as the highest recorded value plus a safety margin of for example 50%, i.e.,

$$\bar{L} \approx 1.5\max(L_i),\ i = 1, \ldots, n_{presimulation}.^{19} \tag{9.3}$$

As losses appear in all scenarios it is not necessary that $n_{presimulation}$ is very large to get a sufficiently good estimate of $\bar{L}$. In the tests I describe in section 9.3, I used 100 scenarios in the presimulation and it worked splendidly, regardless of whether the power system was divided in two, six or seven areas.

To fully utilise the benefits of stratified sampling we also need to estimate $\bar{U}_W$ and $\bar{U}_{WG}$. Unlike $\bar{L}$ these stratification parameters are a lot more difficult to obtain. They are in the highest degree depending on both available generation capacity, $\bar{W}_{tot}$ and $\bar{G}_{tot}$, as well as available transmission capability, $\bar{P}$, which forces us to calculate separate values of these stratification parameters for each combination of available generation and transmission

19. I have not systematically studied how large the safety margin should be—the number 50% is just a qualified guess, which has turned out to work well in practice.

resources. Besides, transmission congestion is a comparatively rare event, which not at all appear in every scenario. Therefore, a lot more scenarios are required the presimulation before adequate estimates of $\bar{U}_W$ and $\bar{U}_{WG}$ are obtained. Consequently, there is a risk that many scenarios of the presimulation are selected from the parts of the scenario population which later turns out to be of rather small significance. This is a waste of work. I noted this problem of estimating $\bar{U}_W$ and $\bar{U}_{WG}$ first when I was writing this chapter of the dissertation and I have not had time to test software to overcome the predicament. I will therefore have to restrict myself to a short reasoning around the alternatives available:

- **Forbearance.** The first alternative is to swallow the annoyance and accept that the variance reduction will not be as large in system with transmission limitations.[20] The possibility to use stratified sampling in the main simulation, as well as the other variance reduction techniques will eventually produce a certain gain of efficiency.
- **Preliminary stratification.** The consequences of not knowing $\bar{U}_W$ and $\bar{U}_{WG}$ are that the limits between scenarios of type II and III as well as between type IV and V will not be known. What we can do is to introduce a number of preliminary limits. When a few batches of the main simulation have been run, then we may see which of these preliminary limits best corresponds to the real. The remaining preliminary limits are removed (i.e., the branches of the strata tree separated by these limits are merged) and the main simulation continues. Using this method we get a short presimulation without stratified sampling, a semi-long preliminary main simulation with slightly hampered efficiency, as we do not have enough information to define optimal strata. Finally we have a continued main simulation, where the stratification can produce maximal benefits.
- **Other variance reduction techniques.** The efficiency of the presimulation might be enhanced by replacing stratified sampling by some other variance reduction technique. To estimate $\bar{U}_W$ and $\bar{U}_{WG}$ it is required that transmission congestion occurs in the selected scenarios. These situations are most common when the load is geographically lopsided compared to the generation capacity (cf. figure 9.4). It could therefore be possible to use importance sampling to increase the probability of having such scenarios in the presimulation. This can be done without deteriorating the estimate of $\bar{L}$, because the transmission losses

20. “Whether ’tis nobler in the mind to suffer the slings and arrows of outrageous fortune” as it has been beautifully written—although the subject was a little bit more dramatic than the fine-tuning of an electricity market simulation…

are also larger when the transmission lines are heavily loaded; hence, better estimations of $\bar{L}$ too are obtained when extreme geographical distributions are overrepresented among the samples.

Stratification

After the presimulation we have access to estimates of the stratification parameters $\bar{U}_W$, $\bar{U}_{WG}$ and $\bar{L}$; thus, we can build a strata tree according to the principles described in section 9.2.2. The strata can be defined from the strata tree and the scenarios of the presimulation are sorted accordingly.

Pilot Study

In the pilot study it is possible to use stratified sampling, but there is still not enough information to determine the optimal distribution of samples according to the Neyman allocation; the number of scenarios per stratum must be chosen in some other way. As the Neyman allocation is the best way to divide the scenarios between the strata, we want to use as few scenarios as possible in the pilot study. It is however necessary that the number of scenarios is large enough to guarantee that we find all strata where the variance is larger than zeros, because otherwise the simulation will risk to suffer from the cardinal error (see section 8.2.5).

The most straightforward method is to quite simply allocate a predetermined number of scenarios to each stratum. Most simple is to use exactly the same number of scenarios in each stratum (so-called proportional allocation), but since we have a certain knowledge about which properties the scenarios have in different kinds of strata, it would be a waste not to use this knowledge. For example, we know that in stratum of type I both *TOC* and *LOLO* are always equal to zero; hence, *ETOC* and *LOLP* can be calculated theoretically for a stratum of type I (or—if someone absolutely wants to sample—one single sample is sufficient). In other types of strata there are diverging units which have to be identified in the pilot study in order to avoid cardinal error in the main simulation, and then a larger number of scenarios is required. Exactly how many scenarios it takes varies between different systems, because it depends on the proportion of diverging units (cf. table 8.2). This proportion is of course not known before the system has been simulated, but based upon the characteristics of the stratum type it is yet possible to get an approximate understanding of an appropriate sample allocation in the pilot study. I suggest some simple rules of thumb in table 9.3. Generally speaking, it can be said that it is possible to be a little bit more generous concerning the number of scenarios per stratum in the reduced stratification, because there

***Table 9.3** Rules of thumb concerning sample allocation in the pilot study.*

Type	Approximate number of scenarios		Comments
	Complete stratification	Reduced stratification	
I	None	A dozen	Both *TOC* and *LOLO* area homogeneous in the stratum.
II	A few hundreds	A few thousands	Here it is necessary to collect as many scenarios as is required to find at least one case where *TOC* > 0 due to transmission congestion.
III	Some dozens	A few hundreds	Here it is necessary to collect as many scenarios as is required to find at least one case where *TOC* > 0 due to transmission losses.
IV	Two	Some dozens	Here it is only *TOC* which varies, but *TOC* is on the other hand very strongly correlated to the total load. As it is very unlikely that two scenarios have exactly the same total load (actually its impossible if complementary random numbers are used) then two scenarios is sufficient to avoid cardinal error.
V	A few hundred	Some thousands	Here it is necessary to collect as many scenarios as is required to find at least one case where *LOLO* = 1 due to transmission congestion.
VI	A few dozens	Some hundreds	Here it is necessary to collect as many scenarios as is required to find at least one case where *LOLO* = 1 due to transmission losses.
VII	None or a few dozens	A dozen	If there are no transmission limitations in the system then both *TOC* and *LOLO* are obvious, because all available generation capacity will be utilised, but still there will be loss of load. If there are transmission limitations then it is possible that some generation capacity will remain unused, causing *TOC* to vary slightly. However, if these variations cannot be detected using more than a few dozens scenarios then they are probably negligible.

are not that many strata.

Another method of choosing the number of scenarios in the pilot study is to utilise the fact that we know in which strata the result variables are homogeneous and heterogeneous respectively, which makes it possible to use a sort of dynamic sample allocation. If we know for a certain stratum that *TOC* is homogeneous and *LOLO* is heterogeneous then we should continue to randomise scenarios for this stratum until at least one diverging sample of *LOLO* has been encountered. Similar procedures can be applied to the other strata. If desirable, it is also possible to have a lower limit to the number of scenarios created in each stratum.

The dynamic sample allocation is not very practical if the number of strata is large and the proportion of diverging units is small in several strata, because each of these strata will then require thousands of scenarios before a diverging unit has been found. Dynamic sample allocation in the pilot study is therefore not recommended if the complete stratification is used. In the reduced stratification, where there are at most a dozen strata or so, the method is very efficient, as far as I have perceived in practical tests.

Main Simulation

The main simulation is divided in batches, where each batch includes a fixed number of scenarios.[21] Each batch starts by checking the stopping rule. If the conditions are fulfilled, the simulation is terminated and the results are presented. Otherwise, more scenarios have to be studied. As there is an estimate of the variance for each stratum and for each of the result variables, it is just to calculate the optimal distribution of the samples according to the Neyman allocation. However, a dilemma is that the optimal allocation normally differs between different result variables. It can then be questioned whether or not it would be more efficient to simulate each result variable one by one, in order to use the optimal scenario distribution with respect to each result variable. The answer is however that this is not the case. When analysing a scenario there is practically no time to be saved by calculating just one result variable and ignore the others. The only sensible option is then to let every scenario generate samples of all result variables. By collecting these samples, they may of course as well be utilised; it is not possible to loose efficiency, but there is on the other hand a possibility of obtaining some gains (cf. the example in figure 9.10).

There are several possible alternatives to find compromise allocations when there are more than one result variables. I recommend the following simple algorithm:

- Calculate the Neyman allocation with respect to each result variable.
- Calculate the mean of these allocations.
- Compare the calculated mean allocation to the number of scenarios which already have been studied in each stratum.

In the last step it may turn out to be impossible to allocate the desired number

21. I have not studied how many scenarios there should be in each batch—my experience is that it does not make any significant difference if each batch comprises 100 scenarios or 1 000 scenarios.

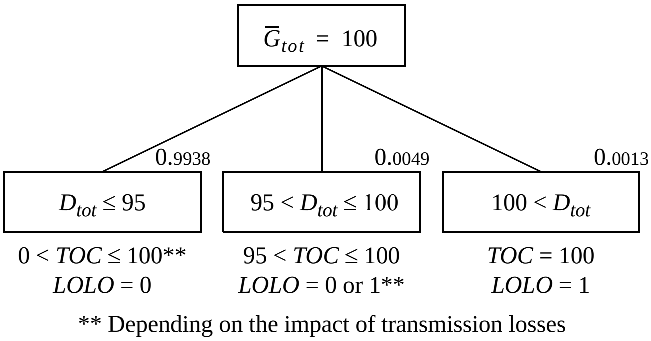


Figure 9.10 *The benefit of simulating several result variables simultaneously. The example shows a system with only one power plant, which has an installed capacity of* 100 MW, *never fails and can generate electricity to a cost of* 1 ¤/MWh. *The total load is* $N(70, 10)$*-distributed and the maximal losses,* $\bar{L}$, *are* 5 MW. *There are no transmission limitations in the system. The figure above shows the strata tree of this little system. Each branch corresponds to a stratum, and the stratum weights are equal to the node weights of the lower level in the tree.*

Concerning the ETOC estimate, stratum 3 is completely uninteresting, as the power plant is always operated at its installed capacity in these scenarios, causing TOC to be completely homogeneous in this stratum. In the other two strata TOC may however vary, but as stratum 1 has both higher stratum weight and variance, the Neyman allocation with respect to TOC will allocate an overwhelming majority of the samples to stratum 1. Concerning the LOLO estimate, both stratum 1 and 3 are uninteresting, because LOLO is homogeneous in both strata. The Neyman allocation with respect to LOLO will therefore direct all samples to stratum 2. A compromise between the two allocations would be to distribute the scenarios rather evenly between stratum 1 and 2.

Assume that it takes 100 scenarios in stratum 1 and 10 in stratum 2 in order to obtain a sufficiently good estimate of ETOC, while it takes 100 scenarios in stratum 2 in order to correctly estimate LOLP. If both system indices are estimated in separate simulations then in total 210 scenarios would be required. If they are estimated concurrently it will be sufficient with 200 scenarios. The efficiency gain from treating the result variables simultaneously is due to the fact that even though the scenarios which the compromise allocation directs to stratum 2 mainly are necessary for the LOLP estimate, they will also have a certain value for the ETOC estimate.

of scenarios to each stratum. The strata that have already been allocated too many scenarios should of course not be given any more. Concerning the remaining strata, we should try to get as close as possible to the mean allocation.[22]

It is possible to use a weighted mean when calculating the compromise allocation, in order to give some result variables a larger impact on the sample allocation than others. The weights do actually not even have to be constant during the course of the simulation; it is for example possible to reduce the weight to zero for the result variables where we have already estimated the expectation value with desired accuracy (i.e., which fulfils the stopping rule), so that this result variable no longer will influence the sample allocation.

I have not tried to make any detailed studies of how the choice of compromise allocation in the main simulation affects the accuracy of the final result. In those cases where I have tried simulating the same system using several alternative methods, I have barely been able to notice any differences. The Neyman allocation is fortunately corresponding to a rather flat optimum, which means that small deviations from the optimal distribution do not necessarily have any larger importance [127].

After each batch of new scenarios the estimates of *ETOC* and *LOLP* are updated as well as the variance within each stratum. Then the above procedure is repeated for the next batch.

9.2.4 Other Variance Reduction Techniques

Stratified sampling is not the only variance reduction technique that can be applied to simulation of electricity markets. Fortunately, the other techniques are somewhat more straightforward to apply. In this section I will summarise how the other variance reduction techniques may be used. I will also present my opinion about which techniques seem most appropriate to utilise. The choice of variance reduction technique is both about the efficiency of the method and how well it can be combined with other techniques. In many cases it is possible to combine several techniques, because they achieve an efficiency gain by using different kinds of information known in advance, but there are some exceptions, which I will describe in more details below.

Complementary Random Numbers

Applying complementary random numbers to simulation of electricity mar-

22. For example, assume that the desired allocation is 50, 50 and 100 scenarios per stratum, and that 60, 38 and 52 scenarios respectively have been studied in each stratum already. In this case stratum 1 has obtained 10 “extra” scenarios, which inevitably means that in total 10 scenarios must be missed in strata 2 and 3; in total we would like to allocate the second stratum 12 scenarios and the third should be given 48, but there are only 50 scenarios to be distributed (the total number of scenarios should be 200 and we have already studied 150). As a suggestion, we can distribute the deficit so that the relative deficit is more or less the same in both strata, i.e., allocate 9 new scenarios to stratum 2 (in total 47 scenarios, i.e., a deficit of 6%) and 41 new scenarios to stratum 3 (in total 93 scenarios, i.e. a deficit of 7%).

kets is not very difficult, because the method only affects the random number generation and does not require any other measures. A special consideration is however that it usually takes more than one random number to generate a scenario, because there are several scenario parameters. When applying complementary random numbers one set of original scenario parameters and one set of the corresponding complementary random numbers are obtained. All of these values can then be combined into 2^s complementary scenarios, where s equals the number of scenario parameters. An example of this is given in table 9.4. In order to at least somewhat restrict the number of complementary scenarios, it is appropriate to apply complementary random numbers only to the primary scenario parameters.

Table 9.4 *Complementary scenarios for three scenario parameters.*

$\overline{G},\ \overline{P},\ D$	$\overline{G},\ \overline{P},\ D^*$	$\overline{G},\ \overline{P}^*,\ D$	$\overline{G},\ \overline{P}^*,\ D^*$
$\overline{G}^*,\ \overline{P},\ D$	$\overline{G}^*,\ \overline{P},\ D^*$	$\overline{G}^*,\ \overline{P}^*,\ D$	$\overline{G}^*,\ \overline{P}^*,\ D^*$

The prerequisite for complementary random numbers to be efficient is as earlier mentioned that there is a negative correlation between an observation based on a particular random number and an observation originating from the complementary random number. If there is a correlation (positive or negative) between a certain scenario parameter and a certain result variable, it is likely that there will be a negative correlation between the original scenario and the complementary scenarios.

It is obvious that there should be a clear correlation between the operation cost and the load; low loads result in low operation costs and high loads will yield high operation costs. Applying complementary random numbers to the load should therefore be efficient. There is also a certain correlation between the available generation and transmission resources and the operation cost; the more resources that are available the more likely it is that they can be dispatched in an efficient manner, thus the operation cost should be reduced.

Concerning the risk of power deficit it is also apparent that there is a correlation between load and power deficit, because power deficit will primary occur during peak load periods. There is of course also a correlation between power deficit and failures in generation or transmission resources. But since *LOLO* is a binary variable with a very asymmetrical distribution—in the vast majority of the scenarios there will not be any loss of load—we cannot expect that complementary random numbers will provide much benefits (cf. the discussion of the benefits of complementary random numbers in section 8.2.1.).

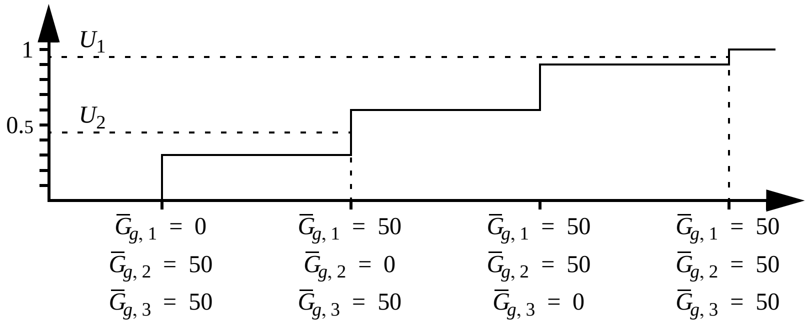
There are no problems with combining complementary random numbers and stratified sampling, but the choice of stratification strategy has a certain impact on the use of complementary random numbers. In a complete stratification there is only one scenario parameter which can vary within each stratum, namely the load. In a simulation using the complete stratification there will thus only be one complementary scenario. If the reduced stratification is used instead then there will in practice be two scenario parameters: load and "strata tree branch" (each branch represents a certain generation capacity and—where appropriate—transmission capability, point of time, or whatever scenario parameters included in the strata tree). Let us for the sake of simplicity assume that each branch corresponds to a unique state of the available generation capacity $\overline{G}_{tot}$. The possible levels of the load depend on the value of $\overline{G}_{tot}$, which therefore must be randomised first. Given $\overline{G}_{tot}$ it is now possible to determine the interval to which D_{tot} should belong. Now it is possible to randomise a first value of the total load, D^i_{tot}, and calculate its complementary random number, $D^i_{tot}*$. Then we study the complementary random number of the available generation capacity, $\overline{G}^*_{tot}$, which in many cases will take us to another branch of the strata tree. If that is the case then we will have another possible interval of D_{tot}; thus, D^i_{tot} and $D^i_{tot}*$ cannot be used. We must therefore randomise a second value of the total load and calculate its complementary random number. In total, we get four complementary scenarios, but we have used three original random numbers, as shown in table 9.5. If stratified sampling had not been used then two original random numbers would have been sufficient, because $\overline{G}_{tot}$ and D_{tot} would then have been randomised independent from each other.

Table 9.5 *Complementary scenarios with and without stratified sampling.*

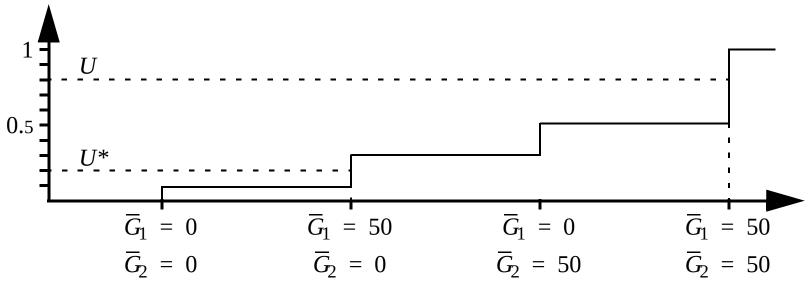
No stratified sampling	Reduced stratification
$\overline{G}_{tot}, D_{tot}$	$\overline{G}_{tot}, D^i_{tot}$
$\overline{G}_{tot}, D^*_{tot}$	$\overline{G}_{tot}, D^i_{tot}*$
$\overline{G}^*_{tot}, D_{tot}$	$\overline{G}^*_{tot}, D^{ii}_{tot}$
$\overline{G}^*_{tot}, D^*_{tot}$	$\overline{G}^*_{tot}, D^{ii}_{tot}*$

Dagger Sampling

This variance reduction technique is appropriate to guarantee that events which have low probability will appear to a sufficient extent in the selected samples. Examples of such events with low probabilities could be failures in power plants or transmission lines. I have however not considered dagger sampling in simulation of electricity markets, because there are other variance reduction techniques which are better suited, but not easily combined with dagger sampling. If we look at the literature in this field, it is hard to find any examples of dagger sampling applied to power system analysis, which to a certain degree may be taken as a confirmation of my opinion.



a) Scenarios randomised by dagger sampling. In each of the two power plants one random number is used to generate three samples of the available generation capacity. The first random number, U_1, generates three values of the available capacity in power plant 1, $\overline{G}_1$, and U_2 generates three values of $\overline{G}_2$. All in all, the two random numbers yield three scenarios, where $\overline{G}_1 = 50$ and $\overline{G}_2 = 50$ in the first scenario, $\overline{G}_1 = 50$ and $\overline{G}_2 = 0$ in the second scenario, and finally $\overline{G}_1 = 50$ and $\overline{G}_2 = 50$ in the third scenario.



b) Scenarios randomised by complementary random numbers. Here, one random number generates a scenario where $\overline{G}_1 = 50$ and $\overline{G}_2 = 50$. By studying the complementary random number U, another scenario is obtained in which $\overline{G}_1 = 50$ and $\overline{G}_2 = 0$.*

Figure 9.11 *Randomizing available generation capacity. In this example there are two power plants which both have an installed capacity of* 50 MW *and are available* 70% *of the time. Dagger sampling is applied in panel a, and the state is randomised separately for each power plant. In panel b complementary random numbers are used instead; the state of both power plants is now randomised simultaneously.*

Complementary random numbers do not work well for asymmetrical distributions (which could speak in favour of dagger sampling), but if a common probability distribution of the available generation capacity is used, then it is far less asymmetrical than the distribution of the available generation capacity in a single power plant. Thus, the variance reduction of complementary random numbers is improved. The difference between randomizing the state of each power plant separately and randomizing the total generation capacity is shown in figure 9.11.

A disadvantage of dagger sampling is that it is difficult to combine with stratified sampling. If the complete stratification is used then each stratum corresponds to a unique state of the generation capacity and then there is no need to apply neither dagger sampling nor complementary random numbers. In the reduced stratification there are more than one state of the generation capacity in each stratum, but it is not guaranteed that all possible states are allowed in a certain stratum, because some branches of the strata tree might have been excluded due to overlapping or low probability.[23] It might be possible to modify the dagger sampling method in order to exclude those states that are not allowed in a stratum, but it seems much simpler to determine a probability distribution of the total generation capacity for each stratum and then apply complementary random numbers.

Another contradiction between stratified sampling and dagger sampling is that when applying the Neyman allocation for stratified sampling it will most likely force us to truncate the dagger sampling cycles. I cannot straight off determine how this will affect the efficiency of dagger sampling, but it is hardly beneficial.

Thus, my conclusion is that dagger sampling is not especially appropriate for simulation of electricity markets, as the method is difficult to combine with stratified sampling and similar gains can be achieved by applying complementary random numbers in a proper way. Possibly, dagger sampling could be useful in the presimulation (where stratified sampling cannot be applied) or for such scenario parameters which are not included in the strata tree (for example secondary scenario parameters).

Control Variates

To make it possible to use this method, we must find some appropriate control variate, the expectation value of which can be calculated analytically. Concerning simulation of electricity markets, there is an analytical simulation method: probabilistic production cost simulation (PPC).[24] This simulation method uses a model which completely disregards the transmission system. Moreover, it is assumed that the load is not price sensitive[25] and that each

23. Cf. section 9.2.2.

power plant has a variable generation cost which is directly proportional to the power generation. The power plants are dispatched in strict variable cost order. Finally, it is assumed that load and failures in power plants are independent random variables. In some cases it is however possible to manage correlations between available generation capacity and load by studying different time periods separately, in a similar manner as correlations between scenario parameters are managed in a strata tree.[26]

Thus, the model can be described as an ideal electricity market with one area and price insensitive load. The scenario problem of such a model is simply to cover the load, while minimizing the operation costs:[27]

$$\text{minimise} \quad \sum_{g \in \mathcal{G}} \beta_{Gg} G_g + \beta_U U \tag{9.4}$$

$$\text{subject to} \quad \sum_{g \in \mathcal{G}} G_g - U = D_{tot} \tag{9.4a}$$

$$0 \le G_g \le \overline{G}_g, \qquad \forall\, g \in \mathcal{G}, \tag{9.4b}$$

$$0 \le U. \tag{9.4c}$$

From the solution to (9.4) we get

$$TOC = \sum_{g \in \mathcal{G}} \beta_{Gg} G_g, \tag{9.5a}$$

$$LOLO = \begin{cases} 0 & \text{if } U = 0, \\ 1 & \text{if } U > 0. \end{cases} \tag{9.5b}$$

The result variables of the PPC model thus only include G_g, U, TOC and $LOLO$. The PPC model can only generate control variates for these result variables! If we want to estimate some other system index than EG, $EENS$, $ETOC$ or $LOLP$ then control variates cannot be used.

