Competition and Welfare:

The Norwegian Experience

Domestic air fares 2000 – 2006
jan 98 = 100

- Subindex for air fares
- Consumer price index
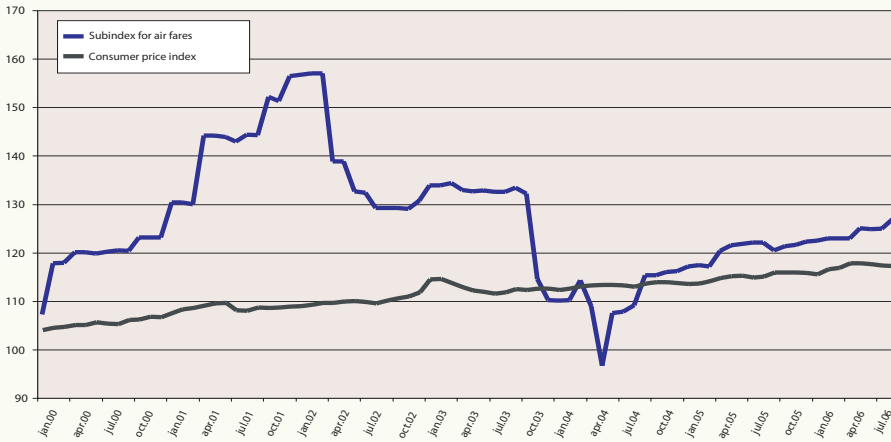
Edited by Lars Sørgard

Konkurransetilsynet
Norwegian Competition Authority

Preface

In many industries, in Norway as well as in other countries, we observe that competition plays a more important role today than in the 70s or in the 80s. Has such a transition towards competition improved welfare? This is the broad question we ask in this book. We describe the experience in seven Norwegian industries, and we draw some lessons from what we observe.

This book is financed by the Ministry of Government Administration and Reform, and we appreciate very much their support. We would also like to thank all the authors who have contributed; without you we would not have a book at all. The views expressed by the authors are not necessarily shared by neither the Norwegian Competition Authority nor the Ministry of Government Administration and Reform.

Bergen, November 2006

Knut Eggum Johansen
Director General

COMPETITION AND WELFARE: THE NORWEGIAN EXPERIENCE

Table of content:

The Contributors

Torstein Bye is director of research at the research department of Statistics Norway, and a part-time professor of Energy Economics at the Norwegian University of Life Sciences. Most of his scientific work is in the field of energy and environmental economics, general-equilibrium modelling and microeconometrics.

Dag Morten Dalen is professor in economics at BI Norwegian School of Management. His main fields of research and interest are industrial organization, regulatory economics and contracting.

Øystein D. Fjeldstad is professor in economics and Telenor Chair of International Strategy and Management at BI Norwegian School of Management. His main research is on models of value creation, cooperation and competition in knowledge and network industries. In recent years his empirical focus has been on telecommunication and banking.

Øystein Foros is associate professor in economics at the Norwegian School of Economics and Business Administration. His main research fields are industrial organization, regulatory economics and telecommunication.

Einar Hope is professor emeritus in economics at the Norwegian School of Economics and Business Administration. He was director general of the Norwegian Competition Authority 1995-99. His main research fields are energy economics, industrial organization and competition policy.

Tommy Staahl Gabrielsen is professor in economics at the Department of Economics, University of Bergen. His fields of interest and research are industrial organization in general with special focus on vertical relations and pricing in oligopolistic markets.

Hans Jarle Kind is associate professor in economics at the Norwegian School of Economics and Business Administration. His main research fields are international economics, industrial organization and tele- and media economics.

Espen R. Moen is professor in economics at BI Norwegian School of Management and research fellow in CEPR (Centre for Economic Policy Research). His main research interests are within labour economics and industrial organization, particularly search theory, contract theory, and regulation. He has analyzed optimal regulation of several sectors, including telecommunication, agriculture, electricity and publishing.

Christian Riis is professor in economics at BI Norwegian School of Management. His main research fields are industrial organization, contract theory and auctions.

Frode Steen is professor in economics at the Norwegian School of Economics and Business Administration, and a research fellow in Centre for Economic Policy Research (CEPR). His main research field is empirical industrial organization, with empirical studies of markets such as the airline market, the cement market and the grocery market.

Steinar Strøm is professor in economics at the University of Oslo, Norway (1985-) and the University of Turin, Italy (2003). Research fields are microeconometrics, labour supply and taxation, pension and retirement, and industrial organization.

Lars Sørgard is chief economist at the Norwegian Competition Authority, and a former professor in economics at the Norwegian School of Economics and Business Administration. His fields of interest and research are industrial organization, competition policy and competition strategy.

Helge Østbye is professor at the Department of Information Sciences and Media Studies at University of Bergen. His main research fields are Norwegian and European media structure and media policy, as well as recent media history.

Lars Sørgard

Some lessons for public policy

1.1 An introduction

From text books in economics we learn that competition can be a powerful way to ensure an efficient allocation of resources. Competition helps bringing prices down to reflect relevant costs, forces firms to reduce inefficiency and thereby lower costs, and promotes R&D and thereby the introduction of new and better products.

However, we also learn about market failures. There are numerous reasons why competition does not work in a proper way from the society's point of view. For example, barriers to entry may imply that the existing firms are able to exploit market power. This may lead to an inefficient use of resources, and the consumers can be hurt by such a market failure through high prices and lower quality products. In fact, firms can behave in a strategic way by, for example, erecting barriers to entry and thereby earn high profits. This is illustrated in Michael Porter's book 'Competitive Strategy' from 1980, which can be seen as a recipe for how firms could act strategically in order to earn high profits.

The prospects for earning profits are in itself a good thing, though. Firms should make an effort to produce products with a high quality at low costs, and the prospects for earning profits is an important incentive for doing so. On the other hand, in some instances firms can also earn profits by setting a high price and producing low cost products with a correspondingly low quality. How do we ensure that firms have the right incentives seen from the consumers' and society's points of view when they strive for earning higher profits?

Competition can thus be a powerful instrument for enhancing welfare, but market failures such as strategic behavior by firms may be detrimental to welfare. Deregulation – interpreted in a broad way to encompass all shifts towards a more market oriented regime in an industry – has therefore a potential for improving welfare. But how do we avoid those possible market failures? The obvious answer is to have a proper regulation in force. This could be either ex ante or ex post regulation. Ex ante regulation is about how to design the market structure and to directly regulate firms' behavior. For example, the auction design in the spot market for electricity and the price regulation in telecom can be seen as ex ante regulation. Ex post regulation is about how to intervene against firm's behavior. Competition policy is the most important regulatory instrument in that respect. It enables the authorities to intervene against price conspiracies and the abuse of dominant positions as well as banning proposed mergers between firms.

Although theory can guide us towards the optimal deregulation, the optimal mix between competition and regulation must be tailor-made to the industry in question. Do we succeed in finding the right balance between competition and regulation? In most industries it turns out that there are lessons to be learnt. Some failures do take place and industries do not operate in an optimal way. In some cases the problem is lack of public intervention, while in other cases the problem is too detailed regulation or simply the wrong type of regulation.

The purpose of this book is to describe the experiences from some Norwegian industries, and to draw some lessons. There are lessons to be learned both from observing things that went wrong and cases that are considered to be successful. In this introduction, we first briefly describe the seven industries that are presented in this book. Thereafter, we discuss some of the lessons that can be drawn from these experiences for public policy.

1.2 Seven Norwegian industries

In **chapter 2** Torstein Bye and Einar Hope describe the liberalisation of the Norwegian market for **electricity**. In 1991, an Energy Act changed the industry quite dramatically. The state-owned vertical integrated Statkraft was split into two entities, one generating company (Statkraft) and one transmission company (Statnett). The transmission company was subject to natural monopoly regulations with a common carriage principle, while Statkraft had to compete with other generating companies on an open spot market that was established. The Norwegian electricity market was subsequently integrated with the Swedish, Finnish and the Danish markets.

Prior to the liberalisation we observed inefficiencies in this industry. There were no mechanism that spurred the firms to minimise costs, no open spot market that could ensure that all the available water was used for producing electricity, and there existed large differences between the prices paid by different consumer groups, creating a large potential for misallocation of resources. The authors argue that the Energy Act led to better resource allocation and benefits for the consumers. In particular, the spilling of water observed prior to liberalisation vanished and prices of electricity fell. The differences in prices between consumer groups diminished, leading to a better allocation of resources. Investments declined, making reallocation to more profitable investment alternatives in other industries possible.

In **chapter 3**, Dag Morten Dalen and Steinar Strøm describe the changes in the **pharmaceutical** industry following the new pharmacy law in 2001. Until 2001 the industry was heavily regulated, including a detailed regulation of the location of pharmacies and where drugs could be sold as well as various kinds of price regulation of drugs. The new law implied a less restrictive regulation. In particular, restrictions on ownership of pharmacies were abolished and just after the law was passed the restrictions on location of new pharmacies were also removed. It is pointed out that the new pharmacy law was beneficial to the consumers. The number of pharmacy stores increased rapidly, opening

hours are longer, and a selection of over-the-counter drugs are now available in super-markets and gas stations.

The new law triggered a large number of acquisitions in this particular industry. A fragmented retail structure was soon replaced with a market structure with three vertical integrated pharmacy chains. In other markets this could raise concerns about market power, but this is not necessarily so in this industry. First, higher concentration leads to more buyer power which can trigger more competition between producers. Second, the concern about higher retail prices is not that important in this industry since low retail prices in general should be ensured by regulation and not by retail competition. On the other hand, there are numerous ways to implement a price regulation. The authors argue that the chosen price regulation in this particular industry is to the benefit of the retail chains, and that there are other ways to price regulate that to a larger extent would benefit consumers and the society.

In **chapter 4** Øystein Fjeldstad, Espen R. Moen and Christian Riis describe and discuss the development in the **telecommunications** market. As in many other countries, the telecom sector in Norway has gone through a remarkable transition over the last decade. The liberalisation has closely followed the time schedule set by the European Union. Open Network Provision principles have been applied to both fixed and mobile services since 1993. The fixed-line telephony market was opened for competition in 1998. The state-owned telephone monopolist Telenor (then Televerket) was turned into a state owned limited liability enterprise in 1994, deregulated in 1998 and listed and partly privatised in December 2000. In the mobile market we have an asymmetric duopoly with Netcom and Telenor as nation-wide operators.

The authors find that the Norwegian telecommunication market is well developed by international standards, with high levels of communication service adoption, a high penetration ratio for mobile communication, and with fairly low prices. The basic idea behind the regulatory regime in Norway has been to stimulate competition in end user markets through wholesale market regulation. The regulator has made an effort in striking a balance between on the one side forcing prices down through regulation and on the other hand allowing agents to establish commercial contracts. The authors argue that the overall picture is that Norwegian Post and Telecommunication Authority has been ambitious in supervising, but not in sanctioning, network owners with market power.

In **chapter 5** Øystein Foros, Hans Jarle Kind and Helge Østbye describe the media market, with a particular emphasis on the deregulation of the **radio** market. In Norway, as in many Western European countries, both radio and TV broadcasting were initially public service monopolies. Gradually this changed, and some chosen private firms were allowed to enter this market. However, those channels were not allowed to strive for pure commercial goals. The government imposed some requirements, for example concerning which types of programs a channel could send. In both the TV and the radio market the entrants were selected through beauty contests. In the radio market the channel Kanal 24 won such a beauty contest in 2002 as the fourth FM network, with a start up from

January 2004. They replaced P4, whose ten year license on the fourth FM license expired in January 2004. Later on P4 won the contest for the fifth FM license, implying that Kanal 24 and P4 competed head to head from January 2004.

It is shown that cultural aspects, such as a goal to preserve diversity, did heavily influence the way this industry was regulated. However, it is also pinpointed that deregulation and even higher market concentration has ambiguous effects on diversity as defined from an economic perspective. The authors argue that competition from nationwide commercial radio channels partly explains the diversity within the public channel, with three highly diversified channels. It turned out that Kanal 24 did not succeed in capturing a large audience. One main reason for this was that P4 made some clever strategic moves at the time of the entry of Kanal 24. It is argued that P4 exploited the fragmentation of responsibility between different regulatory bodies. More specifically, P4 was allowed to start broadcasting in the fifth network one week before Kanal 24 entered, and by doing so P4 continued to serve their listeners while Kanal 24 entered into a 'cold' FM network. P4 succeeded in maintaining their position as the largest non-public radio channel in Norway, and Kanal 24 had chosen a profile closer to P4's profile than they would have chosen if they had anticipated P4's reentry. Furthermore, the authors argue that a merger between P4 and Kanal 24 – which has been proposed – could lead to increased diversity. On the other hand, in a democracy it would certainly raise political concerns if one owner had more or less monopolized the nation-wide commercial radio market.

In **chapter 6** Tommy Staahl Gabrielsen discusses the development in the **grocery** sector during the last two decades. In contrast to the previous mentioned industries, the grocery sector has not been heavily regulated and there has not been any major changes in the regulatory regime. Despite this, we have observed a rather dramatic structural change in this particular industry from the early 80s and onwards. The retail sector was in the early 80s very fragmented, consisting of many small and independent retail outlets. The producer and wholesale markets, on the other hand, were heavily concentrated. The next decade a large number of retailers joined various retail chains, and we also observed closer vertical integration between retail chains and wholesalers. Currently, four retail groups control more than 98 % of the grocery market, and the wholesalers have lost their previous powerful position as price setters. Moreover, each retail group typically sets prices at a national level with limited scope for local competition.

A major question is whether the observed structural change has been beneficial for consumers and the society. The author concludes that the restructuring did create considerable efficiency gains. For example, the wholesale activity now is probably organised in a much more cost-efficient way than before, and the large retail chains has exploited its market power to trigger more fierce competition between producers. No doubt, at least parts of these gains have been passed on to consumers through lower prices. However, the question is whether the pendulum has swung too far so that we at present are stuck with a structure that dampens competition and blocks innovative activities. The author argues that it might be the case that we have gone too far towards a rather limited product range, a development that can be beneficial for producers and retail chains but detri-

mental to the interest of the consumers. Moreover, there are several characteristics and practices in the market that may facilitate collusion between the major retail chains.

In **chapter 7** Frode Steen and Lars Sørgard present a long run perspective on the **cement** market. At the end of World War I three firms entered the Norwegian cement market, which triggered a price war on cement. Consequently, these three firms formed a price cartel in 1923. The cartel lasted until 1968 when the firms merged to monopoly, followed by a gradual reduction in capacity the next decades. The price cartel operated a joint sales office for domestic sales. Sales in the domestic market were allocated according to capacity; the larger a firm's capacity the larger the fraction of domestic sales. This market sharing mechanism produced some potentially perverse incentives for each firm. By investing in capacity each firm could increase its sale in the cartelized domestic market, and then sell the remaining production in the export market.

In the 20s and 30s it operated as a normal price cartel, with no capacity expansion effect. However, after World War II this changed dramatically. Each of the three firms made large investments in capacity. From each firm's perspective this made sense, since each firm captured a larger share of the monopolised domestic market than they otherwise would have done. From a joint perspective this did not make sense for the firms, and we observed a large excess capacity in the Norwegian cement industry. The capacity that exceeded the domestic consumption could be used to produce for exports, but the price in the world market was low and did not cover total costs. The price cartel had then a two-fold negative impact on welfare. First, it had a traditional price raising effect in the domestic market. Second, it led to overinvestment in capacity and thereby to a cost increase.

Finally, in **chapter 8** Frode Steen and Lars Sørgard describe the development in the **airline** industry since the deregulation in 1994. From April 1994, domestic firms were allowed to enter on the largest domestic airline routes in Norway. The airlines SAS and Braathens were the only firms active in the market until summer 1998, when Color Air entered four domestic routes. Color Air exited the market 14 months later, after incurring losses amounting to 400 MNOK. In 2001, Braathens became a failing firm, and it was acquired by SAS. In October 2001, we had then returned to monopoly in the Norwegian airline industry. In the summer of 2002, the airline company Norwegian entered on four domestic routes, and later on expanded to other domestic routes as well as to direct routes to Europe from several Norwegian airports.

The authors argue that it took eight years before the deregulation was successful. During the initial phase SAS and Braathens competed on capacities, and consumers did only to a limited extent benefit through lower prices. For large corporate customers this changed in the late 90s and onwards, when they were offered large rebates on the list prices. After the return to monopoly in 2001, a ban on SAS' frequent flyer program in the domestic market triggered entry by Norwegian. This entry led subsequently to price competition between SAS and Norwegian. From October 2003 and onwards, price competition became quite fierce, and Norwegian suffered losses. The new competition act came into

force in May 2004, and approximately at that time period SAS partly reversed its aggressive price policy in the Norwegian market.

1.3 What have we learnt?

The seven industries are no doubt distinctly different, making it difficult to draw some general lessons. However, although ambiguities definitely exists, some tendencies can be observed. Each industry can in different ways be used to pinpoint some important lessons.

Lesson 1: Price competition tends to reduce costs

Deregulation has in various ways led to competition, and in some industries price competition. In those industries where we have observed price competition, it seems as this kind of competitive process has in most cases led to costs savings. There are three examples from the industries in question:

- In the electricity market, liberalisation led to lower prices. The return on new investments became very low, and return on capital in this industry became lower than in other industries. Thus, in the period after 1991 we have not seen any significant investments in new generation capacity in the Norwegian electricity industry. Moreover, spilling of water – which amounted to 5-6 % of annual production – was eliminated after 1991. To illustrate the importance of such an elimination of waste, an investment of such a magnitude would cost approximately 2 BNOK annually. There has thus been large costs savings to the benefit of the society following the liberalisation of the power market.
- In the airline market, SAS implemented in 2002 a large cost saving plan. The plan was no doubt triggered by the prospects for tougher competition from low-cost airlines. Four years later the company had saved 14 BNOK annually, and it plans to save additionally 2 BNOK annually. This will imply that the firm's unit cost has fallen with approximately 30 %.
- In the grocery sector the buyer power on the retail level triggered competition between producers. The retail chains negotiated directly with the producers, and the role of the wholesalers changed. In particular, we observed quite large changes in the logistics of the grocery sector. It is difficult to measure the exact effect of such a change, but even small changes will have large absolute effects in such a large sector. For example, a 2-4% cost reduction in this sector leads to approximately 2-4 BNOK in savings annually.

On the other hand, firms might compete along other dimensions. For example, firms might compete on capacities rather than prices. In two of the seven industries in question we observed competition on capacities:

- In the cement market, the domestic price cartel triggered investments in capacities by the firms wanting to get a larger fraction of the domestic market. The capacity excee-

ded what was needed in the domestic market, and exports led to losses compared to the costs of investing in capacity.

- In the airline industry, in the first period after deregulation SAS and Braathens competed mainly on the number of flights and not on prices for business passengers. This led to excess capacity, and on some routes we observed a load factor below 50 %. This means that more than every second seat was idle, which, to our knowledge, is a larger excess capacity than in any other airline market.

As those two last examples indicate, competition can lead to higher costs. Competition on prices might alleviate such a problem. When prices are rather low, the incentives for overinvestment in capacity to capture market shares are dampened. This indicates that a policy towards price fixing lead not only to lower prices, but also to cost savings for the society (see Lesson 3).

Lesson 2: Competition typically generates benefits for consumers

As indicated, competition may take various forms. Price competition is typically to the benefit of consumers. In several of the industries in question we have experienced lower prices following deregulation or other structural changes:

- In the electricity market, the liberalization led to reduced electricity prices throughout the 90s. The consumers benefited from the excessive investments in capacity during the 80s, since the spot market set a price so that all available water was used to produce electricity. Even rather small price cuts will have large effects for the consumers. If for example prices went down 5-10 øre/kWh for all consumers except the power-intensive industry (which had regulated prices), consumers would save approximately 4-8 BNOK annually.
- In the airline industry, the entry of Norwegian in 2002 led subsequently to rather large price reductions for airline passengers in the domestic maket and subsequently in the traffic to and from Norway and Europe. If price fell with 10-15 %, then consumers would save approximately 1,5-2 BNOK annually.
- In the grocery sector, the increased buyer power of retail chains forced the producers to compete for shelf space. Lower prices to the retail chains were at least partly passed on to final consumers. For each %-age reduction in prices consumers save approximately 1 BNOK annually.

However, price is not the only dimension. Various aspects concerning quality are also affected by a transition towards more competition. Some of the examples from the industries in question illustrate how competition may affect quality:

- In the pharmaceutical industry, the transition to a new law triggered the establishment of new pharmacies, as well as longer opening hours. In addition, some drugs are sold at supermarkets and gas stations. No doubt, this increased the consumers' options concerning when and where to buy drugs.
- In the radio market, the entry of a commercial channel had an impact on the publicly

owned channel. After NRK was challenged by the nation-wide commercial channel P4, we observed more diversity within NRK. They introduced three channels with distinctly different profiles.

- In the airline industry, we have observed a larger number of routes being served by more than one operator and even new routes being established. For example, the entry of Norwegian in the domestic market made it possible for this company also to offer direct flights from several Norwegian cities to the continent. This development did not crowd out SAS' routes. On the contrary, we observe that SAS as well has increased the number of direct flights to the continent.

- In the telecommunication sector, the regulation of wholesale prices made it possible for retail sellers to enter the network. Competition on the retail level made regulation of end user prices redundant, and in the mobile sector we experienced entry of both mobile service providers and Mobile Virtual Network Operators (MVNO). This development contributed to the introduction and growth of new products.

Lesson 3: The existence of a competition law matters

Since we do have a competition law in Norway, it is very difficult to know what would have happened if no such law had existed. However, the experience from the cement industry may illustrate the potential effect of such a law:

- In the cement market a price cartel was established in the early 20s. It led to higher prices in the domestic market, and triggered large investments in capacity after World War II. It is estimated that if we have had competition rather than cartel in the year 1968 – the last year before a merger to monopoly – then society would in this year alone in this particular industry increase welfare with approximately 300 MNOK in 2006 prices. In comparison, the annual budget for the Norwegian Competition Authority is 83 MNOK.

If our current competition law had been in place already in the early 20s, there would probably not been any price cartel in the cement market. This would have had a two-fold positive effect on welfare. First, prices in the domestic market on cement would have been lower. Second, investments in capacities would have been much lower after World War II and the firms and the society would have saved costs.

If a competition law had been in force in the early 20s, we would not have any knowledge about neither the price cartel nor the overinvestment in capacities. This illustrates that the existence as such of a competition law might have an impact on firms' behavior that we cannot anticipate without undertaking a complicated counterfactual analysis.

The experience from the airline industry may also indicate that the competition law as such might make a difference.

■ In May 2004 the new competition act entered into force. According to this act, abuse of a dominant position is prohibited. The Norwegian Competition Authority was concerned about predatory behavior by SAS, in particular low prices on routes where the company competed head to head against Norwegian. In October 2003, SAS implemented price cuts, and in May 2004 they made additional price cuts. But after May 2004 prices again raised. This price increase might partly be explained by the new competition act coming into force, as well as by the activity by the Norwegian Competition Authority. In June 2004, the Norwegian Competition Authority made a dawn raid at SAS, to confirm or reject if there had been any abuse of a dominant position by SAS. An important question, though, is whether the pendulum has swung too far. There is a risk that a dominant firm may reverse its price policy more than what is needed not to violate the law, and if so this will not be in the interest of the consumers.

Moreover, we could also think about the hypothetical situation in the grocery sector with no competition law.

■ Some years ago there were speculations concerning a possible merger between the discount chains Rema 1000 and Rimi. Such a merger would no doubt have raised concerns from a competition policy perspective. This raises a hypothetical question: in the absence of a competition law, would we have observed an even higher seller concentration in the grocery sector, a development that would have been detrimental to the welfare of consumers?

Lesson 4: The initial regulatory regime might be decisive for the outcome

In some of the industries in question there has been a sudden transition. Some of the experiences show that the initial market situation was not optimal and resulted in some unexpected effects:

■ In the airline industry, the two active airlines prior to the deregulation had been given the exclusive right to serve particular routes. In addition, however, each of them was seven years before the deregulation allowed to have limited number of flights on the other airlines' until then exclusive route. It turned out that this made it possible with a smooth transition to a deregulated regime. Each firm had a 50 % market share, and incentives not to compete on prices were absent, because that could trigger a price war.
■ In the radio market, Kanal 24 won the beauty contest and was supposed to replace P4 when their ten year license ended at the end of 2003. However, some clever strategic moves by P4 changed the game. P4 continued to serve its listeners, and Kanal 24 would probably have chosen a profile more differentiated from P4's profile if they had anticipated this. The lack of coordination between various regulatory bodies made it possible for P4 to make these moves, and thereby probably resulted in less diversity than what we otherwise would have had. If the authority had made it clear earlier that Kanal 24 would face a rival, and most probably P4, then we might have seen more diversity.

- In the telecommunication industry, the state-owned monopoly was deregulated and partly privatised without any prescriptions for the organisation of the local loop. For example, separating out the local loop for fixed telephony would be very difficult to do after the partly privatization was done. In the electricity industry, on the other hand, separating out of the main grid (in the state-owned company Statnett) was done prior to liberalisation.

Moreover, in some cases it was possible to reverse the initial deregulation regime, which explains why the deregulation produced some positive effects for the consumers:

- In the airline industry, the symmetry between SAS and Braathens came to an end when Braathens became a failing firm and was acquired by SAS in 2001. Then the abolishment of SAS' frequent flyer program enabled Norwegian to enter the market. An entrant with by definition zero market shares, combined with the ban of the loyalty program, led to price competition rather than competition on capacities as was the case in the first years after the deregulation in 1994.
- In the pharmacy industry, the initial plan was to limit the number of pharmacies in cities due to a fear that excess entry in cities would lead to few pharmacies in rural areas. When the law was passed this restriction was removed, and it led to the large increase in the number of pharmacies. This implies that a partly reversal of the initial regulatory regime was decisive for the benefits for the consumers, who could benefit from increased options concerning where to buy drugs.

Lesson 5: The market design matters

As indicated by the above mentioned lesson, the regulatory regime at the time of deregulation might have a large impact on the outcome. This calls for a particularly careful design of the regulatory regime following deregulation. The experience from the electricity industry illustrates the importance of market design:

- When the new energy act was adopted in 1991 the structure of the industry changed dramatically. The ownership of the main grid was transferred to a separate entity, and regulated as a monopoly activity. The production activity was not regulated, but competition was used as an instrument to produce an efficient output. Later on, the market was expanded to a Nordic market, which reduced the concerns about high concentration in each national market.

The market design in the electricity industry seems to be an example of successful market design. In the pharmacy industry, we observe a choice of regulatory regime that was not tailor-made to the vertical structure in that particular industry:

- In the pharmacy industry, we observed vertical integration between retailers and wholesalers. With integrated pharmacy chains it would be natural with a direct price regulation of retail prices. One way to do so would be to set prices equal to observed generic prices in other countries. Instead a de-escalation model has been implement-

ed, where the pharmacies can keep the gains obtained from a reduction in prices from the producers.

Moreover, the experience from the pharmacy industry may also illustrate the importance of how one particular regulatory rule may have unanticipated effects along other dimensions:

■ If price regulation is not strict so that pharmacies may earn a rather high price-cost margin, this will influence the entry of pharmacies. New pharmacies will enter until profits for a newcomer is zero. In that respect a rather soft price regulation may have a two-fold effect. It will lead to high end user prices, as well as more entry of pharmacies. More pharmacies are beneficial for the consumers, but at a certain point we will have excess entry on the margin: the benefits for the consumers from the last pharmacies entering the market are outweighed by the costs associated with this entry.