If the control variate method should produce good results then it is absolutely necessary that the analytical expectation value is calculated with sufficient accuracy. Although PPC is an analytical method the calculations involve integrals complex enough to force us to use numerical methods. It is above all the load that is approximated by a discrete variable, in spite of being

24. This method was first presented in [45] and [48] respectively, and has later been further developed by several authors. The method is described in detail in most text books on power system planning, e.g. [11, 31, 32, 34].
25. Although it is possible to avoid this assumption by adding power plants corresponding to load reductions (cf. section 3.2.1).
26. See section 9.2.2.
27. Compare to the general electricity market model described in section 3.2.2.

a continuous variable in reality. In this approximation a certain step length is used; the shorter the step length the more accurate will the approximation become, but on the other hand will the numerical integration become more time consuming. Hence, it is necessary to make a trade-off between accuracy and which computation time is acceptable. In practice, it is however not a major problem that the PPC simulation may require a few minutes. It should also be remembered that it is not certain that a change in the simulation input affects the PPC model. As seen above there are very few scenario parameters and model constants included in the PPC model and it is only when any of these is modified that a new PPC simulation has to be performed.

Correlated Sampling

The problem I described in section 1.1, i.e., to study how a change in an electricity market affects system indices, sounds cut out for applying correlated sampling; we have an original system and a slightly changed version of the same system. The difference in output of the two system is almost certainly small; hence, the result variables from the two systems will be positively correlated. For example, if an investment has been made in increased transmission capability then the operation cost in the two systems will behave in a similar manner; there will be higher operation costs in peak load periods than in low load periods, but the difference will be slightly lesser in the reinforced system.

In spite of these good conditions I have not used correlated sampling for simulation of electricity markets and this for two reasons. My first objection is that although it might be interesting to estimate the difference of two systems, it is probably also desirable to learn about the absolute values of the system indices in both systems, but unfortunately correlated sampling does not provide any variance reduction to individual system indices.

The other objection is that correlated sampling is hard to combine with stratified sampling (and as already mentioned, I consider stratified sampling to be a very efficient method for electricity market simulation). When simulating a single system the strata tree is based upon comparing the system resources to the demand, which may be done in a systematic way. If correlated sampling is applied there are however two different systems (or even more) each having its own set of resources and demand. In these cases the objective is to identify the scenarios where the system behaves differently. This requires that all scenario parameters have the same probability distribution in both systems, as it otherwise would be impossible to calculate the node weights of the strata tree.[28] If that would be the case then it may actually sometimes be possible to identify partitions where it can be expected that the system behaves differently. An example of this is given in figure 9.12. It might be possible to formulate rules specifying the kinds of differences that

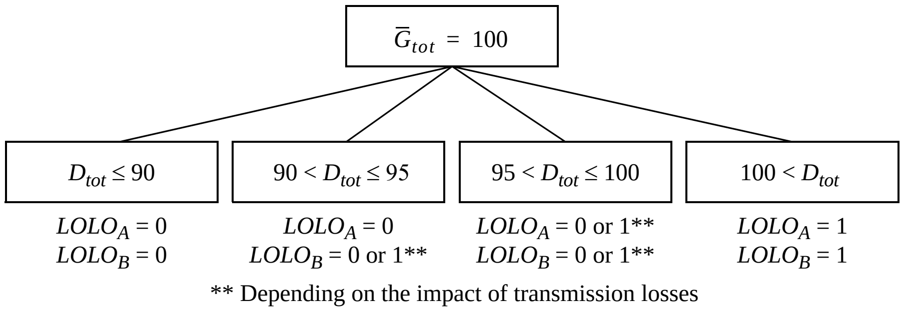


Figure 9.12 *Strata tree for correlated sampling. The difference between system A and B respectively is that the losses are larger in system B. Assume that* $\bar{L} = 5$ *in system A and* $\bar{L} = 10$ *in system B. Concerning TOC the systems should differ in all scenarios in the first three branches. Concerning LOLO it is only in the two middle branches where the systems may differ.*

can be managed by a strata tree and how the strata should be adjusted, but I have not yet done such an analysis.

Importance Sampling

The choice of importance sampling function is of course extremely important when applying importance sampling to simulation of electricity markets. I have only made a few simple experiments with this variance reduction technique and therefore cannot say anything about the design of importance sampling functions. The explanation to my lack of interest in importance sampling is that I find stratified sampling much easier to apply efficiently. The two steps stratification and sample allocation are simple to solve when simulating electricity markets (the stratification by using a strata tree and the samples are distributed according to the Neyman allocation). I find it hard to see that even larger efficiency gains could be achieved by importance sampling and it then seems unjustified to put any effort into finding good importance sampling functions.

However, importance sampling could be useful in the presimulation, when stratified sampling cannot be utilised. There may also occur situations where it could be a good idea to combine stratified sampling and importance sampling. As described earlier it is only the total load which is included in the strata tree, whereas the geographical distribution of the load is treated as a

28. How should for example the node weight be defined in a node where $D_{tot} \leq 1\ 000$ if the load is $N(1\ 000, 100)$-distributed in one system and $N(2\ 000, 200)$-distributed in the other?

secondary scenario parameter. But the more extreme the geographical distribution is the larger is the risk for transmission congestion. These extreme distributions are important, but comparatively rare. Importance sampling could be used to decide how the total load should be divided between the areas. Exactly how to do this, i.e., which importance sampling function to use, is something I have not yet studied. Further research is necessary in this field.

9.3 SOME SIMPLE TEST SYSTEMS

To evaluate how efficient different variance reduction techniques are, I have performed a number of simulations of several test systems. In this section I intend to present some illustrative examples.

Before I describe the test systems more closely I must however briefly describe the software I have used in my test runs, because it has to some extent been limiting the choice of test systems and evaluation criteria. Most of the calculations have been performed in Matlab, as the programming is swift and simple—most mathematical functions are already available. As soon as pure syntax errors have been removed from a Matlab program it is also very reliable. There are however also some disadvantages with Matlab; primarily that the calculations may become slow and memory consuming. To simulate large multi-area systems it is necessary to use the NNP algorithm (or some other algorithm for solving non-linear optimisation problems) and implementing the NNP algorithm in Matlab would result in a too slow and inefficient program. When I wrote the final version of the NNP algorithm I therefore chose to use the programming language C. This causes some problems when performing a Monte Carlo simulation, because most of the software running the statistical analysis is written in Matlab and has to communicate with an electricity market model written in C. This communication could of course be automatised, but to avoid spending too much time on programming, I have chosen to transfer data manually. The nuisance of this solution is that a simulation requires human supervision, which makes it too time consuming to perform many long simulations.

If two-area electricity markets are studied instead, things become much simpler. In such a small system it is easy to find the rules describing in which order the power plants will be dispatched considering transmission losses and limitations (sort of a NNP “light” algorithm). By that means the whole simulation can be performed in Matlab, thus making it possible to start a program which runs twenty simulations of one million scenarios each, take a week’s holiday and get back just in time when the results are available.

Hence, for practical reasons I have mostly stuck to two-area systems when investigating the efficiency of simulation of short scenarios. This does however not at all mean that the conclusions only should be applicable to smaller systems, because the derivation of the strata tree method is in no way based on assuming that a two-area system is simulated.

Apart from simplifying the data processing, studies of two-area systems have yet another advantage, more precisely that an exact value of *LOLP* can be calculated with a reasonable work effort (see appendix D). This is of course valuable if the same system is simulated with two strategies and the resulting *LOLP*-estimates are very different.

9.3.1 Two-area Test Systems

To demonstrate the size of the possible efficiency gains which can be achieved by using variance reduction techniques, I have designed a small two-area system. This system I have then varied in different ways. Each variant has been simulated a large number of times using different combinations of variance reduction techniques. In those cases when stratified sampling has been used I have included a presimulation of 100 scenarios to estimate the stratification parameter $\bar{L}$. Then follows the stratification and a pilot study. The sample allocation of the pilot study is stated in the result table of each test system. I have chosen to just identify scenarios of type I-VII (cf. section 9.2.1) and disregard the more unusual scenario types, as they are so rare that they have no significant impact on the final results.

The accuracy of a Monte Carlo simulation depends, as shown in chapter 8, both on the number of samples and usage of any variance reduction techniques. To facilitate the comparison of different variance reduction techniques I have eliminated the number of samples by seeing to that the number of scenarios in presimulation, pilot study and main simulation amount to about 10 000 scenarios in each simulation.

The simulation results are compiled in separate result tables for each variant of the system. As we are dealing with compiling the results of hundreds of different simulations, it is not strange if these tables seem somewhat confusing. The principle is however simple: each combination of variance reduction techniques has been simulated in 20 separate runs using different seeds for the random number generator. These 20 runs yield a series of separate estimates of *ETOC* ($etoc_1, \ldots, etoc_{20}$) as well as a series of separate estimates of *LOLP* ($lolp_1, \ldots, lolp_{20}$). The result of these 20 runs are described in a row of its own in the result table. The values stated in the table are

$$\text{lowest} = \min_i etoc_i,$$

$$\text{mean} = \frac{1}{20}\sum_{i=1}^{20} etoc_i,$$

$$\text{highest} = \max_i \ etoc_i,$$

and

$$\text{variance} = \frac{1}{19}\sum_{i=1}^{20}(etoc_i - \text{mean})^2.$$

Corresponding definitions are applied also to the *LOLP* estimates. Notice that the variance stated in the result tables thus does no relate to the variance calculations which are part of a single simulation (for example when calculating confidence intervals or sample allocation according to Neyman), but is just a measure of how well concentrated the results of the 20 runs are.

Comparing the results of two different combinations of variance reduction techniques takes some consideration. Most important is that the mean of the 20 estimates is close to the theoretical value. (Concerning *ETOC* I have not calculated any theoretical value for comparison, but the results are on the other hand quite unanimous, so that should not be a problem.) If a method should be considered good then the lowest and highest estimates should not differ too much from the theoretical value and the variance should be low.

As shown in other parts of this dissertation, cardinal error is a problem which must be accounted for and a simulation method cannot be considered very efficient if it is prone to be subject to cardinal error. In this sort of simulation, cardinal error almost exclusively causes the final estimate of *ETOC* or *LOLP* to be less than the true value. If the lowest estimate of the 20 trials is significantly lesser than the mean of all 20 runs, or if it even is equal to the result of a PPC simulation, then this is an indication that cardinal error has occurred.[29] If cardinal errors are frequent it is usually also reflected in the

Table 9.6 *Data of the base case.*

	Area 1	Area 2
Thermal power plants		
Installed capacity [MW|	2 × 30	5 × 6
Operation cost [¤/MWh]	50	100
Availability [%]	99	95
Transmission line		
Transmission capability [MW]	Unlimited	
Loss function	$L = 2{\cdot}10^{-3}P^2$	
Availability [%]	100	
Load (not price sensitive, no correlations)		
Mean	20	30
Standard deviation	3	4

mean, which becomes noticeably lesser than the theoretical value.

The Base Case

In the base case there are only thermal power plants and the transmission line between the areas has unlimited capacity. Detailed data about the base case are given in table 9.6 and the simulation results are compiled in table 9.7.

The results clearly show that complementary random numbers have a favourable effect on the *ETOC* estimate, but this effect is lesser when complementary random numbers are combined with a control variate for *TOC*. Concerning the *LOLP* estimate, the complementary random numbers do not have any larger consequences, neither positive nor negative ones; the differences that appear are likely to be caused by random fluctuations as a result of just comparing 20 estimates. These results are thus in accordance to the analysis made in section 9.2.4.

The control variate method has different impacts on *ETOC* and *LOLP* respectively. The former system index is significantly improved when a control variate is used, regardless of which other variance reduction techniques are used. This is not the case concerning a control variate for *LOLO*, which only provides improved accuracy when stratified sampling is not used. The explanation of this circumstance is simple. When calculating *LOLO* in a PPC model it is only possible to identify loss of load due to insufficient generation capacity. However, these scenarios are also simple to identify using a strata tree (cf. figure 9.2) and therefore a control variate for *LOLO* will not supply any new information when combined with stratified sampling. Yet, there is no distinguishable disadvantage if the control variate method nevertheless is applied to *LOLO*.

Concerning stratified sampling, we see that the *LOLP* estimates are significantly improved when this method is applied. However, at the same time the accuracy of the *ETOC* estimates are slightly decreased, but this negative effect is so small that it should not discourage the usage of stratified sampling—the relative error is after all 0.5% in all runs! The reason why stratified sampling only improves the *LOLP* estimate is that the strata tree method is most appropriate for defining strata in a duogeneous population. In the base case *TOC* is however a continuous, heterogeneous random variable and such are most appropriate to stratify using a classic stratification strategy as the "cum $\sqrt{f(y)}$-rule" and its relatives.[30] It could very well be possible to combine

29. Since the PPC model neglects the losses, we know that the Monte Carlo simulation should result in a higher operation cost and larger risk of power deficit. If this is not the case then it is likely that the Monte Carlo simulation missed that part of the scenario population where the transmission losses are really important.
30. See [127], section 5A.7, or [55].

Table 9.7 *Result of simulating the base case.*

Control variate *TOC*	Control variate *LOLO*	Complementary random numbers	Estimate of *ETOC* [¤/h]				Estimate of *LOLP* [%]			
			Lowest	Mean	Highest	Variance	Lowest	Mean	Highest	Variance
No stratified sampling										
			2623	2629	2635	8.4	0.120	0.172	0.240	$1.3\cdot10^{-7}$
		✓	2627	2629	2632	3.2	0.090	0.160	0.270	$2.1\cdot10^{-7}$
	✓		2623	2629	2635	8.4	0.160	0.168	0.180	$6.2\cdot10^{-9}$
	✓	✓	2627	2629	2632	3.2	0.160	0.165	0.180	$4.7\cdot10^{-7}$
✓			2628	2629	2630	0.1	0.120	0.172	0.240	$1.3\cdot10^{-7}$
✓		✓	2628	2629	2630	0.1	0.090	0.160	0.270	$2.1\cdot10^{-7}$
✓	✓		2628	2629	2630	0.1	0.160	0.168	0.180	$6.2\cdot10^{-9}$
✓	✓	✓	2628	2629	2630	0.1	0.160	0.165	0.180	$4.7\cdot10^{-7}$
Complete stratification (Sample allocation in the pilot study: IV - 2. VI - 32. VII - 0)										
			2622	2630	2641	21.1	0.161	0.163	0.166	$3.8\cdot10^{-10}$
		✓	2628	2629	2631	0.6	0.161	0.164	0.167	$3.6\cdot10^{-10}$
	✓		2617	2629	2637	21.0	0.161	0.163	0.165	$3.3\cdot10^{-10}$
	✓	✓	2628	2629	2632	0.6	0.161	0.164	0.166	$3.6\cdot10^{-10}$
✓			2628	2629	2630	0.3	0.161	0.163	0.166	$4.0\cdot10^{-10}$
✓		✓	2628	2629	2630	0.2	0.161	0.164	0.166	$3.4\cdot10^{-10}$
✓	✓		2628	2629	2630	0.3	0.161	0.163	0.165	$3.4\cdot10^{-10}$
✓	✓	✓	2628	2629	2630	0.2	0.161	0.163	0.166	$3.5\cdot10^{-10}$
Reduced stratification (Sample allocation in the pilot study: IV - 64. VI - 512. VII - 16)										
			2617	2628	2639	24.4	0.164	0.166	0.168	$7.5\cdot10^{-11}$
		✓	2626	2629	2631	2.7	0.165	0.166	0.168	$4.4\cdot10^{-11}$
	✓		2617	2628	2639	24.4	0.164	0.166	0.168	$7.5\cdot10^{-11}$
	✓	✓	2626	2629	2631	2.7	0.165	0.166	0.168	$4.7\cdot10^{-11}$
✓			2627	2629	2630	0.3	0.165	0.166	0.167	$3.5\cdot10^{-11}$
✓		✓	2628	2629	2629	0.2	0.164	0.166	0.168	$6.4\cdot10^{-11}$
✓	✓		2627	2629	2630	0.3	0.165	0.166	0.167	$3.2\cdot10^{-11}$
✓	✓	✓	2628	2629	2629	0.2	0.164	0.166	0.167	$6.1\cdot10^{-11}$
Facts about the base case			Theoretical value of *LOLP*: 0.166%. *LOLP* according to PPC model: 0.160%. *ETOC* according to PPC model: 2521 ¤/h. Number of possible states for the available generation capacity: 18.							

Table 9.8 *Result of simulating the hydro power system.*

Control variate *TOC*	Control variate *LOLO*	Complementary random numbers	Estimate of *ETOC* [¤/h] Lowest	Mean	Highest	Variance	Estimate of *LOLP* [%] Lowest	Mean	Highest	Variance
No stratified sampling										
			49.8	55.3	64.0	11.15	0.110	0.162	0.230	$1.0\cdot10^{-7}$
		✓	38.7	54.4	62.9	31.97	0.050	0.159	0.220	$1.8\cdot10^{-7}$
	✓		49.8	55.3	64.0	11.15	0.160	0.168	0.180	$6.2\cdot10^{-9}$
	✓	✓	38.7	54.4	62.9	31.97	0.160	0.164	0.180	$3.6\cdot10^{-9}$
✓			54.6	55.3	56.0	0.15	0.110	0.162	0.230	$1.0\cdot10^{-7}$
✓		✓	53.8	55.7	57.2	0.93	0.050	0.159	0.220	$1.8\cdot10^{-7}$
✓	✓		54.6	55.3	56.0	0.15	0.160	0.168	0.180	$6.2\cdot10^{-9}$
✓	✓	✓	53.8	55.7	57.2	0.93	0.160	0.164	0.180	$3.6\cdot10^{-9}$
Complete stratification (Sample allocation in the pilot study: I - 2, III - 32, IV - 2, VI - 32, VII - 0)										
			54.5	55.3	56.3	0.17	0.161	0.165	0.167	$4.5\cdot10^{-10}$
		✓	54.9	55.4	55.8	0.05	0.161	0.164	0.167	$2.7\cdot10^{-10}$
	✓		54.5	55.3	56.3	0.17	0.162	0.165	0.167	$2.5\cdot10^{-10}$
	✓	✓	54.9	55.4	55.8	0.05	0.161	0.164	0.167	$3.3\cdot10^{-10}$
✓			55.1	55.3	55.6	0.02	0.160	0.164	0.165	$3.0\cdot10^{-10}$
✓		✓	55.1	55.4	55.6	0.02	0.160	0.163	0.166	$4.4\cdot10^{-10}$
✓	✓		55.1	55.3	55.5	0.02	0.160	0.163	0.165	$3.3\cdot10^{-10}$
✓	✓	✓	55.0	55.3	55.6	0.02	0.160	0.163	0.166	$4.3\cdot10^{-10}$
Reduced stratification (Sample allocation in the pilot study: I - 64, III - 512, IV - 64, VI - 512, VII - 16)										
			53.7	55.1	55.8	0.37	0.165	0.166	0.167	$1.8\cdot10^{-11}$
		✓	54.8	55.4	56.2	0.13	0.165	0.166	0.168	$3.7\cdot10^{-11}$
	✓		53.7	55.1	55.8	0.37	0.165	0.166	0.167	$2.1\cdot10^{-11}$
	✓	✓	54.8	55.4	56.2	0.13	0.165	0.166	0.168	$3.5\cdot10^{-11}$
✓			55.1	55.4	55.7	0.03	0.165	0.166	0.168	$5.1\cdot10^{-11}$
✓		✓	55.0	55.3	55.7	0.03	0.164	0.166	0.168	$8.0\cdot10^{-11}$
✓	✓		55.1	55.4	55.7	0.03	0.164	0.166	0.167	$5.1\cdot10^{-11}$
✓	✓	✓	55.0	55.3	55.7	0.03	0.164	0.166	0.168	$8.2\cdot10^{-11}$
Facts about the hydro power system			Theoretical value of *LOLP*: 0.166%. *LOLP* according to PPC model: 0.160%. *ETOC* according to PPC model: 43.6 ¤/h. Number of possible states for the available generation capacity: 18.							

the two stratification methods, but this not anything I have studied any further.

Moreover, it can be seen that the reduced stratification is more efficient than the complete. The difference is due to the fact that the complete stratification in total uses more scenarios in the pilot study[31] and as these scenarios are not distributed according to the Neyman allocation, some efficiency will be lost.

Hydro Power without Reservoirs

This system is identical to the base case, except that the largest thermal power plants (i.e., the two units in area 1) have been replaced by two just as large hydro power plants with no reservoirs, so-called run-of-the-river units. For the sake of simplicity, it is assumed that the water flow passing by these hydro power plants always is sufficiently large to allow the technically available generation capacity to be fully utilised. The only difference compared to the base case is thus that the operation cost is going to be far less, because the variable operation cost of hydro power is assumed to be negligible; we get $TOC = 0$ whenever the load can be covered by hydro power only. The results of simulating the hydro power system are displayed in table 9.8.

The consequence of using complementary random numbers is not as clear here as in the base case. It appears that complementary random numbers have clear negative impact on the *ETOC* estimates when stratified sampling is not used, whereas the impact on the *LOLP* estimates still is neutral. The explanation should be that a majority of the scenarios result in $TOC = 0$; thus, there will be a weaker correlation between the complementary scenarios. In combination with stratified sampling the pattern is however similar to that of the base case (although slightly less apparent), i.e., complementary random numbers have a positive effect, but it is less important when a control variate for *TOC* is used. As it is recommended to use stratified sampling, there is no reason to advise against the usage of complementary random numbers for this kind of systems.

The control variates works almost exactly the same in the hydro power system as in the base case. Also stratified sampling produces more or less the same effects as in the base case, except that now stratified sampling also provides a clear improvement of the *ETOC* estimates. The reason for this is that in the hydro power system *TOC* is divided in a homogeneous part (those scenarios where $TOC = 0$) and a diverging part (where $TOC > 0$), which means that the advantages of the strata tree method can be fully utilised.

31. Admittedly there are less scenarios per stratum in the complete stratification, but this does not compensate the large number of strata.

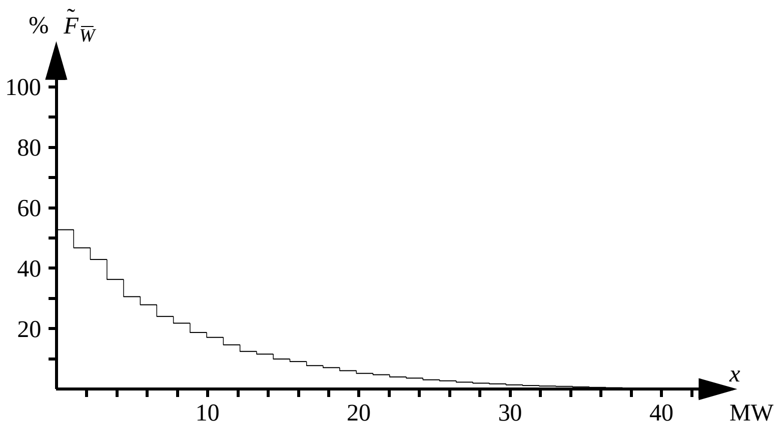


Figure 9.13 *Duration curve of the available generation capacity of the wind farm in the wind power test system.*

Wind Power

All the power plants of the base case are present in this system, but it has also been added a wind farm. The duration curve of the available wind power generation capacity is shown in figure 9.13.[32] The available wind power generation capacity, $\overline{W}$, is actually a continuous random variable, but it does not follow any standard probability distribution. Therefore a discrete approximation of the continuous distribution has been used, so that it will be simpler to generate random numbers of $\overline{W}$. Besides, a discrete approximation is necessary when building the strata tree.