It is extremely hard to tell whether there is excess entry or not. However, the authorities should at least be aware the potential problem. It is an argument for not going too far in the direction of having a soft price regulation. In the telecommunication, the authorities had to strike another kind of balance:

■ A tough regulation of a network provider would reduce the excessive pricing associated with the network owner's market power. In addition, a tough regulation of wholesale prices would make a large scope for entry of service providers and thereby new products being introduced. On the other hand, such a tough regulation might dampen the network operator's incentives to develop their own content and to invest in its own network.

Lesson 6: Rule of reason concerning competition policy

Last, but not least, it is important to take into account the idiosyncratic features of each industry. There is no such thing as one correct way to deregulate an industry. The regulatory regime following deregulation must be tailor-made to fit the industry. In addition, the competition policy must apply a rule of reason approach. One example is the question concerning market concentration. The acceptable degree of market concentration will typically vary between different industries. Some examples from the mentioned industries may illustrate this:

■ In the pharmacy market, the end user prices are regulated. This implies that high seller concentration should not be a large concern when we consider end user prices. On the other hand, high seller concentration implies that each pharmacy chain has a strong bargaining position towards producers of drugs. Due to this we should not be concerned about the fact that only three pharmacy chains dominates the market, but rather look upon this concentration as beneficial since it can trigger competition between producers.
■ In the grocery sector, the situation is distinctly different as the end user prices are not

regulated. Then there is reason to be more skeptical about high seller concentration. Four retail chains dominate the Norwegian grocery industry. Although this has led to increased buyer power and forced producers to compete for shelf space, the flip side of the coin is that high seller concentration might dampen competition in the end user market. Due to this fact, the competition authority has expressed its concern about the high seller concentration in this industry. In contrast, the competition authority has at present no concerns about the high seller concentration in the pharmacy industry.

Moreover, in other settings high market concentration might be beneficial even without any arguments concerning buyer power. The radio market illustrates the potential for beneficial concentration:

■ In the radio market, it turned out that the two nation-wide commercial channels P4 and Kanal 24 had chosen a rather identical profile. Coordination between those two channels might lead to a change. In particular, they could jointly be better off by choosing different profiles. In such a respect, higher concentration might lead to larger diversity.

Put differently, we have to apply an effect-based approach when we make competition policy decisions.

Torstein Bye and Einar Hope

Electricity market reform—The Norwegian experience

2.1 Introduction

Following the enactment of the new Energy Act in 1990, which laid the legal foundation for Norway's electricity market reform, Norway was one of the first countries to deregulate and liberalise its electricity sector. The main motivation for electricity market reform was an increasing dissatisfaction with the performance of the sector in terms of economic efficiency in resource utilisation, particularly with regard to investment behaviour, which caused capacity to exceed demand considerably (see Section 2.2). Simultaneous market liberalisation initiatives in other pioneering countries, such as New Zealand and the UK, increased awareness of the need for electricity reform, and influenced its design and implementation. This was particularly the case within the Norwegian Ministry of Finance, which initiated the reform, together with the Ministry of Oil and Energy.

The market reform should be considered against the background of the structure and functioning of the electricity system before liberalisation (Hope et al. 1992), (Hope 2000, chapter 7), Bye and Strøm (1987), and Førsund and Kittelsen (1994). The generation of electricity in Norway is almost exclusively based on hydropower. When the reform was launched, there were about 70 power-producing companies and 230 network owners in the system. There was some vertical integration between power generation and the network, particularly at the regional and local levels, but many power producers were not integrated. The largest of them, Statkraft, accounted for approximately one-third of total generation. About 85 per cent of the electricity system was publicly owned either by local, regional or state-authorities. The power production capacity of the energy-dimensioned hydro system in 1991 was approximately 108 TWh in a normal precipitation year, of which the energy-intensive industries consumed approximately one-third. Annual production could vary considerably from year to year because of the stochastic nature of water inflow to the hydro system.

On the consumption side, around 90 per cent of power was sold on long-term contracts, defined as contracts for 'firm power'. Those contracts were negotiated individually and were predominantly bilateral, nonstandardised contracts between buyers and sellers. Power producers were obliged to deliver power within their concessionary areas and to cover their firm power contract obligations through contracts with other power producers. However, the lack of an organised secondhand market for contracts made most of the electricity market inflexible. In addition, electricity prices and other contract terms were generally set by administrative or political decree. For example, the basic price charged by the state-owned company Statkraft, known as the Statkraft price, was part of

the annual regulation of the company determined by the Norwegian Parliament. Besides, the Statkraft price functioned as a price signal to the rest of the market.

Because of the stochastic nature of hydropower production, a market for occasional or interruptible power developed. In 1972, this market was formally organised as a spot market in a power exchange, or pool, among the power producers, known as 'Samkjøringen'. Spot market transactions were carried out at a market-clearing price on an hourly basis determined by bids sent in by the generators to the power pool based on expected excess demand and supply schedules. This wholesale, producer-based spot market, on average comprising approximately the 10 per cent of annual power production not included in contracts, met its objectives efficiently. The market is interesting as a forerunner to the design of the organised pool market system produced by the Norwegian electricity market reform. Besides it represented a 'training ground' for market participants in market-based transactions for almost 20 years prior to the market reform in 1991. Thus, because of the occasional power spot market experience, the learning-by-doing curve for market-based operations was not as steep in Norway as in most other countries that also implemented power market liberalisation.

The rest of the chapter is organised as follows: In Section 2.2, we provide a brief review of the relevant background to deregulation. In Section 2.3, we describe the main elements of the market reform, while in Section 2.4, we discuss some market design issues. In Section 2.5, we describe market developments following deregulation. In Section 2.6, we discuss how effectively the new market dealt with extreme supply-side shortages in 2002–2003. In Section 2.7, we discuss some aspects of competition and regulatory policies in relation to the electricity sector, while Section 2.8 lists some remaining challenges for the market and regulatory system. Section 2.9 concludes the article.

2.2 The background to deregulation

During the regulation period, all investments in production and transmission capacity were subject to cost reimbursement. This was implemented either through direct market prices, cross-subsidisation between utilities,[1] or direct public subsidies.[2] There was no direct link between market prices (since there was no functioning market) and investment or between market prices and operating cost efficiency. The government, when determining its budget, set the following year's prices in the electricity market.[3] The government equated prices to average costs until 1979, from when it set prices equal to long-run marginal costs (LRMC). It used LRMC as a price criterion rather than an

---

1 This is made possible by the increasing marginal cost of expansion in firms.
2 This is achieved through either capital subsidy or relaxed requirements on the rate of return on investment.
3 In fact, the government set prices for state-owned companies. However, the municipalities and counties, which own almost all the rest of the power producing capacities, followed. Since the energy-intensive manufacturing industry had long-term contracts, they were exempted. See Bye et al. (1999).

investment criterion.[4] The total costs were automatically covered without any concern about cost minimisation. The market then functioned as a cost reimbursement system and provided no incentives for utilities to be cost effective. In a competitive market the market sets the price and no investments will take place unless the total cost is less than the price, i.e. firms will focus on minimisation of the cost. Moreover, since the transmission network is a natural monopoly, regulation is required for welfare maximisation. During the regulation period, while cost minimisation (given output) was pursued, output maximisation was also used to ensure an adequate supply. In addition, the central government and municipality authorities set different prices for different consumers,[5] which created inefficiencies and welfare losses in the electricity market.

2.2.1. Inefficiencies in production

There was no systematic evaluation of potential inefficiencies in production before deregulation of the electricity market, except an evaluation of imbalances between capacity and demand. Statistics illustrate excess capacity problems. During the late 1980s, between 5 and 6 per cent of the inflow of water to the reservoirs was spilt annually (even in normal inflow years). The prices set by the central government restricted demand relative to the capacity of primary energy supply (water inflow). To eliminate excess primary energy supply, producers accepted overflow from the reservoirs despite sufficient available generator capacity to produce more electricity. In a free competitive market, generators would have produced electricity from this water because prices exceeded variable cost. Prices then would have dropped to equate supply and demand, and eventually they would have been too low to stimulate further investments.

Midttun (1987, pages 102–109) outlines the political discussion of investment and pricing that took place in Norway from the 1960s to the 1980s. His main conclusions include the following. (i) Production capacity in state-owned companies has not increased according to increases in marginal cost. (ii) The power price has never been high enough to cover the marginal cost of expansion. (iii) The expansion of capacity has led to excessive investments. According to Midttun, the bureaucracy wanted to equate prices and long-term marginal costs as an investment rule in the early 1960s. However, politicians resisted this until 1979. Then a debate over the discount rate replaced the arguing about low electricity prices.[6] Politicians simply proposed a lower discount rate on investment projects to secure lower prices. They even implemented a different discount rate for the investment decision and the pricing. Midttun also documents substantial cost overruns in state-owned companies that were due to weak financial management. However, some of

---

4 In 1979, the government, in a green paper, St. meld nr 54 (1979–1980) decided that the electricity price level in the long run should reflect long-run marginal cost. The escalation period continued until 1985.

5 The purpose was to distribute some of the local natural resource value to local consumers or to support energy-intensive industries and the local labor market.

6 A lower discount rate increases the value of future income, which makes the project profitable at a lower price.

the blame must be assigned to increasing environmental concerns, political intervention, changes to plans and development delays. Costs overran by 57 per cent on average. Project planning focused on technical issues rather than economic issues.

2.2.2. Inefficiencies in transmission and distribution

Transmission networks are regional natural monopolies. Moreover, electricity distribution incurs large fixed costs and exhibits decreasing marginal operating costs. This suggests the following socioeconomic challenges: (a) The optimal private business price exceeds the optimal socioeconomic price. This produces suboptimal demand for transmission services. (b) The private price mark-up may cause profitability to exceed the normal rate of return on capital.[7] (c) The mark-up may cause technical or economic inefficiency in resource utilisation. Although public regulation is important, regulation appeared inefficient.

Kittelsen (1993, 1994) and Førsund and Kittelsen (1998) used production frontier analysis to test for inefficiencies in network distribution companies. They estimated total annual efficiency losses to be between 0.16 and 0.27 billion USD. This amount constitutes 25 per cent of the total resources used for electricity distribution per year. They found no evidence that mark-ups exceeded those necessary to cover cost inefficiencies. That is, they found no evidence of monopoly profits. Hence, distribution networks used their monopoly power to be cost inefficient rather than profitable.

There is no documented research on inefficiencies in the central grid.

2.2.3. Inefficiencies in the market

In a perfectly competitive market, one would expect different consumers to pay approximately the same price for a homogenous good. Power at the wholesale level at a specific time is close to being a homogenous good. Average reported prices for different consumers may be based on different types of contract (incorporating factors such as risk, security of supply, time of use, power and energy). However, during the regulation period, there was little risk of power shortages because a primary objective of the power suppliers was to ensure deliveries at any time.[8]

Bye and Strøm (1987) did a backward calculation of prices at the power plant (a homogenous good) from statistics on purchaser prices (including transmission and taxes) for different consumers. Table 2.1 reports their results. Calculated prices for the energy-intensive manufacturing industry were between one-third and one-half of the prices for

---

7 For ease of exposition, we assume the same risk in all alternative projects. However, in practice it is necessary to adjust for uncertainty differences.

8 There was a compulsory delivery standard in each region.

services and households. This indicates substantial price discrimination, which reduces social welfare. Differences in prices between households and services were less substantial. The averages cover over large differences between regions for the same consumer group. In the power plant regions, (net exporting) prices were kept low for local customers at the expense of those in net importing regions, where prices were high.

Table 2.1. Power prices—net of taxes and transmission fees. Current prices. Cent/kWh

	1978	1979	1980	1981	1982	1983
Households	2,33	2,71	3,08	3,01	3,09	3,56
Services	2,71	3,00	3,42	3,34	3,45	3,85
Other Manufacturing	2,37	2,65	2,94	2,92	3,07	3,48
Pulp and paper	1,26	1,38	1,66	1,57	1,68	1,51
Power-intensive industries	0,82	0,89	1,07	1,01	0,96	1,05
Weighted average	1,81	2,03	2,35	2,33	2,41	2,71

Source (Bye and Strøm, 1987) transformed to current cent/kWh

Bye and Strøm (1987), Bye and Johnsen (1991) and (Bye 1991) estimated the implicit annual efficiency losses because of this price discrimination at between 0.55 and 0.66 billion USD (1987). This represents three times the loss in the distribution network described above. Since the calculations assume identical firms within a sector, the calculated efficiency gains are biased downwards.

2.2.3. The main elements of Norwegian market reform

Based on the Energy Act of 1990, the main elements of the Norwegian electricity market reform were as follows.

- With regard to market design, it was decided to build on the established spot market model for trade in interruptible power, while organising it as a regular spot market, incorporating demand. The market was, in principle,[9] open immediately to all potential buyers, including households. Initially, the market was organised as a separate legal entity within the transmission company, Statnett, and was termed the Statnett Market.
- Common carriage principles requiring access to the network system on a transparent and nondiscriminatory basis facilitated market-based trade.
- The deregulation implied a split of the dominant, state-owned and vertically integrated company, Statkraft, into two separate legal entities: the generating company, Statkraft SF, and the transmission company, Statnett SF. The other vertically integrated power companies were separated into generating or trading divisions and network divisions for accounting purposes. Separate legal identities were not required.

---

9 Small consumers had to pay a relatively high access fee when changing contracts in the first four to five years.

■ The network companies were subject to natural monopoly regulations designed to achieve economic efficiency in network operations. The Norwegian Water Resources and Energy Directorate (NVE), applied a network specific yardstick and rate-of-return regulation. In 1997, a model of income-frame regulation replaced the rate of return regulation.

■ The market liberalisation reform was implemented without changes in ownership, because privatisation of the power sector was politically unacceptable. This contrasted with the UK, where privatisation was implemented before market liberalisation. There, privatisation was considered a prerequisite for successful electricity market reform from an economic efficiency perspective (see e.g. Newbery (1999)).

Statnett, the transmission system operator (TSO), began trading on the spot market for power (the day-ahead market) in 1991, when the NVE introduced the rate-of-return regulation of network owners. In 1993, a financial forward market was established for the delivery of traded contracts. In 1994, this was replaced by a continuous trading system, and standardised financial futures contracts were introduced.

To facilitate trade in the retail market while avoiding investment in expensive metering equipment for retail customers, load-profile demand measurement was introduced in 1995. In 1997, fees for consumer switching were also eliminated to stimulate consumer switching and market competition. In 1998, the Norwegian Competition Authority introduced a price information system for retail prices from power suppliers to improve market transparency. The time allowed for consumer switching was reduced to one week.

In 1996, a common Norwegian-Swedish power market was established to become the first intercountry integrated power market in the world. Nord Pool [10] took responsibility for power exchange for the common market from Statnett Market.[11] The Swedish transmission company, Svenska Kraftnät, became co-owner of the Nord Pool exchange with Statnett. In 1998, Finland formed an independent price area on the Nord Pool power exchange. Denmark integrated into the Nordic system in 2002, since when there has been a common Nordic integrated electric power market (excluding only Iceland).

2.4 Market design and market operations

A complete market-based power system should be equipped with markets for the following five basic requirements or functions: (a) markets for trade in electricity; (b) markets and instruments for risk hedging in accordance with risk preferences; (c) short-term markets for production capacity [12] and the balancing of supply and demand; (d) markets for

---

10 See http://www.nordpool.com.

11 See http://www.Statnett.com.

12 Capacity markets are needed to provide two kinds of services (see section 4.3). These are: (1) the instantaneous balancing of supply and demand to prevent system breakdowns or fallouts; (2) accounting for deviations between planned production according to the supply and demand schedules at the time when the price is determined and production needed to meet demand at the time of delivery.

investment in new capacity; and (e) markets for trade in environmental energy products (such as green and eventually white certificate markets). Nord Pool has organised markets for functions (a), (b), and, so far, part of (e). Function (c) is generally handled by the transmission system operators in the individual countries. There are hardly any organised markets for (d).

Nord Pool is a non-mandatory power pool that organises approximately 40 per cent of the total trade in electricity in the Nordic power market. The rest is organised on the basis of bilateral contracts. Nord Pool's share in total trade on the organised spot market is a useful indicator of the liquidity of the market. This is discussed in relation to the volume of trade in organised financial markets in section 2.4.2.

Nord Pool also performs the functions of contract clearing and settlement. Nord Pool established a new environmental market for electricity certificates for renewable energy production (green certificates) in 2004.

2.4.1. Nord Pool's spot market—Elspot

Elspot is a contract spot market on which electricity is traded on a daily basis for physical delivery the following day (a day-ahead market), with full obligation to pay. The bidding procedures are essentially the same as those adopted by Statnett Market in 1991. Market participants place bids in the Pool one day in advance for the next 24 hours of the following day. The Pool then aggregates the bids and prices for each hour based on individual supply and demand for price formation in the market. The Nord Pool system price is the market equilibrium price for each hour. The spot-market system price functions as a reference price for Nord Pool's financial markets and the bilateral markets in the Nordic system. Currently, some 280 participants trade daily on the Nord Pool spot market.

The system price is determined without taking into account potential capacity constraints in the transmission network system. If calculation of the system price indicates that the power flow between two or more network areas, decided in advance by the system operator, exceeds capacity limits in the transmission grid, area prices are determined. A capacity fee, defined as the difference between the system price and the area price, is then calculated. The transmission system operators in the Nordic countries set the capacity fee as an integral part of their operation of the system. Thus, the system operators are obliged to use the price mechanism in the spot market when adjusting power flows during periods of capacity constraints between bidding areas (see Subsection 2.4.3 below).

2.4.2. The markets for derivatives at Nord Pool

The types of contract traded on Nord Pool's financial markets comprise electric power derivatives and electricity certificates. The financial derivatives, like futures, forwards, options, and contracts for differences are developed to help the market participants' handling of risk in a volatile market.

The reference price for those derivatives is the spotmarket system price for the total Nordic electric power market. The maximum trading time horizon is currently four years. For all derivatives, the principle of cash settlement applies. There is no physical delivery of electricity on those contracts.

The futures contracts are standardised contracts for a given quantity of power at a certain price in a specified time period. The forward contracts are typically nonstandardised. While the time horizon for futures over time is reduced from three years to between eight and nine weeks, forward contracts apply to periods of up to four years. Thus, the market seems to favor short-term futures near the due date and long-term forward contracts near the end of the time horizon. This may be because of the difference in margin calls between futures and forwards. The market settles futures daily on a market-to-market basis, which requires a considerable cash commitment up-front. By contrast, forwards only require cash collateral during the delivery period.

The option contracts traded at Nord Pool adopt the European convention that contracts can only be exercised at the stipulated exercise date. Options combined with futures and forwards offer interesting strategies for risk hedging and risk management in electricity power trading. They also allow greater flexibility in contract portfolio composition and administration.

Contracts for difference (CFDs) were introduced to allow market participants to hedge against the price area risk. As already mentioned, the system operators determine area prices that differ from the prevailing system price when there are capacity constraints in the transmission network. Futures and forward contracts cannot hedge against this price area risk. Therefore, CFDs were introduced to enable perfect hedging even when the market is split into two or more price areas.

In 2004, Nord Pool began the trading of electricity certificates in Sweden, on contracts involving physical delivery. In February 2005, Nord Pool also began trading in carbon emissions by using European Union Allowances (EUAs). Hence, it became the first deregulated market in Europe to trade in and clear such contracts.

The volume of trade in financial derivatives markets is currently about five times the volume of physical trade in the spot market. This ratio is used as an indicator of market liquidity and of how efficiently markets are functioning. The ratio is now increasing following a decline in 2003.

2.4.3 The balancing markets—capacity markets

Capacity markets are required to balance supply and demand in an electric power system to avoid system breakdowns or delivery fallouts. The following three major types of imbalance between supply and demand, which create the need for capacity markets or balancing mechanisms, can be distinguished. (a) Deviations can arise between the planned supply and demand schedules on which prices are determined in the day-ahead

market and the actual demand schedule prevailing at the time of delivery within each hourly time section in the spot market. (b) Price deviations can arise because of transmission capacity constraints. (c) There can be imbalances or interruptions because of stochastic fallouts of generation or power line capacity. We focus primarily on (b).

If the power flow between two areas exceeds transmission capacity, the price is reduced relative to the system price in the surplus (low-price) area and is increased in the deficit (high-price) area. This continues until the power flow matches the transmission capacity constraints; that is, supply meets demand within each area. The system operator is responsible for this capacity regulation on the grid, when capacity constraints arise.

However, within the Nordic system, different principles and methods are applied to balance capacity. In Norway, transmission capacity problems are resolved by the price mechanism in the spot market according to the principle of delineation of price areas described above. This is the responsibility of the Regulating Power Market, which is operated by the Norwegian system operator, Statnett. Statnett divides the country into four geographical bidding areas and stipulates the maximum transmission capacity between these areas. Every week, based on data from Statnett, Nord Pool then informs all market participants of the bidding areas that apply for the following week. The number of price areas depends on grid conditions and the relationship between supply and demand in the system. Because of reduced investment in transmission capacity relative to demand, capacity constraints have gradually become more binding. This implies that price area delineations have become more persistent.

Sweden and Finland form one bidding area in the spot market, while Denmark is divided into two. In Sweden and Finland, the counter-purchase principle is applied to manage internal transmission bottlenecks. Counter-purchasing involves system operators in Sweden and Finland paying for the downward regulation of production in the surplus area and upward regulation in the deficit area until the capacity constraint is eliminated. The cost of counter-purchases is financed by tariffs on power production. The balancing mechanism used for Sweden and Finland is known as Elbas.

The Regulating Power Market in Norway is organised as a bidding market in which a 15-minute time span applies to price determination. For imbalances, which cannot be handled within this period, Statnett can impose downward or upward capacity regulation on market participants at short notice (less than 15 minutes). Initially, market participants comprised of a relatively small number of large power producers with considerable regulating capacity. Now, however, the market has been opened to participants from the demand side. These include firms in power-intensive industries and other large consumers that are willing and able to regulate their power consumption if the price in the Regulating Market gives them an incentive to do so.

2.4.4 The retail market

The organised markets at Nord Pool are wholesale markets for the common, integrated Nordic electric power market. The retail markets are largely national markets because of national retail market regulations. However, integrated retail markets are also being developed.

Retail market competition in Norway has been stimulated by regulatory measures to increase market transparency and consumer switching. This switching has resulted from the abolition of switching fees and the establishment of a price information system for the retail market by the Norwegian Competition Authority (NCA) in 1998, as mentioned in Section 2. In 2003, after the 2002 inflow shortage and subsequent escalating prices in the market, the NCA developed the information about differences in the firms' contracts further to reduce transaction cost and improve end user market competition. Note, however, that producers only compete on the electric power price. In Norway, this price accounts for roughly one third of the total end-user price. The remainder consists of the grid-user price and public taxes and fees, which each account for approximately one-third of the total price.

The obligation to report retail prices to the NCA applies to around 170 suppliers, of which 50 to 60 operate regularly in the market. The number of consumers switching suppliers has increased steadily since the retail market was opened in 1995. During the first quarter of 2005, around 65,000 household consumers changed supplier, which represents 3 per cent of all households. In April 2005, 25 per cent of household consumers used a power supplier other than the dominant supplier in the area. However, the absolute number of consumers switching suppliers is not necessarily an appropriate indicator of increased competition. What matters is whether the number is sufficiently large to cause suppliers to set prices competitively.

Approximately three-quarters of Norwegian retail consumers have entered into some form of variable retail-price contract (such as a spot-market contract or a standard variable power-price contract). This exposes them to some variations in the Nord Pool system price on the wholesale spot market. By contrast, in Sweden, 80 per cent of retail consumers pay a fixed price. This difference may have arisen because Norway depends totally on hydroelectric power, whereas Sweden only depends on hydroelectricity for 30 to 40 per cent of its total production. Consequently, price volatility has traditionally been higher in Norway than in Sweden. In a fully integrated market, however, price volatility should converge. Tradition, contract types, and risk preferences may also explain the difference in contracting behaviour.

2.5 The development of the market following deregulation

Deregulation of the electricity market was expected to lower investment, reduce and equalize prices between consumers, lower net tariffs, and raise the rate of return on investment.

2.5.1 Prices

In a virtually completely hydro-based electricity market, we would expect increasing long-run marginal costs because of a scarcity of resources. Given this background, we would expect higher prices in the long run because the investment rule for private investors implies that prices equal long-run marginal costs. However, in the short or medium run, excess capacity implies that prices equal short-term marginal cost plus any shadow price on capacity restrictions.

In the introduction, we indicated that during the regulation period, the public power sector invested in excess capacity because of compulsory short-term supply rules and optimistic forecasting of demand. However, at the same time, it attempted to equate prices to long-run marginal costs. How is it possible to have excess capacity when long-run prices reflect long-run marginal costs? There are three obvious reasons. (a) Energy-intensive industries, which consume one-third of capacity, paid prices corresponding to one-quarter to one-third of long-run marginal costs; that is, instead of increasing capacity further, electricity consumption should be reallocated to equalize prices between consumers. Prices would fall below long-term marginal costs and investment would cease. (b) Excess production in relation to domestic demand was sold on an international market in the form of occasional power (the forerunner of the Nord Pool exchange market) at low prices. Instead of increasing capacity further, this production could have been allocated to Norwegian consumers by lowering domestic prices. Instead Sweden and Denmark benefited from this approach. (c) In the late 1980s and early 1990s, almost 5 per cent of the inflows to reservoirs were spilt [13] during the periods of spring snow melting and autumn rain. [14] This water could have produced electricity given that generator capacity was sufficient, which it turned out to be. Prices would have fallen and demand would have increased. Prices would eventually have been too low to stimulate further investment; that is, existing capacities would have been sufficient for many years.

After the deregulation, the regulatory authorities inspect the spillage of water from the reservoirs in the Norwegian hydropower system. Thus, previous excess capacity competes in the market, and electricity prices become lower than long-run marginal cost in the short and medium run. This persists until demand increases and production capacity constrains growth. Then prices increase and stimulate further investment. Deregulation of the market also puts downward pressure on prices by generating an expected efficiency gain in terms of operating costs and investment costs in capacity per MWh.

---

13 Source: Statistics Norway: Report 90/1.
14 Norwegian power capacity is almost completely based on hydroelectric power.

Moreover, as discussed in the Introduction, there is a major increase in market efficiency; that is, greater price equality between consumer groups.

Figure 2.1. Spot and consumer prices, Nøre/kWh fixed 2002 prices
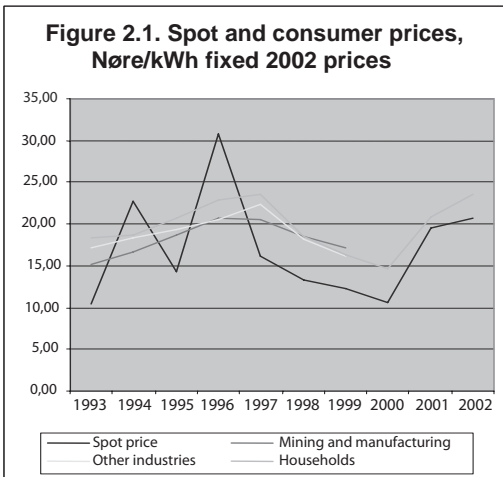
Source: Statistics Norway and NordPool

Figure 2.1 shows changes in the real spot price and average prices (2003-prices) among consumer groups. First, the spot price is low in comparison to the end-user prices prevailing in 1993. This is mainly because of excess capacity and the splitting of the market. Neither the end-user market nor the spot market were fully developed after two years of deregulation.[15] Second, there is almost no correlation between the spot price and end-user prices after four to five years of deregulation, although there was an increasing trend in all prices. In this period, end-user prices were almost identical among consumer groups, which suggests that the market eventually functioned as expected.