The major difference between this system and the former is that there are considerably more possible states of the available generation capacity of the system, which results in much larger strata trees. Another difference is that the installed capacity of the system is larger, which causes *LOLP* to decrease compared to the base case and the hydro power system. The diverging scenarios (that is scenarios where $LOLO = 1$) are more rare, which makes proper choice of strata even more important. Finally, this system can be seen as a mixture of the base case (which was completely dominated by power plants with non-negligible operation costs) and the hydro power system (which was dominated by power plants having negligible operation costs); here there is a certain probability that the wind power can cover the load, but in most cases

32. It can be worth noting that the model corresponds to a wind farm with comparatively poor wind speed conditions.

Table 9.9 *Results of simulating the wind power system.*

Control variate *TOC*	Control variate *LOLO*	Complementary random numbers	Estimate of *ETOC* [¤/h]				Estimate of *LOLP* [%]			
			Lowest	Mean	Highest	Variance	Lowest	Mean	Highest	Variance
No stratified sampling										
			2396	2407	2419	29.5	0.040	0.103	0.170	$1.1\cdot10^{-7}$
		✓	2395	2406	2418	33.2	0.050	0.105	0.240	$2.1\cdot10^{-7}$
	✓		2396	2407	2419	29.5	0.099	0.105	0.119	$3.6\cdot10^{-9}$
	✓	✓	2395	2406	2418	33.2	0.099	0.102	0.109	$2.0\cdot10^{-9}$
✓			2406	2406	2407	0.1	0.040	0.103	0.170	$1.1\cdot10^{-7}$
✓		✓	2406	2407	2407	0.1	0.050	0.105	0.240	$2.1\cdot10^{-7}$
✓	✓		2406	2406	2407	0.1	0.099	0.105	0.119	$3.6\cdot10^{-9}$
✓	✓	✓	2406	2407	2407	0.1	0.099	0.102	0.109	$2.0\cdot10^{-9}$
Multiple state nodes										
			2344	2386	2408	496.5	0.080	0.092	0.107	$1.1\cdot10^{-8}$
		✓	2344	2388	2408	530.7	0.081	0.091	0.106	$1.2\cdot10^{-8}$
	✓		2334	2386	2412	715.2	0.099	0.101	0.103	$1.3\cdot10^{-10}$
	✓	✓	2345	2389	2410	631.9	0.099	0.100	0.102	$8.0\cdot10^{-11}$
✓			2403	2406	2408	1.5	0.080	0.093	0.106	$1.0\cdot10^{-8}$
✓		✓	2404	2406	2408	0.9	0.081	0.090	0.119	$1.2\cdot10^{-8}$
✓	✓		2403	2406	2408	1.1	0.099	0.101	0.102	$1.2\cdot10^{-10}$
✓	✓	✓	2404	2406	2407	1.0	0.099	0.100	0.101	$7.3\cdot10^{-11}$
Reduced stratification (Sample allocation in the pilot study: I - 64, III - 512, IV - 64, VI - 512, VII - 16)										
			2395	2408	2419	31.3	0.102	0.103	0.104	$3.2\cdot10^{-11}$
		✓	2399	2405	2416	25.2	0.103	0.104	0.105	$1.9\cdot10^{-11}$
	✓		2395	2408	2419	31.3	0.102	0.103	0.104	$3.0\cdot10^{-11}$
	✓	✓	2399	2405	2416	25.2	0.103	0.103	0.105	$1.7\cdot10^{-11}$
✓			2405	2406	2408	0.2	0.103	0.104	0.105	$1.9\cdot10^{-11}$
✓		✓	2406	2406	2407	0.1	0.102	0.104	0.105	$2.3\cdot10^{-11}$
✓	✓		2405	2406	2408	0.2	0.103	0.104	0.105	$1.8\cdot10^{-11}$
✓	✓	✓	2406	2406	2407	0.1	0.102	0.103	0.105	$2.3\cdot10^{-11}$
Facts about the wind power system			Theoretical value of *LOLP*: 0.104%. *LOLP* according to PPC model: 0.099%. *ETOC* according to PPC model: 2300 ¤/h. Number of possible states for the available generation capacity: 666.							

thermal generation will be required too.

The simulation results are compiled in table 9.9. No results of the complete stratification are recorded, as this stratification strategy would require more than 33 000 scenarios just in the pilot study; apparently, the method is inefficient compared to reduced stratification. The complete stratification has been replaced by stratifications based on the idea of multiple state nodes. Stratified sampling using this method is more efficient than not using stratified sampling at all, but the results clearly shows that the multiple state nodes are not as efficient as the reduced stratification and there is an imminent risk of cardinal errors.

Besides that, the conclusions of the wind power system are the same as for the hydro power system.

Transmission Limitations

To study the impact of transmission limitations I have used a variant of the base case, where I have restricted the transmission capability between the areas to 25 MW. As I have not yet solved the problem of estimating the stratification parameters $\overline{U}_W$ and $\overline{U}_{WG}$ I have taken myself the liberty of cheating a little bit, and calculate them theoretically, which is a reasonable task in a two-area system. The reason for this trickery is to indicate the efficiency gains which *could* be achieved if the problem of determining the stratification parameters was solved for systems having more than two areas. Therefore, I present three variants of the reduced stratification strategy. In the first I have ignored the unknown stratification parameters and merged scenarios of type IV and V into one single stratum instead. In the second variant I use the theoretically calculated values of $\overline{U}_{WG}$. The theoretical values of $\overline{U}_{WG}$ are based upon the worst case when the whole total load is located in one of the areas, while the load in the other area is zero. This scenario is of course *extremely* unlikely; hence, it might be worthwhile to modify $\overline{U}_{WG}$, by for example assuming that the least possible load in any one area amounts to the mean load minus three standard deviations. This method of calculation has been used in the third variant.

The results of the system with 25 MW transmission capacity are shown in table 9.10. If we start by analysing *ETOC*, we find that regardless of which variance reduction technique we use there is no problem to detect the cost increase due to the transmission limitations; the *ETOC* estimates are roughly 2 890 ¤/h, compared to about 2 630 ¤/h in the base case.

Concerning the *LOLP* estimates the accuracy—as usual—primarily depends on which stratification strategy is used. Apparently those stratifications which do not differentiate scenarios of types IV and V respectively do not convey any larger efficiency gain compared to giving up stratified sampling altogether. It is neither an improvement to use the exact theoretical values of

Table 9.10 *Results of simulating the system with transmission limitation.*

Control variate *TOC*	Control variate *LOLO*	Complementary random numbers	Estimate of *ETOC* [¤/h] Lowest	Mean	Highest	Variance	Estimate of *LOLP* [%] Lowest	Mean	Highest	Variance
No stratified sampling										
			2882	2890	2898	18.2	0.130	0.189	0.260	$1.4\cdot10^{-7}$
		✓	2887	2890	2892	1.8	0.130	0.176	0.280	$2.1\cdot10^{-7}$
	✓		2882	2890	2898	18.2	0.160	0.184	0.210	$1.5\cdot10^{-8}$
	✓	✓	2887	2890	2892	1.8	0.160	0.180	0.210	$2.5\cdot10^{-8}$
✓			2885	2889	2894	4.1	0.130	0.189	0.260	$1.4\cdot10^{-7}$
✓		✓	2887	2889	2891	1.7	0.130	0.176	0.280	$2.1\cdot10^{-7}$
✓	✓		2885	2889	2894	4.1	0.160	0.184	0.210	$1.5\cdot10^{-8}$
✓	✓	✓	2887	2889	2891	1.7	0.160	0.180	0.210	$2.5\cdot10^{-8}$
Reduced stratification not separating strata of types IV and V (Sample allocation in the pilot study: IV/V - 512, VI - 512, VII - 16)										
			2880	2892	2902	32.2	0.166	0.183	0.207	$2.0\cdot10^{-8}$
		✓	2886	2890	2893	3.9	0.166	0.181	0.199	$1.1\cdot10^{-8}$
	✓		2879	2891	2899	35.1	0.166	0.180	0.207	$1.6\cdot10^{-8}$
	✓	✓	2884	2890	2893	4.9	0.166	0.180	0.207	$1.6\cdot10^{-8}$
✓			2883	2888	2891	4.6	0.166	0.182	0.209	$2.1\cdot10^{-8}$
✓		✓	2885	2889	2892	2.3	0.166	0.183	0.231	$3.3\cdot10^{-8}$
✓	✓		2883	2888	2891	4.6	0.166	0.181	0.208	$2.1\cdot10^{-8}$
✓	✓	✓	2885	2889	2892	2.3	0.166	0.183	0.231	$3.3\cdot10^{-8}$
Facts about the system with 25 MW transmission capability			Theoretical value of *LOLP*: 0,180%. *LOLP* according to PPC model: 0,160%. *ETOC* according to PPC model: 2521 ¤/h. Number of possible states for the available generation capacity: 18.							

(continues next page)

$\bar{U}_{WG}$; the best results are obtained when using the slightly modified values of $\bar{U}_{WG}$, where the most unlikely load distributions have been neglected. This is in a way good news, because it implies that it may actually be an advantage that $\bar{U}_{WG}$ is estimated from a presimulation and therefore by practical reasons neglect the most unlikely distributions of the total load. This should mean that it is possible to keep the number of scenarios low in the presimulation, without compromising the quality of the stratification. However, further studies are necessary before any reliable conclusions can be drawn.

Table 9.11 (cont.) *Results of simulating the system with transmission limitation.*

Control variate *TOC*	Control variate *LOLO*	Complementary random numbers	Estimate of *ETOC* [¤/h]				Estimate of *LOLP* [%]			
			Lowest	Mean	Highest	Variance	Lowest	Mean	Highest	Variance
Reduced stratification based on theoretically calculated values of $\bar{U}_{WG}$ (Sample allocation in the pilot study: IV - 64, V - 512, VI - 512, VII - 16)										
			2880	2892	2904	30.3	0.166	0.183	0.227	$2.5 \cdot 10^{-8}$
		✓	2884	2891	2898	10.8	0.166	0.176	0.203	$1.7 \cdot 10^{-8}$
	✓		2880	2890	2904	29.6	0.166	0.178	0.197	$8.8 \cdot 10^{-9}$
	✓	✓	2884	2891	2898	12.6	0.165	0.176	0.202	$1.6 \cdot 10^{-8}$
✓			2885	2889	2894	4.8	0.166	0.179	0.200	$1.1 \cdot 10^{-8}$
✓		✓	2886	2890	2893	3.3	0.166	0.178	0.205	$1.9 \cdot 10^{-8}$
✓	✓		2885	2889	2894	5.4	0.166	0.180	0.199	$1.2 \cdot 10^{-8}$
✓	✓	✓	2886	2890	2893	4.4	0.165	0.175	0.203	$1.2 \cdot 10^{-8}$
Reduced stratification based on theoretically calculated (but slightly modified) values of $\bar{U}_{WG}$ (Sample allocation in the pilot study: IV - 64, V - 512, VI - 512, VII - 16)										
			2880	2888	2897	25.5	0.173	0.180	0.187	$1.7 \cdot 10^{-9}$
		✓	2886	2889	2892	2.0	0.172	0.180	0.191	$2.3 \cdot 10^{-9}$
	✓		2880	2888	2897	25.5	0.173	0.180	0.187	$1.7 \cdot 10^{-9}$
	✓	✓	2886	2890	2892	2.1	0.171	0.180	0.186	$1.7 \cdot 10^{-9}$
✓			2885	2889	2895	5.7	0.175	0.181	0.187	$1.6 \cdot 10^{-9}$
✓		✓	2886	2890	2893	2.6	0.175	0.181	0.202	$3.8 \cdot 10^{-9}$
✓	✓		2885	2889	2895	5.7	0.175	0.181	0.187	$1.5 \cdot 10^{-9}$
✓	✓	✓	2886	2890	2893	2.6	0.175	0.180	0.186	$1.4 \cdot 10^{-9}$
Facts about the system with 25 MW transmission capability			Theoretical value of *LOLP*: 0,180%. *LOLP* according to PPC model: 0,160%. *ETOC* according to PPC model: 2521 ¤/h. Number of possible states for the available generation capacity: 18.							

Conclusions

The most important conclusion of the above described test systems is of course that the three variance reduction techniques complementary random numbers, control variates and stratified sampling all produce efficiency gains when used for simulation of short scenarios. Moreover, it works excellent to combine the three methods. It can be noted that different methods are of different importance for different result variables. The estimate of *ETOC* is above all improved by using a control variate with respect to *TOC*. Complementary random numbers also primarily improve the *ETOC* estimate, but in combination with control variates, the impact of complementary random numbers is almost negligible.[33] I can however not see any major harm of using the method, and as it also saves a certain amount of random number generation efforts, I find no reason to refrain from using complementary random numbers even in those cases when the benefits of the method are hard to notice. Stratified sampling improves the *ETOC* estimates of systems, where there is an element of power plants having negligible operation costs.

The method which is most beneficial when it comes to improving the *LOLP* estimate is without doubt stratified sampling. A control variate with respect to *LOLO* will be good for the *LOLP* estimate, but this benefit is only noticeable when stratified sampling is not used (unless multiple state nodes are used, but that is as mentioned earlier not what we should do). In this case too, we may say that though it does not produce any gain, neither does it make any harm; as we still need to determine the control variate with respect to *TOC*, hardly no extra work is required to determine the control variate with respect to *LOLO* at the same time.

In table 9.12 a compilation of the results from simulating the test systems is shown. The table also includes some simulation runs, where a stopping rule based on relative tolerance (see section 8.1) has been used rather than determining the number of scenarios in advance. The results shows that the efficiency gain is considerable when all the tested variance reduction techniques are combined. Compared to simple sampling about 99% less scenarios are required to obtain an equivalent accuracy of the estimates. In the system with the transmission limitation, this efficiency gain was only obtained if the best possible stratification could be used—how to determine this stratification in an efficient manner remains to be solved.

Strangely enough, it does not seem as if the efficiency gain depends on the absolute values of *ETOC* and *LOLP* respectively, which otherwise could have been expected. In [47] it is shown that in simple sampling a reliability index (like *LOLP*) requires more sample for a given accuracy as the value of the index decreases. This conclusion does not seem to apply to electricity market simulation using variance reduction techniques.[34]

I have not shown to which extent the above described test system simula-

---

33. In some cases complementary random numbers in combination with a control variate for *TOC* produces a very small improvement, whereas in other cases there is a very small deterioration. This should be regarded as random deviations and indicate that the method more or less lacks impact on the final result.
34. Here I have to add a reservation, as it probably takes studies of more systems before any definitive conclusions can be drawn. Another alternative is that whoever is inclined examines the question by calculating $Var[m_{TOC}]$ and $Var[m_{LOLO}]$ analytically—I have tried and found that the calculations become horrible even for a system with just one area.

Table 9.12 *Overview of the test system results.*

Variance reduction	Estimate of *ETOC* [¤/h]				Estimate of *LOLP* [%]				Number of scenarios per simulation			ρ
	Lowest	Mean	Highest	Variance	Lowest	Mean	Highest	Variance	Lowest	Mean	Highest	
Base case												
	2628	2629	2629	0.1	0.159	0.166	0.172	$1.2 \cdot 10^{-9}$	1 000 000			
✓	2628	2629	2629	0.2	0.164	0.166	0.167	$6.1 \cdot 10^{-11}$	10 000			
✓	2628	2629	2630	0.3	0.164	0.166	0.167	$5.9 \cdot 10^{-11}$	2 822	3 688	5 948	0.2
Hydro power												
	54.6	55.2	56.0	0.13	0.159	0.166	0.172	$1.3 \cdot 10^{-9}$	1 000 000			
✓	55.0	55.3	55.7	0.03	0.164	0.166	0.168	$8.2 \cdot 10^{-11}$	10 000			
✓	54.4	55.3	56.0	0.10	0.165	0.166	0.169	$1.1 \cdot 10^{-10}$	2 608	3 508	4 722	0.2
Wind power												
	2406	2406	2407	0.3	0.100	0.103	0.108	$5.8 \cdot 10^{-10}$	1 000 000			
✓	2406	2406	2407	0.1	0.102	0.103	0.104	$2.3 \cdot 10^{-11}$	10 000			
✓	2405	2407	2409	0.8	0.102	0.103	0.105	$6.4 \cdot 10^{-11}$	2 380	3 427	5 385	0.2
Transmission limitation 25 MW												
	2888	2889	2890	0.2	0.171	0.181	0.188	$1.6 \cdot 10^{-9}$	1 000 000			
✓	2885	2889	2892	2.3	0.166	0.183	0.231	$3.3 \cdot 10^{-8}$	10 000			
✓[a]	2886	2890	2893	2.6	0.175	0.180	0.186	$1.4 \cdot 10^{-9}$	10 000			
✓[a]	2887	2890	2895	5.1	0.173	0.181	0.193	$1.4 \cdot 10^{-9}$	3 839	6 116	7 943	0.2

a. Using theoretically calculated stratification parameters.

tion runs resulted in confidence intervals which really included the true value. The reason is simply that I did not think of adding a function investigating this question to my test software. A rough survey[35] yields that it seems like a 95% confidence interval really includes the true value in about 95% of the simulation runs. An important exception are those simulations that are subject to cardinal error; cardinal error means that the variance of the estimate has been undervalued, which results in too small confidence intervals. By using an appropriate stratification strategy cardinal errors should however be possible to avoid. Therefore, it seems reasonable to assume that the estimates really are normally distributed around the true value; hence, (8.4) can be applied.[36]

35. It would be to the least exhausting to manually compile statistics from about 7 000 simulations I have performed…

9.3.2 Kigoma Revisited

In my licentiate thesis [7] I used Kigoma region in western Tanzania as an example to demonstrate Monte Carlo simulation of an ideal electricity market. In those simulations I used complementary random numbers and stratified sampling. The stratified sampling was unfortunately rather clumsily performed—I used multiple state nodes for the available generation capacity, where each node had two child nodes in the load level of the strata tree. The limit between these child nodes I simply put half-way between the highest and lowest available generation capacity of the parent node.

To verify that my new simulation strategies also are efficient for systems with more than two areas, I have brushed up two of the systems from [7] and simulated them again. The first of these systems was referred to as “case 2” in [7] and consisted of seven areas. The most important power plant is a hydro power plant, but there are also diesel generator sets in each load centre. The other system, “case 3a”, is almost the same, but the hydro power plant has been replaced by a wind farm next to one of the load centres; this system therefore has just six areas. A regional transmission grid of 33 kV connects the areas. The transmission on the lines are confined by thermal limits, which however are so large compared to the loads in the system, that there in practice are no transmission limitations. For further details about the two cases, please refer to [7].

Results

As I already have described, my software needs somewhat more human assistance to simulate systems with more than two areas, and I have therefore not simulated every system several times with different seeds for the random number generator, but I have performed one single simulation of each system. The results are displayed in table 9.13.

Before I discuss the accuracy of the recorded simulation runs, I would like to briefly explain some partial results. In the old simulations I restricted myself to estimating $\tilde{F}_{Un}$ (i.e., the duration curve of the unserved power within each area) and calculated the risk of load shedding in each area as $LOLP_n = \tilde{F}_U(0)$. Since I never calculated any confidence intervals of the

36. My results are thus contradicting the claim by Suhartono et al. in [62] that *LOLP* estimates should be considered gamma-distributed. The difference might be that the precision of Suhartono’s simulation are far less than in my test systems; in Suhartono’s system the result is $0.034\% \le LOLP \le 0.195\%$ (confidence level 90%) compared to for example $0.102\% \le LOLP \le 0.106\%$ (confidence level 95%) for a typical simulation of the wind power system.

Table 9.13 *Results of simulating Kigoma region.*

	Case 2		Case 3a	
	Old simulation (from [7])	New simulation	Old simulation (from [7])	New simulation
ETOC [USD/h]				
Monte Carlo	16.46 ± 0.16	16.62 ± 0.03	557.12 ± 2.46	557.83 ± 1.00
PPC	14.32	14.28	556.48	553.90
LOLP [%]				
Whole system				
Monte Carlo	0.064 ± 0.002	0.073 ± 0.001	4.132 ± 0.099	4.192 ± 0.015
PPC	0.065	0.065	4.191	4.153
Kigoma	0.055	0.050 ± 0.014	3.798	3.773 ± 0.561
Kasulu	0.015	0.025 ± 0.016	0.824	0.399 ± 0.561
Kibondo	0.017	0.055 ± 0.014	0.973	1.232 ± 0.857
Uvinza	0.011	0	0.729	0.199 ± 0.407
Number of scenarios	15 360	604	30 720	1 932

duration curves, the $LOLP_n$ estimates lack confidence intervals in the old simulations. In the new simulations I chose to add $LOLO_n$ as individual result variables (which however were not given any consideration when calculating compromises of the Neyman allocation; neither were they included in the stopping rule). By that means I obtained confidence intervals for $LOLP_n$ too, but as these estimates were given no priority, the precision is not exactly something to be bragging about. Yet estimations with large uncertainties can be used to high-light tendencies, which is quite clear in case 3a. As can be seen from the result table, an overwhelming majority of the load shedding occurs in Kigoma. From that it can be concluded that if there is a reserve power plant, the generation capacity, availability and operation cost of which are independent of its location, it will be of most benefit if it is built in Kigoma.

How efficient are then these simulations? As each system only has been simulated once, it is of course impossible to compare the efficiency of the different methods by studying the spread of the *ETOC* and *LOLP* estimates respectively. It is however possible to get an idea about the precision by studying the confidence intervals of each estimated system index. Apparently, the confidence intervals are much smaller using the new simulation method, even though the number of studied samples is about 95% less.[37]

37. In this context it may be noted that the new simulation used a stopping rule with a relative tolerance of 20% for the operation cost and the system-wide risk of load shedding. In the old simulations I made a personal judgement when the sampling procedure “looked” like it had converged and the confidence interval had shrunk to a sufficiently small interval.

Even more important than the size of the obtained confidence intervals is of course whether or not the estimates are correct or not. Here, we have to be content with an assessment of the plausibility of the results, because it would be almost mental abuse to force somebody to theoretically calculate *ETOC* and *LOLP* for systems of this size. If we start by case 2—the hydro power system—we see that both methods produce about the same estimate of *ETOC* and there is no reason to believe that this value should be incorrect. However, the *LOLP* estimates do differ, which apparently depends on cardinal error in the old simulation—the estimated value of *LOLP* is less than the value obtained by PPC simulation.[38] The older simulation has clearly completely disregarded load shedding due to the transmission losses (i.e., scenarios of type VI).

The simulation of case 3a follows the same pattern; both methods yield similar estimates of *ETOC*, whereas the *LOLP* estimate of the older simulation is subject to cardinal error. The most interesting about this case is that it illustrates another important practical aspect, namely that the result of the PPC calculations depends on the precision used in the numerical calculations. The PPC results of the old case 3a simulation differ from the new results and I am convinced that the new results are correct. In the new simulation I performed the PPC simulation repeatedly, successively increasing the accuracy of the calculations, until the resulting values of *ETOC* and *LOLP* no longer were affected by reducing the step length of the PPC simulation.

Finally it can be concluded that the new simulation method does not need more scenarios to produce reasonably accurate estimates of *ETOC* and *LOLP* in these six- and seven-area systems than was required in the corresponding two-area system. Actually, the number of scenarios is less in the larger systems, but considering that each system has only been simulated once, it is hard to draw any conclusions from that circumstance.

Some Practical Observations

Although Kigoma region is quite a small system to simulate (compared to for example simulating the whole Nordic electricity market), yet it is so much larger than the two-area systems described in section 9.3.1, that it provides interesting insight about the problems which have to be addressed when simulating real systems. It is self-evident that it takes quite a lot of work to perform a Monte Carlo simulation, as it involves a large number of scenarios to be analysed. What should not be forgotten is that there are several other time consuming steps, before the Monte Carlo simulation even can be started.

38. It is the irony of fate that this simulation was subject to cardinal error, because it was when simulating this particular system that I first discovered the phenomenon cardinal error—however only concerning *ETOC*…

The first problem is of course to collect the data of the electricity market model. I have only used Kigoma region to test the simulation method and make no claim to provide an exactly correct picture of the possibilities for electrification of the region; hence, I have been able to take the liberty of using rough estimates and pure conjecture; nevertheless it took a few days work to compile at least fairly reasonable input.

Among the preparations a little bit more affiliated to the simulation itself, there are three work intensive steps: determining the probability distributions of the scenario parameters, building the strata tree and performing a probabilistic production cost simulation. If too much time is spent on building the strata tree and running the PPC simulation there is a risk that a major part of the efficiency gain of using stratified sampling and control variates respectively is lost. To avoid that kind of trouble the simulation software should be written in such a manner that as many partial results as possible are saved and can be reused.

The probability distribution of some scenario parameters can be determined at the same time as the strata tree is built; both tasks are about enumerating each possible state of the scenario parameters and calculating the probability of this state. This is certainly not a difficult task, but it is however extensive. In for example case 3a above the available generation capacity can assume 16 848 states. Each of these states can occur during four different time periods of the day (with different probability distributions of the load during the four periods). In each state we need to separate about four partitions of the D_{tot}-axis. The final result is a strata tree comprising no less than 271 223 branches! Inevitably it takes some time to process all these branches. It would therefore be desirable if building a strata tree could be skipped before each new simulation and an old strata tree used instead. This is not possible if there have been any changes to the probability distribution of the scenario parameters (although some parts of the old strata tree might be reused), but if the change only applies to some model constant then hopefully the same stratification may be used again. If the transmission grid is reinforced so that the losses decrease then $\bar{L}$ is affected—and hence basically the whole strata tree—but if the change is small, not much efficiency will be lost by not updating the stratification.

The probabilistic production cost simulation is time consuming due to the number of convolutions which have to be performed. The time consumption increases exponentially when the precision requirements are increased of the PPC calculations. As mentioned above it is important that the PPC simulation results are accurate, as the control variates otherwise will produce biased results, and it may therefore be necessary to perform several PPC simulations with different accuracy, in order to control that there are no numerical errors that disturb the results. In this way the PPC simulation will be even more time consuming. However, it can be utilised that the PPC model does not consider any other factors than the power plants available and the load; hence, there is no reason to make a new PPC run if a new simulation should be made, where some other property of the system has been changed.

Chapter 10

LONG SCENARIOS

In an electricity market with energy storage facilities or non-negligible time constants (e.g. long start-up times in thermal power plants) it is necessary to consider scenarios, where the scenario parameters vary over time. I refer to this as simulation of *long scenarios*. Long scenarios are in many ways more difficult to manage compared to short scenarios; therefore, I chose after a few simple experiments (which are described in [5]) to focus on systems with short scenarios—it seemed to be proper to learn how to walk (i.e., simulate short scenarios) before trying to run (i.e., simulate long scenarios). I have not had time to perform a systematic analysis of how to simulate electricity markets with long scenarios in an efficient way. The objective of this chapter is therefore to show similarities and differences between short and long scenarios.

10.1 SCENARIO PARAMETERS

A scenario was defined in section 1.1 as a situation with given conditions for the electricity market. A scenario is represented mathematically as a fixed outcome for each of the scenario parameters (random variables with known distribution) that appear in the electricity market model. Exactly which scenario parameters are necessary to define a short scenario is of course depending on which model has been chosen. The scenario parameters of a short scenario—available generation capacity, available transmission capability and load—are needed to define long scenarios too. Besides, there are other scenario parameters related to the energy limited power plants: inflow to energy storage facilities as well as start and final contents of the energy storage facilities. The main difference between a short and a long scenario is however that in the former all scenario parameters are constant for the whole duration of the scenario, whereas in long scenarios they may vary over time.

Below follow further details about the probability distributions of different scenario parameters. However, I limit the discussion to such scenario parameters which are needed for simulation of ideal electricity markets. The scenario parameters introduced in some non-ideal electricity markets generally follow one of the patterns described below; for example, emission caps cause thermal power plants to obtain similar properties as energy limited power plants. A more detailed study of the scenario parameters of non-ideal electricity markets will have to be a future project.

Available Generation Capacity and Transmission Capability

Concerning short scenarios I suggested that a probability distribution of the primary scenario parameter total available generation capacity should be determined. The idea was that there is some knowledge about how the total generation capacity affects the result variables of a short scenario and this knowledge may then be used in variance reduction techniques as complementary random numbers and stratified sampling. Unfortunately, it is not possible to define a similar primary scenario parameter in long scenarios, because it is pointless to sum the available capacity in different power plants during different time periods.[1] When randomizing the available generation capacity it must also be considered that the generation capacities in two consecutive time periods are strongly correlated. If a power plant is available in period t then it is likely that it is available in period $t + 1$ too, because failures are rare events.

Therefore, when randomizing the available generation capacity in a long scenario another method must be used than for short scenarios. In thermal power plants and energy limited power plants it can be assumed that in every period the available generation capacity can have one of two possible values: installed capacity (when it is available) or zero (when it is unavailable). In order to randomise the available generation capacity in a long scenario we start by determining the initial state and then we randomise how long time it will take before the power plant changes its state. This approach requires that the following properties are known about the power plants:

1. If we for example have a short scenario where $\overline{G}$ = 100 MW and D = 80 MW then we know that there will not be any load shedding (unless losses or transmission congestion cause problems). If we in a long scenario know that $\overline{G}$ *in average* is 100 MW and D *in average* is 80 MW then this does not say anything about the risk of load shedding, because we need to consider how the scenario parameters vary over time. If the long scenario has two periods and s $\overline{G}$ = 150 MW in the first, and 50 MW in the second, while the load is 120 MW and 40 MW respectively, then everything is all right. But if $\overline{G}$ = 100 MW in both periods, we get load shedding during 50% of the scenario duration. It is in other words impossible to predict the properties of a long scenario just from a summation of the available generation capacity over all time periods, which means that such a scenario parameter cannot be used for any variance reduction technique.

Definition 10.1. The failure rate, λ, states the probability that an available power plant will fail at a certain moment. The failure rate relates to the *MTTF* (Mean Time To Failure) by

$$\lambda = \frac{1}{MTTF}.$$

Definition 10.2. The repair rate, μ, states the probability that the repairs of an unavailable power plant are completed at a certain moment. The repair rate relates to the l *MTTR* (Mean Time To Repair) by

$$\mu = \frac{1}{MTTR}.$$

Definition 10.3. The availability, *p*, states the probability that a power plant can be operated. The availability relates to failure and repair rates by

$$p = \frac{\mu}{\mu + \lambda}.$$

To determine the initial state of the power plant the following trivial probability distribution is used:

$$f_{\overline{G}g}(x) = \begin{cases} p & x = \hat{G}_g, \\ 1 - p & x = 0, \\ 0 & \text{all other } x. \end{cases} \tag{10.1}$$

The time to next failure and repair respectively is commonly assumed to be exponentially distributed,[2] i.e., $TTF \in E(\lambda)$ and $TTR \in E(\mu)$.[3] When the initial state of the power plant has been determined, a series of *TTF* and *TTR* is generated until the state of the power plant has been determined for each period of the scenario. An example of the result of this method is shown in figure 10.1.

The available generation capacity of a non-dispatchable power plant during a long scenario is more complicated to randomise, because it depends both on whether or not the power plant is technically available and some weather

2. There are also other possible distributions, where running-in problems and aging are considered. For further details about this kind of modelling, please refer to [19, 47].
3. Notice that if X is $E(\lambda)$-distributed then the expectation value $E[X] = 1/\lambda$. Thus, *MTTF* and *MTTR* correspond to the expectation values of *TTF* and *TTR*. Methods for generation of exponentially distributed random numbers are described in appendix C.

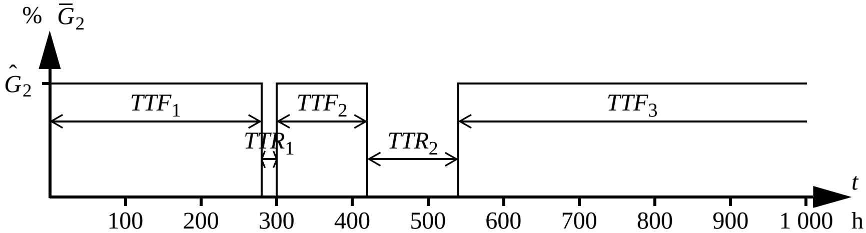


Figure 10.1 *Example of available generation capacity of a power plant in along scenario. The power plant in the example has MTTF* = 500 *and MTTR* = 40 *respectively, which corresponds to an availability of about* 93%. *In this case the scenario starts by the most likely state of the power plant (i.e., it is available) and it fails and repairs twice before the scenario ends after* 1 000 *hours.*

depending factor. To generate a realistic series $\overline{W}_1, \ldots, \overline{W}_T$ it is necessary to use similar methods as for load and inflow, which I will describe below.

The available transmission capability between two areas is partly limited by the requirements of voltage stability and partly by the number of available transmission lines. The voltage stability requirement is approximately represented in a multi-area model (cf. section 3.2.1), which means that the transmission capability in the model only depends on whether or not the lines interconnecting the areas are available or not. Each of these lines has its own *MTTF* and *MTTR*; the availability is treated using exactly the same methods as for thermal power plants.

Load and Inflow

Both load and inflow show periodical patterns, the exact nature of which is depending on socio-economic factors (e.g. people's work hours) as well as meteorological factors (e.g. seasonal variations); some examples are given in figures 10.2 and 10.3 respectively. The periodical pattern is however just a part of the varying load and inflow; there is a considerable amount of random events, too. Somehow the combination of periodical patterns and purely random events must be imitated when generating time series of $D_{n,\,t}$ and $Q_{r,\,t}$. The periodicity results in correlations both between different time periods and between different areas of the power system. Therefore, it is no easy task to generate realistic probability distributions. The following two alternatives are available:

- **Historical data.** A simple solution to the problem is to use historical data. However, access to large amounts of measurements is necessary in order to produce correct results, as the data base should include as many potential outcomes as possible.

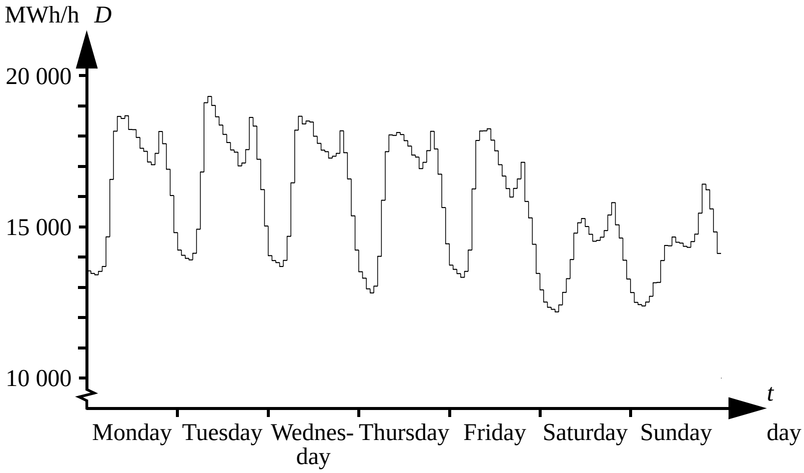


Figure 10.2 *Example of load cycle. The figure shows the electricity consumption in Sweden during the period 29/9 to 5/10, 2003. A typical daily cycle can be seen, with one peak in the morning, another peak in the afternoon and low load during the night. There is also a weekly cycle, as the daily cycle is somewhat different during workdays and weekends.*

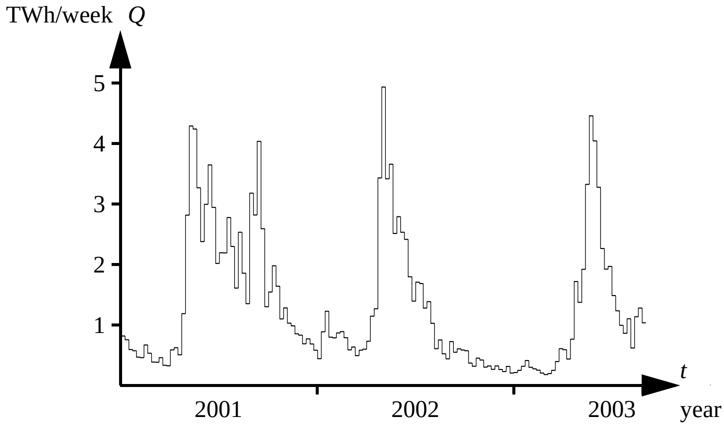


Figure 10.3 *Example of inflow cycle. The figure shows the inflow per week in Sweden from January 1, 2001 up to August 25, 2003. The pattern is similar from year to year: not much inflow during the winter, a big peak when the snow starts melting in May and June.* However, the total inflow can vary widely from year to year. During 2001 the autumn was extremely rainy (with several large floods), whereas the autumn in 2002 was extremely dry.*

* *This refers just to the rivers of northern Sweden. The general climate conditions in Sweden are not* that *harsh…)*

- **Stochastic process.** This method requires that we design a stochastic process, which generates output with approximately the same statistical properties as historical data. The advantage of this method compared to directly using historical data is that the scenario population includes such scenarios that are possible, but not yet recorded in the historical data. The disadvantage is of course that it takes a more or less extensive work effort to design a proper stochastic process.[4]

Regardless of how the time series are generated it is possible to introduce primary scenario parameters, D_{tot} and Q_{tot} respectively. Given a probability disturbing of those, a time series can be scaled so that $\Sigma_t\Sigma_n D_{n,\,t} = D_{tot}$ and $\Sigma_t\Sigma_r Q_{r,\,t} = Q_{tot}$. The scaling works similarly and has similar consequences as when a geographical load distribution in a short scenario is scaled to match a certain total load (cf. section 9.1).

Initial and Final Contents of Energy Storage

The reason why we at all study long scenarios rather than the more comfortable short scenarios is that we want to model how the decisions of the players do not depend on the system state at a certain occasion, but is also depending on the earlier decisions of the players and their perception of the future. The dilemma is that their decisions do not just depend on events within the limited period of time included in the scenario. Figure 10.4 is an attempt to illustrate the difference between the time perspective of the players in the electricity market and of the simulation.

In the model all earlier decisions are primarily represented by the amount of energy stored in the system at the beginning of the scenario; i.e., $M_{r,\,0}$ represent the history of the system. Accordingly, the future is represented by the energy that is stored at the end of the scenario, i.e., $M_{r,\,T}$. These scenario parameters must be chosen carefully, because they are correlated to the load, D_{tot}, as well as the inflow, Q_{tot}. If the inflow is large and the load is small in a particular scenario then it would be reasonable that the energy storage facilities are quite empty at the beginning of the scenario, because the players have probably prepared for the energy surplus by storing less in the storage facilities. Moreover, the final contents of the scenario should have an increased likelihood of lesser values; if the scenario is followed by a "normal scenario" (i.e., a scenario with inflow and load closer to the mean) or even another "surplus scenario" (i.e., continuing high inflow and low load) the players will be less inclined to store energy. Meanwhile, there is also a possibility that the

4. For that matter, it is not unthinkable that a systematic method could be developed to adjust the parameters of a stochastic process to match a series of load or inflow data. Please refer to [126] for further details on this topic.

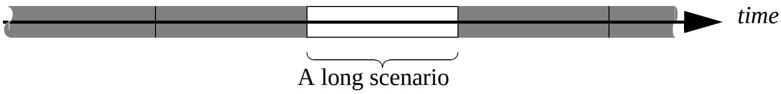


Figure 10.4 *Modelling of time. The players modelled in a long scenario see time as a straight line—they have a certain, known history and they anticipate a certain future (in accordance to their forecasts). Both history and future will affect the decisions of the players.*

The time perspective is different in a simulation; a scenario is just a limited part of the time line. Assuredly, the scenario must be proceeded by other scenarios and it will be followed by further more scenarios, but when analysing a single scenario, we do no have any exact knowledge of these scenarios—the history and future is a grey, unknown mist, which somehow must be modelled in order to simulate how the players of the electricity market relates to their earlier decisions and their forecasts.

scenario is followed by a “deficit scenario” (where inflow is low and load high) and then the energy storage facilities should be well filled at the end of the scenario.

The reasoning above indicates that it is a major challenge to choose the probability distribution of the initial and final contents of energy storage facilities. Further research is required in this field before an acceptable solution can be presented. Some insight in the subject are given in for example [25]. I myself have been thinking about two options, which could be worthwhile to study closer:

- **Dynamic probability distribution.** If the predetermined final contents of the energy storage facilities are replaced by benefit functions (see section 3.2.1) then an iterative process can be used to obtain an appropriate probability distribution of the initial contents. We start by letting the storage facilities be half-filled or some other arbitrary value. After simulating a number of scenarios, we can estimate the duration curve of the final contents of the storage facilities. As initial and final contents should have the same probability distribution, this duration curve can be used as a new estimate of the probability distribution of the initial contents. Then some further scenarios are simulated according to the new distribution, and a duration curve of the final contents is estimated, etc. Hopefully, this process will converge to an appropriate probability distribution
- **Very long scenarios.** If a scenario comprises a very long time period (compared to how long time it takes to empty and refill

the energy storage facilities of the system) then it is reasonable that both load and inflow will be approximately equal to their expectation values, which means that the contents at the end of the scenario should be equal to the contents at the beginning. In this kind of scenarios it is reasonable to assume that all scenarios start and end with half-filled energy storage facilities (or some other value.).[5]

10.2 POSSIBILITIES FOR VARIANCE REDUCTION

To create a variance reducing effect it is necessary to have some information about the system to be studied. As described in the previous section, the probability distribution of the scenario parameters have partly different properties in short and long scenarios respectively, which means that the advance information about the properties of a long scenario is somewhat different compared to simulation of short scenarios. Below I will discuss how this affects the different variance reduction techniques.

Complementary Random Numbers

Complementary random numbers are based on the existence of a negative correlation between the values of the result variables of a scenario and its complementary scenarios. Of the primary scenario parameters we have in a long scenario it is only the total load and the total inflow that clearly have correlations to the result variables. It is likely that the operation cost is high if the total load is high, whereas if the load is low, the operation cost should be low, too. A large inflow mean more power generation in the energy limited power plants and as these power plants have negligible or small variable costs, the operation cost should be low. A small inflow will for the same reason probably result in higher operation costs. Both D_{tot} and Q_{tot} are thus correlated to *TOC*. Complementary random numbers can therefore be used when simulating long scenarios to randomise D_{tot} and Q_{tot} by forming complementary scenarios in the same manner as for short scenarios (see section 9.2.4).

Besides, there is also a connection between load, inflow and load shedding. If the load is high and/or the inflow is low, there is a risk that the electricity market becomes subject to energy deficit, i.e., the stored energy is not sufficient to cover the remaining load when thermal and non-dispatchable power plants operate at their maximal capacity. The consequence is that although

5. I used a similar idea in my first experiments with long scenarios; see [5].

the available generation capacity is continuously larger than the load, at some point load shedding will be necessary, because the available capacity in the energy limited power plants cannot be utilised due to empty storage facilities. Thus, energy deficit inevitably results in power deficit, but the players of the electricity market have some control of when load shedding will occur. If desirable, a minor part of the load could be continuously disconnected, which would give *LOLO* = 100%, or it would be possible to choose larger disconnections for example one night per week, which would result in a value of *LOLO* around 5%. The exact relation between energy deficit and power deficit is thus dependant on how to ration the electric energy that actually can be produced. The correlation between load, inflow and load shedding may therefore vary depending on the rules of the electricity market.

Dagger Sampling

I could hardly see any use for dagger sampling when simulating short scenarios, because I find it easier to use complementary random numbers for a common probability distribution of available generation capacity. In long scenarios, we do not use such common probability distributions, because we must generate a sequence of states for each power plant, which is exactly what dagger sampling does. However, the sequence obtained by dagger sampling will have at most one period where the power plant is not available. Such a sequence is only useful if *MTTF* is so large that it is extremely unlikely that two errors would occur in one power plant during one dagger sampling cycle, while *MTTR* must be so small that it is almost certain that a failing unit will be repaired before next time period. It should be fairly unusual that these conditions are fulfilled and therefore dagger sampling seems to have a very limited use for simulation of long scenarios; the only possibility I see is to use the method when randomizing the initial state of the power plants and transmission lines. Whether or not this would produce any efficiency gain is something I do not dare to predict—this question has to be answered by practical experiment.

Control Variates

As for short scenarios the results of probabilistic production cost simulation (PPC) may be used as control variates for a Monte Carlo simulation. In PPC an optimal energy value is determined for each energy limited power plant [11, 32]. This energy value, ν_r, is then compared to the generation cost of the thermal power plants when determining the merit order of the power plants. It is not considered that energy storage facilities have limited storage capacity. A PPC model of a long scenario can be written as

$$\text{minimise} \quad \sum_{t=1}^{T} T_t\Big(\sum_{g \in G} \beta_{Gg} G_{g,t} + \sum_{r \in R} \nu_r H_{r,t} + \beta_U U_t\Big) \tag{10.2}$$

$$\text{subject to} \quad \sum_{g \in G} G_{g,t} + \sum_{r \in R} H_{r,t} - U_t = D_t, \qquad t = 1, \ldots, T, \tag{10.2a}$$

$$0 \le G_{g,t} \le \bar{G}_{g,t}, \qquad \forall\, g \in G, t = 1, \ldots, T, \tag{10.2b}$$

$$0 \le H_{r,t} \le \bar{H}_{r,t}, \qquad \forall\, r \in R, t = 1, \ldots, T, \tag{10.2c}$$

$$0 \le U_t, \qquad t = 1, \ldots, T, \tag{10.2d}$$

From the solution of (10.2) we can calculate

$$TOC = \sum_{t=1}^{T} T_t \sum_{g \in G} \beta_{Gg} G_{g,t}, \tag{10.3a}$$

$$LOLO_t = \begin{cases} 0 & \text{if } U_t = 0, \\ 1 & \text{if } U_t > 0, \end{cases} \tag{10.3b}$$

$$LOLO = \frac{\sum_{t=1}^{T} LOLO_t T_t}{\sum_{t=1}^{T} T_t}. \tag{10.3c}$$

The values of the result variables of the PPC model (10.2) are then compared to the result of the electricity market model of the Monte Carlo simulation; the principles are the same as for short scenarios.

Correlated Sampling

My primary objection to using correlated sampling for simulation of short scenarios was that the method in many cases is hard to combine with stratified sampling. Concerning long scenarios stratified sampling is used in a completely different manner (see below) and there should not be any difficulties in using both methods simultaneously. This is of course an advantage, as correlated sampling is well suited to study small differences between different systems, i.e., exactly the kind of studies necessary to determine the value of an investment (cf. section 1.1). An objection is however that it would still be interesting to know the absolute values of the system indices in both cases and not just the differences.

Stratified Sampling

In chapter 9 I showed that stratified sampling is an excellent method to make Monte Carlo simulations of electricity markets with short scenarios more efficient. The strata definitions were based on the fact that knowing the available resources it would be possible to predict how the system will behave for different levels of the total load. Such predictions are unfortunately not possible for long scenarios.[6] However, as I concluded in the above discussion of complementary random numbers, it is possible to make predictions about the primary scenario parameters total load, total inflow and the result variables operation cost and loss of load. It could therefore be possible to build a strata tree with two levels below the root, where one level specifies values of D_{tot} and the other one states values of Q_{tot}. The resulting strata from such a strata tree would however not be as homogeneous as the strata we get when building a strata tree for short scenarios; hence, the efficiency gain would be less.

It therefore seems as if stratified sampling is far less useful for long scenarios compared to short ones. There is however a brand new problem to address when simulating long scenarios, namely that the size of the scenario problem can become so large that the analysis of single scenarios becomes troublesome time-consuming.[7] But there is a possibility that the calculation time can be reduced by using stratified sampling.

The idea is that a scenario problem can be divided in a long-term problem and a number of short-term problems, where the duration of one period of the long-term problem corresponds to the total duration of a short-term problem. In the short-term problems we use the same period length as in the original scenario problem. The long-term problem yields an approximate solution to the scenario problem and this solution is further refined in the short-term problems. Depending on how large optimisation problems we can manage, it might become necessary to divide the analysis of a scenario in more than two steps; cf. figure 10.5. In the following reasoning I will however assume that there is only one long-term problem and several short-term problems, without any intermediate levels.