Since 1997, the Nord Pool market expanded when Sweden and Finland deregulated their markets and joined NordPool.[16] The fee on contract switching for small consumers introduced in 1991 was removed during this period. End-user prices then followed spot prices on a downward trend. Nevertheless, end-user prices remained above the spot price. When the spot price increased in 2000, the gap narrowed.

Figure 2.2. The spot price (NOK/MWh left axis) and hydro power production (TWH-right axis)
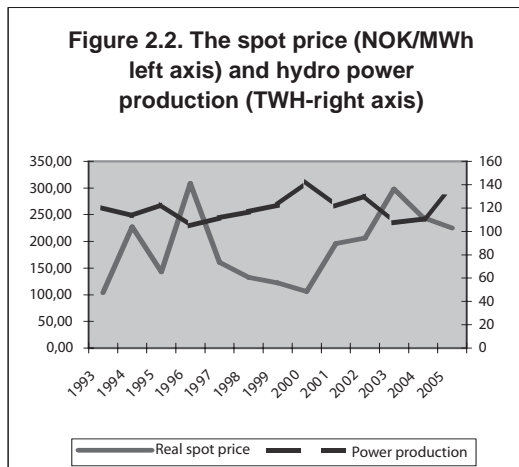
Source: Statistics Norway and NordPool

As expected, figure 2.2 shows that fluctuations in the spot price are negatively correlated with hydropower production. Since demand elasticities are low (see Bye et al., 2003), a modest change in supply may have a large impact on the spot price.

Hence, deregulation did put a downward pressure on the electricity price, seem to have reduced price differentials between consumers and have closed the gap between end-user prices and market equilibrium prices. There is one exception to

---

15 Small consumers had to pay a high access fee when changing contracts in the four or five years following deregulation.

16 Market expansion increased the amount of surplus power and prices fell.

this; the power intensive industries which are not shown in the graph. Power intensive firms signed long term contracts of 40 to 60 years with very low prices and large volumes in the 1950s and 1960s. The full effect of the deregulation will then not take place before these contracts expire in 2008-2011. This could create short term turbulence in the electricity market since they cover large volumes with extremely low prices. Some of the industries will not survive sharply increasing prices on electricity.
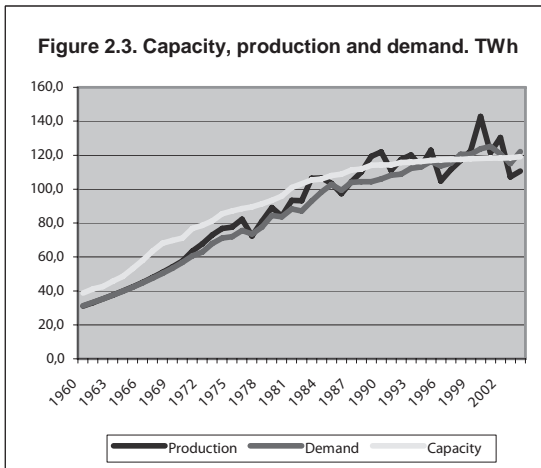
5.2. Investment in power production capacities

Investments in new power producing capacity (TWh) were at a high level during the regulation period in the 1960s and the early 1970s, despite the fact that production capacity well exceeded demand, see figure 2.3 and 2.4. However, production followed demand, which implied either annual water runover or abnormal low annual inflow. During the 1970s, investments escalated and kept capacity well over demand. Eventually, production also exceeded demand and Norway became a net exporter of power on an annual basis until six years after the deregulation in 1991. Since 1997, production and capacity have been lower than demand, except in 2002–2003, when inflows were well below normal[17]. Demand exceeded production, prices increased dramatically and demand responded.
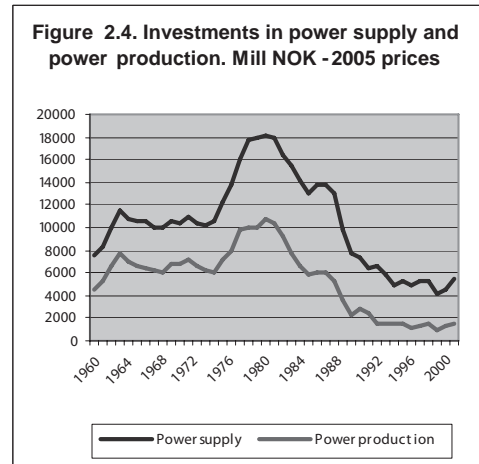
Investments in new production capacity began to fall in the early 1980s, see figure 2.4, long before deregulation. This was mainly because of a sharp increase in the marginal cost of expansion, see figure 2.5, and a continuing increase in environmental concerns. These concerns made expansion politically unacceptable. After deregulation, investment continued to fall and reached a low level. As demand increased and Norwegian capacity was restricted, prices started to increase. However, when Sweden, Finland, Denmark and other northern European countries deregulated, excess capacity in these countries kept prices low and imports to Norway high.
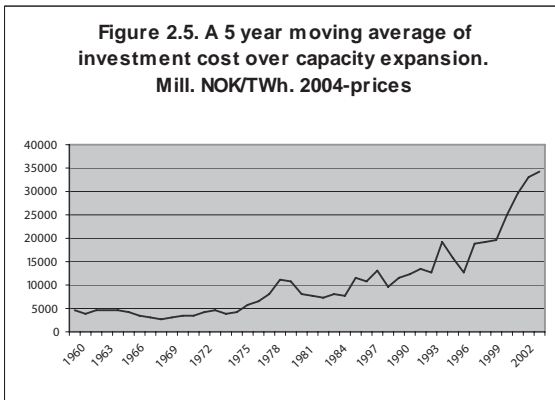
---

17 'Normal' refers to the average over the period 1970–1999.

Figure 2.3. Capacity, production and demand. TWh

Source: Statistics Norway and NVE

Figure 2.4. Investments in power supply and power production. Mill NOK - 2005 prices

Source: Statistics Norway

Figure 2.5. A 5 year moving average of investment cost over capacity expansion. Mill. NOK/TWh. 2004-prices

An important issue is whether deregulation provides sufficient incentives for investors to participate in expanding the electricity sector in Norway. This issue is high on the public agenda and skepticism is widespread. However, conclusions are often confused with the potential effect of persistent regulation of energy resources because of environmental concerns. Politicians seem unanimous in blocking new investment in large hydropower plants, gas-fired power plants and nuclear and other thermal plant technologies. Hence, there are restrictions on the supply side. Consequently, prices tend to increase. The only feasible alternatives seem to be renewable technologies based on, for example, wind, biomass, solar energy and wave power. Since these technologies are costly, the market prices are still not sufficiently high to stimulate investment in the absence of strong financial support.

2.5.3 Rate of return on power production

Excess capacity relative to demand may imply low capacity utilisation (before 1991) or low prices (after 1991) and consequently a low rate of return on investment. A low rate of return normally hinders investment in new capacity. Unlike in the manufacturing industry, the rate of return in the power sector in Norway has been low since the 1960s, see figure 2.6, but still investments have been high see figure 2.4.[18] In a regulated sys-

---

18 In much of the manufacturing industry in Norway, the rate of return varies because of international business cycles (in, for example, aluminum, ferro alloys, chemicals, and pulp and paper).

tem, this may happen, but the development is in conflict with normal optimization of welfare in a competitive market. The low rate of return in the power sector lasted for several years after the deregulation in 1991. However, the rate of return has recovered since 2000 because hardly any investment in new capacity has taken place while demand has increased by, on average, between 1 and 1.5 % per year for the last 10 to 15 years. Besides, the environmental cost and primary energy cost have increased in the competing thermal power plants, cfr. taxes, permit prices, and gas and coal prices.

Figure 2.6. Rate of return in manufacturing and power production. Percent. 1963-2004

Source: Statistics Norway

In a hydropower-based electricity production system, one can expect the average rate of return to be higher than in other industries (such as manufacturing). This follows since scarce water fall resources and increasing marginal cost of expansion, see figure 2.5, imply many cheap plants that should harvest an economic rent.[19] Scarcity is not yet a problem, but potentially will become one as the market develops.

2.5.4 Investment in networks

In the 1950s and early 1960s, much of the power production capacity in Norway expanded through 'combined' projects. Comparative advantage has often been related both to the access of low cost natural resources and non-tradability of these resources. Due to regulation and/or lack of transmission capacities for electricity, almost no broad international market for electricity existed 50 years ago. The comparative advantage in power production then channeled through energy intensive manufacturing industry. Power projects were implemented and energy-intensive manufacturing were developed at the same location. These developments took place on the western coast of the country. This minimized the investments in network capacity, see figure 2.7.

Figure 2.7 Investments in network capacity. Mill NOK - 2005 prices
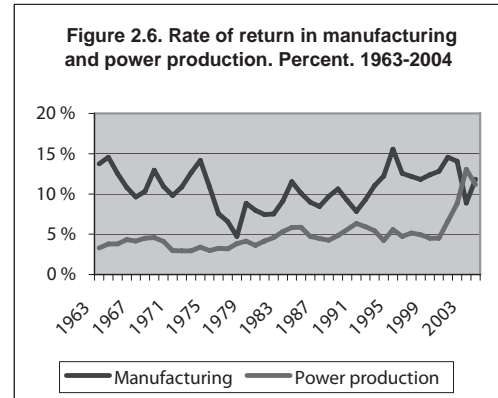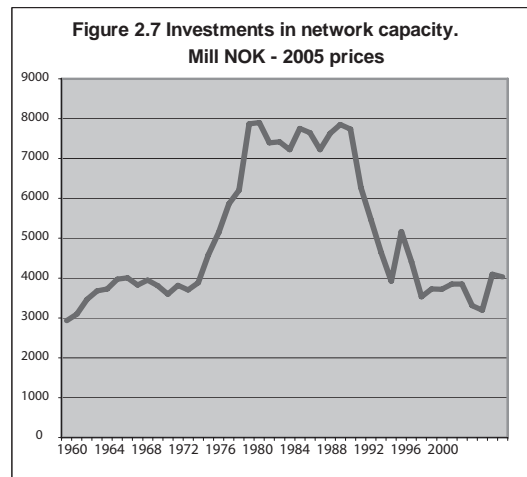
Source: Statistics Norway

19 New investment will not occur unless prices cover the long-run marginal cost. This applies unless there is a backstop technology or unlimited import capacity at fixed prices.

Over time, general services and household energy demand grew, mainly in the densely populated areas in the east of the country.[20] This increased the general need for expansion of the transmission networks, which in turn was self enforced by the increased opportunity cost for electricity, both domestically and internationally, created by these investments. Comparative advantage for aluminum melting faded and comparative advantage for electricity deliveries increased. To some extent, favorable policy measures tried to counteract the fading of comparative advantage for manufacturing (cfr Bye and Holmøy (2006)).

The location of hydropower in the west increased the need for transmission capacity from west to east and, to a certain extent, from north to south. As higher fuel oil prices in the 1970s (introduced by OPEC in 1973–1974 and again in 1978–1979) and increasing environmental concerns triggered a sharp increase in sulfur taxes and regulation of the sulfur content in oil products, massive substitution of fuel oil with electricity took place in the heating sector. Along with aggregate economic growth, this raised demand for capacity investment in the distribution network. Once this large infrastructure project had been completed, investment in network infrastructure capacity decreased.

These developments could explain the sharp increase in network investment that occurred in the early 1970s and the sharp decrease between 1988 and 1993, see figure 2.7. However, this is just part of the story. The decrease in investment also coincided with the debate that took place before and during deregulation of the Norwegian electricity market in 1991. While power production and trade were exposed to competition, a new regulatory regime was introduced for transmission and distribution network companies in 1991. Rate-of-return regulation and 'yardstick competition' reduced network tariffs. The network reimbursement policy was replaced by yardstick competition to secure cost efficiency and an improved social investment strategy in capacity. When the profitability of network investment fell, investment also declined. Investment in networks increased again in 2002–2003 because of the upgrading of existing networks and the installation of new capacity designed to alleviate temporary network constraints.

2.5.5 The end user market

In the wholesale market trade is divided between bilateral contracts and trade on the pool. In the end user market trade is on fixed price contracts, variable price contracts and spot price contracts. The price in the spot price contract equals the wholesale price on the pool with an administrative firm specific mark-up (either a fixed monthly or annual rate). The variable price contract is subject to dismissal a week in advance and the price may be changed in fourteen days, i.e. there are some, however, sometimes unclear link between the variable price contract and the expected spot price on the pool. The expected spot price on the pool should, according to theory, equal the future price adjusted for

---

20 Demand from the residential sector was five times higher in 2004 than in 1960, while demand from manufacturing industries was only about twice as high. The residential sector used almost the same amount of electricity as did the manufacturing industry in 2004.
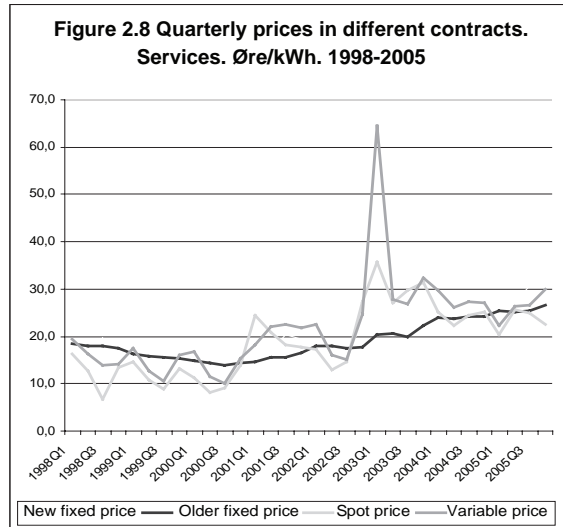
water storage cost and uncertainty. The variable price contract is the default contract, i.e. if the customer does not actively choose contract type he will face the standard variable price contract. The fixed contract price should reflect the futures and the forwards on the pool and include a risk-sharing element.

To reduce the information and transaction costs in the market each nationwide seller has to report their offer price in the end user market for each contract type to the Norwegian Competition Authority, which again reports the entire price information set for each firm on their website.

Figure 2.8 reports the development of the average price in each contract type for the services sector, as a representative sector, for the last eight years. In average, the spot price and the older fixed price contracts are the cheapest one for the customer. The same yield for the residential sector and the manufacturing industry although not reported here. In periods, the variable contract price has been way above the spot price and the fixed price. If this prevails, you may question if the market functions very well. However, it seems as if the price gap is narrowing over time as the

Figure 2.8 Quarterly prices in different contracts. Services. Øre/kWh. 1998-2005
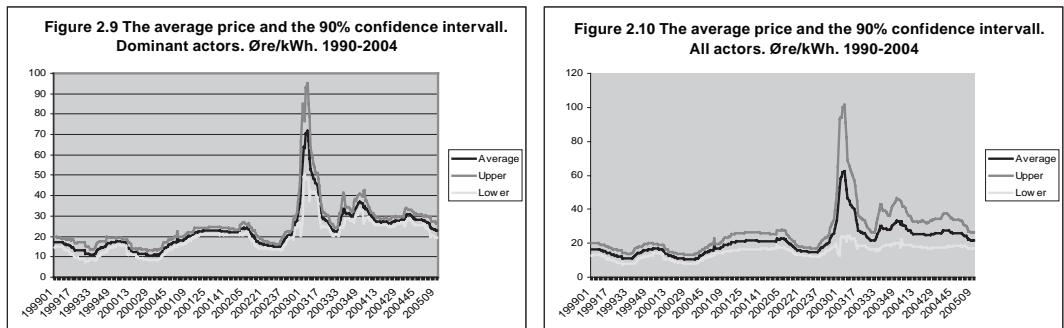
effect from the 2002 lack in inflow event has passed away. Evaluated on average then, the market seems to function reasonably well. Bye et al (2003), however, pointed out that the standard variable price contract might create a market problem in being the default contract. The variable price contract pretends to give the consumer some price insurance, but the 2002-2003 event showed that the variable price contract was even worse for the consumer than the spot price contract measured over the whole period, i.e. frequent contract changes were necessary to gain from this contract. The consumers do not seem to be fully aware of the properties of this contract. Since switching is limited, the spot price contract would probably be a better default contract with respect to market efficiency. Bye et al (2003) also advocated a development of the standard variable contract to include both a variable and fixed price contract element. This would offer the consumer some flexibility in risk handling of the volatile electricity prices.

Statistical averages may conceal large heterogeneity in prices and possible market failures. More than 160 firms offered standard variable contracts in 1999. Because of bankruptcy and acquisitions the number has decreased to 90 in 2005. In figure 2.9-2.10 we report the development of price differences week by week in the standard variable contract type among these firms. There are some interesting aspects concerning competition. First of all we note that the differences between firms have increased over time, especial-

ly after the inflow event in the autumn of 2002. Besides, the price differences are huge. The most important question then is why do not customers change supplier? Does this really reflect an efficient market? In mid 2005 the upper price is approximately 50 per cent higher than the lower price, i.e. the cost increase in a household budget term is 0,7 per cent (the budget share of electricity is approximately 1,5 per cent). Is this too low compared to the transaction cost involved in switching? The second observation is that the average price for dominant firms is higher than the average price for all firms, but the price differences among dominant firms are less than the price differences among all firms. One explanation may be that consumers express loyalty to local firms. Another may be a segregation of the market; one owner may separate the firm into two units, one supplying loyal consumers the other serving market oriented cost minimising customers. A third reason may be that transaction costs are still too high to defend switching. Further research on this issue is required.

Figure 2.9 and 2.10

Figure 2.9 The average price and the 90% confidence intervall. Dominant actors. Øre/kWh. 1990-2004

Figure 2.10 The average price and the 90% confidence intervall. All actors. Øre/kWh. 1990-2004

It seems as if there still are some efficiency problems with the end user market.[21] Bye et al (2003) advocated that the default contract should be the spot price contract. This could probably help to overcome some of the possible efficiency problems raised in the end user market.

2.6. A market under stress—a real test

The Nordic electricity market was exposed to an extreme primary energy shock between 2002 and 2003. A short-term shortage of precipitation and inflow sharply increased prices and led to vigorous discussion of the functioning of the deregulated market when exposed to such extreme situations. Policies that could help to relieve these so-called 'infirmities' in the market were discussed. However, Bye (2003) showed that the market functioned remarkably well; producers tried to optimize the value of water, as expected; electricity trade followed anticipated patterns; and consumers responded as predicted by theory. (See Bye et al., 2003, and Fehr et al., 2004).

---

21 See also von der Fehr, Amundsen and Bergman (2005).

During the period of regulation of the Norwegian electricity market, a security-of-supply rule was introduced for investment decisions. According to this rule, there should be enough capacity at any time to satisfy demand. However, the primary supply (precipitation and inflow) is stochastic in a power system dominated by hydropower. Demand also changes because almost 90 per cent of heating is electric and outdoor temperatures vary substantially. The any-event security-of-supply rule thus implied excess capacity in normal situations and water spillage in above-normal inflow conditions. When deregulation was introduced, prices and investment fell as shown above. Firm-specific profit maximisation reduced excess capacity over time. The tightening of the market and the stochastic supply and demand eventually increase price fluctuations. If the rains fail, as they did in the autumn of 2002, prices have to increase to clear the market. A vertical supply curve and small elasticities of demand may even imply escalating prices.

The fundamental profit-maximisation problem for a hydropower producer is to optimise the value of stochastic inflows of water over time (Førsund, 1994). The reservoir is an instrument for equalising prices over time. Under simplifying assumptions, it leads to the maximisation of profit from the production of power. If enough producers compete efficiently, if no transmission constraints occur, if there is sufficient storage capacity, and if precipitation, inflow and weather are normal, then equalising prices between periods also optimises the social surplus. If there are no constraints, this optimising strategy generates equal prices for electricity over time and place. In practice, these assumptions are simplistic, but sufficient for the reasoning that follows, except for the inflow uncertainty. For further discussion of a more complex system, see Førsund (1994), Bye (2003) and Førsund and Hoel (2005).

In the Norwegian hydropower system, water typically flows into the reservoir during the snow-melting period from early May to mid July and in the rainy season from mid September to late October. The high-demand period is winter, from October to April, while demand is low in summer, from May to August. In a hydropower system, profit-maximising behavior involves equalising the value of water (the price of power) between periods. To do so, the storage capacity of the reservoirs is used.

In the spring of 2002, since the inflow to the hydro reservoirs exceeded the normal level, production increased and prices decreased. The water level was above normal. Producers had the incentive to produce to avoid an overflow in the rainy autumn season. However, the autumn rains did not come, which resulted in a 20 TWh (17 per cent of Norway's annual production) inflow shortfall within 6 weeks, relative to the normal inflow for this period. The probability of this happening was approximately 0.5 per cent. Prices in the spot market increased to an all-time high level (and quadrupled on average within two months). Over a period of 12 months, average spot prices increased by almost 50 per cent. Demand fell by about 5 per cent, despite many manufacturing companies having fixed-price and volume contracts. Some companies even sold power back to the electricity companies under these fixed contracts.[22]

---

22 These contracts represented a combination of price and volume contracts.

During this period, physical rationing of power was discussed because of a possible draining of the reservoirs during winter. Some focused on a possible malfunction of the market (because of abuses of market power, irrational behavior by new firms, and the inadequacy of the market for dealing with extreme events). Politicians threatened to reregulate the market and proposed several measures for dealing with extreme situations. They were primarily motivated by public and media focus on the possibility of rationing and severe price effects on the income distribution.

At the request of the Minister of Administrative Affairs, Bye et al. (2003) evaluated the event and concluded that the market functioned as expected and that the market dealt with the extreme almost perfectly. The historic rate of return in power production explains low investment in production capacity and is not a consequence of malfunctioning or the abuse of market power. Moreover, between 2002 and 2003, expectations of futures prices (contract prices for hedging two or three years ahead) were low despite the high prices specified in physical contracts. Thus, short-term prices did not justify an expansion of productive capacity. High prices simply reflected a water shortage and the need to stabilise water values over time, which reflected great uncertainty. The water balance in the summer of 2002 was well above normal. This put downward pressure on prices to increase demand and generate a water balance that was low enough to accommodate the autumn rains. Because the rain failed and the water balance fell, the market had to adjust to restore the water balance in the spring of 2003. Since imports were restricted by the transmission capacity, domestic prices had to rise.

Although the market seems to have functioned well, Bye et al. (2003) identify issues for further study and follow-up by the competition authorities. One issue is the future design of contracts. The market seems to have been competitive despite the fact that transmission was restricted between Norway and other countries almost 60 per cent of the time during the winter of 2002–2003. However, there seems to have been a problem because of price differences in the contract market, both in the wholesale market and in the retail market (see Fehr et al., 2005).

2.7 Competition and regulatory policies

The regulatory policy for the electric power sector consists of a competition policy for electricity markets and a regulatory regime for network activities. In both cases, economic efficiency has been defined as the policy objective for the Norwegian regulatory system for the electricity sector.

2.7.1 Competition policy

On the basis of a new Competition Act of 1994, the Norwegian Competition Authority (NCA) began to investigate competition policy problems and issues in the liberalised Norwegian electricity markets after 1991. Because there was considerable overlapping legal competence between the Competition Act and the Energy Act, and, consequently, uncertainty about the division of labour and responsibility between the NCA and the sec-

tor-specific regulator, the NVE, with regard to enforcement of the Acts, the NCA initiated a process of clarifying the respective roles of the two agencies. This process ended up with a written agreement on cooperation, exchange of information and enforcement rights and responsibilities between the NCA and the NVE. Later, a similar agreement was signed with the financial regulator, the Norwegian Securities and Exchange Authority, in relation to the financial derivative markets.

Market transparency
On the market design side, Norwegian competition policy has been concerned with improving market transparency through various measures; first and foremost the design and implementation of the retail-price information system. This is discussed in Section 2.4 above. The price information system seems to have performed quite well as a stimulant to competition in the retail market, however, some problems seem to remain with respect to the default contract. The NCA is to our knowledge the only competition authority that operates such an information system for electricity markets to increase market transparency. The Nord Pool market is not fully transparent in the sense that bids are considered private information. In Nordic Competition Authorities (2003) Wolak argues for releasing such information. This could, however, be controversial in a competition context, since the number of members of the pool is relatively small, which may open for collusion when information is open.

Market dominance and market power
The prime concern in Norwegian competition policy in relation to the electric power markets has, however, been with the potential for exercising market power to the detriment of competition, resulting from unilateral or collective market dominance.[23]

Because of the hydro structure of the Norwegian electricity sector, with a fairly high number of independent producers, market concentration and market dominance have been less of a competition issue than in most other countries. The largest producer of hydropower in Norway, Statkraft, a state-owned company, produced around 30 per cent of Norway's electric power, prior to the deregulation of the industry. However, much of its output was delivered to the energy-intensive manufacturing industry on the basis of long-term contracts. Statkraft's share of the remainder of the market, the competitive part, was less than 15 per cent. Private firms provided about 10 per cent of Norway's production capacity, while municipalities and counties supplied the rest.

Following deregulation, many of the companies under local-government ownership were turned into limited-liability firms. Larger regional power companies were established, partly through acquisitions and mergers among local-government entities. The state-owned company, Statkraft, also grew through mergers, acquisitions and the purchase of shares in other large and small power companies. This was to some extent encouraged by

---

23 For a comprehensive study of market dominance and market power in electric power markets, see Hope (2005).

politicians, who focused on Norway as part of a larger Nordic integrated electricity mar-ket, and even of an international/European market, in which Statkraft would not be a dominant player. Thus, the "national champion" argumentation could be used with regard to Statkraft, without endangering competition, it was argued from some political circles. The NCA followed this merger and acquisition development closely, particularly with regard to competition in regional markets, i.e, when transmission was constrained and the dominant firm(s) could potentially exercise market power, but did not actually intervene in any of the cases it investigated.

Table 2.5 Market concentration index in the Nordic power market.

	HHI	+ incentives	+ control
Sweden	0,2893	0,2923	0,2988
Finland	0,1766	0,2037	0,3005
Norway	0,1634	0,1980	0,3325
Nordic	0,0892	0,0989	0,1138

Source: Nordic competition authorites (2003).

Bye et al. (2003) report a Hirschman–Herfindahl concentration index for the Norwegian market based on direct ownership of 0.1634. One additional index that incorporates inac-tive but incentive-based cross-ownership is 0.1980. A third index that controls for demand [24] and incentive-based cross-ownership is 0.3325. They concluded from the tra-ditional measure (0.1634) that the Norwegian market remains not concentrated. However, if we take into account cross-ownership, the market is reasonably concentrated (0.3325). For the whole Nordic region, they found a cross-ownership, incentive-based index of 0.1138, which suggests an unconcentrated market.

The relevant issue is whether the Nordic market is an integrated market or a regionalised market. Hourly data on area prices indicate the scale of transmission constraints and allow a calculation of the scope of the relevant market. In 2001, the Nordic market was fully integrated 51.8 per cent of the time and regionalised otherwise, based on calcula-tions for seven Nordic regions. The most populated area in Norway, the south, was clas-sified as a separate area less than 10 per cent of the time, while the northern part of the country was a separate area nearly 20 per cent of the time. Thus, the issue of market power is relevant to observe from a competition policy perspective.

Mergers and acquisitions

As the electric power market has become more concentrated through mergers and acqui-sitions, the NCA has tightened its market surveillance and enforcement policy to prevent

---

24 That is, the index controls for demand according to ownership share.

the abuse of market power. The Statkraft acquisition of Agder Energi (a large producer in the South of Norway) is an illustrative case. With reference to a regulation in the new Competition Act of 2004, the NCA has issued obligations to several major power companies to report mergers and acquisitions to the Authority.