Already the division into several analysis with different time perspectives in most cases results in an efficiency gain, because it is generally faster to solve several small multi-area problems than one large.[8] But there is also a

6. Cf. footnote 1.
7. For example, consider an electricity market divided in five areas, where there are in total four equivalent energy limited power plants. If each scenario in this electricity market comprises a year divided in one hour periods then we get $5 \cdot 8\,760$ load balance constraints and $4 \cdot 8\,760$ energy balance constraints, i.e., in total 78 840 constraints. Such large problems can be solved rather quickly if a linear model is used, but otherwise there is—to say the least—sweaty work ahead…

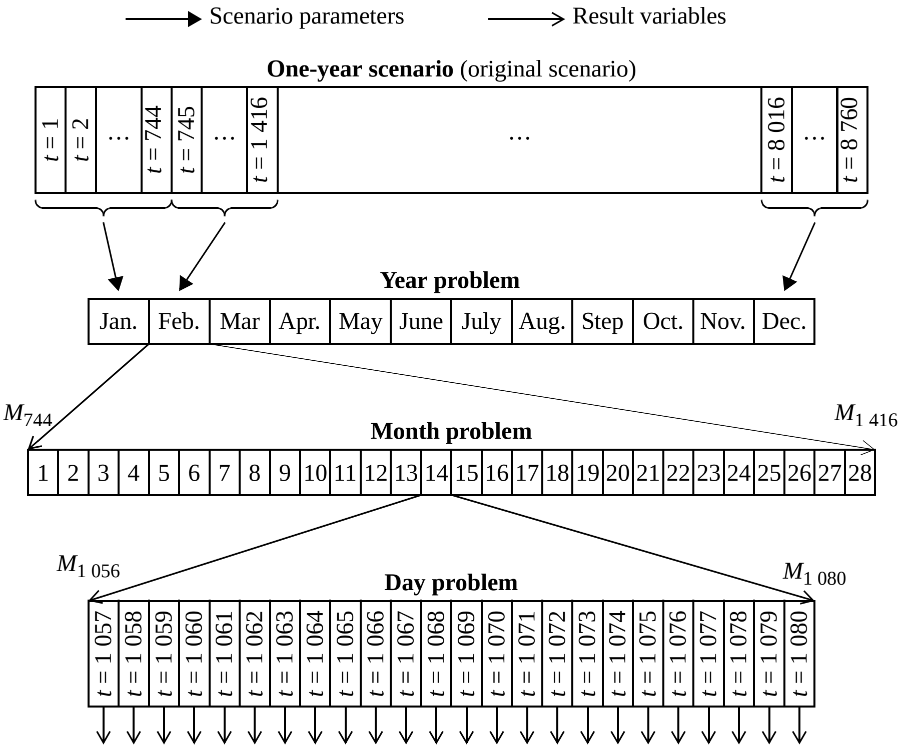
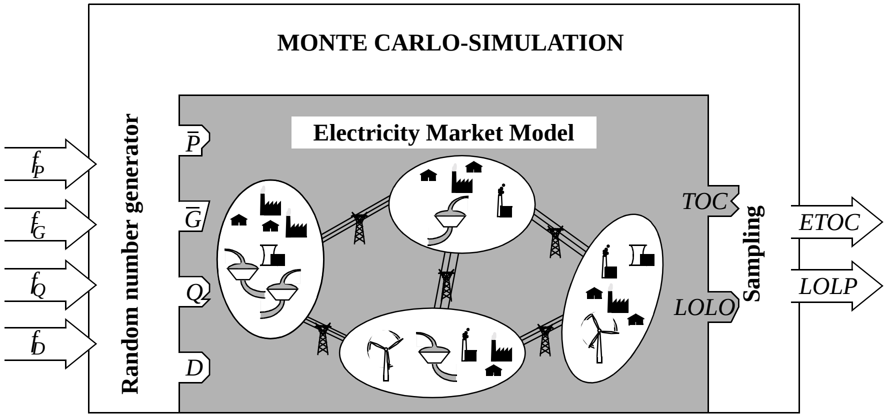


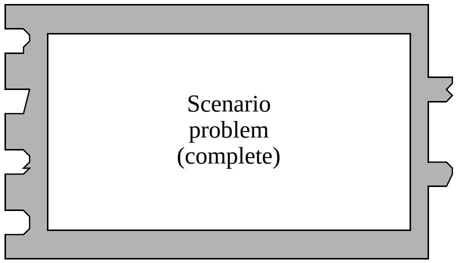
Figure 10.5 *Analysis of a one-year scenario. We start from a scenario divided in 8 760 hours. From this scenario we first define a year problem by taking monthly averages or monthly sums of the scenario parameters of the original scenario problem. The solution to this problem provides initial and final contents for the energy storage facilities in each of the month problems. In the month problems each period corresponds to a day, and mean values of the original problem are used here too (not illustrated in the figure). The results of the month simulations give initial and final contents of the energy storage facilities in the 365 day problems. The scenario parameters in the day problems are directly taken from the original problem. The solution of the day problems are used to determine TOC, LOLO and other result variables.*

Thanks to the division into smaller problem we will not have to solve a multi-area problem comprising 8 760 time periods; we will have 378 smaller problem instead, where the largest one comprises 31 time periods.

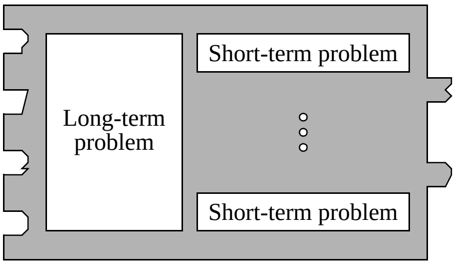
8. Cf. the practical experiments of [135].



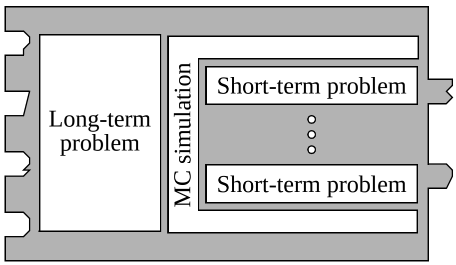
a) Desired electricity market model.



b) Here a certain scenario is analysed as one single optimisation problem, where each period is one hour long.



c) Here a certain scenario is analysed by first solving an overall long-term problem and then several more detailed short-term problems. One period in the long-term problem is as long as the duration of a whole short-term problem. Each period in the short-term problems is one hour.



d) Here the same basic idea is used as in panel c, but rather than solving all short-term planning problems, we restrict ourselves to a number of random samples. In this case Monte Carlo simulation is used to calculate TOC and LOLO.

Figure 10.6 *Monte Carlo simulation of long scenarios. The figure illustrates how the same electricity market model can be treated in several ways when analysing individual scenarios. Which solution is chosen does not affect the electricity market simulation, as the three pieces of a puzzle in panel b-d all fits into the same Monte Carlo machinery in panel a.*

Figure 10.7 *Example of division of a long scenario. The model constants and scenario parameters are shown in the figure to the right. Panel a shows the values of the result variables when the whole scenario problem is solved at once. Panels b-d show the result when the same scenario is divided in long- and short-term problems. The capacity limit in the sixth hour is missed in the long-term problem, which has difference consequences depending on how information is transferred to the short-term problems. If reservoir limousines are used (panel c) then too much water is allocated to hour 4-6. All the water will be utilised, but not in an optimal manner, which causes the TOC to be slightly higher than in the original scenario. Using water values (panel d) the missed limitation results in overestimated water values; hence, too much water is saved for future use. If the reservoir contents is calculated according to the solutions of the short-term problems then we find* 8 MWh *too much water is left after hour 9. The lesser hydro power generation has been replaced by increased thermal generation, which results in too high TOC compared to the original problem.*

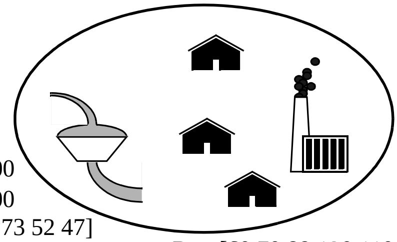
Scenario problem (original)	Hour 1	2	3	4	5	6	7	8	9
Load, D [MWh/h]	60	70	89	120	110	130	120	90	81
Inflow, Q [MWh/h]	57	42	45	58	55	63	73	52	47
Reservoir contents [MWh]									
At beginning of hour, M_{t-1}	100	138	151	148	127	113	96	90	93
At end of hour, M_t	138	151	148	127	113	96	90	93	100
Generation [MWh/h]									
Hydro power, H	19	29	48	79	69	80	79	49	40
Thermal, G	41	41	41	41	41	50	41	41	41
Electricity price, λ [¤/MWh]	141	141	141	141	141	150	141	141	141
Water value, ν [¤/MWh]	141	141	141	141	141	141	141	141	141

a) Solution to the original scenario problem. TOC = 45 774 ¤.

Long-term problem	Hours 1-3	4-6	7-9
Load, D [MWh/h]	73	120	97
Inflow, Q [MWh/h]	144	176	172
Reservoir contents [MWh]			
At beginning of hour, M_{t-1}	100	151	93
At end of hour, M_t	151	93	100
Generation [MWh/h]			
Hydro power, H	31	78	59
Thermal, G	42	42	42
Electricity price, λ [¤/MWh]	142	142	142
Water value, ν [¤/MWh]	142	142	142

b) Solution to the long-term problem. *(continues next page)*

Figure 10.7 (cont.)



$\overline{H} = 80 \quad M_0 = 100$

$\overline{M} = 200 \quad M_9 = 100$

Q_t = [57 42 45 58 55 63 73 52 47]

$\overline{G} = 100$

$C_G(G) = 100G + \frac{1}{2}G^2$

D_t = [60 70 89 120 110 130 120 90 81]

Short-term problem (reservoir limit)	Hour 1	2	3	4	5	6	7	8	9
Load, D [MWh/h]	60	70	89	120	110	130	120	90	81
Inflow, Q [MWh/h]	57	42	45	58	55	63	73	52	47
Reservoir contents [MWh]									
At beginning of hour, M_{t-1}	100	139	153	151	129	110	93	88	92
At end of hour, M_t	139	153	151	129	110	93	88	92	100
Generation [MWh/h]									
Hydro power, H	18	28	47	80	74	80	78	48	39
Thermal, G	42	42	42	40	36	50	42	42	42
Electricity price, λ [¤/MWh]	142	142	142	140	136	150	142	142	142
Water value, ν [¤/MWh]	142	142	142	136	136	136	142	142	142

c) *Solution to the short-term problems when using initial and final reservoir contents from the long-term planning problem. TOC* = 45 790 ¤. *Values from the solution to the long-term problem are shaded.*

Short-term problem (water value)	Hour 1	2	3	4	5	6	7	8	9
Load, D [MWh/h]	60	70	89	120	110	130	120	90	81
Water value, ν [¤/MWh]	142			142			142		
Generation [MWh/h]									
Hydro power, H	18	28	47	78	68	80	78	48	39
Thermal, G	42	42	42	42	42	50	42	42	42
Electricity price, λ [¤/MWh]	142	142	142	142	142	150	142	142	142
Inflow, Q [MWh/h]	57	42	45	58	55	63	73	52	47
Reservoir contents [MWh]									
At beginning of hour, M_{t-1}	100	139	153	151	131	118	93	88	92
At end of hour, M_t	139	153	151	131	118	101	88	92	100

d) *Solution to the short-term problems when using water values from the long-term problem. TOC* = 46 906 ¤. *Values from the solution to the long-term problem are shaded. Notice that the reservoir contents are calculated* after *solving the short-term problems.*

possibility of making even larger efficiency gains, because most periods in a long scenario will show large resemblance. Rather than solving all short-term planning problems, we could restrict ourselves to a number of randomly chosen samples—in other words, Monte Carlo techniques may be used to estimate *TOC* and *LOLO* in a long scenario.[9] The difference between the three options solving the whole scenario problem at once, dividing it into several problems and solving all of them, and to use Monte Carlo techniques is illustrated in figure 10.6.

Stratified sampling should be useful for Monte Carlo simulation of a long scenario. Exactly how to use stratified sampling depends on how it is chosen to transfer information form the long-term and short-term problems. Here there are two alternatives.

The first is to transfer initial and final contents of the long-term problem to the short-term problems; the contents of the energy storage in the beginning of one period in the long-term problem becomes the initial state of the corresponding short-term problem (cf. figure 10.5).This means that the long-term and short-term problems have exactly the same structure—the only difference is that the short-term problems have shorter time periods. Using this kind of information transfer allows stratified sampling to be used for choosing which short-term problems should actually be solved. If there are large seasonal differences it might for example be possible to consider each month as one stratum.

The other method to transfer information is to use energy values. The dual variables of the energy balance constraint (3.14b) correspond to the value of the stored energy. When solving the long-term problem we thus get energy values for each energy storage and for each period. As a period in the long-term problem corresponds to a short-term problem we may say that the long-term problem produces one energy value for each storage and short-term problem, $\nu_{r,\,k}$. In the short-term problem we may now choose to neglect the energy balance constraints and treat the energy limited power plants as thermal power plants instead. The operation cost in the energy limited power plant is then assumed to be $C_{Hr,\,t}(H_{r,\,t}) = \nu_{r,\,k}H_{r,\,t}$. In this case we will have a long-term problem in the same shape as the original scenario problem, but with longer period length and short-term problem in the same form as (10.2). The finesse is that without the energy balance constraint each short-term problem will disintegrate to a number of one-period problems. Thus, when

9. Thus, we use Monte Carlo techniques to solve a subproblem of a Monte Carlo simulation. To avoid confusion it is very important to differ between Monte Carlo simulation of an electricity market (or *electricity market simulation* if a shorter designation is preferred), i.e., the problem of estimating the system indices *ETOC* and *LOLP* of a certain electricity market, and Monte Carlo simulation of a long scenario (or *scenario simulation*), i.e., the problem of estimating the result variables *TOC* and *LOLO* in a certain long scenario.

using energy values we may transform a long scenario of T periods to T short scenarios—and for sampling of short scenarios there are huge efficiency gains to be claimed! Hence, the long-term problem is solved first and then the resulting short scenarios are sorted into a strata tree. Based on the strata tree, we may define strata according to the principles described in the chapter on short scenarios.

When long scenarios are divided into several subproblems we introduce several error sources. The long period length of the long-term problem may cause incorrect mean values and overlooked limitations (see section 3.2.1), which means that the information transferred to the short-term problems will be slightly incorrect, as illustrated in figure 10.7. If we also choose to use Monte Carlo techniques to determine *TOC* and *LOLO* we will also add errors to the estimates. The question is if the benefit of the division—significantly shorter calculation time for analysis of a single long scenario—is large enough compensation for these disadvantages. The results in [5] indicated that the estimation errors from a Monte Carlo simulation of a long scenario do not have any larger importance for the final result of the electricity market simulation, but closer studies will be necessary to finally conclude how useful stratified sampling is to simulation of long scenarios.

Importance Sampling

When I was discussing importance sampling of short scenarios in section 9.2.4, I advocated that importance sampling was an alternative to stratified sampling, but the former method was to be preferred. A possible application of importance sampling was however to determine how to distribute the primary scenario parameter D_{tot} in the different areas of the system—this problem cannot easily be dealt with using stratified sampling. The same judgement is valid for long scenarios, where both D_{tot} and Q_{tot} must be distributed not just over different areas, but also over different time periods. Solving this task requires further studies of appropriate importance sampling functions.

Chapter 11

CLOSURE

When I started my Ph.D. project, it was not known beforehand that Monte Carlo simulation was a practical method to analyse modern electricity markets. As stated in my problem definition (see section 1.1), a simulation may not take days or—horrible thought—weeks to perform, if the simulation technique should be of any practical use. In a Monte Carlo simulation, this means that the number of scenarios necessary to obtain a sufficiently accurate result must be kept on a reasonable level. What constitutes a reasonable number of scenarios depends on how fast a single scenario can be treated; if it takes about a minute to calculate the result variables of a scenario, we may accept perhaps a few hundred scenarios—but if we can process hundred scenarios per second instead, it becomes possible to allow the simulation to include up to a million scenarios.

With efficient usage of variance reduction techniques, it hardly takes millions of scenarios to produce useful results, but rather between a thousand and ten thousand scenarios. Neither does it take very long time to evaluate a scenario. For example, it took my version of the NNP-algorithm[1] less than half a minute to process thousand scenarios when I was simulating Kigoma region, i.e., a multi-area model of an ideal electricity market having six or seven areas and twenty-one power plants.[2] When simulating larger power systems or non-ideal electricity markets, the computation time of each scenario will probably increase somewhat. Long scenarios will also require a larger work effort per scenario, which however might be compensated by a reduction of the total number of scenarios per simulation—closer studies of variance reduction techniques for long scenarios will tell how that turns out. Nonetheless, I am fairly certain that the simulation time will not be a problem, thanks

1. A program written in C, which was run on a 433 MHz Alpha processor. It should be noted that all matrix operations in this program was performed by subroutines which I had written myself. A professional programmer would undoubtedly write more efficient code and by that means cut the computation time significantly.
2. See [7] for more details.

to the increasingly fast computers of today, as well as further development of both optimisation algorithms and variance reduction techniques.

The first conclusion of my work is thus that the Monte Carlo technique is well suited for simulation of electricity markets. Below, I will give a further summary of my results concerning how to perform electricity market simulations, and give some suggestions of interesting issues for future research.

11.1 SIMULATION OF ELECTRICITY MARKETS

The objective of a simulation is to predict how an electricity market will behave for a certain set of resources, a certain demand and a certain market design. Each electricity market will face an infinite number of scenarios and therefore we need a simple way of summarizing the behaviour of the electricity market in all these scenarios, so that it is possible to get an overview of the simulation results. Thus, we may define a number of system indices, which are equal to the mean result of all possible scenarios.

I differentiate between two kinds of electricity market simulations: static and dynamic. In the former, the set of possible scenarios is constant. This means for example that no new power plants are built and that the load follows the same probability distribution all the time. However, in a dynamic electricity market simulation, the scenario population may change over time—the change may even be depending on what has happened earlier in the electricity market.

Another important aspect when simulating electricity market is how time is represented. I differentiate between two groups of models, namely those with short scenarios and those with long scenarios. In a short scenario, it is assumed that all the players of the electricity market make their decisions based only on the conditions of the present moment; the behaviour of the electricity market is thus independent of that has happened earlier in the system and of the players' expectations about the future. A short scenario is in other words an on-the-spot account of the electricity market. A long scenario on the other hand, comprises a certain time period, during which the conditions of the electricity market (for example the state of the power plants or the load) will have time to change. Thus, in a long scenario, we have the possibility to follow how the players of the electricity market adjust to changes.

Two questions are of particular importance when performing a Monte Carlo simulation of an electricity market. Firstly, it must be possible to find out how the electricity market will behave in a certain scenario, and secondly, we must choose the scenarios to study closer. Let me separately summarise my results concerning these two issues.

11.1.1 Electricity Market Models

It is the electricity market model which tells us how the electricity market will behave in a certain scenario. In practice this means that we solve an optimisation problem, which I refer to as the scenario problem.[3] Different electricity market models are thus separated by the different structures of the scenario problem.

In this dissertation I have developed a basic model—the ideal electricity market. Moreover, I have compared this basic model to the conditions in a real electricity market, described different market designs and strategies which follow from the differences, and I have suggested models. The analysis of the differences between ideal and non-ideal electricity markets has primarily been focused on three aspects of the electricity market: environment issues, forecast uncertainties and grid costs.

The Ideal Electricity Market

The basic model of an electricity market is what I refer to as an ideal electricity market. I have provided an extensive definition of the conditions which have to be fulfilled in order to consider an electricity market as ideal. This definition is useful in several ways. Firstly, it is straightforward to formulate the scenario problem of an ideal electricity market. Moreover, the ideal electricity market represents the optimal resource utilisation, which means that it can be used as a benchmark when evaluating a real electricity market. A good understanding of the conditions of an ideal electricity market is also valuable when analysing the operation of a real electricity market. Finally, the ideal model can be useful to simulate simple electricity markets (preferably such with short scenarios, where the assumption of perfect information is fairly reasonable).

The Environmental Impact of the Electricity Market

In an ideal electricity market, all players consider the environmental impact of their actions, as the costs of damages to the environment are directly included in the cost and benefit functions of the players. Unfortunately, reality far to seldom works like this. The reason is partly that damages to the environment can be hard to value or that the relation between a certain human activity and a certain environmental damage is unclear, and partly because it

3. In some cases we will rather have an optimisation problem of each player (so-called player problems) instead of one single scenario problem. The player problems are related to each other by one or more balance constraints which apply to the whole market.

is sometimes too simple for unscrupulous or ignorant players to neglect the environment. The players can however voluntarily reduce the environmental impact of the power system, for example if consumers require a specification how the electricity they purchase has been generated, and that they then refrain from buying generation which does not pay its environmental costs.

Generally, the initiatives taken by the players of the electricity market are insufficient, and authorities or international organisations have to intervene and prompt the development by introducing different rules. I have described most environmental rules which can be found in electricity markets and presented basic models of them. An interesting observation when comparing different rules is that they all have a potential to result in a resource utilisation which is most beneficial to the society. The prerequisite is however that the authorities choose correct values of various control parameters; hence, the question is not which rules are the best, but for which parameters it is most easy to find an optimal value.

Consequences of Uncertain Forecasts

The players in an ideal electricity market have perfect information and therefore never have to make a decision without being able to predict the consequences of their actions. This is not the case in a real electricity market, and I have identified a number of areas, where the players of the reality due to forecast uncertainty will behave differently compared to what they would have done in an ideal electricity market. The exact consequences of forecast uncertainties depend on the context—above all the time perspective in the planning where the forecast is used. If desirable, detailed models can be used, which simulate the planning process of the players, but the risk is that such models are unnecessarily complicated, without actually producing very much extra precision.

My recommendation is to use simplified models, which only simulate the consequences of forecast uncertainty, rather than simulating the entire planning process. (An example of this idea is to introduce a random deviation of the players valuation of stored energy.) The disadvantage of this kind of simplified models is that we assuredly get rid of complex calculations when solving the scenario problem, but on the other hand we are faced by the challenge of collecting appropriate input to the model.

Grid Costs

A grid is a public good, which makes it difficult to build and operate the grid in a fair manner. I have focused on the operation costs of the grid, which basically consist of the costs of the electric losses and rationing costs during those times when a part of the grid has reached is capacity limit. I have described different methods to divide the cost of losses among the grid users and suggested models to be used in electricity market simulation. The methods for congestion management which I have modelled are restricted to the market based solutions. I have also briefly discussed possible market dynamic of the grid tariffs.

11.1.2 Monte Carlo Techniques

The basic idea of a Monte Carlo simulation is seductively simple: samples are chosen until a sufficiently clear picture of the studied population has been obtained. The difficulty is to make the simulation as efficient as possible, i.e., to produce good results using as few samples as possible. Using various variance reduction techniques, it is possible to use information known in advance to increase the efficiency. Different variance reduction techniques are based on different statistical relations and hence use different kind of advance knowledge. I have described six well-known variance reduction techniques and demonstrated which kind of advance knowledge they use (and I have made some small contributions of my own concerning stratified sampling). I have also studied to which extent the six variance reduction techniques are applicable in electricity market simulation using short and long scenarios respectively.

Short Scenarios

The properties of a short scenario are quite easy to predict, because it is possible to compare directly the available generation capacity and transmission capability to the demand. Using stratified sampling, it is possible to divide the scenario population in various parts (strata) and concentrate the sampling to those scenarios, where the properties are harder to predict. This is most of all important when estimating the risk of power deficit, *LOLP*.

Another variance reduction technique which is efficient for short scenario is control variates. The difference between the model used in the analytical method of probabilistic production cost simulation (PPC) and a multi-area model of an ideal electricity market is mostly that in the latter, the operation costs will be slightly higher, due to the losses. The estimate of the expected operation cost, *ETOC*, is therefore much better if we start from the PPC model and then just estimate the extra cost caused by the losses.

I have in practical tests shown that the simulation method I suggest produces a significant efficiency gain. In the test systems recorded in chapter 9 equally good or more accurate results are obtained when using variance reduction techniques, even though the number of scenarios is 99% less compared to simple sampling. Very accurate results—the relative error is about ±0.1% concerning *ETOC*[4] and ±2% for *LOLP*[5]—are obtained by using some thousands of scenarios.

Long Scenarios

It is harder to predict the properties of a long scenario than for a short scenario, which makes it more difficult to find efficient methods to utilise variance reduction techniques. Besides, it is difficult even just to determine the probability distribution of all inputs in a long scenario—it is primarily the start and final contents of energy storage facilities, which cause problem. I have therefore not performed any extensive, practical survey of how to simulate long scenarios, but what I have presented in this dissertation is a theoretical analysis. I point out the different properties of long and short scenarios and reason about the impact on the efficiency of various variance reduction techniques. I also make a suggestion of how the theory of short scenarios can be utilised, by breaking down each long scenario into a population of short scenarios.

11.2 FUTURE WORK

The number of issues in this dissertation which I have either chosen or been forced to refer to future studies are legion and it would be little rewarding to enumerate them all here. I will rather present some general project proposals, where each project includes research around some questions which have not been answered in this dissertation.