In 2002 the NCA intervened against the acquisition of Agder Energi by Statkraft. However, the decision was appealed by Statkraft to the Ministry of Labour and Administration, to which the NCA is subordinated. Although the Ministry fully agreed with the NCA's analysis of competition, it allowed the merger to go through, albeit with some modifications that were supposed to sustain competition in the relevant markets.

Later on the same year, the NCA imposed a resale order on a complete takeover of Trondheim Energiverk (TVE) by Statkraft, mainly on the basis of the stated effects on competition of transmission constraints in the area. In late 2005, the Ministry again agreed fully with the NCA's analysis of competition, but still allowed the merger to be completed. Just before that decision a change of government had taken place, when the former central/liberal coalition government was substituted by a social democratic/leftist coalition government.

In both cases, the Ministry has emphasised dynamic competition and efficiency arguments in its decisions, e.g. a greater potential for innovation from increased size of firm and the transfer of skills between the merged entities. It has also called upon investment in transmission capacity to overcome the adverse effects on competition of transmission constraints in the specific cases. The NCA has also become more preoccupied with the implications of transmission capacity constraints for market competition, as such constraints have gradually become more binding in the Norwegian electricity network.

Market monitoring and surveillance ex ante **versus control and enforcement** ex post
Generally, it is difficult to prove the actual abuse of market power in electric power markets, especially in a hydropower system in which the primary energy source, and, implicitly, the total production from a reservoir, is determined by the inflow of water. Given the special market properties of an electric power system it is not only market concentration as traditionally understood that matters; any plant on the margin in a restricted price area, even a small firm, may be in a position to abuse market power.

This raises the question whether a system of market monitoring and regulatory oversight ex ante of the electric power markets should be developed as an integral part, or at least as a supplement, to the competition control and enforcement ex post system. In such a system the Competition Authority could cooperate with other relevant agencies and institutions to develop the necessary data and information for market monitoring and surveillance, and for the actual oversight of the markets. For further details and suggestions, see Hope (2005).

The organisation of competition policy

Under the present system, the parties to a competition case can appeal a decision by the NCA directly to the Ministry to which the Authority is subordinated. As mentioned above, the Ministry has altered the decisions of the NCA in recent competition cases within the electric power sector. This raises important questions and issues, both in terms of the independence of the Authority in relation to the political sphere, as well as the competition policy analysis performed by the NCA and the Ministry, respectively. In most cases, the Ministry typically states that it agrees fully with the competition analysis of the Authority, but then in the end brings in arguments of a competition policy nature that overturn the decision, e.g. more emphasis on the scope for increased innovation by increased firm size through mergers, the potential for the transfer of competence and skills within a larger organisation, the expected empolyment effects of a merger, etc. The negative long-term effects on regulatory independence, trustworthiness of public agencies, and regulatory foresight in decision-making are serious and should be considered by the political system in Norway.
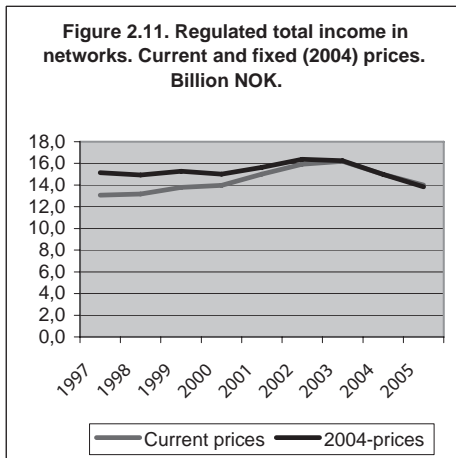
2.7.2 Regulatory policy

With the introduction of the new Energy Act, rate-of-return regulation was introduced for network companies. In 1997, this regulation was replaced by income regulation. An important aspect of income regulation is the efficiency rate, which is specific to the distribution network and, *ceteris paribus*, reduces annual allowable network-specific income. Given capacity and utilisation, the net tariff is reduced by this efficiency rate, which is based on both yardstick competition and a catching-up-period rule.[25] On average, the network tariff in Norway was expected to fall by about 20 per cent between 1997 and 2005.

Figures 2.11-2.12 show changes in total income and the network tariff in this period. Income was on an upward trend before 2003, when it fell. Over the whole period, real income fell by 1.5 per cent, which is less than the fall in the efficiency rate. This was mainly because of an increase in transmission capacity, as income per transmitted kWh fell by about 18 per cent in this period. Operating costs drove tariffs up, while the fall in interest rates reduced them.

Because regulation is more sophisticated and because supply and demand are stochastic, transmission tariffs and regulated income per transmitted unit may behave differently in the short run. However, in the longer run, transmission and distribution networks must pay back excess income and may add accumulated, but insufficient, income to future regulated income. The regulatory regime allows this adjustment to take time (several years). According to the regulatory authority, this fully explains the increase in tariffs over the last three or four years.

---

25 The catching-up period in this rule is the period that firms are allowed for catching up with the leading firm.

Figure 2.11. Regulated total income in networks. Current and fixed (2004) prices. Billion NOK.

Figure 2.12 Net-tariffs. 2004-prices. Nøre/kwh

Source: Statistics Norway and NVE
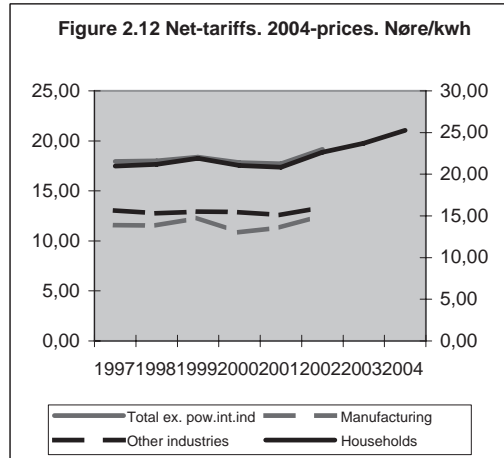
Source: Statistics Norway and NVE

As tariffs had previously been low, tariffs had to increase to make up the income short-fall. Since precipitation and inflow were low in 2002 and 2003, and the resulting high prices in the market reduced demand substantially, income regulation resulted in higher tariffs per transmitted and distributed kWh. According to the regulatory authority, *ceteris paribus,* tariffs are expected to fall over the next two years.

In the longer run we should expect interest rates, operating costs and the spot-market price (the price of transmission losses) to increase. This may offset the downward bias that is due to the yardstick efficiency gain measure.

Another important issue is whether the new regulatory regime provides sufficient incentives to invest in infrastructure capacity in this sector. This is a widely debated issue in Norway and represents a further regulatory challenge.

A new regulatory regime for network companies was, as mentioned, introduced in 1997. It was essentially a revenue-cap incentive mechanism,[26] but it contained elements of rate-of-return, price-cap and yardstick regulation. The regulatory period was defined for intervals of five years, but revisions could be undertaken during the period. An annual income frame was set *ex ante* by the regulator (the NVE) for each network company, of which there were 225, mainly distribution companies. The transmission company, Statnett, was regulated on the same basis. The revenue cap was based on the total cost coverage of network activities, including a stipulated rate of return on invested capital. This rate was set at 8.3 per cent for the first year. An efficiency improvement factor was defined for each network owner, based on a data envelopment analysis (DEA) of the efficiency improvement potential for each company. For the first regulatory year, the efficiency factor was set at 2 per cent of the income frame for all network owners. However, it was subsequently modified to depend on the efficiency position of each

---

26 In Norway, this is referred to as 'income frame regulation'.

owner in relation to the DEA-efficiency frontier. The highest annual efficiency requirement has been a factor of 4.5 per cent.

Although the regulatory regime was supposed to be evaluated and revised in 2001, following the first regulatory period, it was extended on more or less the same basis for the five-year period from 2002 to 2007. The NVE has commissioned much research and consultation on the design of the new regime to be implemented in 2007. An important and challenging issue facing the new regulatory model is the design of an incentive mechanism for optimal investment in the network that enables the market-based electricity system to function efficiently.

The need for harmonisation of national rules and regulations within the integrated Nordic electric power market has gradually become more pressing, to improve the efficiency and functionality of the common power market and network system. E.g., different methods and rules for the handling of transmission constraints operate side by side, from country to country, in the Nordic system, as do different regulatory regimes for network entities. There is an increased awareness of the negative effects on competition and resource allocation of the lack of harmonisation, calling also for a closer integration of competition and regulatory policies, both within and across countries in the Nordic area.

2.8 Some challenges

The Norwegian electricity market is integrated into a larger Nordic market. In a comprehensive EU-financed research project on European electricity reforms, known as SESSA,[27] the Nordic electric power market model was suggested as a potential benchmark for market organisation and the efficient functioning of electric power markets.[28] However, even if the Norwegian and Nordic electric power markets and their regulatory systems have performed reasonably well in terms of competition and economic efficiency, there is scope for improvement. Some issues and challenges in this context may be listed as follows.

1. *Market dominance and market power.* Investigations by competition authorities and research studies have not documented instances of the exercise of market power to the detriment of competition in the Nordic power market, either unilaterally or collectively. However, market power is a recurring issue in the debate on the Norwegian and the Nordic markets. This is partly because of the characteristics of electricity as a commodity in market terms and partly because of the increase in market concentration following restructuring of the market through mergers and acquisitions between electric power firms, as discussed in Section 2.7. The issue of market power suggests the need for the design of a system for market monitoring and regulatory oversight of the system, as argued by Hope (2005).

27 For documentation, see the SESSA webpage, www.sessa.eu.com.
28 See Hope and Singh (2006).

2. *Design and operation of investment markets.* The Nordic market has performed rea-
   sonably well in terms of the efficient operation of a market system within a more or
   less given capacity, because the built-up excess capacity prior to the market reform
   has not, so far, necessitated any new investment to speak of in capacity expansion.
   Thus, the market has still the optimal investment or capacity dimensioning test to
   pass. Generally speaking, an overall investment planning system for the Nordic elec-
   tric power system is not in place and investment markets for optimal investment
   within the integrated Nordic market are to a large extent lacking. Generally, this also
   applies to the Norwegian market.
3. *Network integration and system operation.* The Nordic transmission network system
   remains decentralised in the sense that national transmission companies are responsi-
   ble for the operation of, and investment in, the national network, and for system
   operation. Cooperation between transmission companies takes place on a voluntary
   basis through NORDEL. As mentioned above, the regulation of network companies
   and the handling of network constraints are not harmonised on a Nordic-wide basis,
   which results in potential inefficiencies in the functioning of the power markets. A
   common, independent transmission system operator for the integrated Nordic market
   is also lacking.
4. *Integration of the Nordic market with the European electricity market at large.*
   Economic efficiency could be increased if the Nordic market were more closely inte-
   grated with the European electricity market. Although insufficient transmission
   capacity limits such integration, transmission investment is planned. For example, an
   subsea cable between Norway and the Netherlands is being developed. The more
   mature Nordic market in terms of market organisation, competition and regulation,
   may promote power market liberalisation in Europe.

2.9 Summary

During the regulation period, investment in production and transmission capacity in the
electricity market was subject to cost reimbursement, through either direct prices in the
market, cross-subsidisation between utilities or direct public subsidies. There was no
direct link between market prices and investment or between market prices and operating
cost efficiency. Several studies report substantial inefficiencies in the production, trans-
mission, distribution and market distribution of electricity.

The new deregulated market was intended to build on the principles applied in an
already existing spot market for interruptible power. Vertically integrated power compa-
nies were split into divisions on an accounting basis. A derivate market was opened to
deal with hedging against uncertainty. Introducing common carriage and securing access
to the grid on a transparent and nondiscriminatory basis opened up the electricity net-
work. The network companies were subject to regulation, the objective of which was to
increase economic efficiency.

Deregulation produced the following benefits:

- Electricity prices fell.
- Prices between consumer groups became more equal, except for the power intensive industry, but this will eventually happen when long term contracts expire in 2008-2011
- Investment declined in both production and transmission capacity which made a relocation of capital to more profitable investments possible.
- Over time, the return on capital in the hydro power increased, although still lower than we should expect in the long run which also explains the remaining low investment pattern in the electricity industry.
- The spilling of water, which was a real waste, also vanished. For instance in 1992, the large inflow to the hydro reservoirs produced electricity prices as low as 2 øre/kWh (0,3 cent).
- The stochastic electricity market is occasionally tested by extreme events, particularly on the supply side. However, the market seems to have handled these events well.

In addition, market concentration by mergers and acquisitions has been allowed, which has increased the opportunities to exercise market power, especially as the market has become more regionalised because of transmission constraints. However, market power does not seem to have been abused so far, but market concentration is a potential problem.

Some challenges also remain with respect to the design and operation of investment markets, network integration and system operation, and integration of the Nordic and European electricity markets.

References

Bye, T. (1991): 'Elforsyning og markedet for elektrisk kraft', *NOU 1991:28 - Mot bedre vitende?,* pp 67-71. Electricity supply and the electricity market, *NOU 1991:28 - Against better knowledge, pp 67-71.*

Bye, T. (2003): 'A Nordic Energy Market under Stress', *Economic Survey,* No.4, pp 26-37.

Bye, T., N. H. M. v. d. Fehr, Christian Riis and Lars Sørgard (2003): *Kraft og makt - en analyse av konkurranseforholdene i kraftmarkedet (Electric Power and Power - a study of competition in the power market),* Report of an Expert Group appointed by the Norwegian Ministry of Labour and Administration.

Bye, T., M. Hoel and S. Strøm (1999): Et effektivt kraftmarked konsekvenser for kraftleverende næringer og regioner. *An efficient power market - consequences for power intensive industries and regions,* søs 102, Oslo: Statistics Norway.

Bye, T, and Einar Hope (2005): "Deregulation of electricity markets. The Norwegian experience", *Economic and Political Weekly,* Vol XL No 50, December 10-16, 2005, also as Discussion Paper 433 of Research Department, Statistics Norway.

Bye, T. A. and T. A. Johnsen (1991): 'Effektivisering av kraftmarkedet (Efficiency gains in the power market)', *Rapporter 91/13,* Statistics Norway.

Bye, T. A. and S. Strøm (1987): 'Power prices and power demand (kraftpriser og kraftforbruk)', *Sosialøkonomen,* No.4, pp 19-29.

Fehr, N. H. M. v. d., E. S. Amundsen, and L. Bergman. (2005): 'The Nordic market: Signs of stress?' *The Energy Journal,* Special Issue on European Electricity Liberalisation, 2005.

Førsund, F. (1994): 'Optimal running of a hydropower system (Driftsoptimalisering i vannkraftsystemet)', *SNF-rapport 29/94,* Oslo.

Førsund, F. A. and S. A. C. Kittelsen (1998): 'Productivity development of Norwegian electricity distribution utilities', *Resource and Energy Economics,* No. 20, pp 207-224.

Hope, E. (2000): *Studier i markedsbasert kraftomsetning og regulering (Studies in electric power markets and regulation),* Bergen: Fagbokforlaget.

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

Hope, E., L. Rud, et al. (1992): 'Markets for electricity in Norway', in E. Hope and S. Strøm (editors): *Energy markets and environmental issues: A European perspective,* Oslo: Scandinavian University Press.

Hope, E. and B. Singh (2006):"Harmoniting effective regulation", in Team. Michel Glachant and Francois Lévêque (eds): *Achieving European Union electricity markets in 2009? Assessment and proposals*. EU SESSA.

Kittelsen, S. A. C. (1993): 'Stepwise DEA. Choosing variables for measuring technical efficiency in Norwegian electricity ditribution', *Memorandum no.6/93,* Department of Economics, University of Oslo.

Kittelsen, S. A. C. (1994): 'Efficiency and regulation in Norwegian electricity distribution. (Effektivitet og regulering i norsk elektrisitetsdistribusjon)', *SNF Report no 3/1994,* SNF Foundation for Research in Economics and business Administration, University of Oslo.

Midttun, A. (1987): *Segmentation, institutional lags and manufacturing adaptation; Norwegian power capacity expansions political economy during 1970s and the 1980s. (Segmentering, institusjonelt etterslep og industriell omstilling: norsk kraftutbyggings politiske økonomi gjennom 1970- og 1980 årene),* PhD, Uppsala University, Sweden

Newbery, David M. (1999): Privatization, restructuring, and regulation of network utilities. The MIT Press.

Nordic Competition Authorities (2003): A powerful competition policy. Towards a more coherent competition policy in the Nordic market for electric power. *Report 1/2003.*

Dag Morten Dalen and Steinar Strøm

The pharmaceutical market in Norway

3.1 Introduction

In January 1996 a commission was appointed by the Norwegian government and asked to prepare an Official Norwegian Report concerning the structure of the Norwegian pharmaceutical market, ranging from ownership to price regulations. The commission delivered its report in January 1997 (NOU 1997:6)[1]. The report contained many proposals with regard to restructuring the Norwegian pharmaceutical market, with varying majority votes behind each of the proposals. Most of the proposals backed by the majority in the committee were later implemented through a New Pharmacy Law of March 1st 2001.

There were several reasons why the Norwegian government asked for a report on the structure and functioning of the pharmaceutical market. In the first place, the structure of the market, location of the pharmacies and the way of selling drugs, together with formalities related to entry and exit of Norwegian pharmacies, had been kept the same for decades, or even centuries. The license to open and operate a pharmacy was given by a government agency (Statens Helsetilsyn) to persons with a cand.pharm degree. In the period 1980 to 1996 15 of 300 licenses, varying in value with regard to the location of the pharmacy, were given to government officials with relations to the offices regulating the pharmacies. The procedure followed for the approval of pharmacy licenses implied that there should be a certain geographical distance between the pharmacies even in densely populated areas. Thus, the decisions regarding the entry, location and ownership of pharmacies were based on administrative decisions with loose connection to changing market conditions.

Second, the selling of non-prescribed drugs was restricted to be done through pharmacies. However, there could be good reasons for letting these drugs be sold outside pharmacies, say, in supermarkets and at gas stations.

Third, there was a growing concern among politicians that public expenditure on prescribed pharmaceuticals could place a steadily increasing and non-sustainable burden on public finances. The Norwegian Health System offers statutory public health insurance, and close to 70 percent of total drug expenses are covered by the insurance scheme. These expenses have been increasing rather rapidly due to an increasing ageing population and entry of new and more expensive drugs. Figure 3.1 shows total annual sale of pharmaceuticals in Norway during the period 1980-2005 (in 2005-prices):

---

1 S. Strøm was the chairman of the commission.

Figure 3.1: Annual sale of pharmaceuticals in Norway during the period 1980-2005 (in 2005-prices). Source. Norwegian Association of Pharmaceutical Manufacturers (LMI)
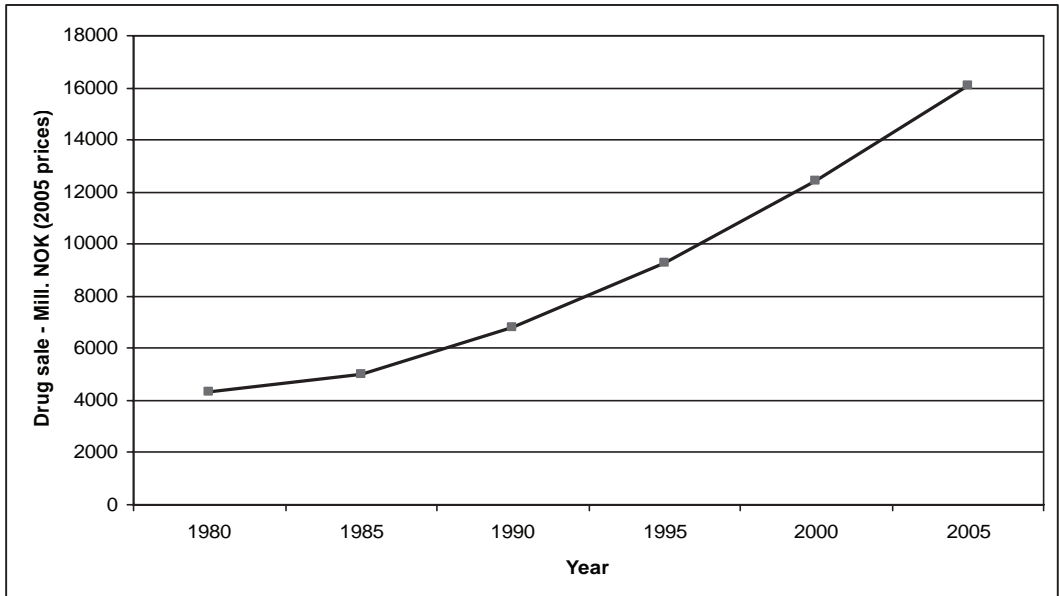
In 2003 the share of total costs on reimbursable medicine paid by the patient amounts to 8.5 percent in Norway, compared to 31.4 percent in Denmark and 22.1 percent in Sweden.[2] Due to these differences one should expect that price competition is stronger in the neighboring Scandinavian countries than in Norway. Various price regulations had been tried before the New Pharmacy Law was passed in 2001, and none proved to be very successful in terms of reducing the prices paid by the patients and/or covered by the tax-payers. There are many reasons for this and they are all related to weak price responses.

Third-party financing of prescribed drugs imply that the patient, or rather his or her agent, the medical doctor, has weak incentives to choose the cheapest drug among drugs with the same chemical substance and/or the same treatment effects. The more the tax-payers cover the costs, the weaker one should expect the price responses to be. Moreover, the medical doctors need not be perfect agents for their patients. Ties to the pharmaceutical industry may hinder otherwise better choices for the patients and/or for the tax-payers. During the work of the commission referred to above, a GP in the city of Oslo was asked by the commission to register all contacts with the pharmaceutical industries (NOU 1997:6). It turned out that during 45 days in the fall of 1996 the GP received 84 surface mails from 33 different pharmaceutical companies, 19 invitations to meetings and conferences, 6 invitations to what was termed "gastro-events", 22 invitations to lec-

---

2 LMI. Facts and figures 2006.

tures, 18 dinner invitations and 6 travel proposals of which 2 were to other countries. Thus, to enhance price responses based on decisions made by the patients and medical doctors is a difficult task. To increase price responses, a better alternative may be to induce the pharmacies to choose cheaper drugs among otherwise equal alternatives. This was also one of the main issues discussed in the commission of 1996.

In the next sections we will present and discuss the most important aspects of the New Pharmacy Law. Most of the focus will be on price regulations and their impact on drug expenses.

3.2 The New Pharmacy Law

With the New Pharmacy Law of March 1st 2001, restriction on ownership was abolished. It was not longer required that the owner of the pharmacy should have a cand.pharm. degree. The practice of a government agency giving licenses to open and operate pharmacies to individuals with this type of education, and in some cases also with a career in government offices and agencies, came to an end. Also restrictions on the location of new pharmacies were abolished, at least after the law was passed.

The new law allowed wholesalers to own the retailers, but pharmaceutical companies and medical doctors were not allowed as owners or shareholders in firms taking part in the distribution of pharmaceuticals (wholesalers as well as retailers). The new law implied a substantial liberalization with respect to entry into distribution of pharmaceuticals and ownership. Soon after the law was passed many of the pharmacies were sold to (three) wholesalers.

When the law came there was a cap on the opening of new pharmacies in central areas of Norway, in cities and in other urban areas. The reason for this was that the government feared that too many new pharmacies would be opened in central Norway with the implication that pharmacists working in the rural Norway would move to the central part of Norway. Due to this restriction the price that the new owners, the wholesalers, had to pay for pharmacies in central Norway was rather high and implied a handsome transfer of money to the previous owners of the pharmacies. However, as pointed out by the majority in the commission referred to above, there were many means to maintaining pharmacy services in rural Norway at the same level as before the liberalization of the pharmacy market. Thus, these fears of a reduction in the service levels in rural Norway were exaggerations, and a cap on the number of new pharmacies in central Norway was not needed. Around the dates when the new law was passed in March 2001 one of the wholesalers, NMD, was in contact with the government and promised to service rural parts of Norway in case of problems with the distribution of pharmaceuticals in these parts of the country, provided that restrictions on the entry of new pharmacies in central Norway were lifted. The government said yes to this proposal and removed the restrictions on entry of new pharmacies in central Norway. At the time when this was known other wholesalers had already started to negotiate take-over contracts, including the price, with pharmacies in central Norway. In these early stages of the ownership changes

in central Norway the new owners paid a higher price than later equilibrium prices implied.

The agreement between the government and the wholesaler NMD to service rural parts of Norway in case of delivery problems was later extended to another wholesaler, APOKJEDEN (Apotek 1). This agreement of servicing rural Norway has only been enforced three times and the agreement expired February 28th 2004. For more details see Brekke et al (2003).

3.3 Ownership and pharmacy availability.

Almost immediately after the new pharmacy law was passed three wholesalers bought up most of the privately and pharmacist owned pharmacies. Almost overnight the pharmacy structure in Norway, with private owners, regulated entry determined by government officials with licenses to operate pharmacies given by these civil servants to pharmacist, with no price competition and local monopolies, was replaced by three retail chains. By now, somewhat more than 90 percent of total retail sales of pharmaceuticals are covered by these three retailer chains, each of them owned by three wholesalers. These three retail chains and wholesalers in Norway are:

1. Retailer chain: Apotek 1, wholesaler: Apokjeden; main shareholder Tamro OYJ, which is the biggest wholesaler of pharmaceuticals in the Nordic countries,
2. Retailer chain: Vitusapotek and Ditt Apotek, wholesaler: NMD AS; owned by the German wholesaler company Celesio,
3. Retailer chain: Norges Apoteket and Alliance apotekene, wholesaler Holtung; owned by the British wholesaler Alliance UnicChem Plc.

In addition to the retail chains there are pharmacies in hospitals and a very few still privately owned pharmacies. Their share of total sales amounts to less than 10 percent. The new market structure of the Norwegian pharmacy sector implies horizontal integration at the retail level and vertical integration across wholesalers and retailers.

Compared to the structure prior to the passing of the new law, this new market structure has implied an increased competition with regard to location and quality of service. First, the total number of pharmacies has increased. At the time when the new law was passed, there were around 370 pharmacy outlets in Norway. One year later it had increased to around 480, and in the end of 2005 there were 560 pharmacies. 93 percent of these pharmacies are either owned by or have entered purchasing agreements with one of the three vertically integrated pharmacy chains Apotek 1, Vitus and Alliance. Most of the increase has taken place in cities and other urban areas. In the same period the number of inhabitants per pharmacy has declined from 11 281 to 8754, which implies that pharmacy coverage has moved towards the coverage in the US (around 5000 inhabitants per pharmacy) and continental Europe (less than 5000 inhabitants per pharmacy), NOU 1997: 6. Second, opening hours have increased, on average by 2 hours per week. Taking into

account the increase in the number of new pharmacies, the total number of "pharmacy hours" has increased from around 17000 at the end of 2000 to 23000 at the end of 2003 (ECON, 2004). Self-service pharmacies have become more common and by now above 80 percent of the pharmacies are organized as self-service pharmacies. Availability of pharmacy services, including shorter waiting time, has increased rather strongly as a result of the new market structure, and this change was also one of the most important reasons for making the change in the market structure, see LOV 2000-0602, 39 and the discussion of these issues in NOU 1997:6.

Moreover, despite the high increase in pharmacy service availability, the total use of labor in retailing is approximately the same as before the change in market structure in 2001, ECON, 2003. The number of employees with a cand.pharm degree per pharmacy has declined, while the number of employees with lower education has increased a little.

Thus the change in market structure that the Law of March 2001 allowed for has improved considerably the availability of pharmacy services, in particular in the densely populated areas of Norway, with less use of scarce resources in each pharmacy than before.

Most likely as a consequence of the new law, and proposed by the commission of 1996, non-prescribed pharmaceuticals, in particular some pain-killers in certain small packages, have become available outside pharmacies. From November 1st 2003, supermarkets and gas-stations were allowed to sell these drugs.
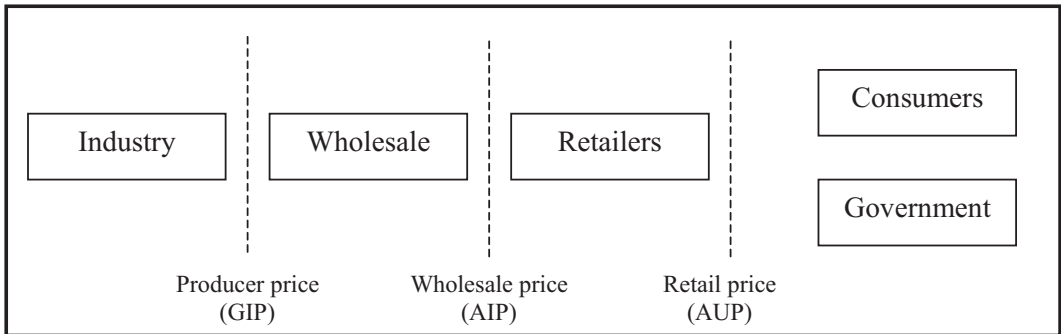

3.4 Price regulations

In Norway, regulations of prices on pharmaceuticals were introduced as early as in 1928. One of the goals was to ensure equal prices on the same drugs in all parts of Norway. In the beginning, prices at all stages in the distribution chain were controlled. Thus, prices and margins at the wholesaler and retailer level were all fixed.

As of today, pharmaceutical prices and margins in pharmacies are based on Government Regulations of December 16th 1994, which were implemented in January 1995. In order to identify how price regulation interacts with competition it becomes critical to describe the exact type of regulation. The regulation scheme consists of a base model that involves all prescription drugs, and a generic drug model that regulates prices on pre-scription drugs that no longer have patent protection.

Given the current market structure, several prices are present.

Figure 3.2: Prices in the pharmaceutical market

Industry	Wholesale	Retailers	Consumers
			Government

Producer price (GIP)   Wholesale price (AIP)   Retail price (AUP)

This new regulation scheme of 1995 abolished the regulation of the producer price (i.e. the wholesalers' purchasing price – in Norwegian termed GIP). Instead two price caps were introduced, one on the retailers' purchasing price (in Norwegian termed AIP) and then one on retailers' margin when selling the drug to the patient – in effect determining a maximum retailer price (in Norwegian termed AUP). In a new Government Regulation of December 22nd 1999, the cap on the retailers' purchasing price (AIP) was changed a little and it now implies that the cap on the price shall reflect the average of the three lowest prices on similar drugs (same chemical substance) in some selected European countries like Sweden, Finland, Denmark, Germany, the UK, the Netherlands, Austria, Belgium and Ireland. The Regulation of 1999 mentioned that to some (minor) extent and in certain cases one should pay attention also to prices on drugs sold in the Norwegian market and with the same treatment effect and/or to the cost of producing and distributing drugs. For practical purposes, however, the Regulation of 1999 implied that maximal prices on drugs in the Norwegian pharmaceutical market are set according to an international reference pricing system. In the Government Regulation of 1999 there is no mentioning of how low drug prices may make it difficult for the pharmaceutical industries to cover their research and development costs.

The regulation of the retailer prices in Norway (AUP) is a drug-specific cap on the retailer margin in terms of a percentage increment of the cap on the retailers' purchasing price and a fixed amount in NOK per drug. The maximum of the retailers' price on a certain prescribed drug (max AUP) equals the maximum of the retailers' purchasing price (max AIP) plus the regulated max retailer margin per drug. Prices on non-prescribed drugs are not regulated.

The argument for the regulation of prices at the retailer level, both the retailers' purchasing price and the drug specific retailer margin, is lower prices and lower burdens on taxpayers. With low price and/or no price responses among the consumers (patients/doctors) the Government fears that prices would have rocketed.

With the regulation of December 1994 a Benefit-sharing-model was introduced. This model states that if a retailer buys drugs at lower prices than the regulated max retailer purchasing price, then the pharmacy and the customer (the patient) shall split the difference between the regulated max retailer price (max AUP) and the retailer price that follows from adding the regulated max retailer margin to the actual price that the retailer paid the wholesaler. The idea behind this model is to strengthen the price responses among some of the agents in the market, the pharmacies, and to give them an economic incentive to choose cheaper drugs among those available in the market.

However, as explained above, the New Pharmacy Law of 2001 paved the way for wholesalers being the owners of retailers. This vertical concentration in the distribution of drugs, combined with regulation on the retailers purchasing price, which is an internal price in the vertical market structure, made the benefit-sharing model counterproductive.

To illustrate this problem, let

AUP= the actual retailer price of a prescribed drug,
Max AUP= the regulated maximum retailer price
AIP= the actual retailers' purchasing price
Max AIP= the regulated retailers' purchasing price,
GIP= the wholesalers' purchasing price
RM= the retailers' margin
RWM =the vertical owned retailers'/wholesalers' margin.

Moreover, let NOK 20 be the maximum amount in NOK in the regulated retailers' margin and let the percentage increment on purchasing price allowed for in the regulation of the retailers margin be 8 %. For expository reasons we will assume that AUP=max AUP, which there are many reasons to expect also will be the actual outcomes in the market.

According to the benefit-sharing model, we then get

$$(1) \quad MaxAUP = AUP = 20 + 1.08\frac{MaxAIP + AIP}{2}$$

Because RM= AUP-AIP, we get

$$(2) \quad RM = 20 + 0.54MaxAIP - 0.46AIP$$

and

$$(3) \quad AUP = RM + AIP = 20 + 0.54MaxAIP + 0.54AIP$$

Thus, if the retailer and the wholesaler were two different agents and with no common owner, the benefit-sharing-model and a regulation of the retailers' purchasing price as in the regime of December 1994 would work. For each NOK the retailer is able to negotiate down the actual purchasing price, AIP, the retailer earns 0.46 NOK and the patient

and/or taxpayer gets a benefit of 0.54 NOK. Thus, there was a good argument for having the benefit-sharing-model, provided that there was no ownership link between the retailer and the wholesaler. To what extent actual retailer prices would be affected by this benefit-sharing-model depend on the bargaining power of retailers relative to wholesalers. If several independent wholesalers competed in the market, one could expect that wholesale-margins would have reflected wholesalers' costs. In addition, with generic drugs available, a combination of benefit-sharing regulation and independent retailers could put downward pressure on producer prices (GIP).

The New Pharmacy Law of 2001 resulted almost immediately in a change in the ownership structure in the distribution of drugs. As outlined above, three wholesalers got ownership control of almost the entire retailer segment of the market. With vertical ownership in this market and keeping the benefit-sharing-model, together with regulating the retailers' purchasing price, we now get that

(4) $RWM = (AUP - GIP) = 20 + 0.54\ MaxAIP + 0.54 AIP - GIP$

Thus, the vertical owned retailer/wholesaler would now benefit from having a purchasing price AIP at the maximum level, that is AIP=Max AIP. This entity would then have no incentive to lower the actual AIP. Of course, the retailer/wholesaler company would have an incentive to search for cheaper products in the producers' markets and to negotiate down its purchasing price GIP.

In the Commission that delivered its report in 1997, a majority argued that the government should go back to the regulation scheme prior to 1995, in which the wholesalers' purchasing price, GIP, was regulated and set in negotiations between the industry and the government. Doing this and continuing with the benefit-sharing-model one could induce the retailer/wholesaler to search for cheaper drugs to the benefit of the retailer/wholesaler and the patient/tax-payer. With international and large wholesalers one should also expect that the bargaining power of this new agent could match the bargaining power of the pharmaceutical industry.

With the New Law of Pharmacy of 2001 the government liberalized the Norwegian pharmacy market, increased the availability of pharmacy services and created a structure in the distribution of drugs that was more cost-efficient than before and with the potential of matching the bargaining power of the pharmaceutical industry. But by keeping the regulation of the retailers' purchasing price, AIP, and not switching back to the old regime before 1995, in which the wholesalers' purchasing prices were regulated, the incentives to reduce drug prices and search for cheaper generics vanished.

3.5 Regulation and competition

Of great concern for the authorities was the need to induce the agents in the market to choose drugs that could reduce prices paid and reimbursed by the government. During the years prior to the new law was passed, several price reforms had been implemented,

with the purpose of encouraging the agents operating in the market to switch to cheaper drugs when available, i.e. to switch to generic drugs. Parallel imports were introduced in 1998, and pharmacists were allowed to keep 50 percent of the savings if they were able to obtain lower prices on these drugs (Benefit-sharing model). In 1999 doctors were required to prescribe generics with the lowest price. However, this recommendation and the reference price system used proved not sufficient to increase physicians' awareness of prices.

Within the current regime, competition in the retail market is only related to location and quality (such as opening-hours and staffing). There are no reasons for the patients to choose pharmacies on grounds related to prices.[3] As long as the patient accepts the drug proposed by the chosen pharmacy (if options are available at all), the co-payment by the patient will be zero or negligible. Without changing the role of co-payment, the demand-side cannot trigger downward pressure on prices. Instead, regulation has tried to give pharmacies incentives that could compensate for the lack of a price-responsive demand side.

For innovative drugs that still enjoy patent protection, the pharmacies will not be in position to put downward pressure on prices. Price regulation makes sure that the prices on these drugs are in line with the prices offered in other comparable countries. Prospects for triggering price competition are better for older drugs, where generic substitutes are available in the market. Much of the recent focus – both by the regulatory authority and the Norwegian Competition Authority – has been on the ability of the market to benefit from generic competition. There is a common understanding that the current regime has succeeded in creating price competition among generic producers. These producers now meet large pharmacy chains with strong incentive to reduce their purchasing prices (GIP).[4] However, getting lowered wholesale purchasing prices transferred into lower retail prices has proved much harder. Below we review some of the actions taken by the government to benefit from generic competition.

Generic substitution

Following up a proposal given by the majority in the commission referred to above, generic substitution was introduced in March 2001, with the purpose of making the actual choice of drugs less dependent on the physicians' prescription policy. For a sample of about 100 drugs (chemical substances), pharmacies were permitted to substitute a cheaper generic (if available) for a branded drug, independent of which producer the physician had prescribed. However, the physician and/or the patient had the right to prevent this substitution, but then the physician had to state so explicitly on the prescription. This

---

3 Not all drug expenses are covered by the insurance scheme. These patients will have incentives to buy at lower prices. However, since close to 70 percent of total drug expenses in Norway are covered by the scheme, the behavior of the insured patients will be important for the pharmacies.

4 Systematic evidence on how these prices have developed is not available. There is no official reporting on GIP.

new policy gave the pharmacy an important influence on whether a branded drug or a generic drug should be chosen. Because there is no direct link between the chains' purchasing prices (GIP) and the retail price, the pharmacy chains were actually free to set the retail prices on the generic substitutes – as long at the prices stay below the maximum retail price. The only reason why one should expect reduced retail prices from generic substitution in such a situation, is that patients need to see a significant lower price to accept another product (producer) than the one chosen by the doctor on the prescription. Looking at the prices (AIP), mechanisms like this seem to be in force. Dalen (2003) looked at monthly data over the period 1995-2002 for all drugs subjected to generic substitution regulation. A simple regression with log(price) as the dependent variable gave the following result:

Table 3.1: Regression with log(price) per DDD (defined daily doses) as dependent variable

	Estimates	S.E
Generic drug (Yes=1)	-0.152***	0.025
Generic substitution (YES=1)	-0.049***	0.016
N (number of firms producing generics)	-0.022***	0.006
Generic drug*Generic subst.	-0.084***	0.027
N*Generic drug	-0.007	0.007
T (Trend)		YES
Number of observations		7119
R^2	0.92	

***:Significant at ≤1% level

We see from the generic drug dummy that even before the introduction of generic substitution in 2001, the generic drug had a substantial lower price compared to the original patented drug (brand name). Generic substitution increased the price difference. There is nothing in the price cap regulation scheme that forces the pharmacy chains to operate with these lower prices. They are free to set prices equal to the maximum price. Nor are the lower prices on generic drugs explained by strong retail price competition among the pharmacy chains. Instead, these prices can reflect that lower prices need to be offered in order to move the patients towards (by substitution) the chain's preferable drug. The chain's preferable drug will be the drug with the most favorable wholesale purchasing prices.

Despite the fact that generic drugs are sold at lower prices compared to brand name drugs, market shares of brand name drugs are still high. In 2001 more than 50 percent of the drug units sold, for which generic substitution is allowed, were the more expensive original patented drugs. In 2004 this share was reduced to 40 percent. The Benefit-sharing model introduced before the new law, together with vertical integrated whole- and retail sellers and the continuation of regulating the purchasing price of the pharmacies

(AIP), proved to be counterproductive in creating a more lively market for generic drugs and to enhance both price and demand responses.

With respect to the substitution of brand name drugs by generic drugs, a regulation in the Pharmacy Law is of particular interest. According to the Law, wholesalers are obliged to carry any pharmaceutical products in demand in the Norwegian market and they must be able to satisfy demand anywhere in Norway within 24 hours. Hence, if a pharmacy places a delivery order, the wholesaler must be able to meet the demand without delay. For patented drugs, this places the pharmacy chain in a very weak bargaining position vis-à-vis the producer. The wholesaler's selling price is capped by regulation (Max AIP), while the wholesaler's purchasing price is unregulated. Pharmaceutical firms producing patented drugs will often have a portfolio of products, some off-patent and some on-patent. A firm with such a portfolio could easily punish a chain for entering a purchasing agreement with a cheaper generic producer, just by increasing the prices on patented drugs.

What might protect the chain from "retaliation", however, is parallel import of brand name drugs. For sufficiently high purchasing prices, the wholesaler will find it profitable to switch to imported versions of the original brand name drug.

Index price - incentives

Since the market proved to be slow in switching to cheaper generic drugs after the introduction of generic substitution, the government introduced a new price regulation scheme[5]. As explained above, one of the main weaknesses with the base model had been that no direct link existed between the wholesalers' purchasing prices and the retailers' price caps. With a regulatory change in March 2003 such a link was established for a selection of drugs that had been subject to generic substitution. A retail price cap, termed "*index price*", on six different drugs (chemical substances) was introduced. The index price defined the maximum retail price. When a pharmacy sold a reimbursable drug to a patient, the pharmacy would get the index price (minus the patient's co-payment) from the National Insurance Scheme. The index price was set equal to the average of reported wholesale purchasing prices on that drug, plus a pre-determined distribution (wholesale and retail) margin.

Again, let NOK 20 be the maximum amount in NOK in the regulated retailers' margin and let the percentage increment on purchasing price allowed for in the regulation of the retailer's and wholesaler's margin sum up to 18 %. Then the index price can be written as

(5) $\qquad I = 20 + 1.18\,AGIP$

---

5 Due to a proposal by D. M. Dalen.

AGIP is the average wholesaler purchasing price, given by

(6)
$$AGIP = \frac{\sum_i GIP_i \cdot Q_i}{\sum_i Q_i}$$

Here GIP_i is the wholesaler's purchasing price from producer i, and Q_i is the number of units sold by producer i.

The index price was updated every third month. The reimbursed price is now independent of the pharmacy chain's actual purchasing price. If a pharmacy selects a producer with a price exceeding the average of the three lowest prices, the net margin of the integrated retailer-wholesale pharmacy firm drops, whereas a retailer selecting a producer with a lower producer price experiences an increase in his net margin. This incentive to negotiate and sell cheaper generic drugs, will bring down reimbursable prices in the future. In this way, the index price model is equivalent to yardstick competition regulation.

Since the index price is calculated with a percentage based retail chain margin (18 % in total) it has been held that the pharmacy chains do not have an incentive to negotiate lower prices (despite the short term effect) since this just brings down future average purchasing price. With three retailer chains in the market, that will not be the case unless the chains are able to coordinate their behavior in the market. If a pharmacy chain negotiates a lower purchasing price, this will only have a long-term impact (after updating) on the average price of about 1/3 of the reduced purchasing price.[6] Within relevant ranges of percentage mark-ups, this will be sufficient to sustain a long-term negative relationship between the purchasing price (GIP) and profit.

In terms of aggregated profit for all three chains, however, the long-term profit is positively related to the purchasing price. This is a standard prisoner's dilemma situation in which the dominant strategy for each chain is to negotiate a low purchasing price, but where there are mutual benefits to be shared by coordinating to high purchasing prices:

---

6 If we assume the pharmacy chains have equal market shares. Note that the patient's co-payment will always be independent of a chain's purchasing prices.

Figure 3.3: Incentives in the index price scheme. Numbers indicate profit.

For the purpose of illustration, if the two pharmacies accept a high purchasing price, they would earn 10 each in profit. However, whatever purchasing price the other pharmacy has obtained, the best strategy for the pharmacy will be to negotiate low prices. When both pharmacies have a dominant strategy in getting low prices, they will end up in a situation in which they earn a profit of 6 each.

With only three pharmacy chains present in the market, one might fear that the benefit of coordination could sabotage the index price scheme. Coordination is facilitated by the relatively concentrated producer industry, and the fact that these have mutual interest with pharmacy chains in supporting high prices. The number of producers is often in the range of 4-5 to 10 when generic drugs are available.

Index price - experiences

Dalen et al (2006) have studied the competition between generics and branded drugs after the introduction of the index price. An empirical model with two components was developed. First, a demand model in which doctor-patient's choices follows from a discrete choice structure with random utility function is estimated. Second, assuming that the drug producers set prices non-cooperatively to maximize profit and adhering to the estimated price elasticities, a time-conditioned measure of market power – the Lerner index – for each product is derived.

The dataset was provided by the Norwegian Social Insurance Agency, and covers monthly observation of the six chemical substances included in the *index price* system. The data are collected at 22 pharmacies in Norway in the period 1998-2004. The sample of pharmacies is considered to be representative for the sale of drugs in Norway. The main variables reported by the pharmacies are volume of sale, both in retail value (NOK) and

number of defined daily doses (DDD) for each product (as of Nov 2005 1USD is about NOK 6.80). These are used to calculate the prices per DDD and market shares of each product within the submarket (chemical substance).

The chemical substances subjected to index price regulation are listed in Table 3.2.

Table 3.2: Sample of drugs

ATC-code	Generic name	Indication	Brand name producer	First entry of generica
A02BC01	Omeprazol	Ulcer	AstraZeneca	November 2001
C09AA02	Enalapril	High blood pressure	MSD Norge	October 2000
C09AA03	Lisinopril	High blood pressure	AstraZeneca, MSD Norge	November 2000
N06AB04	Citalopram	Depression	H.Lundbeck	May 2002
R06AE07	Cetirizin	Allergy	UCB Pharma	February 2002
R06AX13	Loratadin	Allergy	ScheringPlough	April 2002

Modelling demand with a discrete choice structure, with random utility function, gives multinomial logit choice probabilities, which can be shown to yield the following log-odds rate regression:

$$(5) \qquad \ln \frac{\varphi_{imt}}{\varphi_{1mt}} = g_{mi} + a_0 (P_{imt} - P_{1mt}) + a_1 GD_{imt} \left[\frac{A_{imt}}{A_{1mt}} P_{imt} - P_{1mt} \right] + a_2 GD_{imt} \tau_t + v_{imt}$$

Here φ_{1mt} is the probability that drug i in market m at time t is chosen. Drug 1 is the original brand name drug. The log of the market share of drug i relative to the market share of drug 1 can be used as a proxy for the log of the relative choice probabilities. P_{imt} is the price of drug i in market m at time t. GD_{imt} is a dummy variable. $GD_{imt} = 1$ if drug i in market m is a generic drug. A_{imt} equals the number of months drug $i \neq 1$ has been on the market. Thus the ratio of A_{imt} to A_{1mt} measures the "market age" of drug i relative to the "market age" of the branded product. If the drug has been in the market since 1998, the ratio is equal to one. For younger drugs, the ratio takes values less than one. Total price response for the branded drug is captured by the direct effect a_0, where-as for the generic drugs "the market age" of the drug is allowed to matter. If, for given prices, generic drugs experience increased demand as market age increases, a_1 will be positive. The index price scheme was introduced March 2003. In order to identify the effect on demand for generic drugs, this new policy is represented in the log odds ratio equation above by the variable τ_t, which takes the value 1 for t=March 2003 and the following months, and 0 otherwise. v_{imt} is a random variable assumed to be white noise.

If the index price succeeded, a_2 will be positive. For given prices, generic drugs will experience a positive shift in demand. Table 3.3 shows the estimation results:

Table 3.3: Estimates of the demand equation

Coefficients	Estimates without instrument variables	t-values	Estimates with instrument variables	t-values
a_0	-0.8324	-6.34	-2.1125	-6.03
a_1	0.2607	3.52	-0.1378	-1.47
a_2	1.1594	8.12	2.3964	10.63
$\bar{a}_m$ (fixed effect estimation)	-0.1113	-1.18	-0.7697	-5.52
R^2	0.2034	0.1816		
Number of observations		1159		

When the model is estimated without using instruments we observe from Table 3.3 that the direct effect of prices on demand is negative and significant. However, all pharmaceuticals have side effects. These side effects are likely to be known by the medical doctors. The less negative these side effects of a certain drug are, the more likely it is that this drug will be chosen. The producer knows this. Hence drug with less negative side effects may get a higher price in the market. As econometricians we do not have this knowledge and we do not observe the side effects either (qualitative effects of the drugs). To us these side effects may be present in the random terms in the demand equations (5) above. Due to the pricing strategy of the firm a correlation may arise between the price and the error term in the demand equations. Ignoring this correlation when estimating the model may yield biased estimates and we would expect that price responses are underestimated. What we wrongly get as weak price responses may be due to the fact that more expensive drugs are just bought because of some unobserved drug characteristics. In order to account for this possible endogeneity bias we have instrumented the price in the demand equation (5) above. The ideal instruments should be correlated with the price but not with the unobserved qualities of the drugs. From Table 3.3 we clearly see that when the prices are instrumented[7], the direct price responses become numerically much stronger.

The impact of "market age" on demand is significantly positive when no instruments are used, but disappears when instruments are used. The impact of the index price on demand for generic drugs is significant and positive in both models. Using instruments, the effect becomes stronger.

These results are in line with the results found by Brekke et al. (2003). With only a few months of observations after the introduction of the index price, they are able to identify a significant price reduction of 4 % on generic drugs due to the new scheme.

---

7 See Dalen, Strøm and Haabeth (2006) for a discussion on instruments.

3.6. Discussion of recent policy developments

In 2003 and 2004 several evaluations of generic competition were undertaken.[8] The Ministry of Health concluded from this that generic competition had not succeeded in bringing down drug prices as much as one could have hoped for. Several comparisons with other Nordic countries revealed that the Norwegian prices were much higher. In March 2004, the drug Citalopram, an anti-depression drug, was sold in Norway, Sweden and Denmark at these prices:

Table 3.4: Prices on Citalopram in March 2004. NOK (20 mg, 98 pills)

| Brand name Norway	Index price Norway	Generic price Norway	Generic price Sweden	Generic price Denmark	Generic price Finland
694	665,5	602	132-234	495-516	326-333

Source: Ministry of Health

In addition, the Ministry referred to problems with regulatory schemes that use information on wholesalers' purchasing prices. These do not follow standardized price reports (e.g. by IMS Health). Instead the wholesalers were asked to report their prices directly to the government. The interaction between wholesalers and producers are more complex than just entering an agreement to deliver drugs at a certain price. The contracts may involve payment for other services such as information and supervision, and storage. With the index price scheme there are economic incentives to channel a wholesalers' profit over to these other unregulated services. Although this will be termed illegal according to the law, a regulatory rule vulnerable to such practice will call forth higher monitoring costs for the authorities.

In January 2005 the index price scheme was replaced with a new price regulation scheme that abandoned the use of economic incentives to bring down pharmaceutical prices after patent expiration. The new scheme – called the *de-escalation* model – consists of a predetermined reduction of the reimbursed price when generic alternatives become available. The pharmacies are instructed to have the drug available at the reimbursable price.

The regulated reimbursable price is based on the maximum retail price of the patented drug before generics become available. Let *AUP** be the price before the patent expires and generic drugs enter the market. The de-escalation model determines the reimbursable price according to the following rule:

---

8 Brekke et al (2003) financed by the Norwegian Pharmacy Association, Dalen (2003) financed by the Norwegian Competition Authority, Dalen and Strøm (2004) financed by the Norwegian Medicines Agency, and ECON (2004) financed by the Ministry of Health.

Figure 3.5: The de-escalation model
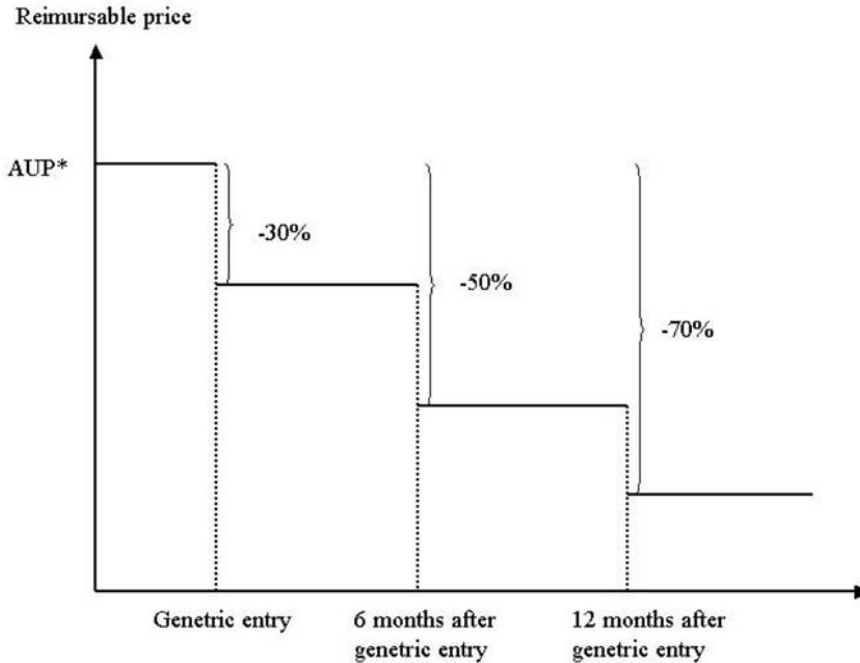
For drugs with annual sale above 100 million NOK prior to generic entry, the second price reduction is set equal to 40 percent (instead of 50 percent), and the third is set equal to 50 percent (instead of 70 percent).

The model gives the pharmacy chains strong incentives to lower their purchasing prices. The model does not prescribe any future price reviews based on the development of these prices. All cost savings – in terms of reduced purchasing prices – are kept by the pharmacies themselves. This scheme nicely illustrates the trade-offs that sometimes have to be made in regulation of prices. Maximum incentives to minimize costs are obtained by offering fixed prices. However, in order to be credible, these prices must be set at sufficiently conservative levels. If the government is too eager in reducing the cost of drug reimbursement – by setting the post-generic prices at very low levels – the pharmacies will report economic problems, which in turn will make it necessary for the government to increase the prices.

When such a scheme can be enforced without protests from the pharmacy chains, there are good reasons to expect the predetermined prices to be pleasantly higher than the purchasing prices.[9] Generic drugs are relatively cheap to produce, and the pre-generic prices reflect the cost of undertaking R&D-investments to innovate the drug in the first place. For many drugs, a price drop of 50 or 70 percent may still keep prices above the cost of producing generic drugs.