Further Development of the Monte Carlo Technique

There are several challenges to study more closely, in order to develop the simulation technique to manage all sorts of electricity markets. The major problems are of course managing transmission limitations and how to create the largest possible variance reduction when simulating long scenarios. We cannot be satisfied with the Monte Carlo technique, until such systems can be

4. Except for the test system with a large share of hydro power; in that case the *ETOC* is so small that even a minor error in absolute terms result in a relative error around 2%.
5. With the exception of the test system including transmission limitations; even when using calculation methods which can only be applied to two-area systems, the relative error is still ±6%. Further research on simulation of transmission limitations is necessary.

mastered, because they appear in several practical applications (more about that below). There are also other challenges, which are not as fundamental, but nevertheless as interesting. I am for example thinking about how to simulate really large systems, where the strata tree becomes so huge that the computational effort to administrate the tree structure becomes troublesome in itself (see section 9.2.2). When I analysed the scenario population in section 9.2.1, I assumed that only available generation capacity, available transmission capability and load had any importance to *LOLO* and *TOC*. In a non-ideal electricity market, there might be other factors which come into play (for example, power plants with negligible operation costs might not be fully utilised due to presence of market power), which might require that the strata have to be defined differently.

Finally, there are also several minor details, where I have more or less arbitrarily chosen a solution without a closer analysis. Examples of this kind of detail issues are the choice of compromise when the Neyman allocation produces conflicting results for different result variables (see section 9.2.3) and determining correct confidence intervals for the estimate of a system index (cf. sections 8.1 and 9.3.1). A systematic analysis of which solution is the best will maybe not result in a radical reduction of the simulation time, but any achieved improvement is of course valuable. And if it would not be possible to point out any possibilities to gain efficiency then that too would be a valuable conclusion—then we would definitely know that no further studies are necessary in this particular issue.

Methods for Data Collection

In my research I have studied general designs for models of ideal and non-ideal electricity markets. I have however not performed any closer studies of how to determine numerical values of the involved model constants or how to identify the probability distributions of the scenario parameters. Obviously, without proper input even the most sophisticated electricity market model is useless; consequently, it is extremely important to develop methods to collect the necessary data and process them to fit into the desired electricity market model.

The technical data needed in the model (for example generation capacity in individual power plants, availability and generation costs) generally cause no major problems, although in today's restructured electricity market they are to some extent considered company secrets. It might however be difficult to identify the price sensitivity of the consumers, the demand for eco-labelled electricity, etc. It is probably hard to develop a general method to collect this kind of data, but even simple rules of thumb could doubtless be of value for the unhappy engineer, who has been assigned the task of compiling inputs.

In a Monte Carlo simulation it is desirable that the scenario problem is as small as possible (in order to keep the solution time short) and it is therefore appropriate to use a multi-area model. The idea of a multi-area model is to merge smaller units into larger, “equivalent” units; several power plants with similar properties can be combined into one equivalent power plant, a group of buses in the grid can be merged into an area, etc. If the model should not loose accuracy, it is important that the behaviour of the equivalents is as close as possible to the original units. It would of course be helpful if standard algorithms were developed to identify the optimal parameter values of the equivalents.

Market Dynamics and Applications

I have in this dissertation not built any dynamic electricity market models. In the future it is probably a necessary further development. It would be appropriate to combine studies of how to simulate market dynamics with case studies around some interesting issue.

An example of an application for market dynamics is to study the supply of reserve power plants.[6] It is not certain that in a competitive electricity market there will be enough reserve power to keep the risk of power deficit (*LOLP*) on a level which maximises the benefit to the society. In theory it is profitable to invest in reserve power plants [17]. However, for natural reasons a reserve power plant is only operated during shorter time periods and with very irregular intervals (years may pass between two occasions when a reserve power plant is needed), which might make investors judge the income as too uncertain to be attractive (cf. [33]). It might therefore be necessary to design rules which either subsidise reserve power plants or in some other way make investments in reserve power more attractive. A good survey of existing and conceivable methods is found in [18]. These rules should be possible to formulate as different dynamic electricity market models.

Another question involving market dynamics is about the various rules to reduce the environmental impact of the electricity market. I have in this dissertation shown that the different rules have various impacts in the short run, but also that they have market dynamic effects. To fully evaluate different environment legislations, these market dynamic effects must of course be modelled.

6. The term reserve power plant refers to power plants which are only dispatched during extreme peak load situations or when several other power plants are unavailable. The boundary between reserve power and base power (i.e., such power plants which are more continuously dispatched) is rather indefinite.

Rural Electrification

Access to electric power is probably one of the factors which has the greatest importance to our standards of living; it is hard to imagine how modern health care, communications, industrial production and education could work without electric power. Yet about 1.6 billion people completely lack access to electricity and most of them live in poor, rural areas in developing countries [148].

Research about Monte Carlo simulation of electricity markets can in several ways be beneficial when studying the possibilities of electrifying an area. For example, Monte Carlo simulation can be used to determine which costs and which reliability different options result in, as demonstrated in for example the study of Kigoma region in my licentiate thesis [7]. But to be able to evaluate all options which might be of interest, it is necessary to be able to simulate long scenarios in an efficient way, because we in many cases would like to evaluate the possibility to use dispatchable hydro power or isolated systems supplied by photovoltaics or wind power in combination with batteries. Moreover, we must be able to manage market dynamics, because the load in a recently electrified area can be expected to increase, but the rate of increase probably depends on the electricity prices paid by the consumers.

Monte Carlo simulation can also be used to study how different market designs influence the rate of expansion in a country which is only partly electrified. If the electricity market is restructured, will it then be profitable for investors to build grids in rural areas and which prices will they charge the consumers? Is it preferable to refrain from restructuring and make a vertically integrated power company responsible for all rural electrification? These kinds of questions are further examples of simulations which require modelling of market dynamics.

Grid Tariffs to Cover Cost of Losses

I was surprised when I started studying how the cost of losses was transferred to the grid users; I simply took it for granted that in centralised electricity market some kind of optimal power flow algorithm was used in the central power pool (which means that the losses are considered when deciding which power plants should be dispatched) and that bilateral electricity markets used feed-in tariffs, which reflect the marginal losses of the grid. However, it turned out that many electricity markets use post allocation of transmission losses, which I always considered as a rather inadequate method, since it is quite arbitrary.

It would be interesting to study how a single system would be affected if we switched between the different methods. How would the total surplus be affected? How important are the patterns of the electric power flows?[7] The most relevant would be to study a real system, but if it turns out to be difficult to collect the necessary data then a fictitious—but realistic—test system would at least indicate the size of the problem. It would for example be possible to get an idea of how much the social cost would increase if improper grid tariffs were chosen. It would also be possible to estimate how large impact the choice of grid tariffs could have for the surplus of individual players.

7. For example, the Nordic feed-in tariffs are partly based that the power flow in the system is generally quite predictable, because a large part of the generation resources are in the northern part of the system, whereas most of the consumption is in the south. If the power flows often and quickly switched direction between different parts of the system, feed-in tariffs could possibly become a less attractive alternative.

Appendix A
NON-LINEAR OPTIMISATION

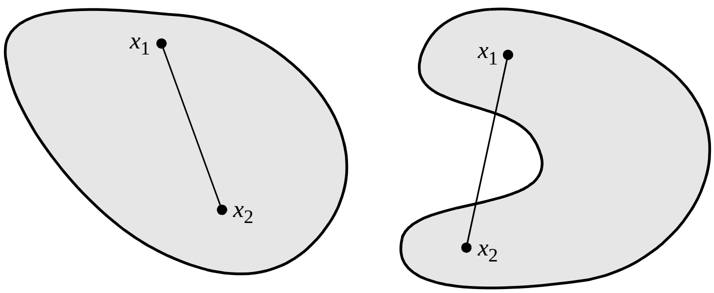
In this appendix a short summary is given of the most important definitions and theorems concerning non-linear optimisation. More detailed descriptions are found in [123, 125, 131, 132] and other textbooks on mathematical programming.

Convexity

The notion convexity is central to the mathematical programming, because convex problems do not have any local optima.

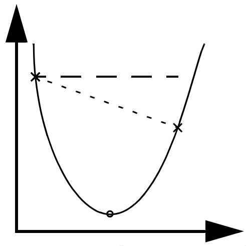
> ***Definition A.1.*** A set X is convex if for every $x_1 \in X$, $x_2 \in X$ and real number $\alpha \in [0, 1]$ it holds that $\alpha x_1 + (1 - \alpha)x_2 \in X$.

The definition has a simple geometrical interpretation: if it is always possible to draw a line between two arbitrarily chosen points within the set and all points on the line also belong to the set, then the set is convex. The idea is illustrated in figure A.1.

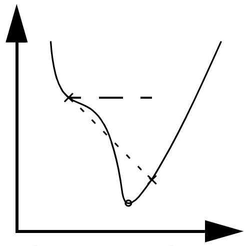


Convex *Not convex*

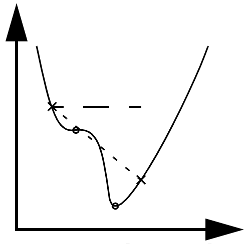
Figure A.1 *Examples of convex sets and not convex sets.*



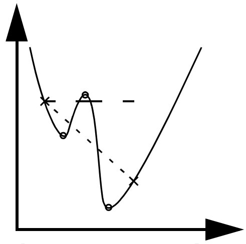
a) Convex, pseudoconvex and quasi-convex function.



b) Pseudoconvex and quasiconvex, but not convex function.



c) Quasiconvex, but neither convex nor pseudoconvex function.



d) Neither convex, pseudoconvex nor quasiconvex function.

Figure A.2 *Illustration of various types of convexity.*
× *Arbitrarily chosen point*
∘ *Extremum point*
- - - - - - *Line for testing convexity*
– — – *Line for testing quasiconvexity*

Proposition A.2. The intersection of any number of convex sets is also a convex set.[1]

Definition A.3. A function f is convex on a convex set X if for every $x_1 \in X$, $x_2 \in X$ and real number $\alpha \in [0, 1]$ it holds that

$$f(\alpha x_1 + (1 - \alpha)x_2) \leq \alpha f(x_1) + (1 - \alpha)f(x_2).$$

The function is said to be strictly convex if the left hand side is strictly less than the right hand side for all $x_1 \neq x_2$ and $\alpha \in (0, 1)$.

In the one-variable case the definition of a convex function means that a line which connects two points of the curve $f(x)$ will never be below the curve (cf. figure A.2a).

Definition A.4. A function f is concave on a convex set X if $-f$ is convex.

1. Cf. [123], p. 35.

Proposition A.5. If f_i, $i = 1, \ldots, n$, are convex functions on the convex set X then the sum $f_1 + \ldots + f_n$ is also a convex function on X.[2]

Proposition A.6. If f is a convex function on a convex set X then the set $C = \{x: x \in X, f(x) \le c\}$ is a convex set for each real number c.[3]

Sometimes it is not necessary to require convexity, but it is sufficient to assume the weaker notions of pseudoconvexity or quasiconvexity instead.

Definition A.7. A function f is pseudoconvex on a convex set X if for all $x_1 \in X$, $x_2 \in X$ such that $\nabla f(x_1)^T(x_2 - x_1) \ge 0$ it holds that $f(x_2) \ge f(x_1)$.

Definition A.8. A function f is pseudoconcave on a convex set X if $-f$ is pseudoconvex.

Definition A.9. A function f is quasiconvex on a convex set X if for all $x_1 \in X$, $x_2 \in X$ and real number $\alpha \in [0, 1]$ it holds that

$$f(\alpha x_1 + (1 - \alpha)x_2) \le \max \{f(x_1), f(x_2)\}.$$

Definition A.10. A function f is quasiconcave on a convex set X if $-f$ is quasiconvex.

Pseudoconvexity means that if we have an extremum point x_1, i.e., $\nabla f(x_1) = 0$, then $f(x_1) \le f(x_2)$ for all $x_2 \in X$, which means that x_1 is a global minimum for $f(x)$ on X. In both figure A.2a and A.2b there is only one extremum point and it corresponds to the global minimum in the interval. But in figure A.2c and A.2d there are at least one extremum point which is not a global minimum, which means that the functions in these figures are not pseudoconvex.

In the one-variable case the definition of quasiconvexity means that if we choose two arbitrary points and draw a horizontal line from the point where $f(x)$ is largest, the line will never be below the curve on the interval between the two points. In figure A.2 this requirement is fulfilled for all examples except in panel d.

Finally, it can be noted that nice convex functions are also pseudoconvex and quasiconvex:

Proposition A.11. A convex function which is differentiable is also pseudoconvex.[4]

Proposition A.12. A convex function is also quasiconvex.[5]

2. Cf. [123], p. 80.
3. See [123], lemma 3.1.2.
4. Follows by the definitions A.3 and A.7.

Optimality Conditions

Consider a general optimisation problem in the following form:

$$\text{minimise} \quad f(x) \tag{A.1}$$

$$\text{subject to} \quad h_i(x) = 0, \qquad i = 1, \ldots, n, \tag{A.1a}$$

$$g_j(x) \le 0, \qquad j = 1, \ldots, m. \tag{A.1b}$$

The optimality of this problem is given by the following two theorems:

Theorem A.13. Let λ_i denote the dual variable of the constraint $h_i(x) = 0$ and let μ_j denote the dual variable of the constraint $g_j(x) \le 0$. Let A be the index set for the active inequality constraints, i.e., the j for which $g_j(x) = 0$.[6] The necessary optimality conditions can then be written

$$\nabla f(x) + \sum_{i=1}^{n} \lambda_i \nabla h_i(x) + \sum_{j \in A} \mu_j \nabla g_j(x) = 0, \tag{A.2a}$$

$$h_i(x) = 0, \qquad i = 1, \ldots, n, \tag{A.2b}$$

$$g_j(x) = 0, \qquad j \in A, \tag{A.2c}$$

$$g_j(x) < 0, \qquad j \notin A, \tag{A.2d}$$

$$\mu_j \ge 0, \qquad j \in A, \tag{A.2e}$$

$$\mu_j = 0, \qquad j \notin A.^7 \tag{A.2f}$$

Theorem A.14. If f is pseudoconvex, g_j is quasiconvex for $j \in A$, h_i is quasiconvex when $\lambda_i > 0$ and quasiconcave when $\lambda_i < 0$ then the necessary conditions are also sufficient for an optimal solution.[8]

Most optimisation problems in this dissertation can be seen as special cases of (A.1), where the constraints mostly consist of equality conditions, but where there also is an upper and lower limit respectively for each optimisation variable, i.e., problem in the following form:

$$\text{minimise} \quad f(x) \tag{A.3}$$

$$\text{subject to} \quad h_i(x) = 0, \qquad i = 1, \ldots, n, \tag{A.3a}$$

$$\underline{x}_j \le x_j \le \overline{x}_j, \qquad j = 1, \ldots, m. \tag{A.3b}$$

5. Follows directly from definitions A.3 and A.9.
6. The set A is generally referred to as the *active set*.
7. A proof is provided in for example [123], p. 162f.
8. A proof is provided in for example [123], p. 164f.

To apply theorems A.13 and A.14, we rewrite (A.3) as

$$\begin{aligned} &\text{minimise} && f(x) && && \text{(A.4)}\\ &\text{subject to} && h_i(x) = 0, && i = 1, \ldots, n, && \text{(A.4a)}\\ & && l_j(x) = \underline{x}_j - x_j \le 0, && j = 1, \ldots, m, && \text{(A.4b)}\\ & && u_j(x) = x_j - \bar{x}_j \le 0, && j = 1, \ldots, m. && \text{(A.4c)} \end{aligned}$$

The optimality conditions can thus be written as

$$\nabla f(x) + \sum_{i=1}^{n} \lambda_i \nabla h_i(x) + \sum_{j \in A_l} \mu_j \cdot (-1) + \sum_{j \in A_u} \nu_j \cdot 1 = 0, \qquad \text{(A.5a)}$$

$$\begin{aligned} &h_i(x) = 0, && i = 1, \ldots, n, && \text{(A.5b)}\\ &x_j = \underline{x}_j, && j \in A_l, && \text{(A.5c)}\\ &x_j > \underline{x}_j, && j \notin A_l, && \text{(A.5d)}\\ &\mu_j \ge 0, && j \in A_l, && \text{(A.5e)}\\ &\mu_j = 0, && j \notin A_l, && \text{(A.5f)}\\ &x_j = \bar{x}_j, && j \in A_u, && \text{(A.5g)}\\ &x_j < \bar{x}_j, && j \notin A_u, && \text{(A.5h)}\\ &\nu_j \ge 0, && j \in A_u, && \text{(A.5i)}\\ &\nu_j = 0, && j \notin A_u. && \text{(A.5j)} \end{aligned}$$

The j:th row in (A.5a), which corresponds to the optimality condition of x_j becomes

$$\frac{\partial f(x)}{\partial x_j} + \sum_{i=1}^{n} \lambda_i \frac{\partial h_i(x)}{\partial x_j} - \mu_j + \nu_j = 0. \qquad \text{(A.6)}$$

Now notice that j cannot simultaneously belong to both A_l and A_u, because that would imply that $x_j = \underline{x}_j = \bar{x}_j$, but if the upper and lower limit of x_j coincide then x_j would be a constant and not an optimisation variable. Hence, either μ_j or ν_j must equal zero, while the other dual variable is equal to an arbitrary non-negative value. Therefore, (A.6) can be simplified to an inequality condition in those cases when an optimisation variable is equal to its upper or lower limit.

Another observation is that the functions $l(x)$ and $u(x)$ are linear; thus, they fulfil all convexity conditions. Hence, it only depends on the objective function, $f(x)$, and the equality constraints, $h(x)$ whether the problem is convex or not.

The above results can be summarised as the following theorem:

Theorem A.15. If $f(x)$ is pseudoconvex and $h_i(x)$ is quasiconvex when $\lambda_i > 0$, and quasiconcave when $\lambda_i < 0$, then the necessary and sufficient optimality conditions for a problem in the form (A.3) can be written as

$$\begin{cases} \dfrac{\partial f(x)}{\partial x_j} \geq -\sum_{i=1}^{n} \lambda_i \dfrac{\partial h_i(x)}{\partial x_j} & \text{if } x_j = \underline{x}_j \\ \dfrac{\partial f(x)}{\partial x_j} = -\sum_{i=1}^{n} \lambda_i \dfrac{\partial h_i(x)}{\partial x_j} & \text{if } \underline{x}_j < x_j < \overline{x}_j \qquad j = 1, \ldots, m \\ \dfrac{\partial f(x)}{\partial x_j} \leq -\sum_{i=1}^{n} \lambda_i \dfrac{\partial h_i(x)}{\partial x_j} & \text{if } x_j = \overline{x}_j \end{cases}$$

and

$$h_i(x) = 0, \qquad i = 1, \ldots, n,$$

$$\underline{x}_j \leq x_j \leq \overline{x}_j, \qquad j = 1, \ldots, m.$$

Appendix B
RANDOM VARIABLES

In this appendix, a short overview is given of the most important definitions and theorems concerning random variables. More detailed descriptions are found in [128, 129] and other textbooks on probability theory and random variables.

In the description below it is generally assumed that we are considering a continuous random variable X. If X is discrete instead, all integration should be replaced by summation instead.

Probability Distributions

The result of a random experiment is referred to as an *outcome*. In each trial, there is a set of possible outcomes, which is generally called the *sample space*. A random variable can be seen as a function, which associates each outcome to a numerical value. The probability distribution of the random variable states how likely different outcomes are. The probability distribution can be described in several ways:

Definition B.1. The probability that an observation of X belongs to a set $\mathcal{X}$ is given by the density function $f_X(x)$, i.e.,

$$P(X \in \mathcal{X}) = \int_{\mathcal{X}} f_X(x)dx.$$ [1]

Definition B.2. The probability that $X \leq x$ is given by the distribution function $F_X(x)$, i.e., $F_X(x) = P(X \leq x)$.

Definition B.3. The probability that $X > x$ is given by the dura-

1. It can be noted that concerning discrete random variables, $f_X(x)$ is often referred to as the probability function, because in this case $f_X(x)$ is exactly equal to $P(X = x)$. Except for this difference in the interpretation of $f_X(x)$, probability functions and density functions have the same properties.

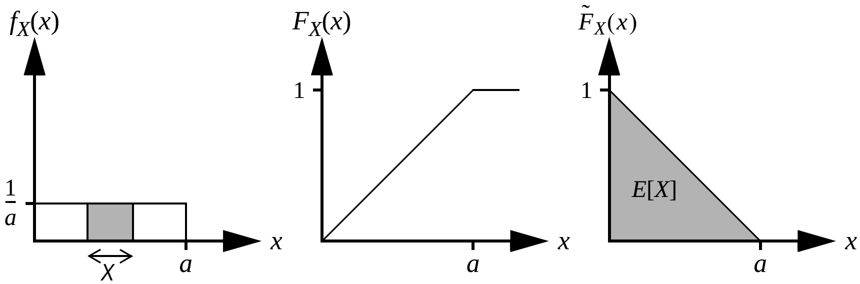


Figure B.1 *Example of density function, distribution function and duration curve of a random variable. The variable X is uniformly distributed on the interval* [0, *a*].

tion curve $\tilde{F}_X(x)$, i.e., $\tilde{F}_X(x) = P(X > x)$.[2]

Theorem B.4. The density function, distribution function and duration curve are related to each other according to

$$\tilde{F}_X(x) = 1 - F_X(x) = 1 - \int_{-\infty}^{x} f_X(t)dt = \int_{x}^{\infty} f_X(t)dt.$$ [3]

If we take the sum of two random variables, the result is a new random variable, having its own distribution function. The following applies to independent random variables:[4]

Theorem B.5. The density function of the sum, *Z*, of two independent random variables *X* and *Y* is given by convolution. The convolution formula of independent discrete random variables is written

$$f_Z(x) = \sum_i f_X(i)f_Y(x-i)$$

and for independent continuous random variables, we have

$$f_Z(x) = \int_{-\infty}^{\infty} f_X(t)f_Y(x-t)dt.$$ [5]

2. Duration curve is my own designation for this kind of function. Other authors may use terms as cumulative distribution function or reliability function.
3. The theorem follows directly by the definitions B.1-B.3.
4. See the section ***Correlations*** on page 266.
5. A proof is provided in [129], p. 70 and. p. 113 respectively.

Expectation Value and Variance

The density and distribution functions defined above describe a random variable in detail. In many cases, we have a larger practical use of an approximate description, in the form of the expected mean value. This information is obtained using the expectation value:

> ***Definition B.6.*** The expectation value of a random variable X is given by
>
> $$E[X] = \int_{-\infty}^{\infty} x f_X(x)dx.$$

Sometimes, we also need a measure of how much a single sample can deviate from the mean. This measure is either stated by the standard deviation, which is defined as the square root of the variance. The variance in its turn is defined as follows:

> ***Definition B.7.*** The variance of a random variable X is equal to the expected quadratic deviation from the expectation value, i.e.,
>
> $$Var[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2 = \\ = \int_{-\infty}^{\infty} (x - E[x])^2 f_X(x)dx.$$

The following two theorems describe how expectation value and variance are affected by linear operations:

> ***Theorem B.8.*** If a random variable X is multiplied by a scalar a then the expectation value and variance of aX are given by
>
> $$E[aX] = aE[X],$$
> $$Var[aX] = a^2\,Var[X].$$[6]

> ***Theorem B.9.*** The expectation value and variance of the sum of two random variables X and Y are given by
>
> $$E[X + Y] = E[X] + E[Y],$$
> $$Var[X + Y] = Var[X] + Var[Y] + 2Cov[X, Y],$$
>
> where $Cov[X, Y]$ is the covariance between the two variables (see definition B.11).[7]

One of the reasons that duration curves are useful is that the expectation value of the corresponding random variable can be computed by studying the

6. A proof is provided in [128], §28-29.
7. A proof is provided in [128], §28-29.

area below the duration curve:

> ***Theorem B.10.*** If $\underline{x}$ is the least possible outcome of a random variable X then the expectation value of X is given by
>
> $$E[X] = \underline{x} + \int_{\underline{x}}^{\infty} \tilde{F}_X(x)dx.$$ [8]

When studying power systems, $\underline{x}$ is often equal to zero, which, means that the expectation value equals the area of the part of the duration curve in the first quadrant. This situation is illustrated in figure B.1.