---

9 The de-escalation model was actually proposed by the pharmacy chains.

With the de-escalation model, there is no link between reduced purchasing prices and consumer or insurance prices. The gain obtained from reduced price is entirely kept by the pharmacies. Since pharmaceutical products are standardized products sold all over the world, information about prices in other markets is available to regulators. In such situations, alternative models could combine the benefits of fixed prices with tight control on pharmacy margins. Instead of basing the reimbursement price on uniform price cuts, the price could be set equal to observed generic prices in other European markets. Since these are exogenous to Norwegian pharmacies, the pharmacies will have strong incentives to negotiate low purchasing prices. Different from the de-escalation model, however, the margins collected by the pharmacies will be perfectly controlled by the marks-up set by the government.

With international benchmarking it becomes essential to get comparable information about generic prices. Using average prices, that include the original brand name drug as well, will support excessive prices. Even in markets where generic competition is strong, the brand name drug will often be sold at much higher prices than the generic substitutes.

Several studies have provided information about the price structure in pharmaceutical markets after patent expiration. Grabowski and Vernon (1992) examined the effect of generic entry in the US market on prices for 18 drugs that were first exposed to generic competition during the years 1983 through 1987, and they found that the branded drug price increased by an average of 7 percent one year subsequent to generic entry and 11 percent two years after generic entry. At the same time generic prices continued to fall after first entry. The average generic price two years after entry was 35 percent lower than the first entry price. Frank and Salkever (1997) arrived at similar results when they looked at a sample of 32 drugs that lost patent protection during the early to mid-1980s. Increased competition from generic drugs is not accompanied by lower prices on branded drugs. Their results suggest instead a small price increase on branded drugs. The research by Caves et al. (1991) gave somewhat different results. In their study, the branded drug price declines with the number of generic entrants, but the rate of decline is still small. For the mean number of generic drugs, the brand name price declines by only 4.5 percent. At the same time, generic prices are much lower than the brand name prices. Their results suggest that average generic price is about 50 percent of the branded drug price when 3 generic producers have entered the market.

Being able to track generic prices elsewhere in Europe – and setting the reimbursable price equal to these prices – one could obtain both fixed prices (from the pharmacies' point of view) and prices that track the purchasing prices. In some submarkets this may result in prices higher than the ones that follow from the de-escalation model, but in these cases there might be good economic reasons to support higher prices. In other submarkets, one can hope for lower prices. If competition among generic producers in Europe is functioning, prices paid by the National Insurance Scheme in Norway would reflect the cost of producing and distributing generic drugs.

For pharmaceuticals competition is important in both ends of the products' life cycle. R&D competition among the large international pharmaceutical companies is intensive, and will be so irrespectively of the Norwegian policy choices. When older drugs loose patent protection, regional markets for generic substitutes develop, which may help patients and insurance schemes to bring down costs. Due to the importance of third-party financing of medical costs, regulatory choices have proved important, and will in fact be essential in stimulating price-reducing competition.

3.7 Concluding remarks

Consumers have benefited from the deregulation of the pharmacy market. The number of pharmacy stores increased rapidly, there are longer opening-hours, and a selection of over-the-counter (OTC) drugs have become available in supermarkets and gas stations. Furthermore, since a large majority of drugs are subjected to price regulation, there are no reasons to expect deregulation – as such - to have had a negative impact on prices.

Prior to the New Pharmacy Law of 2001, there were many small, and privately owned pharmacies and a few large wholesalers, and the price regulation scheme capped the pharmacies' purchasing prices and their margins. Almost overnight this ownership structure was replaced with three vertically integrated pharmacy chains. In other markets such concentration could have weakened competition at the expense of consumers and economic efficiency. This is not necessarily the case in the Norwegian pharmacy market, since price regulation of some sort always will be needed to accompany the public health insurance scheme. For generic drugs, competition is certainly an important mean to bring down prices – but the crucial competition is between producers, and not between pharmacy chains. The pharmacy chains should transfer lower producers prices to lower retail prices – but this must be ensured by regulation, and not by retail competition.

In Norway it proved difficult to both stimulate producer competition and to lower retail prices on generic drugs. The main reason for this seems to have been that the price regulation scheme did not adjust to the vertically integrated ownership structure. With integrated pharmacy chains one should instead cap the retail price directly – and if the pharmacy margin is going to be capped as well, it must be based on the pharmacy chain's purchasing prices (producer prices), and not on a transfer price within the chain.

The today's concentrated pharmacy chains can be efficient actors in stimulating competition on generic drugs to the benefit of the public health insurance. Two important conditions must be satisfied. First, the competing producers must be able to affect the retail price of their own product. Second, a producer with a low retail price must be allocated a larger market share. The latter condition can be satisfied in two different ways: Either by law (i.e. generic substitution) – or with use of incentive-based co-payment schemes.

References

Brekke, K. and Straume, O.R. (2003): Pris- og avanseregulering i legemiddelmarkedet. *HEB – University of Bergen*. Rokkan-notat 20-03 (In Norwegian).

Brekke, K, Grasdahl, A, Holmås, T. H., Steen, F, and Sunnevåg, K. (2003): Evaluering av ny apoteklov og indeksprissystemet. *HEB – University of Bergen*. Rokkan-notat 21-03. (In Norwegian).

Caves R.E., Whinston M. D., and M. A. Hurwitz (1991), "Patent expiry, entry and competition in the US pharmaceutical industry", *Brookings Papers on Economic Activity*. Microeconomics, 1- 48.

Dalen, D.M. (2003): Legemiddelmarkedet etter apotekreformen. *HERO - University of Oslo*. Working paper 2003-11. (In Norwegian).

Dalen, D.M. (2003): Pris- og avanseregulering for legemidler. *Frisch Centre*. Report 4/2004. (In Norwegian).

Dalen, D.M, S. Strøm and T. Haabeth (2006): Price regulation and generic competition in the pharmaceutical market. European Journal of Health Economics, Vol 7, No. 3, 208-214.

ECON (2003): Evaluering av apotekloven og indeksprissystemet, Rapport 2004-010. (In Norwegian).

Frank R.G., Salkever D.S. (1997), "Generic Entry and the Pricing of Pharmaceuticals", *Journal of Economics & Management Strategy*. Spring. pp. 75-90.

Grabowski H., Vernon J. (1992), "Brand Loyalty; Entry, and Price Competition in Pharmaceuticals after the 1984 Drug Act", *Journal of Law and Economics 35*.

NOU 1997:6: *Rammevilkår for omsetning av legemidler*. Ministry of Health. (In Norwegian).

Øystein Fjeldstad, Espen R. Moen and Christian Riis

Regulation and competition in the Norwegian telecommunications market

4.1 Introduction

The Norwegian telecommunication market has gone through a remarkable transition over the last decade, caused both by technological development and by the liberalization of the regulatory framework of the sector. As a result, the Norwegian telecommunication market is well developed by international standards, with high levels of communication service adoption. For instance, in mobile telecommunication the penetration rate is as high as 102 percent, with relatively low prices compared to most countries outside the Nordic region.

 Liberalization of the Norwegian telecommunication sector has closely followed the time schedule set by the European Union. The Open Network Provision principles have been applied to both fixed and mobile services since 1993. The fixed-line telephony market was opened for competition in 1998. The state-owned telephone monopolist Telenor (then Televerket) was turned into a state-owned limited liability enterprise in 1994, deregulated in 1998 and listed and partly privatised in December 2000.
The regulator in the market is the Norwegian Post and Telecommunications Authority (NPT), established in 1987 based on the European Commission Green Paper (1987) on telecommunications. NPT is an autonomous administrative body under the Ministry of Transportation and Communication responsible for the administration and implementation of regulations. In addition the NPT conducts a number of activities, e.g. maintenance of an Internet site for electronic communication service price comparisons, in order to stimulate competition.

In this chapter we first give an overview of the development of the Norwegian telecommunication markets, and include two case studies of business practices that illustrate the functioning of the markets. We then take a closer look at the vertical structure of the mobile segment, and argue that the practice of open standards for content providers is likely to have contributed positively to the development of the market. Finally we discuss different modes of regulating the vertical structure of the telecommunication industry as well as the local loop.

4.2 Market overview[1]

In this section we first study the development of fixed line phone services. Then we move on to mobile telephony, and finally to broadband.

Fixed line telephony

High levels of communication service adoption characterize the Norwegian telecommunication market. As of 31 December 2005 Norway had a total of 2 128 997 fixed line phone service connections. PSTN (Public Switched Telephone Network, the analogue public telephone network) still dominates with 1 299 134 connections. The number of fixed line phone subscriptions peaked in 1996 at 2 484 173. During the period 1996-2005 the total number of fixed line subscriptions fell by nearly 14%. By comparison the population grew by 6.2%, and between population counts held in 1990 and 2001 (most recent) the average number of persons per household fell from 2.4 to 2.3 (SSB). Accordingly, the number of fixed line subscription per household fell from 0.96 in 1998 to 0.82 in 2005. Two trends are evident during this period: a number of new households adopted mobile phone service only, and established fixed-line subscriptions were cancelled in favour of individual household members having mobile phones. The number of ISDN connections peaked in 2002 with 810 913 subscriptions, but has since fallen to 621 933.

Telenor is the dominant actor with a 66% market share (down from 69% in 2004) for fixed-line voice in the residential market. The second largest competitor in this market, Tele2, has a 19% market share. In the business market Telenor has a 60% market share (down from 63% in 2004) and the second largest competitor, Ventelo, has 14%. Cable TV Telephony service was introduced in 2000 for 14 325 subscribers. Cable-TV subscriptions peaked at 23 563 in 2003. Internet Telephony on the other hand has risen from a modest 142 subscribers in 2002 to 186 972 in 2005, with the number of subscribers more than tripling during 2005. As of June 2005, there were 36 firms providing Internet telephony service in Norway, with 30 providing service to residential customers and 18 providing service to business customers. At the end of 2005 the largest provider, Telio, had a 35.8% market share.

The prices of long distance telephony decreased sharply between 1994 and 1999 when local and long distance prices were equalized. Prior to deregulation, long distance telephony was priced above long-run marginal costs in Norway, as was common in most other countries. Transmission exhibits strong cost economies of scale. Large-capacity cables, in particular fibre-optic cables, have considerably lower cost per MBit than the low-capacity copper-wire. Furthermore, transmission demand is aggregated from a large number of individual users. The justification for the historical pricing scheme of high long-distance charges was cross-subsidisation between a market with low demand elas-

---

1 Source: The Norwegian Post and Telecommunications Authority's 2005 statistics, unless otherwise cited.

ticity (the long-distance market) to a market with high demand elasticity (the local loop). However, when the market opened for competition this pricing structure allowed for cream skimming by new entrants, which would have strongest incentives to enter the markets with the highest profit margins. The events leading up to uniform prices within Norway are interesting from a regulatory point of view as illustrated by the Enitel bankruptcy (see Case study 1).

> *Case study 1: The Enitel bankruptcy.* In 1996, 7 electric utility companies established the company Enitel with the mission to provide high-speed data and voice transmission via fibre-optic cable spun around power lines. The company filed for bankruptcy in 2001. Enitel's goal was to become Norway's second largest transmission provider. A consulting firm estimated market growth between 200% and 400% and price reductions of 50%. Enitel's strategy was to offer transmission at lower prices than Telenor (www.digi.no, 1998). Enitel started offering their services in February 1999. On 1 July 1999 Telenor introduced uniform pricing for all calls within Norway, thereby eliminating the separate long-distance market. Subscription and termination prices were reduced during the following months (St.prp. nr. 66 (1999-2000)). Subsequently, in 2001 subscription prices were raised by 30%.

> The reason for Enitel's failure may be numerous and heavily intertwined with the burst of the dotcom bubble in 2000, as well as the aborted Telenor-Telia merger. However, with hindsight one may speculate that Enitel misjudged Telenor's likely response to their entry. Although margins in the transmissions market at the time of Enitel's entry were high, there was also considerable transmission capacity available in the existing network. The main effect of Enitel's entry was thus cut-throat competition that brought down prices and eliminated voice transmission as an independent market. Consumers gained short-term by reduced rates. Moreover, investments made in additional capacity ensured a more competitive long-run transmission market in line with European market liberalization intentions. However, Enitel investors suffered major, potentially unnecessary, financial losses, reflecting a waste of resources associated with building excessive capacity in transmission. Price reductions and further investments could also have been achieved by establishing a price cap on transmission that was high enough to provide incentives for additional investments.

Broadband service

The number of Internet residential broadband connections rose from 1000 in 1999 to 887 214 in 2005. The majority of these (79%) are xDSL connections. The growth of Cable TV based Internet has accelerated since 2002, and grew by 49% from 2004 to 2005 to a total of 136 706 subscriptions. The corresponding figures for fibre-based subscriptions are: 129% and 354 571 subscriptions, and for xDSL: 46% and 703 789 subscriptions. The relative ease of xDSL diffusion combined with strong performance increases from xDSL innovations and local loop upgrades have fuelled the internet connection growth. Over time, increased demand for higher bandwidth may reduce the initial attractiveness of the copper wire as the preferred medium. The outcome depends on investments made in technological innovation and upgrades of xDSL, cable and fibre infrastructures.

Mobile communication

Norway was one of the first countries in the world to introduce commercial mobile telephony. Since then, the Norwegian telecommunication market has been first in a number of mobile technology introductions, amongst these the first public WAP portal in 1999 and the first MMS service in March 2002 (Lie 2004).

During 1999 the number of mobile subscribers surpassed the number of fixed-line subscribers. Norwegian mobile phone penetration is among the highest in the world and has surpassed 100% with 4 754 453 total subscribers at the end of 2005. A number of the subscriptions are data subscriptions and machine to machine (M2M) only subscriptions. In the early phases of GSM market development the majority of mobile originated calls terminated in the fixed line network. In the year 2000, revenues from mobile to mobile traffic surpassed mobile to fixed traffic, and in 2005 revenues from mobile to mobile traffic was 2.8 times greater than mobile to fixed traffic.

There are three mobile network operators in Norway. Telenor and Netcom have full coverage. They were licensed as GSM operators in 1991. Teletopia was licensed in 2001 and covers only the capital Oslo. Service to Teletopia customers outside of Oslo is provided by way of a roaming agreement with Telenor. A third GSM license was awarded to Swedish Telia in 2000, but this license was returned when Telia later bought Netcom. The market was opened for mobile service providers in 1999. As of 31 December 2005 Telenor Mobil had 70.5% of the operator market and NetCom 29.5%. The figures have been stable for a number of years.

Norway has among the highest usage of SMS and MMS per capita in Europe. In 2005, the average annual number of SMS messages per user was 963. The total number of both SMS and MMS messages grew by almost 27% from 2004 to 2005. The number of content messages peaked in 2004 with 494 million messages, i.e. 109 content messages per user, and dropped to a total of 375 million messages in 2005. SMS messages accounted for 11% of traffic revenues in 2005.

An important aspect of Telecommunication is the existence of network effects, or increasing returns to scale on the demand side (Rohlfs, 1974, Katz and Shapiro, 1985). The more subscribers a network has, the more attractive it is for new customers to connect to the network. The mobile markets illustrate a number of interesting aspects of competition under network externalities, several of which have been subject to regulatory debate and action. For instance, there are clear direct network effects in SMS-messaging. Until recently, when gateways were established between Internet service and SMS, the mobile phone network constituted an incompatible messaging network.

The total cost of mobile phone usage is together with the other Nordic countries among the lowest in OECD area. Over the period 2000 to 2004 total revenues per subscription rose from NOK 2573 to NOK 3024, while it dropped to NOK 2967 from 2004 to 2005. Revenues per minute fell from NOK 1.47 to NOK 1.16 over the period 2000 to 2004. This revenue increase per subscription is consistent with network effects coupled with elastic demand for mobile communication. There was strong growth in the number of

mobile subscriptions during the period, and although still substantially higher than fixed-line to fixed-line calling, the price of mobile to mobile calling fell. The combined effect was increased mobile to mobile communication.

4.3 Network effects in mobile telephony

The effects of network externalities on optimal business strategies and aggressiveness of competition are thoroughly analysed in the economics literature, see for instance (Laffont, Rey and Tirole, 1998). If two networks are incompatible, network effects imply that volume increases the attractiveness of a network. This effect tends to sharpen competition between rival networks, as it increases the price elasticity of demand. If a firm reduces its prices, it attracts more customers, and thereby becomes even more attractive. If competitors are asymmetric, network effects strengthen the position of the biggest firms. Thus, competing symmetric network operators have an incentive to make their networks compatible in order to soften competition. However, this may not be the case if the network operators are asymmetric, as the largest network may prefer incompatibility in order to take full advantage of its scale advantage.

One may argue that network externalities are particularly important in Norway relative to the cost of mobile telephony. Not only is the GDP per capita in Norway high (as of January 2006 Norway had a total population of 4 640 219 (SSB), and a GDP per capita of USD 40 568. This GDP per capita is second only to Luxembourg in the OECD area (OECD, 2006). In addition, income distribution is even. This combination is favourable for the diffusion of new network services, where value of adoption depends on network size. Furthermore, operators and service providers have aggressively subsidized handsets in order to boost diffusion.

Interestingly, it seems that it took some time before network operators fully acknowledged the effects of network externalities. This is illustrated in our second case study.

> *Case study 2: NetCom's pricing model.* Beginning November 1996, NetCom introduced a pricing model where business customers were charged 50% less for calls to other NetCom subscribers. The rationale was to pass on cost-advantages from bypassing Telenor in the newly deregulated transmission market. A spokesman for the company pointed out that this would benefit NetCom because of stronger ties between the customers and the company, essentially creating a lock-in (*Dagens Næringsliv,* 1996). In January 1997, Telenor Mobile followed suit (VG, 1996) and cut both the price of calling within Telenor Mobile's network and the price of calling from Telenor fixed-line to Telenor Mobile.
>   Introducing a price differential between on-net and off-net calls has the same effect as reducing compatibility between networks. As argued above this gives larger operators a huge advantage and smaller operators a corresponding disadvantage. Since NetCom is much smaller than Telenor, the winner of this pricing strategy was Telenor and the loser NetCom. During 2000 NetCom reduced the price of calls ter-

minating outside their own network considerably (*NetCom Annual Report,* 2000). In 2001 and 2002 they abandoned differential off-net on-net pricing for regular subscribers and prepaid customers respectively (NetCom press release 25 February 2002), whereas Telenor maintained differential pricing. From a network externality perspective an on-net off-net price differential is also unfavourable for larger operators if smaller operators practice uniform pricing. During the period 2000-2003 NetCom gained in market shares from 23.3% to 26.9% despite the entry of new service providers, whereas Telenor's market share dropped from 68.8% to 58.2%. (A true comparison of changes in operator market shares would need to incorporate the two operators' share of new service providers).

4.4 Mobile telephony: vertical structure

The mobile communication market was opened for service providers without their own network in 2000 and mobile number portability was introduced in 2001. The idea was to increase competition downstream, where the end users are served. It is common to distinguish between mobile service providers and Mobile Virtual Network Operators (MVNO). MVNOs are network operators without their own physical infrastructure but with their own functions for adopting services. Service providers, or retail sellers, buy their network services from a network operator. Mobile service operators and MVNOs have made significant inroads into the services markets and total market revenues have increased significantly. These recent market developments mirror those in the other Nordic countries, with the exception of Sweden. By the end of 2005 Telenor Mobile had 57.4% of subscriptions and NetCom had 23.5%.

Sense was the first service provider entrant into the market. The company initially sought to become an MVNO. However, the company went bankrupt while waiting for access to Telenor's network, and was then offered a reseller agreement by Telenor. This agreement came into effect in 1999.

Mobile content is made available by independent content providers via a Content Provider Access (CPA) model. CPA is a technical platform, protocol and business model that enables content providers to deliver content to an operator's subscribers and bill the subscribers for content. Subscribers send an SMS message to one of the operators' SMS servers identifying the content provider. The operators' server exchanges SMS messages with subscribers and communicates with the content providers via the Internet.

From the point of view of content providers and subscribers, the Content Provider Access models of the two major operators, Netcom and Telenor, are quite similar. The operators publish information about the technical standards of their services and their business models. Netcom and Telenor initially chose different protocols for communicating with content providers, but Netcom subsequently adopted the SonicMQ protocol already in use by Telenor. Furthermore, common SMS numbers used by end users to access content are terminated within both the Netcom and Telenor networks. The content providers set their own prices and the operator terms include a predetermined revenue

split. The two operators deal differently with traffic charges: Netcom charges separately for traffic whereas Telenor includes traffic charges as part of their portion of the revenue split. At the time of its introduction in 1999, the business model surrounding content provision in the Norwegian market was unique. The mobile content business was in its infancy and for the most part content was either provided by operators, e.g. directory information or it was resold by operators with operator specific solutions. A common framework for third party content provision with common access numbers and message protocols allowed the content providers to create operator independent products and brands. The adoption of a single communication protocol further simplified the technical development of content providers.

The model chosen is interesting from a system market and network economic perspective. Content is a strong complement to mobile subscription and SMS-traffic, and hence there are indirect network effects (Katz and Shapiro, 1985). The operators have chosen not to provide content that directly competes with the content providers. By this choice, they have refrained from price and R&D squeezes. Alternatively, they could have engaged in price and R&D squeezing in order to increase demand beyond content equilibrium, and thereby attempt to capture extra profits from the complementary network operation activity. Also, operators could have chosen their own specific solutions, so called "walled gardens". This could have limited the total content market and hence also the adoption rate of mobile phone subscriptions and mobile phone traffic. Non-operator specific access numbers, common message protocols and business models may have contributed to a larger and more competitive content marked. There is however, no obvious explanation for why the content marked dropped from 2004 to 2005.

4.5 Philosophy of regulation in Norway

The basic idea behind Norwegian telecommunications regulation has been to stimulate competition in end user markets through wholesale market regulation. To achieve this, PT requires that network operators with significant market power give retail sellers access to their networks at cost oriented prices. The lack of significant entry barriers downstream indicates that fierce competition in the market for service providers should be achievable. Hence, by regulating the bottleneck, direct regulation of end user prices should be redundant.

This idea, to regulate wholesale prices and let end user prices be market determined, is also reflected in the present legislation. According to the Electronic Communications Act, the authorities may impose on a provider with significant market power an obligation to open their networks for service providers and MVNOs (rf § 4.9). The subsequent paragraph states that: "When regulation in accordance with § 4-9 does not succeed, the Authority may impose regulation of end-user services".

Wholesale regulation has its advantages. As network services tend to be more standardized than downstream products, implementation of price regulation of network access may be less costly than regulation of downstream products. Furthermore, one reduces

traditional concerns that the price regulation of end user prices weakens incentives for product innovation and quality improvements.

As a matter of fact, the authorities have been cautious in the implementation of this principle. On one occasion, see the Sense case below, NPT forced Telenor to reduce their access prices. The overall picture is still that NPT has been ambitious in supervising, but not in sanctioning, network owners with market power.

This may reflect that wholesale price regulation is not without costs. Contracts between pure service providers and integrated networks are examples of vertical contracts. It is well established in economics that vertical contracts tend to be efficient. If service providers are more efficient in serving customers, introducing new products or recruiting new customers, there are gains from trade associated with vertical contracts. Hence vertically integrated companies have incentives to sign contracts with service providers in order to realize these gains. This does not rule out that inefficiencies associated with rent extraction may occur; integrated companies may foreclose downstream to protect the market upstream (if building up independent customer bases accommodates upstream entry). However, the benchmark case is that vertical contracts are efficient.

Interfering negotiations between vertically related agents like networks and service providers have two types of costs. First, regulations typically standardize contracts, which can be socially damaging. Vertical contracts are sometimes specifically designed to take various agency issues into account, for instance by protecting relation specific investments. Furthermore, standardization rules out price discrimination. Price discrimination is necessary in order to ensure that fixed costs are allocated to consumers in an optimal way. Historically, price discrimination has also been an instrument for creating profitability when introducing new products and creating new markets – both of which are incompatible with standard regulation procedures.

Secondly, regulations may give integrated firms incentives to foreclose the market downstream. If wholesale price regulation is used as an instrument to eliminate monopoly profit (thus regulated prices are set below retail minus), integrated firms can protect their market power by deterring entry of independent service providers. This can be done by delaying access and making access costly and difficult.

Regulating wholesale prices thus creates dilemmas. No regulation, or mild regulation based on a retail minus principle (under which the network operator keeps its profits of network services) ensures the efficient entry of service providers, in the sense that pure service providers generate profits if and only they are more efficient in retailing than the integrated firm. However, this does not reduce the monopoly problem associated with the network. Furthermore, if the network owner is a monopolist, more of the efficiency gains derived from service providers are channeled to the monopolist. In a situation with some competition, for instance Norway's asymmetric duopoly, more of the efficiency gains are channeled to consumers.

Hard-handed regulation of the network (for instance cost plus regulation) reduces the excessive pricing associated with the network owner's market power. On the other hand, unless the market for service providers is perfect this may lead to excessive entry of service providers, fuelled by profit shifting. Furthermore, the network provider will have incentive to obstruct entry of service providers, and not take advantage of possible gains from vertical relations from fear of being regulated. Finally, reducing the profitability of network owners may reduce the network owner's incentives to invest in the network, and thus harm economic activity in the long run.

The history of Norwegian telecommunications regulation in the late 1990s and the beginning of this century can be described as a strategy aiming at a balance between, on the one side, forcing prices down through regulation, and on the other, allowing agents to establish commercial contracts. The following two examples illustrate this.

Case study 3: Sense and Telenor. Sense and Telenor disagreed during a 2002 renegotiation of the terms. Sense brought the case to the Norwegian Post and Telecommunication Authority. In, July 2002, the NPT ruled that Telenor's wholesale prices should be reduced by 25%, and imposed constraints on Telenor allowing Sense to transfer customers to another operator. The NPT, by regulating the whole-sale price, demanded a cost-plus pricing regulation. Telenor appealed the decision and in March 2004, the Ministry of Communication and Transportation ruled that Telenor Mobile had to reduce their prices by 25%, but that models such as retail minus (efficient component pricing rule) also are compatible with cost oriented pricing, a view supported in the economic literature and by court ruling (Borgarting Lagmannsrett, 2006). A retail-minus system increases competition among resellers for the portion of value creation activities that they perform. Such a system does not address potential monopoly profit problems, which need to be addressed by end-user price caps. Such price caps would in principle have been possible because Telenor Mobile has a dominant position, but the government has intentionally refrained from using them. The importance of predictability reduces the attractiveness of ex-post dramatic changes in regulatory principles.

Case study 4: Telenor and Teletopia The assessment of whether SMS provision is an information processing activity or a networking service is of principle interest with regards to what constitutes relevant cost-basis and the interpretation of cost-oriented pricing principles. In 1998, the company Teletopia proposed to establish an SMS server for the purpose of competitively offering services to Telenor's customers. Telenor rejected their proposal and Teletopia brought the case before the NPT in 1999. Again the NPT favoured the proposal and saw it as being in line with their stated strategy of wholesale regulation. The high number of SMS messages transmitted is of course intimately related to the high diffusion of mobile phones invested in by the operators. Dynamic efficiency concerns imply that these acquisition investments are considered when setting access prices. This was reflected in a letter concerning the case from the Ministry of Transportation and Communication October 2003, determining that retail minus (or minimum efficient component pric-

ing rule) is in accordance with cost oriented pricing. The Ministry comments specifically that voice services and SMS is a joint product, and that the viability of the entrant's service was completely dependent on the customer acquisition costs made by Telenor (Teleplan and Schjødt 2004 and Norwegian of Transportation and Communication REF 03/1482-4 JIE)

4.6 Fixed telephony

The main issue regarding the regulation of fixed telephony is the local loop. The local loop consists of local switches and connection lines that connect these switches to end users (consumers). Barring technological uncertainty about optimal future access technologies, the local loop, "the last mile", constitutes a natural monopoly. When Telenor was deregulated in 1998 and privatised in 2001, Telenor retained ownership of the local loop. Today, a price cap regulates pricing of the local loop. The price cap in turn is calculated using a cost plus model. In contrast with most other price regulation practices in the telecommunication industry, cost accounting is based on replacement costs, not historic costs.