Correlations

If we consider more than one random variable it is in many cases important to know how their probability distributions are related to each other. A very important notion in this context is independent random variables. There are several definitions of the meaning of independent random variables, but briefly speaking, we may say that if X and Y are independent then knowledge of the outcome of X will not provide any information about the outcome of Y and vice versa. If we for example roll a dice twice and let X be the outcome of the first roll and Y the outcome of the second then X and Y are independent.[9] However, if X is the outcome of the first roll and Y is the sum of the two rolls then X and Y are dependent; given the outcome of X, the random variable Y will have completely different probability distributions.

A measure of how two probability distributing are correlated to each other is the covariance:

> ***Definition B.11.*** The covariance of two random variables X and Y is given by
>
> $$Cov[X, Y] = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y].$$

A related measure is the correlation coefficient:

> ***Definition B.12.*** The correlation coefficient between two random variables X and Y is given by
>
> $$\rho(X, Y) = \frac{Cov[X, Y]}{\sqrt{Var[X]Var[Y]}}.$$

The correlation coefficient will always be in the interval $[-1, 1]$. If $\rho(X, Y) = 0$, we say that X and Y are uncorrelated. Independent variables are

8. The theorem follows more or less directly from the definitions of density function, duration curve and expectation value; cf. [7], p. 115.
9. Unless there is something fishy about the dice.

always uncorrelated, but uncorrelated variables are not always independent![10] If $\rho(X, Y)$ is positive, the random variables are positively correlated, which means that if we for example know that the outcome of X was higher than the expectation value then it is more likely that Y too is higher than the expectation value. If the variables had been negatively correlated instead ($\rho(X, Y) < 0$) then a high value of X would have resulted in an increased probability of a low value for Y.

10. Consider for example a $U(-1, 1)$-distributed random variable X and $Y = X^2$. Although Y is a direct function of X (and hence clearly dependent of X), the correlation coefficient is $\rho(X, Y) = 0$.

Appendix C

RANDOM NUMBER GENERATION

Without algorithms to generate random numbers, it would not be possible to perform a Monte Carlo simulation on a computer. Several methods have been developed for generation of so-called pseudorandom numbers from a $U(0, 1)$-distribution. The term pseudorandom numbers is used to emphasise that we do not get truly random numbers, but a pseudorandom number generator is based on a seed. Given the same seed, the random number generator will produce the same sequence of numbers. A good generator will produce a long sequence before it starts repeating itself. Moreover, it is desirable that the distribution of the generated numbers is as close as possible to a uniform probability distribution, and that the correlation between the random numbers is negligible.

Software capable of producing good random numbers is available on all modern computers; therefore, there is no reason to describe the function of a random number generator in detail.[1] However, below I will briefly describe how to transform the $U(0, 1)$-distributed random numbers to random numbers of any desirable probability distribution.

Generating Random Numbers from an Arbitrary Distribution

An ordinary random number generator randomises numbers which are uniformly distributed between 0 and 1, i.e., random numbers from a $U(0, 1)$-distribution. Using the following theorem, we can obtain random numbers from most other distributions:

> ***Theorem C.1.*** If a random variable U is $U(0, 1)$-distributed then the random variable $Y = F_Y^{-1}(U)$ will have the probability distribution $F_Y(x)$.[2]

1. The reader who wants to learn more about random number generator may for example consult [47], section 3.3, or [133], chapter 2.
2. A proof can be found in for example [47], p. 43 or [133], p. 39.

$F_Y^{-1}(x)$ in the theorem is the inverse function of $F_Y(x)$. Hence, the theorem is referred to as the *inverse transform method*. It can be noted, that we might as well use the duration curve as the distribution function.

Generating Exponentially Distributed Random Numbers

Random numbers from the exponential distribution can easily be generated by applying the inverse transform method:

> ***Theorem C.2.*** If U is a $U(0, 1)$-distributed random number then Y is $E(\lambda)$-distributed if Y is calculated according to
>
> $$Y = -\frac{1}{\lambda}\ln U.$$ [3]

If a large number of exponentially distributed random numbers are needed, a logarithm calculation is necessary for each random number generated using theorem C.2. There is an alternative method too, which only requires one logarithm calculation:

> ***Theorem C.3.*** Let $U_1, \ldots, U_{2n-1}$ be independent and $U(0, 1)$-distributed random numbers. Sort the random numbers $U_{n+1}, \ldots, U_{2n-1}$ in ascending order and denote the sorted sequence $U'_1, \ldots, U'_{n-1}$. Then
>
> $$Y_k = \frac{1}{\lambda}(U'_{k-1} - U'_k)\ln\prod_{i=1}^{n} U_i, \qquad k = 1, \ldots, n,$$
>
> are independent and $E(\lambda)$-distributed.[4]

According to [133] theorem C.3 is more efficient than theorem C.2 if the number of generated random numbers, n, is between 3 and 6.

Generating Normally Distributed Random Numbers

The inverse of the distribution function, $\Phi^{-1}(x)$, does not exist for the normal distribution. However, it is possible to use an approximation of the inverse function for the transformation:

> ***Theorem C.4.*** If U is a $U(0, 1)$-distributed random number then Y is $N(0, 1)$-distributed if Y is calculated according to

3. The theorem follows directly from theorem C.1.
4. A proof is provided in for example [133], p. 68ff.

$$Q = \begin{cases} U & \text{if } 0 \le U \le 0.5, \\ 1 - U & \text{if } 0.5 < U \le 1, \end{cases}$$

$$t = \sqrt{-2\ln Q},$$

$$c_0 = 2.515517, \quad c_1 = 0.802853, \quad c_2 = 0.010328,$$

$$d_1 = 1.432788, \quad d_2 = 0.189269, \quad d_3 = 0.001308,$$

$$z = t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}$$

and finally

$$Y = \begin{cases} -z & \text{if } 0 \le U < 0.5, \\ 0 & \text{if } U = 0.5, \\ z & \text{if } 0.5 < U \le 1. \end{cases}$$ [5]

This theorem is referred to as the *approximate inverse transform method*.

Independent $N(0, 1)$-distributed random numbers can then be transformed to random numbers from a general normal distribution. It is also possible to generate random numbers with a particular correlation:

> ***Theorem C.5.*** Let $\boldsymbol{X} = [X_1, \ldots, X_n]^T$ denote a vector of independent, $N(0, 1)$-distributed elements. Let $\boldsymbol{B} = \Sigma^{1/2}$, i.e., let λ_i and g_i be the i:th eigen value and the i:th eigen vector respectively of the matrix Σ, and define matrices as follows:
>
> $$\boldsymbol{P} = [g_1, \ldots, g_n],$$
> $$\Lambda = diag(\lambda_1, \ldots, \lambda_n),$$
> $$\boldsymbol{B} = \boldsymbol{P}\Lambda^{1/2}\boldsymbol{P}^T.$$
>
> Then $\boldsymbol{Y} = \mu + \boldsymbol{BX}$ is a vector of normally distributed random numbers with the expectation value vector μ and the covariance matrix Σ.[6]

From the theorem above, we have that if we only need one random number Y from an $N(\mu, \sigma)$-distribution, then the covariance matrix is $\Sigma = \sigma^2$, which means that we use the transformation $Y = \mu + \sigma X$.

Generating Random Numbers from a Part of a Distribution

When using stratified sampling, it is necessary to be able to direct the out-

5. The theorem is a summary of the method described in [47], p. 48f.
6. The theorem is based on remark 6 in [126], p. 35.

come of a random variable to a specific interval. Concerning uniformly distributed random numbers, this is a trivial task. As random numbers with any other distribution than a $U(0, 1)$-distribution can be generated by transformation, a random number from a certain interval can be generated by first deciding which subinterval of the $U(0, 1)$-distribution will after transformation correspond to the desired interval:

> ***Theorem C.6.*** If U_0 is $U(0, 1)$-distributed then the outcome of the random variable Y, which has the distribution function $F_Y(x)$, will belong to the interval $[a, b]$, if Y is calculated according to
>
> $$\alpha = F_Y(a),$$
> $$\beta = F_Y(b),$$
> $$U = (\beta - \alpha)U_0 + \alpha,$$
> $$Y = F_Y^{-1}(U).$$ [7]

7. The theorem follows directly from theorem C.1.

Appendix D

TWO-AREA POWER SYSTEMS

In a two-area system it is quite straightforward to calculate *LOLP* analytically, even if we consider transmission losses and limitations. Given the available generation capacity in each area, $\bar{G}_1$ and $\bar{G}_2$, the density function of the load, f_D, the available transmission capability, $\bar{P}$, and the transmission losses as a function of transmitted power, $L(P)$, we can calculate the risk of power deficit. As we in practice have several possible states for $\bar{G}_1$ and $\bar{G}_2$, we are forced to repeat this calculation for each state. The final value of *LOLP* is then obtained by weighting the partial results according to their probabilities.

The tricky part of the calculation is to determine $LOLP(\bar{G}_1, \bar{G}_2)$. If we consider the D_1D_2-plane and use the set $\mathcal{L}$ to denote the scenario in which load shedding occurs, then we can calculate the risk of power deficit as

$$LOLP(\bar{G}_1, \bar{G}_2) = \iint_{\mathcal{L}} f_D(x_1, x_2)dx_1dx_2. \tag{D.1}$$

If D_1 and D_2 are independent then we may rewrite (D.1) as

$$LOLP(\bar{G}_1, \bar{G}_2) = \int_{-\infty}^{\infty} f_{D_1}(x_1)\tilde{F}_{D_2}(\bar{D}_2(x_1))dx_1, \tag{D.2}$$

where $\bar{D}_2(x_1)$ is the largest load permitted in area 2 if we should avoid power deficit, provided that the load in area 1 is equal to x_1.

The function $\bar{D}_2(x_1)$ can be divided in two parts. If $x_1 < \bar{G}_1$ then there is surplus capacity to be exported from area 1; in this case the maximal load in area 2 is larger than the area's own generation capacity. How much larger the load can be is determined by $\bar{P}_{12}(x_1)$, which is the maximal export from area 1. If sufficient generation capacity is available then $\bar{P}_{12}(x_1)$ equals the transmission capability between the areas, i.e., $\bar{P}$. Otherwise, the maximal export is equal to the unused generation capacity of area 1, i.e., $\bar{G}_1 - x_1$.

The second alternative is that $x_1 > \bar{G}_1$, which means that the load in area 2

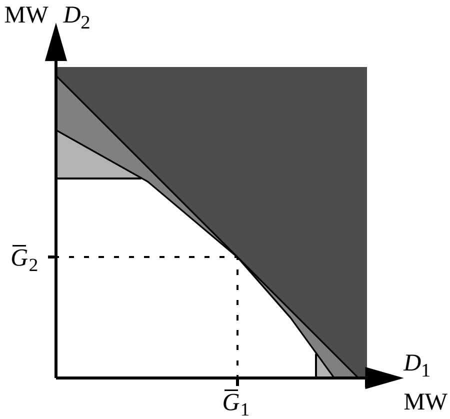


Figure D.1 *The set of scenarios. The shaded areas correspond to the scenarios where load shedding is necessary due to transmission limitations, transmission losses and capacity deficit respectively (cf. figures 9.2-9.4).*

must be less than the area's own generation capacity; there must be a surplus in area 2 to be exported to area 1. The surplus should be large enough to allow the transmitted power, $P_{21}(x_1)$, to cover both the transmission losses and the part of the load in area 1 which cannot be covered by the generation resources within the area. Therefore, the transmitted power is determined by the solution to the equation

$$P_{21}(x_1) = L(P_{21}(x_1)) + x_1 - \overline{G}_1. \tag{D.3}$$

If we have a loss function according to $L = \gamma P^2$ then (D.3) has the solution

$$P_{21}(x_1) = \frac{1}{2\gamma} - \sqrt{\frac{1}{4\gamma^2} - \frac{x_1 - \overline{G}_1}{\gamma}}. \tag{D.4}$$

Finally, we must consider that there is an upper limit to how much we can export from area 2.

All in all, we get the following function:

$$\overline{D}_2(x_1) = \begin{cases} \overline{G}_2 + \overline{P}_{12}(x_1) - L(\overline{P}_{12}(x_1)) & 0 \le x_1 \le \overline{G}_1, \\ \overline{G}_2 - P_{21}(x_1) & \overline{G}_1 \le x_1 \le \overline{G}_1 + \overline{P}_{21} - L(\overline{P}_{21}), \\ 0 & \overline{G}_1 + \overline{P}_{21} - L(\overline{P}_{21}) < x_1, \end{cases} \tag{D.5}$$

where

$$\bar{P}_{12}(x_1) = \min\,(\bar{G}_1 - x_1, \bar{P}), \tag{D.6}$$

$$\bar{P}_{21} = \min\,(\bar{G}_2, \bar{P}). \tag{D.7}$$

and where $P_{21}(x_1)$ is the solution of (D.3).

ABBREVIATIONS AND NOTATION

Units

¤	arbitrary currency unit
A	Ampere
h	hour
Hz	Hertz
kV	kilovolt
kW	kilowatt
kWh	kilowatt-hour
MW	megawatt (= 1 000 kW)
MWh	megawatt-hour (= 1 000 kWh)
TWh	terawatt-hour (= 1 000 000 kWh)

Abbreviations

Eltra	system operator of western Denmark
EMPS-model	multi-area power scheduling model (simulation model developed by SINTEF)
ETSO	European Transmission System Operators (cooperation organisation of the Europeans system operators)
EU	European Union
IEE	The Institution of Electrical Engineers (cooperation organisation for electrical engineers)
IEEE	The Institute of Electrical and Electronics Engineers (cooperation organisation for electrical engineers)
HVDC	High Voltage Direct Current
KRAV	Kontrollföreningen för ekologisk odling (Swedish organisation for eco-labelling of agricultural prod-

	ucts)
KTH	Kungliga tekniska högskolan (Royal Institute of Technology in Stockholm)
MNZD	millions New Zealand Dollars
MSEK	millions Swedish Crowns
NNP	Non-linear Network Programming
NO_x	nitrogen oxides
Nordel	cooperation organisation of the Nordic system operators
PJM	Pennsylvania-New Jersey-Maryland (electricity market in north-eastern U.S.)
PPC	Probabilistic Production Cost simulation
rms	root mean square
SEK	Swedish Crowns
SINTEF	Stiftelsen for industriell og teknisk forskning ved Norges tekniske høgskole (The Foundation for Scientific and Industrial Research at the Norwegian Institute of Technology)
SNF	Svenska Naturskyddsföreningen (Swedish Society for Nature Conservation)
SO_2	sulphur dioxide
SP	Stochastic Programming
SvK	Svenska kraftnät (Swedish system operator and owner of the national grid)
UCTE	Union for the Co-ordination of Transmission of Electricity (cooperation organisation of most European system operators except for the Nordic countries, the British isles and former Soviet Union)

Electricity Market Modelling

Index

$\hat{\Diamond}$	physical limit (for example installed capacity)
$\bar{\Diamond}$	temporary technical limitation (for example available capacity at a certain occasion)
$\Diamond^{\downarrow}$	down-regulation
$\Diamond^{\uparrow}$	up-regulation
$\Diamond^{+}$	certified
$\Diamond_a$	accounting period
$\Diamond_c$	consumer
$\Diamond_e$	electricity product
$\Diamond^F$	trading in the ahead market

$\lozenge_g$	power plant (generally thermal power plants)
$\lozenge_m, \lozenge_n$	area
$\lozenge^P$	capacity reserved for primary control
$\lozenge^R$	actual operation
$\lozenge_r$	energy limited power plant
$\lozenge_t$	time period
$\lozenge_{tot}$	sum for the entire system

Functions (scenario parameters/model constants)

$B_\lozenge(\lozenge)$	benefit function of $\lozenge$
$C_\lozenge(\lozenge)$	cost function of $\lozenge$
$D_\lozenge(\lozenge)$	damage function of $\lozenge$
$E_\lozenge(\lozenge)$	emission function of $\lozenge$
$L_\lozenge(\lozenge)$	loss function of $\lozenge$
$MB_\lozenge(\lozenge)$	marginal benefit function of $\lozenge$, i.e., $dB_\lozenge/d\lozenge$
$MC_\lozenge(\lozenge)$	marginal cost function of $\lozenge$, i.e., $dC_\lozenge/d\lozenge$
$MD_\lozenge(\lozenge)$	marginal damage function of $\lozenge$, i.e., $dD_\lozenge/d\lozenge$
$ME_\lozenge(\lozenge)$	marginal emission function of $\lozenge$, i.e., $dE_\lozenge/d\lozenge$
$ML_\lozenge(\lozenge)$	marginal loss function of $\lozenge$, i.e., $dL_\lozenge/d\lozenge$

Sets (scenario parameters/model constants)

A	set of time periods belonging to a certain accounting period
C	set of consumers
C_D	set of price insensitive consumers
C_Δ	set of price sensitive consumers
G	set of power plants (generally thermal power plants)
M	set of power plants supplying a certain electricity product)
N	set of areas
P	set of interconnections
$\mathsf{P}_{n\rightarrow m}$	set of areas m capable of importing from area n
$\mathsf{P}_{n\leftarrow m}$	set of areas m capable of exporting to area n
R	set of energy limited power plants

Other Scenario Parameters and Model Constants

A	number of accounting periods
D	load (price insensitive) [MWh/h]
E	number of electricity products

k	green certificate quota [%]
Q	inflow to energy storage during a certain time period [MWh]
R	primary control reserve [MW]
T	period duration [h]
$\mathcal{T}$	number of time periods
$\alpha_\Diamond$	constant term of the function $\Diamond$
$\beta_\Diamond$	coefficient of linear term in the function $\Diamond$
$\gamma_\Diamond$	coefficient of quadratic term in the function $\Diamond$

Result Variables and System Indices

CS	consumer’s surplus [¤/h]
EE	expected emissions [ton/h]
EG	expected generation [MWh/h]
$EENS$	expected energy not served [MWh/h]
EL	expected losses [MWh/h]
ENS	energy not served [MWh/h]
EP	expected transmission [MWh/h]
$ETOC$	expected total operation cost (i.e., $E[TOC]$) [¤/h]
G	generation (generally in thermal power plants) [MWh/h]
H	generation in energy limited power plant [MWh/h]
$LOLO$	duration of load shedding during a certain time period [% or h/year]
$LOLP$	risk of power deficit (i.e., $E[LOLO]$) [% or h/year]
M	contents of energy storage at the end of a time period [MWh]
MS	market operator’s surplus [¤/h]
P	transmission [MWh/h]
PS	producer’s surplus [¤/h]
S	spillage from energy storage during a certain time period [MWh]
TS	total surplus [¤/h]
U	unserved power [MWh/h]
TOC	total operation cost [¤/h]
W	generation in non-dispatchable power plant [MWh/h]
Δ	load (price sensitive) [MWh/h]
Γ	purchased and cancelled certificates [MWh/year]
ζ	price of emission rights [¤/ton]
Θ	emission rights purchased by players outside the electricity market [ton/year]

ι	individual emission rights value [¤/ton]
κ	individual certificate value [¤/MWh]
Λ	purchased emission rights [ton/year]
λ	electricity price [¤/MWh]
ν	energy value [¤/MWh]
Ξ	consumption subject to penalty fee [MWh/year]
ξ	certificate price [¤/MWh]
ρ	rebate [¤/MWh]
Ψ	emissions subject to penalty fee [ton/year]
Ω	unused certificates [MWh/year]

Monte Carlo Technique

Random Variables

$Cov[\Diamond_1, \Diamond_2]$	covariance between $\Diamond_1$ and $\Diamond_2$
$E[\Diamond]$	expectation value of $\Diamond$
$F_{\Diamond}$	distribution function of $\Diamond$
$\tilde{F}_{\Diamond}$	duration curve of $\Diamond$
$f_{\Diamond}$	density function of $\Diamond$
$P(\Diamond)$	probability of the event $\Diamond$
$Var[\Diamond]$	variance of $\Diamond$
$\mu_{\Diamond}$	expectation value of $\Diamond$
$\sigma_{\Diamond}$	standard deviation of $\Diamond$

Probability Distributions

$E(\lambda)$	exponential distribution with intensity λ
$N(\mu, \sigma)$	normal distribution with expectation value μ and standard deviation σ
$U(a, b)$	uniform distribution between a and b

Sampling

$a_{\Diamond}$	coefficient of variation for $\Diamond$
L	number of strata
N	number of units in population
$m_{\Diamond}$	estimate of expectation value of $\Diamond$
n	number of samples
$s_{\Diamond}$	estimate of standard deviation of $\Diamond$
ρ	relative tolerance
ω_h	weight of stratum h

Special Stratification Parameters for Simulation of Electricity Markets

$\bar{L}$	maximal total losses [MWh/h]
$\bar{U}_W$	maximal generation capacity with negligible operation cost which cannot be utilised due to congestion [MWh/h]
$\bar{U}_{WG}$	maximal generation capacity which cannot be utilised due to congestion [MWh/h]

Electrical Engineering

f	frequency [Hz]
R	speed-droop characteristics [MW/Hz]

Reliability Analysis

$MTTF$	mean-time to failure [h]
$MTTR$	mean-time to repair [h]
p	availability [%]
TTF	time to failure [h]
TTR	time to repair [h]
λ	failure rate [h^{-1}]
μ	repair rate [h^{-1}]

REFERENCES

Publications of the Author

[1] M. Amelin, “Minkostnadsproblem i nätverk med kvadratiska förluster”, technical report, A-EES-9810, Department of Electric Power Engineering,* KTH, Stockholm 1998.

[2] M. Amelin, “Förstudie till modell av en avreglerad elmarknad”, technical report, A-EES-9812, Department of Electric Power Engineering,* KTH, Stockholm 1998.

[3] M. Amelin, “The Minimum Cost Problem for Lossy Networks”, technical report, A-EES-9901, Department of Electric Power Engineering,* KTH, Stockholm 1999.

[4] M. Amelin & L. Söder, “A Fast Multi-Area Economic Hydro-Thermal Power System Model”, *31st Annual North American Power Symposium Proceedings*, San Luis Obispo 1999.

[5] M. Amelin, “A Study of Monte Carlo Methods for Simulation of Electricity Markets”, technical report, A-EES-9911, Department of Electric Power Engineering,* KTH, Stockholm 1999.

[6] M. Amelin, “A Revision of the NNP Algorithm”, technical report, A-EES-9912, Department of Electric Power Engineering,* KTH, Stockholm 1999.

[7] M. Amelin, “The Value of Transmission Capability Between Countries and Regions”, Licentiate thesis, TRITA-EES-0002, Department of Electric Power Engineering,* KTH, Stockholm 2000.

[8] M. Amelin, “Simulering av icke-ideala elmarknader. Problem och tänkbara lösningar.” Technical report, A-EES-0106, Department of Electrical Engineering, KTH, Stockholm 2001.

[9] M. Amelin & L. Söder, “On Monte Carlo Simulation of Electricity Markets with Uncertainties in Precipitation and Load Forecasts”, *2001 IEEE Porto Power Tech Proceedings*, Vol. 1, Porto 2001.

* Currently *Department of Electrical Engineering.*

[10] M. Amelin & L. Söder, “The Strata Tree: A Useful Tool for Monte Carlo Simulation of Electricity Markets”, *7th International Conference on Probabilistic Methods Applied to Power Systems Proceedings*, Vol. 2, Naples 2002.

[11] L. Söder & M. Amelin, “Efficient Operation and Planning of Power Systems”, fifth edition, course compendium, A-ETS/EES-0308b, Department of Electrical Engineering, KTH, Stockholm 2003.

[12] E. Bompard, P. Correia, G. Gross & M. Amelin, “Congestion-Management Schemes: A Comparative Analysis Under a Unified Framework”, *IEEE Transactions on Power Systems*, Vol. 18, No. 1, February 2003.

General about Electricity Markets and Power Systems

[13] I. Arciniegas, C. Barrett & A. Marathe, “Assessing the Efficiency of US Electricity Markets”, *Utilities Policy*, Vol. 11, No. 2, 2003.

[14] M. Benini, M. Marracci, P. Pelacchi & A. Venturini, “Day-ahead Market Price Volatility Analysis in Deregulated Electricity Markets”, *2002 IEEE Power Engineering Society Summer Meeting Proceedings*, Vol. 3, Chicago 2002.

[15] L. Bergman, T. Hartman, L. Hjalmarson & S. Lundgren, *Den nya elmarknaden*, SNS Förlag, Stockholm 1994.