At the time when Telenor was deregulated, it was uncertain what the importance of the local loop would be in the future. It was a commonly held belief that mobile phones would take over the voice market and that cable TV, or new wireless network technologies would take over the data transmission market. However, improved xDSL technology coupled with investments in upgrading portions of the local loop have considerably improved the efficiency of the local loop as a carrier of both voice and data. As a result, the local loop plays a much more significant role now than was anticipated when Telenor was deregulated.

A recent report (Econ and Oeconomica 2003) suggests that the instituted regulation of the local loop may not have been optimal as it leads to artificially high prices for fixed telephony, resulting in severe welfare effects for the telecommunication industry. More fundamentally, it can be discussed whether it was wise policy to privatise this part of the network in the first place.

The local loop typically refers to bundles of copper-wire where the individual wires terminate in a circuit switch in a residence or an office building. The marginal cost of using a particular wire for voice or low bandwidth data communication is negligible. In a static situation, a price substantially above marginal costs may result in less than optimal usage and give rise to a welfare loss. In most regulated markets with large investment costs and small marginal costs there exists a trade-off between short-term efficiency, often referred to as static efficiency, and long-term efficiency, referred to as dynamic efficiency, where investment incentives are also taken into account. Static efficiency requires low prices in order to utilize capital. However, in order to provide firms with sufficient incentive to make large initial investments, firms should be allowed to set prices sufficiently high to recoup their investment costs. A regulator can strike a compromise between static and

dynamic efficiency by imposing a (global) price cap on the firm in question. The price cap should be set in such a way that the firm in question can recoup its investment costs. If the costs associated with a mark-up are high, and the information problems regarding investments and costs are relatively low, the government may opt for alternative modes of regulation. One alternative is that the government owns and operates the local loops, however, there are costs associated with such a separation and technical difficulties associated with defining the border between the local loop and the rest of the network. The latter problem is compounded by ongoing technical xDSL developments which require the substitution of portions of the copper-wire bundle with fibre optics and the installation of modems and switches closer to subscribers. Another alternative is that the government finances necessary investments, upgrades and replacements of the local loop directly.

There are several reasons why price cap regulations with prices above marginal costs are less attractive for the local loop than they are for other forms of telephony such as mobile telephony: (1) The marginal costs of single or dual ISDN voice calls over a copper-wire is very low. This contrasts with mobile telephony where the capacity, at least in parts of the network, is optimized to fit the demand. As a result, the difference between a price cap and the true marginal cost may be relatively large. (2) Duplication costs associated with parallel local loops are detrimental. Thus, too generous a price cap does not lead to the entry of new competitors or substantial growth in market shares for existing competitors. The local loop is at present an essential facility. (3) The price of fixed telephony directly affects the price of alternative modes of communication, particularly mobile telephony. This is partly because Telenor owns a wide array of different communication infrastructures. For instance, a lower price on fixed telephony could make Telenor more aggressive in the mobile phone markets, thereby increasing competition in this market as well. (4) Although the local loop was traditionally used for telephony, other uses such as broadband must also be considered in today's market. A high access price to the local loop implies a high price on broadband, and an inefficient number of subscribers. Furthermore, inefficient bypass may occur. (5) An actor with monopoly in one infrastructure may have incentive to monopolize another in order to protect monopoly rent in the former. This argument is raised in the literature and may be relevant with respect to the relationship between xDSL- and cable TV-based broadband services.

At the time of deregulation in 1998, a large part of the local loop was already in place, financed by the general population through taxes. However, the operator still incurs costs when maintaining and extending the network and these costs must be financed in some way.

Suppose Telenor had been split before deregulation and privatisation, and the local loop was organized as a separate entity owned by the state. In this case, use of the local loop would be free if maintenance and development of the local network was financed directly through the state budget. Of course this model has its own drawbacks. The main idea behind deregulation and privatisation was to boost efficiency and the ability to innovate, thereby increasing consumer welfare. Keeping the local loop in the hands of the govern-

ment could reduce this positive effect. Another alternative would have been to reimburse Telenor for the cost of upgrading and maintaining the loop directly. Again this has drawbacks, most importantly the lack of incentives for keeping costs under control. Furthermore, there are dead-weight losses associated with public funds.

Today, with a privatised Telenor, initiating changes in the structure of asset ownership could prove costly. The same may also be the case for re-regulation of the local loop. However, a lesson from the Norwegian experience may be that to deregulate and privatise a state-owned monopoly without specific prescriptions for the local loop comes at a cost. Alternative models, like separating out the local loop and maintaining it as a separate entity, should be seriously considered before privatisation. On the other hand the issues involved are more complex than simply making copper-wire available for efficient use. A recent OECD report concludes that whereas the costs of separation are high, the benefits are uncertain and that there is little empirical evidence that benefits are sufficiently in excess of costs. (OECD 2003)

4.7 Final remarks

As mentioned initially, the Norwegian telecommunication market has gone through a remarkable transition over the last decade, caused both by technological development and by the liberalization of the regulatory framework of the sector. Today the Norwegian telecommunication market is well developed by international standards, with relatively low prices.

Over the next decade the sector will certainly face new challenges. First, third generation mobile telephony, or UMTS (Universal Mobility Telephony System) is expected to replace second generation mobile telephony, GSM. Second, internet-based telephony, or IP- (Internet protocol) telephony is expected to replace today's fixed line telephony. These two developments will give rise to different challenges.

Three operators have licences to build UMTS-networks in Norway: Telenor, NetCom and 3. However, investment in 3G telephony has been delayed (Digi.no). This may partly reflect that licensees fear harsh regulation once investments are sunk, such that recoupment may prove difficult. Furthermore, there are strong network effects associated with 3G. Creating a market and launching 3-G telephony may prove costly. Agents may have incentive to free-ride on the efforts of others, and enter the market at a later stage. Finally, the demand for UMTS services is genuinely uncertain, and network operators may execute the option value of waiting. Low investment in 3G telephony would suggest that development is slower than what is socially desirable. At the same time one cannot rule out that the delays may reflect that the private as well as the social value of UMTS is less than expected.

 IP-telephony implies that broadband internet connection is used for telephony over the internet. IP-telephony thus challenges the telecommunication firms' possibility to price

discriminate between broadband and fixed telephony. In addition, the local loop as an essential facility for fixed-line telephony may also be challenged, as broadband can be supplied by Cabel-TV. IP-telephony may thus reduce the profits of the owners of the local loops. As the marginal cost of using copper wire is very low, the likely effect of IP-telephony is a substantial drop in the prices of fixed line telephony.

Regulation of industries with high fixed costs and strong network externalities will always be complex. These factors can reinforce each other and accentuate the conflict between static and dynamic efficiency concerns in regulation. Once an investment has been made either in a fixed asset such a physical copper-wire or fibre optic transmission lines, or in the adoption of a new network service by way of which users can exchange, simple static efficiency concerns dictate pricing at marginal cost. Because of rapid technological development in telecommunication, the magnitude of network investments, in many cases a result of incremental investments in networks not being possible, and the risks associated with the investments, dynamic efficiency is of particular concern.

The main object of regulation is to increase welfare by ensuring sufficient competition. Normally, the object of regulation is not to protect firms, except against predatory action. Ensuring the continued investments that are necessary for an efficient market, however, may require that action is taken to avoid excessive misallocation of investors' funds to obtain simple market changes. There is no reason to protect competitors against moves that may have temporary anti-competitive effects if the long run effect benefits consumers.

References and sources

Borgarting Lagmannsrett, 2006, Reitan Gruppen vs. Telenor 04-102499ASI-BORG/02.

Dagens Næringsliv, 1996, "NetCom-kunder kan halvere prisen", newspaper article by Erik Lahnstein, 6 August 1996 p.9.

Econ og Oeconomica (2003) "Effektiv og bærekraftig konkurranse i tele- og mediemarkedene – hva skal til?", (www.npt.no).

European-Commission (1987) Green Paper on Telecommunication COM (87) 290 Final.

Farrell, J. and M. L. Katz (2000). "Innovation, Rent Extraction and Integration in System Markets." *The Journal of Industrial Economics* 48(4): 413-432.

Katz, M. L. and C. Shapiro (1985). "Network Externalities, Competition and Compatibility." *American Economic Review* 75(3): 424-441.

Laffont, J.-J., P. Rey, and J. Tirole. (1998). "Network Competition: II Price Discrimination." *RAND Journal of Economics* 29(1): 38-56.

Lie, E. (2004). Shaping the future mobile information society: the case of the Kingdom of Norway, International Telecommunication Union (ITU) Document: SMIS/09 - 25 February 2004.

Norwegian Ministry of Transportation and Communication, 2003 "Klage fra Telenor på vedtak om vilkår for Teletopias tilgang for tilknytning av SMS-C", by REF 03/1482-4 JIE.

Norwegian Post and Telecommunication Authority, 2005 Telecommunication Statistics, 8 May 2006.

OECD, 2006, Labour Productivity Levels and GDP per capita - Data: Breakdown of GDP in its Components and Differential in GDP per Capita and their Decomposition 31March 2006.

OECD, 2003, The benefits and costs of structural separation of the local loop, DSTI/ICCP/TISP(2002)13/FINAL.

Ro, Sommersnes & Co. DA – M.N.A., 2001, Innberetning til Asker og Bærum Skifterett sak nr. 01-1775 S: Enitel ASA.

Rohlfs, J. (1974). "A Theory of Interdependent Demand for a Communication Service." *Bell Journal of Economics and Management Science* 5(Spring): 16-37.

SSB (2001) Statistics Norway, Population and Household Count 2001 (www.ssb.no).Stortingsproposisjon nr. 66 (1999-2000): "Om Telenor".

Teleplan og Schjødt, 2004, "Post- og teletilsynets rolle i utviklingen av konkurransesitu-asjonen på teleområdet i perioden 1995-2003", 23 August 2004.

VG 1996, "Billigere å bruke mobiltelefonen", newspaper article by Kristin M. Hauge, 11.11.1996, p.23.

www.digi.no 17 June 1998: "EniTel investerer 1,2 mrd. i telenett".

Øystein Foros, Hans Jarle Kind and Helge Østbye

Norwegian Radio Broadcasting:
From Public Monopoly to Competitive Homogeneity?

> *We settled in for the night my baby and me*
> *We switched 'round and 'round 'til half-past dawn*
> *There was fifty-seven channels and nothin' on*
> *Bruce Springsteen, 57 channels (and nothing's on), 1992*

5.1 Introduction

"57 channels (and nothing's on)" is Bruce Springsteen's reflection on competition and diversity within media markets. Springsteen's lyrical claim that competition does not create media pluralism is to a certain extent consistent with predictions from economic theory. However, it is at odds with the view often put forward by policy makers. Indeed, the wish to create pluralism has typically been the crown argument for allowing - and encouraging - the establishment of an increasingly large number of private and public radio and television channels (see e.g. St.meld nr. 88 1981-82; Ot.prp. nr. 55 1989-90). While the broadcasting market in most countries was heavily regulated in the aftermath of the Second World War, it has now been liberalized in most democracies.

In this chapter we use economic theory and a Norwegian case study to discuss whether competition in radio broadcasting creates diversity. We start out by giving a brief overview of the development of competition within the European broadcasting markets. However, our main focus will be the market for commercial radio in Norway, partly because this represents a very interesting case from a regulatory point of view. Twenty five years ago there existed only one Norwegian radio channel, which was run by the public sector, but a second public radio channel was in the planning process. When it opened in 1984, listeners were offered "more of the same in two channels". Commercial radio channels were not allowed in Norway until 1988, and in 1993 the Norwegian parliament granted a license for the first nation-wide private radio channel. A second nation-wide private license was granted in 2003, seemingly without creating much more diversity. The entrant (Kanal 24) initially chose a program profile close to that of its rival, P4. Below, we argue that this lack of diversity was a consequence of both the regulatory policy and of market forces. In prolongation of this argument, we also provide a short discussion of whether horizontal mergers in media markets may increase rather than reduce diversity.

5.2 From monopoly to competition

5.2.1 Regulation in broadcasting

For forty years following World War II, broadcasting in Western Europe was dominated by a system with national public service monopolies. Most of these were financed completely or partly by a licence fee, and owned by the state. The only exception was Luxembourg, who has never had public service broadcasting (Siune 1986). Great Britain was the first country to depart from the system with a public service monopoly when ITV was introduced in 1954-55. ITV is a nation-wide superstructure producing national television programmes based on contributions from privately owned, regional television companies. The ITV is financed by advertising, but strictly controlled by a regulatory body, the Independent Broadcasting Authority (Tunstall 1986).

Italy gradually developed a commercial television sector from 1976, with legislation lagging behind the actual development (Mazzoleni 1976). Since 1980, almost all European countries have changed to a system where the old public service broadcasting companies have to compete with privately owned, nation-wide radio and television companies (McQuail 1990). In Norway, this process started in 1982 and was completed by the establishment of a commercial sector for radio and television in 1992?93.

There is, however, strict entry regulation into radio and television markets. At least until analogue transmission has been replaced by digital, there will be a shortage of available radio spectrum. The state therefore has to find a way of selecting which companies to grant licences to.

Allocation of licences takes place either through a so-called commercial approach or a broadcasting policy approach. A commercial approach would be to put up an auction and give the licence to the highest bidder. In the broadcasting policy approach, the government specifies some media political goals that a licensee has to fulfil. Interested groups or companies are then invited to submit a tender where the emphasis is put on how well the different propositions fulfil the goals. This is what is often called a «beauty contest». Frequently, criteria from the two approaches are combined.

One problem with beauty contests is that the bidders may have to fulfil goals that make it more difficult to attract large audiences. This certainly poses a problem for the owners. Advocates of beauty contests regularly rely on paternalistic and/or market failure arguments. Historically, the paternalistic motive - that people do not know their own good - has been important. Necessary conditions for the market failure argument to be valid are well expressed by BBC chairman Gavyn Davies:[1]

> *'some form of market failure must lie at the heart of any concept of public service broadcasting [and regulation]. Beyond simply using the catch-phrase that public service broadcasting must "inform, educate and entertain", we must add "inform, educate and entertain in a way which the private sector, left unregulated, would not do". Otherwise, why not leave matters entirely to the private sector?'*

1 Quoted by Armstrong and Weeds (2005).

There is little doubt that there may exist serious market failures in broadcasting that a benevolent regulator in principle could solve. However, a hands-off attitude has become increasingly more prevalent in democratic countries in the media market in general. Presumably, we shall see the same development also when it comes to broadcasting, not least since technological progress reduces the natural barriers to entry (e.g., Armstrong 2005, Armstrong and Weeds, 2005). It should further be noticed that people who prefer channels run according to some public service ideology, are usually already well served by the channels operated by the old PSB (Public Service Broadcasting) companies. In this respect beauty contests tend to reduce differentiation between public and private broadcasting. Beauty contests may also impose a political burden by provoking accusations of subjectivity and favouritism.

Some of the problems with beauty contests are illuminated by the turbulent period in Norway leading up to the launch of the radio channel Kanal 24 on January 1st 2003 and the channel's struggle on the market ever since.

5.2.2 The end of the broadcasting monopoly in Norway

The Norwegian broadcasting system underwent a dramatic change from the early 1980s to the early 1990s. Norsk rikskringkasting, the NRK, had been established as a BBC inspired public (radio) broadcasting monopoly in 1933. Television was formally opened in 1960, and in 1983 the NRK still operated only one radio channel and one television channel.

In 1984 the NRK formally started a second radio channel, P2. The two channels - P1 and P2 - were planned to be very similar – the motto being «more of the same in two channels» (Dahl and Bastiansen 1999:538-539).[2] At that time NRK had already experienced its first competition from other actors, since liberalization of radio had begun in 1982 when experiments with organizational local radio started. Commercials were not allowed, and the stations had to rely on support from organizations and a lot of voluntary work. The number of licences for local radio grew and reached a maximum in 1988, with 488 stations. In 1988 commercials were finally allowed and local radio became established on a permanent basis, although the licence of each of the local radio stations is awarded for a limited number of years only (Halse and Østbye 2003:192-200).

Following ten years of heated political discussion, the political parties made an agreement on the establishment of a second, nation-wide, terrestrial television network, privately owned and independent from the NRK. The new channel – TV 2 – was to be financed by advertising. As a part of this agreement, NRK was allowed to start a third radio channel (P3), while a fourth FM channel was to be private and commercial.

Two groups applied for the licence for the fourth FM channel: one consortium headed by

---

2 This resembles the situation in the UK in the first post World War II period; at this time there was no clear distinction between the two programs offered by the BBC.

the Aller Group, which was heavily involved in local radio, and «P4 – Radio Hele Norge», headed by the Swedish company Kinnevik, which was involved in commercial television in all the Scandinavian countries. Aller's involvement in local radio turned out to be an argument against this consortium, and the licence was awarded to the P4 Group. The licence, which was valid for ten years, included some obligations to run the station according to a vaguely defined public service ideology. P4's program consists mostly of music and news. With AC (Adult Contemporary) music, its main target group is young adults 25 to 40 or 50 years of age. (Halse and Østbye 2003:240-242)

In addition to introducing a third channel, P3, NRK responded to the increased competition on the audience market by giving each channel a distinctive profile. P1 was planned to be a broad radio channel with middle aged and older listeners. The popular regional programs were transmitted via this channel. P2 was planned as a channel for younger adults with higher education. Classical music, culture and current affairs dominate the channel. P3 was to focus on pop music, aiming at a group of listeners that NRK had almost lost to the local radio stations: older teenagers and young adults. NRK's change in program profile strategy from "more of the same in two channels" to segmentation with three distinctive profiles probably had greater impact on the Norwegian radio market than the direct effect of the entry of P4.

Even before P4 started its transmissions, the channel was criticized by the Minister of Culture for lack of ambition in its scheduling (Aftenposten 29/9-1993). In 1996 the Ministry of Culture appointed a Public Service Broadcasting Council (Allmennkringkastingstingsrådet) in order to survey the channels with public service obligations (NRK's radio and television channels, TV 2 and P4). P4, and to some extent NRK's P3, was criticised[3] for not complying with the council's definition of public service broadcasting[4], but, as Table 5.1 shows, the channels were accepted by the listeners.

| Table 5.1: Market share for Norwegian radio channels 1992 and 1995		
	1992	1995
NRK/P1	39 %	41 %
NRK/P2	27 %	5 %
NRK/P3	-	12 %
NRK tot	66 %	60 %
P4	-	27 %
Local radio	30 %	12 %
Other channels	4 %	1 %
- : this channel did not yet exist.		

Source: MediaNorge's databases, with data from TNS Gallup and NRK

---

3 See the Council's annual reports 1997 – 2003. For an overview of the reports, see:
   http://www.smf.no/sw2509.asp

4 See the Council's annual report for 1997:
   http://www.smf.no/graphics/SMF/Rapporter/Kringkasting/akrapp_1997.pdf

With a partial exception of P2, the four nation-wide channels reached the intended age groups (see figure 5.1 with data from 1995). P3 was the most attractive channel for young people. Young adults (20-39) preferred P4, while people over 40 overwhelmingly turned to P1. P2 had approximately the same market share in all age groups from 39 and above. This situation was more or less the same into the early years of the new millennium.

Figure 5.1
Market share in different agegroups 1995
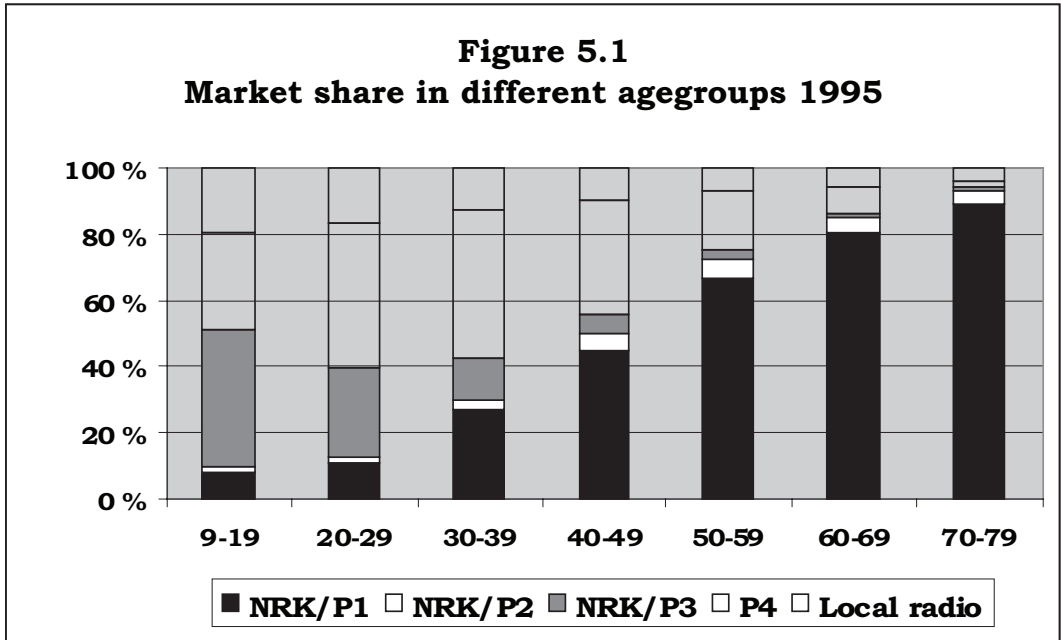
Source: Application

5.2.3 Kanal4/Kanal 24 vs. P4

P4's licence for the fourth FM network expired at the end of 2003. The agreement from 1993 between the Ministry of Culture and the P4 indicated that under normal conditions the holder could expect a renewal of the licence for the next period. But the Ministry decided to go for an open beauty contest. The public service obligations were somewhat more specific than ten years earlier, and the winner would have to pay NOK 160 mill (€ 20 mill.) for the licence.

It was also made known that a fifth FM-network was about to be launched. While the fourth network was estimated to reach more than 90 percent of the population, the fifth network was estimated to cover only 50 percent in early 2004, increasing to 60 percent within the first year. The charge for this network was NOK 90 mill. (11 mill. €).

When the Ministry invited applications for the licence for the fourth FM network, P4 obviously thought that this announcement was a mere formality. However, six groups applied, most of them dominated by well established media companies; among them P4, TV 2 and a new group – Kanal4 – dominated by six newspapers, mainly strong local papers.

It came as a surprise when the Ministry of Culture on December 20th 2002 announced that Kanal4 had been awarded the licence. There were strong reactions from the P4 company and some of its employees. P4 mobilised its listeners, and there were also reactions from leading politicians from opposition parties. Accusations of favouritism flourished. The Ministry stated that it had made its decision on the basis of the applications only. Kanal4 had the most ambitious plans, and could rely on resources from the newspapers that owned the radio channel, thereby creating an exciting radio channel.

P4 referred the matter to the Ombudsman, both in an attempt to overturn the decision, and in order to make public all documents relevant to the case. The Ombudsman required all documents to be made public[5], but he did not change the decision. His final verdict came on June 6th 2003. Three weeks later, it was announced that P4 was awarded the licence for the fifth FM network.

The last six months before the new channels (Kanal4 in the fourth FM network and P4 in the fifth) were due to open, P4 was fighting hard in order to maintain its position among the listeners. P4 claimed that Kanal4's name was too similar to their own, well established name, and that this could confuse the listeners. Just before Christmas, P4 obtained a court decision against their opponent's name, and the new channel had to change its name to Kanal 24 just a week before its program was launched. The licence for the fifth FM network gave P4 the right to start broadcasting in this new FM network from midnight December 31st 2003/January 1st 2004. From the same moment, Kanal4 (now, Kanal 24) was supposed to take over the fourth FM network.

The lines and transmitters for both networks are owned by Norkring, a subsidiary to Telenor (the regulated and partly privatized telecommunication incumbent). By means of some investments in the new network, Norkring was able to increase the coverage of the fifth FM network to 70 percent from the start of 2004, and add another ten percentage points to this on a longer term basis.

When awarding the licences, the Ministry of Culture had obviously intended a continuation of programs in the fourth FM network from one operating company to another. However, while the licences to operate commercial radio channels are regulated by the Ministry of Culture, Norkring is regulated by the Norwegian Post- and Telecommunication Authority (NPT). When Norkring gave P4 permission to start transmitting on the fifth FM network (and close down the transmission in the fourth network) approximately one week before the licence was actually valid, NPT apparently had no objections. Kanal 24 thus had to start its transmissions in a network that had been «cold» for a week; and only a negligible fraction of the audience still had their radio receivers tuned into the fourth network when Kanal 24 started to broadcast.

---

5 All documents are available at:
  http://www.odin.dep.no/kkd/norsk/tema/medier/konsesjon/043061-990015/dok-bn.html
  This article is to a large extent based on analysis of these documents.

The battle between P4 and Kanal 24 to avoid starting up in a cold network indicates that the consumers' switching costs are significantly higher for radio than for television. A large part of the consumers have tuned in to a specific station, and are not zapping from channel to channel for radio in the same way as they do for TV. The higher degree of switching costs is probably one of the key features that distinguish the competition between commercial radio channels from competition between commercial TV channels. Summing up, the two main reasons why Kanal 24 believed they would gain a competitive advantage over P4 turned out not to hold at the end of the day. First, the degree of coverage offered by Norkring was planned to be significantly higher in the fourth than in the fifth network. This expected advantage was jeopardized with the regulated Norkring as a facilitator and, in the end, the difference in coverage offered was much smaller than initially planned. Second, Kanal 24 assumed that it would take over P4's clientele in the fourth network, but P4 started broadcasting in the fifth network before Kanal 24 was on the air, and, consequently, kept most of their listeners. This ingenious way of changing the rules of the game, which gave Kanal 24 a cold start, is probably a main reason for Kanal 24's problems and P4's success.

5.2.4 A change in the ownership of Kanal4/Kanal 24

When Kanal4 applied for the licence of the fourth FM network, the composition of the shareholders was as shown in table 5.2.

Table 5.2: Shareholder of Kanal4 when the application was handed in	
Adresseavisen ASA	20 %
Agderposten AS	11 %
Fædrelandsvennen AS	17 %
Gudbrandsdølen Dagningen AS	11 %
Harstad Tidende Gruppen AS	13 %
Mediehuset Vårt Land AS	11 %
Norsk Telegrambyrå AS	5 %
21st Venture AS	11 %
Source: Application	

Adresseavisen and Fædrelandsvennen are regional newspapers; Agderposten and Gudbrandsdølen Dagningen are major local newspapers. Vårt Land is a national Christian newspaper, but also the hub of a small chain. Harstad Tidende Gruppen is a chain of local newspapers in Northern Norway. Norsk Telegrambyrå is the old, national news agency. 21st Venture was a group of financial investors.

Schibsted, Norway's leading media conglomerate, was an important actor in all media sectors with the exception of radio. Radio was indirectly included in Kanal4's portfolio, as Schibsted was an important shareholder in Adresseavisen, Fædrelandsvennen, Harstad Tidende Gruppen and Norsk Telegrambyrå.