[16] R. Billinton & R. N. Allan, *Reliability Evaluation of Power Systems*, second edition, Plenum Press, New York 1996.

[17] M. Caramanis, “Investment Decisions and Long-term Planning Under Electricity Spot Pricing”, *IEEE Transactions on Power Apparatus and Systems*, Vol. PAS-101, No. 12, December 1982.

[18] G. L. Doorman, “Peaking Capacity in Restructured Power Systems”, Doctoral dissertation 2000:100, Institutt for elkraftteknikk, Norges Tekniske-Naturvitenskapelige Universitet, Trondheim 2000.

[19] J. Endrenyi, *Reliability in Electric Power Systems*, John Wiley & Sons, 1978.

[20] J. J. Gonzalez & P. Basagoiti, “Spanish Power Exchange Market and Information System Design Concepts, and Operating Experience”, *21st International Conference on Power Industry Computer Applications Proceedings*, Santa Clara 1999.

[21] R. S. Hartman & R. D. Tabors, “Optimal Operating Arrangements in the Restructured World: Economic Issues”, *Energy Policy*, Vol. 26, No. 2, February 1998.

[22] P. Kundur, *Power System Stability and Control*, McGraw-Hill Inc. 1994.

[23] D. W. Lane, C. W. Richter Jr. & G. B. Sheblé, “Modelling and Evaluating Electricity Options Markets with Intelligent Agents”, *International Conference on Electric Utility Deregulation and Restructuring and Power Technology Proceedings*, London 2000.

[24] T. Nakashima & T. Niimura, “Market Plurality and Manipulation: Performance Comparison of Independent System Operators”, *IEEE Transactions on Power Systems*, Vol. 17, No. 3, May 2002.

[25] A. Nilsberth, “Evaluation of Electricity Contracts on a Deregulated Market”, Licentiate thesis, TRITA-EES-9901, Department of Electric Power Engineering,* KTH, Stockholm 1999.

[26] A. L. Ott, “Experience with PJM Market Operation, System Design, and Implementation”, *IEEE Transactions on Power Systems*, Vol. 18, No. 2, May 2003.

[27] H. Outhred, “The Competitive Market for Electricity in Australia: Why it Works so Well”, *Proceedings of the 33rd Annual Hawaii International Conference on Systems Sciences*, 2000.

[28] J. Persson & L. Söder, “Validity of a Linear Model of a Thyristor-controlled Series Capacitor for Dynamic Simulations”, *14th Power System Computation Conference Proceedings*, Sevilla, 2002.

[29] E. G. Read, “Transmission Pricing in New Zealand”, *Utilities Policy*, Vol. 6, No. 3, September 1997.

[30] B. Saunders & M. Boag, “NETA: A Dramatic Change”, *Electra*, No. 199, December 2001.

[31] H. G. Stoll, et. al., *Least-Cost Electric Utility Planning*, John Wiley & Sons, 1989.

[32] X. Wang, J. R. McDonald, et al., *Modern Power System Planning*, McGraw-Hill Book Company, London 1994.

[33] C. Vázquez, M. Rivier & I. J. Pérez-Arriaga, “A Market Approach to Long-Term Security of Supply”, *IEEE Transactions on Power Systems*, Vol. 17, No. 2, May 2002.

[34] A. J. Wood & B. F. Wollenberg, *Power Generation Operation and Control*, second edition, John Wiley & Sons, 1996.

[35] “Balansavtal 2004”, Svenska kraftnät 2003. [Available at http://www.svk.se]

[36] “Current State of Balance Management in Europe”, ETSO 20003. [Available at http://www.etso-net.org]

[37] “Derivatives Trade at Nord Pool’s Financial Market”, Nord Pool, November 2003. [Available at http://www.nordpool.no]

[38] “Elavbrottet 23 September 2003 – händelser och åtgärder”, Svenska kraftnät 2003. [Available at http://www.svk.se]

* Currently *Department of Electrical Engineering*.

[39] "Förutsättningar för balansansvariga", arbetsgruppsrapport, Svenska kraftnät 2003. [Available at http://www.svk.se]

[40] "Nordel Annual Report 2002", Nordel 2003. [Available at http://www.nordel.org]

[41] "Slutrapport om schablonavräkning", Svenska kraftnät 1999. [Available at http://www.svk.se]

[42] "Systemdriftavtalet 2002-05-02", Nordel 2002. [Available at http://www.svk.se]

[43] "The Nordic Power Market", Nord Pool, January 2003. [Available at http://www.nordpool.no]

[44] "Årsrapport 2002", Eltra 2002. [Available at http://www.eltra.dk]

Simulation of Electricity Markets and Power Systems

[45] H. Baleriaux, E. Jamoulle & F. L. de Guertechin, "Simulation de l'exploitation d'un parc de machines thermiques de production d'électricité couplé à des stations de pompage", *Extrait de la revue E* (édition S.R.B.E.), Vol. 5, No. 7, 1967.

[46] R. Billinton & A. Jonnavithula, "Variance Reduction Techniques for Use with Sequential Monte Carlo Simulation in Bulk Power System Reliability Evaluation", *1996 Canadian Conference on Electrical and Computer Engineering Proceedings*, Vol. 1, Calgary 1996.

[47] R. Billinton & W. Li, *Reliability Assessment of Electric Power Systems Using Monte Carlo Methods*, Plenum Press, New York 1994.

[48] R. R. Booth, "Power System Simulation Model Based on Probability Analysis", *IEEE Transactions on Power Apparatus & Systems*, Vol. PAS-91, No. 1, January/February 1972.

[49] A. Breipohl, F. N. Lee, J. Huang & Q. Feng, "Sample Size Reduction in Stochastic Production Simulation", *IEEE Transactions on Power Systems*, Vol. 5, No. 3, August 1990.

[50] G. L. Doorman, A. Gjelsvik & A. Haugstad, "Hydro Power Scheduling in Norway, Before and After Deregulation", *IEEE Stockholm Power Tech Proceedings*, Stockholm 1995.

[51] E. D. Farmer, M. J. Grubb & K. Vlahos, "Probabilistic Production Costing of Transmission -Constrained Power Systems", *10th Power Systems Computation Conference Proceedings*, Graz, 1990.

[52] N. Flatabø, A. Haugstad, B. Mo & O. B. Fosso, "Short-term and Medium-term Generation Scheduling in the Norwegian System Under a Competitive Power Market Structure", *International Conference on Electrical Power Systems Operation and Management Proceedings*, Zürich 1998.

[53] B. F. Hobbs & Y. Ji, "A Bounding Approach to Multiarea Probabilistic Production Costing", *IEEE Transactions on Power Systems*, Vol. 10, No. 2, May 1995.

[54] B. F. Hobbs, Y. Ji, C.-W. Chang, K. A. Loparo, J. Jober & M. Ohman, "An Improved Bounding-based Method for Multiarea Probabilistic Production Costing", *IEEE Transactions on Power Systems*, Vol. 11, No. 2, May 1996.

[55] S. R. Huang, "Effectiveness of Optimum Stratified Sampling and Estimation in Monte Carlo Production Simulation", *IEEE Transactions on Power Systems*, Vol. 12, No. 2, May 1997.

[56] M. E. Khan & R. Billinton, "A Hybrid Model for Quantifying Different Operating States of Composite Power Systems", *IEEE Transactions on Power Systems*, Vol. 7, No. 1, February 1992.

[57] C. Marnay & T. Strauss, "Effectiveness of Antithetic Sampling and Stratified Sampling in Monte Carlo Chronological Production Cost Modelling", *IEEE Transactions on Power Systems*, Vol. 6, No. 2, May 1991.

[58] G. C. Oliveira, M. V. F. Pereira & S. H. F. Cunha, "A Technique for Reducing Computational Effort in Monte Carlo-based Composite Reliability Evaluation", *IEEE Transactions on Power Systems*, Vol. 4, No. 4, October 1989.

[59] M. Rios, K. Bell, D. Kirschen & R. Allen, "Computation of the Value of Security", research report, Manchester Centre for Electrical Energy, Dept. of Electrical Engineering and Electronics, UMIST, Manchester 1999.

[60] M. Rios, D. Kirschen & R. Allen, "Computation of the Value of Security. volume II.", research report, Manchester Centre for Electrical Energy, Dept. of Electrical Engineering and Electronics, UMIST, Manchester 1999.

[61] C. Singh & J. Mitra, "Composite System Reliability Evaluation Using State Space Pruning", *IEEE Transactions on Power Systems*, Vol. 12, No. 1, February 1997.

[62] Suhartono, Y. Zoka & H. Sasaki, "The Evaluation of Confidence Limit on LOLP for Multi Area System", *International Conference on Electric Utility Deregulation and Restructuring and Power Technology Proceedings*, London 2000.

[63] J. Valenzuela & M. Mazumdar, "Monte Carlo Computation of Power Generation Production Costs Under Operating Constraints", *IEEE Transactions on Power Systems*, Vol. 16, No. 4, November 2001.

[64] P. D. C. Wijayatunga & B. J. Cory, "Sample Size Reduction in Monte Carlo Based Use-of-system Costing", *Proceedings of the IEE International Conference on Advances in Power System Control, Operation and Management*, November 1991, Hong Kong.

Multi-area Models

[65] J. C. Enamorado, A. Ramos & T. Gómez, "Multi-Area Decentralized Optimal Hydro-Thermal Coordination by the Dantzig-Wolfe Decomposition Method", *2000 IEEE Power Engineering Society Summer Meeting Proceedings*, Vol. 4, July 2000.

[66] D. Streiffert, "Multi-Area Economic Dispatch with Tie Line Constraints", *IEEE Transactions on Power Systems*, Vol. 10, No. 4, November 1995.

[67] G. Quaranta, "Model for Simulation of the Nordic Electricity Market", Master thesis, B-EES-0208, Department of Electrical Engineering, KTH, Stockholm 2002.

[68] G. Warland & A. T. Holen, "Transmission Operating Limits: A Decision Problem in the Environment of Uncertainty", *6th International Conference on Probabilistic Methods Applied to Power Systems Proceedings*, Funchal 2000.

[69] J. Wernérus, "A Model for Studies Related to Deregulation of an Electricity Market", Licentiate thesis, TRITA-EES-9503, Department of Electric Power Engineering,* KTH, Stockholm 1995.

[70] J. Wernérus & L. Söder, "Area Price Based Multi-Area Economic Dispatch with Transmission Losses and Constraints", *IEEE Stockholm Power Tech Proceedings*, Stockholm 1995.

[71] T. Yalcinoz & M. J. Short, "Neural Network Approach for Solving Economic Dispatch Problem with Transmission Capacity Constraints", *IEEE Transactions on Power Systems*, Vol. 13, No. 2, May 1998.

Environment

[72] C. Bernes, *Will Time Heal Every Wound?* Monitor 17, Swedish Environmental Protection Agency, Stockholm 2001.

[73] B. Boardman, J. Palmer, et al., "Consumer Choice and Carbon Consciousness: Electricity Disclosure in Europe", research report, Environmental Change Institute, Oxford 2003.

[74] D. B. Das & C. Patvardhan, "New Multi-objective Stochastic Search Technique for Economic Load Dispatch", *IEE Proceedings—Generation, Transmission and Distribution*, Vol. 145, No. 6, November 1998.

[75] M. R. Gent & J. W. Lamont, Minimum-Emission Dispatch, *IEEE Transactions on Power Apparatus and Systems*, Vol. PAS-90, No. 6, November/December 1971.

[76] M. Hindsberger, M. Hein Nybroe, H. F. Ravn & R. Schmidt, "Co-existence of Electricity, TEP, and TGC Markets in the Baltic Sea Region", *Energy policy*, Vol. 31, No. 1, January 2003.

* Currently *Department of Electrical Engineering*.

[77] C.-M. Huang & Y.-C. Huang, "A Novel Approach to Real-Time Economic Emission Power Dispatch", *IEEE Transactions on Power Systems*, Vol. 18, No. 1, February 2003.

[78] S. R. Huang, K. Y. Hung & Y. H. Chen, "Emission Control Research of Spot Markets for Separate Generation Systems", *IEE Proceedings—Generation, Transmission and Distribution*, Vol. 147, No. 6, November 2000.

[79] *Our Common Future*, report from the World Commission on Environment and Development, Oxford University Press, Oxford 1987.

[80] J. Lemming, "Financial Risks for Green Electricity Investors and Producers in a Tradable Green Certificate Market", *Energy policy*, Vol. 31, No. 1, January 2003.

[81] P. E. Morthorst, "A Green Certificate Market Combined with a Liberalised Power Market", *Energy policy*, Vol. 31, No. 13, October 2003.

[82] P. E. Morthorst, "The Development of a Green Certificate Market", *Energy policy*, Vol. 28, No. 15, December 2000.

[83] R. T. Watson (red.), *Climate Change 2001: Synthesis Report*, Cambridge University Press, Cambridge 2001. [Available at http://www.ipcc.ch]

[84] M. Voogt, M. C. Boots, G. J. Shaeffer & J. W. Martens, "Renewable Electricity in a Liberalised Market – The Concept of Green Certificates", *Energy & Environment*, Vol. 11, No. 1, January 2000.

[85] "Acid Rain Program: 2001 Progress Report", United States Environmental Protection Agency, 2001. [Available at http://www.epa.gov]

[86] "Det här är elcertifikatsystemet", Statens energimyndighet 2003. [Available at http://www.stem.se]

[87] *Official Journal of the European Union*, C 125 E, 27 May 2003. [Available at http://europa.eu.int/eur-lex]

[88] "Key World Energy Statistics", International Energy Agency, 2003. [Available at http://www.iea.org]

Short-term and Long-term Planning in Electricity Markets

[89] L. F. Escudero, J. L. de la Fuente, C. Garcia & F. J. Prieto, "Hydropower Generation Management under Uncertainty via Scenario Analysis and Parallel Computations", *IEEE Transactions on Power Systems*, Vol. 11, No. 2, May 1996.

[90] S. Feltenmark, "On Optimization of Power Production", Doctoral dissertation, TRITA-MAT 97-OS1, Department of Mathematics, KTH, Stockholm 1997.

[91] S. Feltenmark, R. Halldin, J. Holst & J. Rappe, "A model for Seasonal Planning in a Hydro-thermal Power System", research report, TRITA/MAT-00-OS9, Department of Mathematics, KTH, Stockholm 2000.

[92] N. Gröwe-Kuska, H. Heitsch & W. Römisch, “Scenario Reduction and Scenario Tree Construction for Power Management Problems”, *2003 IEEE Bologna Power Tech Proceedings*, Bologna 2003.

[93] K. Høyland & S. W. Wallace, “Generating Scenario Trees for Multistage Decision Problems”, *Management Science*, Vol. 47, No. 2, February 2001.

[94] O. Nilsson, “Short Term Scheduling of Hydrothermal Power Systems with Integer Hydro Constraints”, Doctoral dissertation, TRITA-EES-9703, Department of Electric Power Engineering,* KTH, Stockholm 1997.

[95] M. V. F. Pereira & L. M. V. G. Pinto, “Stochastic Optimization of a Multireservoir Hydroelectric System: A Decomposition Approach”, *Water Resources Research*, Vol. 21, No. 6, June 1985.

[96] L. Söder, “Integration Study of Small Amounts of Wind Power in the Power System”, research report, TRITA-EES-9401, Department of Electric Power Engineering,* KTH, Stockholm 1994.

[97] B. Vitoriano, S. Cerisola & A. Ramos, “Generating Scenario Trees for Hydro Inflows”, *6th International Conference on Probabilistic Methods Applied to Power Systems Proceedings*, Vol. 2, Funchal 2000

Grid costs

[98] L. Bertling, “Reliability Centred Maintenance for Electric Power Distribution Systems”, Doctoral dissertation, TRITA-ETS-2002-01, Department of Electrical Engineering, KTH, Stockholm 2002.

[99] J. Bialek, “Topological Generation and Load Distribution Factors for Supplement Charge Allocation in Transmission Open Access”, *IEEE Transactions on Power Systems*, Vol. 12, No. 3, August 1997.

[100] J. Bialek, M. Hartley & S. Topping, “Average Zonal Transmission Losses”, *Power Engineer*, Vol. 17, No. 5, October-November 2003.

[101] A. J. Conejo, J. M. Arroyo, N. Alguacil & A.L. Guijarro, “Transmission Loss Allocation: A Comparison of Different Practical Algorithms”, *IEEE Transactions on Power Systems*, Vol. 17, No. 3, August 2002.

[102] F. D. Galiana, A. J. Conejo & I. Kockar, “Incremental Transmission Loss Allocation Under Pool Dispatch”, *IEEE Transactions on Power Systems*, Vol. 17, No. 1, February 2002.

[103] C. Linke, “Influence of Connection Tariffs on System Performance”, Master thesis, B-EES-9806, Department of Electric Power Engineering,* KTH, Stockholm 1998.

[104] M. Meisingset & Ø. Breidablik, “A Method to Determine Charging Principles for Losses in the Norwegian Main Grid”, *13th Power System Computation Conference Proceedings*, Trondheim, 1999.

* Currently *Department of Electrical Engineering*.

[105] J. R. Saenz, P. Eguia, J. L. Berasategui, J. Marín & J. Arceluz, "Allocating Distribution Losses to Customers Using Distribution Factors", *2001 IEEE Porto Power Tech Proceedings*, Vol. 1, Porto 2001.

[106] R. S. Salgado, C. F. Moyano & A. D. R. Medeiros, "Reviewing Strategies for Active Power Transmission Loss Allocation in Power Pools", *Electrical Power & Energy Systems*, Vol. 26, No. 2, February 2004.

[107] E. de Tuglie & F. Torelli, "Nondiscriminatory System Losses Dispatching Policy in a Bilateral Transaction-Based Market", *IEEE Transactions on Power Systems*, Vol. 17, No. 4, November 2002.

[108] L. J. de Vries, "Capacity Allocation in a Restructured Electricity Market: Technical and Economic Evaluation of Congestion Management Methods on Interconnectors", *2001 IEEE Porto Power Tech Proceedings*, Vol. 1, Porto 2001.

[109] *Official Journal of the European Union*, L 176, 15 July 2003, p. 4. [Available at http://europa.eu.int/eur-lex]

[110] "Congestion Management in the Electric Power System", from *Nordel Annual Report 2000*, Nordel 2001. [Available at http://www.nordel.org]

[111] "Nordisk tariffharmonisering", working group report, Nordel 2000. [Available at http://www.nordel.org]

[112] "Prislista för stamnätet 2004", Svenska kraftnät 2003. [Available at http://www.svk.se]

Market Power

[113] C. A. Berry, B. F. Hobbs, W. A. Meroney, R. P. O'Neill & W. R. Stewart Jr., "Understanding How Market Power Can Arise in Network Competition: A Game Theoretic Approach", *Utilities policy*, Vol. 8, No. 3, September 1999.

[114] J. Contreras, M. Klusch & J. B. Krawczyk, "Numerical Solutions to Nash-Cournot Equilibria in Coupled Constraint Electricity Markets", *IEEE Transactions on Power Systems*, Vol. 19, No. 1, February 2004.

[115] C. J. Day, B. F. Hobbs & J.-S. Pang, "Oligopolistic Competition in Power Networks: A Conjectured Supply Function Approach", *IEEE Transactions on Power Systems*, Vol. 17, No. 3, August 2002.

[116] A. Halseth, "Market Power in the Nordic Electricity Market", *Utilities policy*, Vol. 7, No. 4, December 1998.

[117] D. S. Kirschen, "Market Power in the Electricity Pool of England and Wales", *2001 IEEE Power Engineering Society Winter Meeting Proceedings*, Vol. 1, Columbus 2001.

[118] M. de Luján Latorre & S. Granville, "The Stackelberg Equilibrium Applied to AC Power System—A Non-Interior Point Algorithm", *IEEE Transactions on Power Systems*, Vol. 18, No. 2, May 2003.

[119] D. A. Lusan, Z. Yu & F. T. Sparrow, "Market Gaming and Market Power Mitigation in Deregulated Electricity Markets", *1999 IEEE Power Engineering Society Winter Meeting Proceedings*, Vol. 2, New York 1999.

[120] J. Villar & H. Rudnick, "Hydrothermal Market Simulator Using Game Theory: Assessment of Market Power", *IEEE Transactions on Power Systems*, Vol. 18, No. 2, May 2003.

[121] Z. Yu, "A Market Power Model with Price Caps and Compact DC Power Flow Constraints", *International Journal of Electrical Power & Energy Systems*, Vol. 25, No. 4, may 2003.

[122] "Appendix E. Standard Market Design and Trading Strategies Encountered in the Independent System Operators", Docket No. RM01-12-000, Federal Energy Regulatory Commission, 31 July 2002. [Available at http://www.ferc.gov/industries/electric/indus-act/smd/nopr.asp]

Mathematics

[123] M. S. Bazaraa, H. D. Sherali & C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*, second edition, John Wiley & Sons, 1993.

[124] J. R. Birge & F. Louveaux, *Introduction to Stochastic Programming*, Springer-Verlag, New York 1997.

[125] S. P. Bradley, A. C. Hax & M. H. Wright, *Applied Mathematical Programming*, Addison-Wesley Publishing Company, 1977.

[126] P. J. Brockwell & R. A. Davis, *Time Series: Theory and Methods*, second edition, Springer-Verlag, New York 1991.

[127] W. G. Cochran, *Sampling Techniques*, third edition, John Wiley & Sons, 1977.

[128] B. V. Gnedenko, *The Theory of Probability*, Chelsea Publishing Company, New York 1962.

[129] G. R. Grimmet & D. R. Stirzaker, *Probability and Random Processes*, second edition, Oxford University Press, Oxford 1992.

[130] H. Kumamoto, K. Tanaka, K. Inoue & E. J. Henley, "Dagger-Sampling Monte Carlo for System Unavailability Evaluation", *IEEE Transactions of Reliability*, Vol. R-29, No. 2, June 1980.

[131] D. G. Luenberger, *Linear and Nonlinear Programming*, second edition, Addison-Wesley Publishing Company, 1989.

[132] S. G. Nash & A. Sofer, *Linear and Nonlinear Programming*, McGraw-Hill, 1996.

[133] R. Y. Rubinstein, *Simulation and the Monte Carlo Method*, John Wiley & Sons, 1981.

[134] L. Råde & B. Westergren, *Mathematics Handbook for Science and Engineering*, third edition, Studentlitteratur, Lund 1995.

[135] T. Sigurd, “Jämförelse av metoder för att minimera kostnaden i nätverk med förluster”, Master thesis, E244, Department of Mathematics, KTH, Stockholm 2001.

Economy

[136] T. Groves & J. Ledyard, “Optimal Allocation of Public Goods: A Solution to the ‘Free Rider’ Problem”, *Econometrica*, Vol. 45, No. 4, May 1977.

[137] M. L. Katz & H. S. Rosen, *Microeconomics*, third edition, Irwin/McGraw-Hill, 1998.

[138] P. D. Klemperer & M. A. Meyer, “Supply Function Equilibria in Oligopoly Under Uncertainty”, *Econometrica*, Vol. 57, No. 6, November 1989.

[139] D. G. Luenberger, *Microeconomic Theory*, international edition, McGraw-Hill, 1995.

[140] H. R. Varian, *Microeconomic Analysis*, third edition, W. W. Norton & Company, 1992.

Newspapers, etc.

[141] T. Ackermann & D. Müller, “Auckland unplugged—or the story of a blackout”, *Electric Light & Power*, Vol. 76, No. 11, November 1998, s. 20-23.

[142] P. Englund, “Det stora huset i världens mitt”, *Brev från nollpunkten*, Bokförlaget Atlantis, Stockholm 1996.

[143] L. Eriksson, “Nättariffen stryper de nya kraftverken”, *Dagens Industri*, January 12, 2004, s. 4.

[144] J. Massy, “Green credits auction: Rising value of green power”, *Wind Power Monthly*, February 2003, s. 44.

[145] K. Ross, “Call for green power”, *Wind Power Monthly*, July 2003, s. 41.

[146] “Kväveoxidavgifter: Marginell minskning av utsläppen”, *Energimagasinet*, No. 4, 2001, s. 17.

[147] *IEEE Power Engineering Review*, Vol. 22, No. 8, August 2002.

[148] “Energy and poverty”, *World Energy Outlook 2002*, IEA 2002. [Available at http://www.iea.org]

INDEX