The commercial television company, TV 2 (of which 33% of the shares are owned by

Schibsted), had also applied for the radio licence, but lost to Kanal4. When it became clear that Kanal4 would face competition from the well established P4, some of the investors got cold feet. In June 2003 – midway between the decision to give the licence to Kanal4 and the start of the broadcasting – the media reported that TV 2 wanted to buy either P4 or Kanal4 (Dagens Næringsliv 5/7-03). TV 2 opted for Kanal4 and bought 34 percent of the shares, with an option of another 17 percent (Aftenposten 1/8-03). In October 2005, TV 2 passed the 50 percent mark, and owns 51.3 percent of the shares (Kampanje 7/10-05). The other shareholders are now Adresseavisen, Agderposten and Fædrelandsvennen.

When Kanal 24 started, it soon became obvious that the channel had difficulties in attracting the predicted audience share. This meant even worse problems in attracting commercials, and the revenue was low. At several stages, the owners have consequently had to increase their investments. Some of the investors have sold their shares, and TV 2 has increased its dominance in Kanal 24. This is particularly visible for the news programs in Kanal 24. During one period, Kanal 24 sent the soundtrack from TV 2's main news program, and cuts from interviews etc. from TV 2 were used in Kanal 24's news programs without reference to an external source. There are no traces of the use of regional and local newspapers as producers of news items for Kanal 24, as was promised in Kanal 24's application and mentioned in the press release from the Ministry of Culture when Kanal 24 was awarded the licence (Messel 2005).

5.2.5 The present Norwegian radio market
From 2003 the Norwegian radio market via analogue, terrestrial transmission has consisted of three channels from NRK, Kanal 24, P4 and numerous local radio channels. The total volume of listening has probably been more or less stable since the introduction of P4 in 1993 (TNS Gallup/NRK via MedieNorge's databases). The opening of Kanal 24 meant that five nation-wide channels now compete on this market. Kanal 24 still has severe problems in reaching the expected market share (see table 5.3).

Table 5.3: Market share for Norwegian radio channels 2000 and 2004		
	2000	2004
NRK/P1	47 %	48 %
NRK/P2	4 %	5 %
NRK/P3	8 %	7 %
NRK total	59 %	60 %
P4	28 %	22 %
Kanal 24	-	5 %
Local radio	12 %	12 %
Other channels	1 %	1 %
- : this channel did not yet exist. Source: MediaNorge's databases, with data from TNS Gallup and NRK		

5.3 Competition in the market and competition *for* the market.

In this section we argue that the development in the Norwegian market for commercial radio to a large extent is consistent with predictions from economic theory. Economic theory shows that competition does not necessary create diversity in media markets. Furthermore, high switching costs may explain why two channels with apparently quite similar program profiles differ greatly in the size of their audiences.

With a market share of no more than five per cent, Kanal 24 is such a minor player that the competitive pressure in the Norwegian radio market is presently relatively weak. Indeed, the battle between P4 and Kanal 24 illustrates nicely how competition in the market may be less important than competition *for* the market in the broadcasting sector, as in many other segments of the information industry. P4 is the winner, largely because it managed to change the order of moves in the game.

5.3.1 Competition, diversity and the choice of program profile

The choice of program profile is a crucial factor in determining the possible success of a media firm. Should it for instance primarily try to attract a young public, a public with high education, women, or special interest groups? Among other things, the answer to this question depends on the program profile of other media firms and whether the firm can achieve higher advertising revenue from one segment than from others.

To keep things simple, assume that people's preferences for radio programmes can be defined in terms of their age and that only people between 15 and 55 years listen to radio.[6] Suppose that there are equally many people in each age group, say 10 000, and that each person listens to one and only one radio channel. There will then be 400 000 radio listeners altogether (10 000 x 40). This is illustrated in Figure 5.2.[7] The youngest people prefer typical youth programs, but the older people are, the more they prefer what we may label adult programmes. A radio channel which is "located" at point 35 thus has a program profile that perfectly matches the preferences of 35 year old people, but is not particularly attractive to the youngest or the oldest.

---

6 This section builds on Kind and Sørgard (2004).

7 This model dates back to the British economist Harold Hotelling (1929), and was in the 1950s used to analyze the consequences of allowing private radio channels to compete against BBC. See e.g. Anderson and Coate (2005) and Anderson and Gabszewicz (2005) for recent formal analysis of the media framework within the Hotelling framework.

Figure 5.2: Choice of program profile.
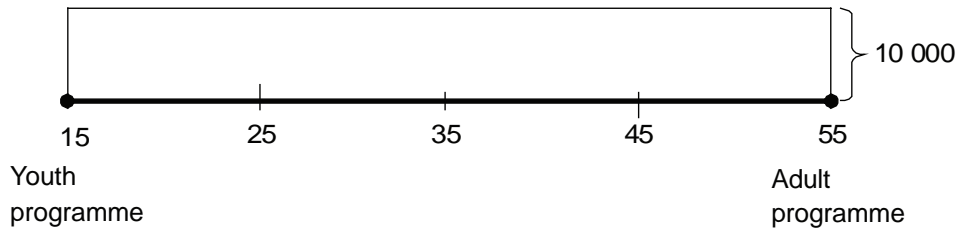
400 000 radio listeners: 10 000 in each age group between 15 and 55.

If there are only two radio channels in the market – call them Kanal 24 and P4 - which locations will they choose? Suppose first that the two radio channels are owned by the public sector, and that the aim of the public sector is to maximize the public's utility of listening to radio. In this case it will be optimal for the public sector to choose program profiles such that the average difference between what the audience actually listens to and what they are offered is as small as possible. With two radio channels, the average difference will be minimized if Kanal 24, say, is located at point 25 and P4 at point 45 in Figure 2. Those between 15 and 35 will then listen to Kanal 24, and those between 35 and 55 to P4. Each radio channel will accordingly have 200 000 listeners, and those who are 25 and 45 years old will be most satisfied.

What if the two channels are owned by a private media house, which behaves as a monopolist in the media market? If each age group is equally profitable on the advertising market, the monopoly would choose the same age profiles on the radio channels as the public sector. This is true even though the monopolist does not care about the public's utility per se. The reason is that no other program profile can generate a larger public and thus a higher advertising income. This is one of the rare examples where a monopolist and a utility maximizing government will generate the same outcome.

But what if the two radio channels are independent, competing firms? Should we still expect to find Kanal 24 at point 25 and P4 at point 45? No. Suppose that Kanal 24 chooses a program profile suited to people who are just slightly younger than 45, where P4 is located. Then everyone younger than 45 will prefer Kanal 24 – giving this channel 300 000 listeners. P4, on the other hand, will have only 100 000 listeners.

P4 will of course be aware of this, and we can see where the story will end: both channels will choose the same program profile, namely the one given by age group 35. Thus, competition does not create diversity. On the contrary, the competing channels will have overlapping program profiles.

This simple model explains why many economists have expressed doubt over whether media competition really creates diversity. The model is certainly a great simplification of reality, but in Norway we actually observed that both Kanal 24 and P4 focused on age groups around 35 years old. This is illustrated by Figure 5.3, which shows Kanal 24's

own judgement of their location relative to the competitors.[8] In addition to age, the figure shows the educational level of the listeners. Also along this dimension the channels offer pretty similar program profiles.

Figure 5.3: Profiles of Norwegian radio channels
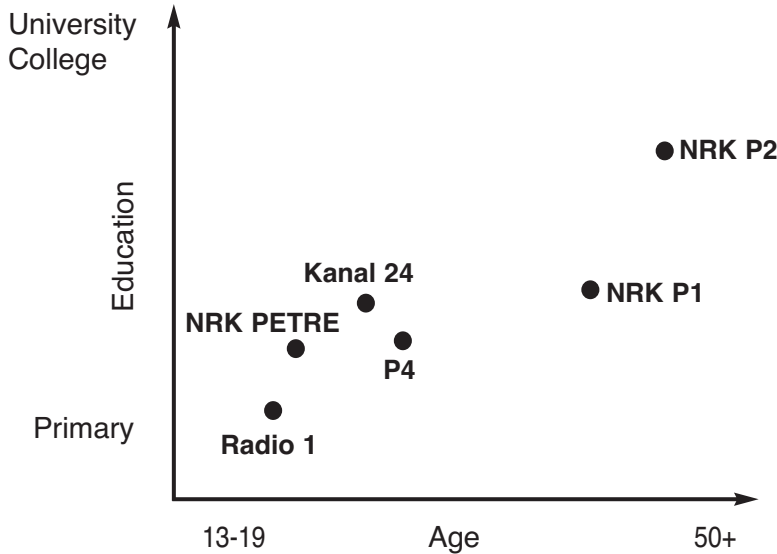
As discussed above, NRK P1 is by far the largest Norwegian radio channel (with a market share twice as large as that of P4 over the last few years). So why did Kanal 24 not choose a program profile close to that of P1 instead?

There are probably two main reasons for this. First, commercial radio channels like Kanal 24 must attract an audience that can generate high advertising income. In this sense people between 30 and 40 years old seem to be particularly interesting, both in Norway and other countries. This age group has a relatively high purchasing power, and is more responsive to ads than older people. Second, empirical analysis indicates that a large share of the radio listeners dislike being interrupted by commercials. Since P1 has no ads, it would probably be very difficult for Kanal 24 to be competitive if it chose a program profile close to that of P1. Why should people listen to Kanal 24 and be interrupted by commercials every now and then if instead they could listen to a similar radio channel without ads?

Kanal 24's choice of location is thus as we should expect from economic theory: close to the biggest commercial competitor. But why, then, has Kanal 24 been so unsuccessful relative to P4 (with market shares of approximately 5% and 22%, respectively)? This is the focus of the next section.

---

8 See http://pub.tv2.no/multimedia/kanal4/archive/00128/Annonsere_p_Kanal_ 128886a.pdf.

5.3.2 Switching costs

People do not zap between radio channels to the same extent as they do between TV-channels; apparently, the switching costs are perceived to be too high to change radio channel several times a day. Thus, people might be perfectly indifferent between two radio channels, e.g. Kanal 24 and P4, *before* turning to one of them. However, after having tuned in to a station, people are willing to switch to the other channel only if they perceive it to be much better. Thus, it is generally hard for an entrant to capture large market shares if switching costs are high.

In most markets with switching costs an entrant must have a significantly lower price or higher quality than the incumbent. Commercial radio does not charge the listeners directly. However, empirical analysis shows that people tend to dislike advertising, which may thus be perceived as an indirect price for listening to commercial radio. In order to steal listeners from the incumbent's clientele, entrants may therefore have to accept a relatively low level of advertising and/or invest more in programming than the incumbent. Both these strategies are obviously costly.[9]

It is well known that firms tend to compete more fiercely for the market than in the market if switching costs are high. To take one example, Microsoft does not face much competition from rivals producing word processors. People are simply not willing to switch to a new word processor unless it is clearly superior to Microsoft Word. However, Microsoft initially had to fight vigorously to take over the "incumbent advantage" from the former industry leader, WordPerfect. In the case at hand, Kanal 24 expected to take over P4's clientele, and thus the incumbent advantage, when it was due to start broadcasting in the anticipated "warm" fourth FM network. On this background, Kanal 24's choice of program profile makes perfect sense. By choosing a profile close to P4 fewer listeners would switch back to P4 (in the fifth FM network) than if Kanal 24 presented a profile more different from P4.

However, as described above, P4 managed to change the rules of the game, and with permission from the regulated Norkring P4 started broadcasting on the fifth FM network approximately one week before the licence was actually valid. Furthermore, P4 was allowed to close down its transmission in the fourth network from the same date. Thus, P4's listeners switched to the fifth FM network, and as a result Kanal 24 had to go on the air in a «cold» network without a clientele of listeners. When the competition in the market started from midnight December 31st 2003, P4 maintained its incumbent position.

If Kanal 24 had anticipated that P4 would keep its clientele, it would probably have chosen a program profile that differed more from that of P4. After a few months on the air Kanal 24 found that it had to restructure its program profile, and in September 2004 it introduced a music and program profile closer to that of NRK's P3 than P4. Whether Kanal 24 will manage to attract a large share of P3's clientele and other listeners remains to be seen, but the success has not been overwhelming so far.

---

9 See Kind, Schjelderup and Stähler (2006) for a formal analysis,

5.4. Concluding remarks

When the Ministry of Culture gave Kanal 24 the right to broadcast in the fourth FM network, it was widely expected that Kanal 24 would take over P4's base of listeners from midnight December 31st 2003. By employing some brilliant strategic moves, P4 managed to turn the game around. This is a textbook example of how to behave in a market with switching costs. It is also a textbook example of a rather unsuccessful regulatory policy, partly caused by fragmentation of responsibility. While the Ministry of Culture obviously aimed at improving the quality of the Norwegian radio market by giving the entrant Kanal 24 some initial advantages, the Norwegian Post- and Telecommunication Authority and the regulated Norkring became useful tools that P4 could use to minimize the threats from the entrant. With a more coherent regulation, it is likely that P4 and Kanal 24 would have ended up with somewhat more differentiated program profiles than they actually did. But both theory and observations from other countries suggest that the broadcasting market's ability to create diversity is limited.

If "57 channels and nothing's on" is the outcome of market forces, which tool may then be used to ensure diversity? Perhaps it is necessary to have non-commercial public service channels? At first glance, such a view is supported by the fact that NRK's three main channels have distinctly different profiles, while those of the competing commercial radio channels Kanal 24 and P4 are quite overlapping. However, NRK did not leave its motto "more of the same in two channels" until it was challenged by a major commercial rival (P4). Hence, it may be argued that commercial radio has lead to more pluralism in the Norwegian radio market, but mainly through the impact on the diversity within NRK. Soon after the battle for the pole position between Kanal 24 and P4, and the tough infancy period for Kanal 24, the two rivals started to check out the opportunities for a merger. The Ministry of Culture poured cold water on these ideas; a license for a second commercial nation-wide radio channel was granted precisely to create more diversity. If both licenses were given to one and the same media firm, the radio market would be back to square one.

Stein Gauslaa, editor of the regional newspaper Agderposten, and one of the initiators behind Kanal 24, claimed that the Ministry of Culture was wrong. Gauslaa instead argued that a merger would give the owners strong incentives to differentiate the program profiles (Dagens Næringsliv, 22 February 2005): aggregate profit cannot be maximized by operating two almost identical radio channels. Hence, he insisted that the Ministry of Culture should welcome such a merger, since that would generate more differentiation. One possible objection against Gauslaa's recommendation is that it could be even more profitable for the merged company to close down one of the channels than to differentiate their profiles. However, this could presumably be prevented through licensing conditions. But in a democracy it would certainly raise political concerns to let one media firm more or less monopolize the nation-wide commercial radio market.

References:

Anderson, S. P. and S. Coate (2005) Market Provision of Public Goods: The Case of Broadcasting, Review of Economic Studies, 72 (4); 947-972.

Anderson, S. P. and J. J. Gabszewicz (2005) The Media and Advertising: A Tale of Two-Sided Markets, Discussion Paper 5223, Centre for Economic Policy Research; Handbook of Cultural Economics (V. Ginsburgh and D. Throsby, eds.), Elsevier.

Armstrong, Mark (2005) Public Service Broadcasting, Mimeo.

Armstrong, Mark and Helen Weeds (2005) Public Service Broadcasting in the Digital World, Mimeo.

Dahl, Hans Fredrik and Henrik G. Bastiansen (1999) Over til Oslo. NRK som monopol 1945–1981, Oslo: Cappelen.

Halse, Ketil Jarl and Helge Østbye (2003) Norsk kringkastingshistorie, Oslo: Det norske samlaget.

Hotelling, Harold (1929) Stability in Competition, Economic Journal 39; 41.57.

Kind, Hans Jarle and Lars Sørgard (2004) Mediekonkurranse: P4 versus Kanal 24 (English title: Media competition: P4 versus Kanal 24), Magma 2(8); 83-92.

Kind, Hans Jarle, Guttorm Schjelderup and Frank Stähler (2006) Taxation of newspapers: less differentiation and higher quality?, Mimeo.

Mazzoleni, G. (1986) Italy, pp. 169–190 in H. J. Kleinsteuber, D. McQuail and K. Siune (eds.): Electronic Media and Politics in Western Europe, Frankfurt: Campus Verlag.

McQuail, Denis (1990) Caging the Beast: Constructing a Framework for the Analysis of Media Change in Western Europe, pp. 313–331, European Journal of Communication (Vol. 5) 1990.

Messel, Kari Wicklund (2005) Kanal 24 – En studie av TV2s oppkjøp av Kanal 24, og de mulige konsekvenser dette kan ha for radiokanalens nyhetsutvalg samt nyhetsmangfoldet i Norge, Bergen: Department of information science and media studies (master thesis).

Ot.prp nr. 55 1989-90: TV 2. Lov om reklame i kringkasting m.v.

Siune, Karen (1986) Broadcasting, point of departure, pp. 44–54 in Denis McQuail and Karen Siune: New Media Politics, London: Sage.

St.prp. nr. 88 1981-82: Om medieutvikling og nye kringkastingsformer.

Tunstall, J. (1986) Great Britain, pp. 110–134 in H. J. Kleinsteuber, D. McQuail and K. Siune (eds.): Electronic Media and Politics in Western Europe, Frankfurt: Campus Verlag.

Tommy Staahl Gabrielsen

Increased retail power in the Norwegian grocery market: To the benefit of consumers?

6.1. Introduction

Since 1980 the Norwegian grocery sector has undergone drastic changes. Around 1980 the grocery industry was dominated by a small number of producers and wholesalers. The vertical structure was a very traditional three-layer structure with little vertical integration. The producer and wholesaler markets were heavily concentrated, whereas the retail market was very fragmented with many small and geographically dispersed and independent retailers. Twenty years later the situation is very different. Since 1980 around 20 000 small retail outlets have disappeared, larger supermarkets have emerged, and almost every retailer is now a member of a retail group. Today four large vertically integrated retail groups dominate the retail and wholesale markets. By vertically integrating the former wholesale and retail level, the price setting power of the wholesalers that traditionally used to be strong is gone.

In contrast to many other sectors in the Norwegian economy, the development in the grocery sector was not induced by market deregulation or active liberalisation by a change in public policy. Instead the actors in the industry largely triggered the restructuring of the industry themselves. One inspiration for the retail sector was the observed popularity of discount chains in continental Europe in the 70s. The present structure emerged through horizontal and vertical mergers and was more or less in place already by the mid 90s.

The main purpose of this chapter will be to try to understand and pinpoint what this relatively dramatic development has meant for the Norwegian society, and especially for Norwegian consumers of grocery products. Has the development benefited Norwegian consumers by offering them better variety of high quality grocery products at low prices, or has the powerful retail chains benefited most? It seems obvious that such changes cannot be made without generating winners and losers, and we will try to point at players that have benefited most. Producers, wholesalers and retailers will benefit if profits increase. Consumers however, will be interested in low priced, high quality products and sufficient innovation of new products to satisfy consumer needs. With heterogeneous preferences consumers as a group will also value product diversity.

Beyond any doubt vertical integration, increased retail concentration and the recent consolidation the last decade have increased retailers' bargaining position towards the producer sector. In the previous regime, the dominant producers and wholesalers could more or less dictate the wholesale terms to the retailers. Now it is the retailers that seem to

have the upper hand. Intuitively, from this we should expect that increased retail bargaining power to lead to better wholesale terms for retailers. If increased retail bargaining power lowers retail costs this potentially could lead to lower retail prices and enhanced consumers' surplus. Also more vertical integration has the potential to alleviate problems with double margins in distribution channels that in turn should result in a downward pressure on retail prices. However, from economic theory we know that there are circumstances under which increased retail power and vertical integration not necessarily will lower retail prices for consumers. Instead increased efficiency in distribution and increased retailer bargaining power may end up as increased retail profits at the expense of the producers in the industry. A discussion of when the effect is likely to be positive to consumers will be an important element of this chapter.

Increased bargaining power on the retailers' part has also brought about some new business practices whose consequences for economic efficiency need to be evaluated. Of particular interest in this respect is the introduction of retailer-owned brands, also denoted as private labels. The private label phenomenon is a relatively new feature of the Norwegian grocery market, and it is expected to grow significantly over the years to come. Also, various types of vertical restraints seem to be more predominant now than before. Of special interest for retail prices and product variety is the use of exclusivity clauses and slotting allowances. There are arguments in the theoretical literature that such restraints may be used to dampen or eliminate competition and thus lead to higher prices and less product variety to the detriment of consumers. We will therefore give these issues a closer treatment.

In the following we start by giving a closer description of the central developments and characteristics of the Norwegian grocery market over the last 2-3 decades. Next we will look closer into the development in prices, product variety and innovation and compare this with the development in other markets. Finally, we will discuss possible explanations for the observed outcome on the Norwegian grocery market, and the consequences for consumers and welfare.

6.2 A brief overview of the Norwegian grocery market

The Norwegian grocery market consists of manufacturers and importers of food and non-food grocery products, wholesalers and retailers. As noted above the wholesale function is largely vertically integrated with retailing for all four retail groups. Below we briefly present the structure of the retail sector, the focus that each group has on discount chains and private labels. Finally we describe the most important players in the producing sector.

6.2.1 The retail sector

Until around 1980 the retail sector consisted of small, independent and geographically dispersed retailers with weak bargaining power. The decisions on product lines and

prices in each retail store were partly dictated by the producers themselves and partly by a handful of powerful wholesalers. Today the situation on the retail market is almost completely reversed. Four important retail groups, Norgesgruppen, Ica Norge as, Coop Norge and REMA 1000 (Rema), together controlling 98% of the retail market for grocery products, now dominate the retail sector. Three of the retail groups are umbrella organisations covering several different profile chains. The fourth group, the Rema group, has a homogeneous profile where all outlets are discount stores carrying a limited number of product lines. The discount profile has gained significant popularity in the Norwegian market, and all major retail groups have at least one discount profile chain. All groups except the Rema group also have other profiles with a larger number of product lines. Hard discounters, which are popular in many European countries, have not yet gained terrain on the Norwegian market. However, this is expected to change somewhat over the next years after the German hard discounter Lidl entered the Norwegian market in 2003.

The table below illustrates the development in market shares for the four leading retail groups over the last decade.

Figure 6.1 Development in retail market shares 1994 – 2004 for the four major Norwegian retail chains. Source: ACNielsen.
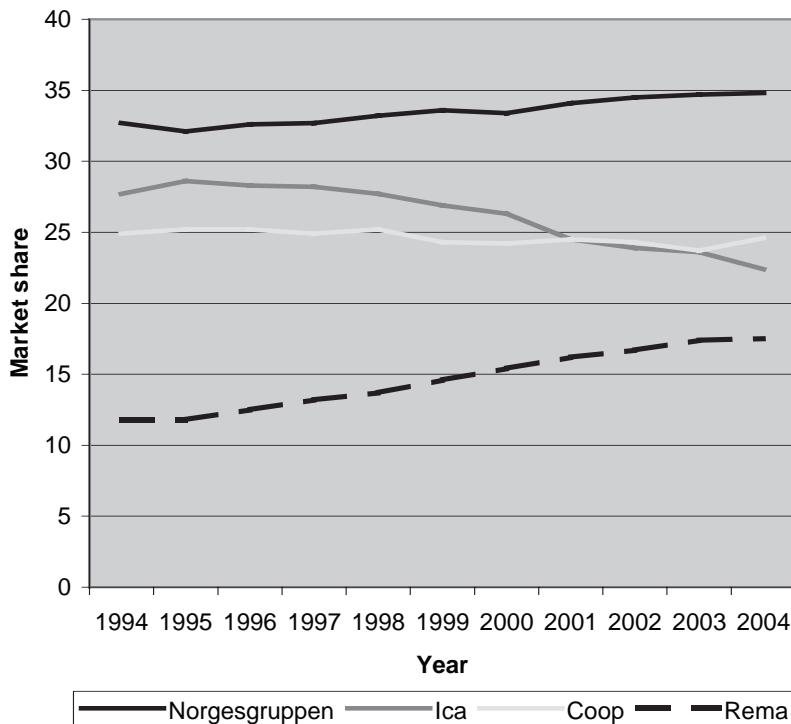
The European trend from the 70s with discount chains were introduced in Norway around 1980, and its main proponent was Rema.[1] Rema originally started out with a very limited product range, around 500-600 product lines, but soon found out that at least 1000 product lines were needed to generate sufficient demand, hence the name Rema 1000. Today a typical Rema outlet has approximately 2500 product lines. Rema's main business idea is mainly copied from the German discounter ALDI with simple and recognisable stores with a relatively limited product line. The chain relies on franchising as its main organisational form. Each retailer carries often one or only a few products in each category and the supply rights are auctioned off to producers on a yearly basis. Prices are mainly set at the central level and remain the same independent of the geographical localization of each outlet. From its initial beginning in 1980 with only three stores, the chain today has 380 stores in Norway and around 6500 employees. Its market share in the retail market has been steadily increasing and is currently close to 18 %.

Norgesgruppen is the largest retail group. It covers around 35 % of the market. This group originated from a wholesaler that subsequently integrated into retailing, and the current organisation was in place in 1994 as cooperation between wholesalers and retailers. The group has several profile chains under its main umbrella, some with relatively high product variety, but also one dedicated discount chain. The group also have a looser cooperation with some associated chains. In total the group handles the wholesale activities for 1751 grocery stores and has a turnover around 35 billion NOK. The central body negotiates with the industry on behalf of all members and associated members, and the product lines in each profile chain are negotiated between the central group and each profile chain. Also in this group consumer prices are set by the central body and thus remain the same in all outlets. Each store pays a monthly service fee to the central group. ICA Norge AS has currently around 23 % market share in Norway. It also comprises several different profile chains, from discounters to chains with larger variety. Also in this group the prices charged by each profile are determined at the central level and therefore are the same at the national level. The group controls 1028 shops and has a turnover close to 23 billion NOK (2003). ICA Norge AS is a subsidiary of ICA AB, a part of Royal Ahold that is the world's third largest retail group. The group employs around 15 000 people in Norway. The present organisation was established through a merger between the former Norwegian group Hakongruppen and the Swedish ICA AB in 1999.

The final group Coop Norge AS is owned by the consumers, and runs 953 retail outlets in Norway. It is a fully integrated group with profile chain within all the major segments. The group has same prices in all outlets as its main principle.

As illustrated above, an extremely high concentration level characterizes the Norwegian retail market for groceries. Table 6.1 below highlights how the Norwegian structure compares with other markets by illustrating the market shares and concentration in a number of EU countries and Norway in 1999.

---

1 Rema 1000 is owned by the Reitan group which is an integrated group involved in distribution and retailing of grocery products. The Reitan group also owns the convenience chains 7-eleven and Narvesen.

Table 6.1. Market shares and concentration for the five largest retail and wholesale groups in the EU and the four largest in Norway, 1999. Source: Dobson et al (2003).

Country	Total share 5 largest	Share of largest	HHI	
Austria	60,2	19,6	880	Asymmetric oligopoly
Belgium+Luxembourg	60,9	23,7	950	Asymmetric oligopoly
Denmark	56,4	21,9	932	Duopoly
Finland	68,5	29	1410	Duopoly
France	56,2	17,8	698	Asymmetric oligopoly
Germany	44	10,9	408	Symmetric oligopoly
Greece	26,8	9,5	166	Non concentrated
Ireland	62,1	18,8	927	Asymmetric oligopoly
Italy	17,6	5,1	71	Non concentrated
The Netherlands	56,2	29,1	1112	Dominant firm
Portugal	63,3	19,5	946	Duopoly
Spain	40,3	17,9	449	Asymmetric oligopoly
Sweden	78,2	36,5	1804	Dominant firm
Great Britain	63	21,1	922	Asymmetric oligopoly
Norway	99,4	33,6	2657	Asymmetric oligopoly

6.2.2 Private label penetration in Norway and Europe

Private labels (PL) are retailer-owned brands that are sold exclusively in profile chains belonging to a specific retail group. All the major retail groups in the Norwegian market have private labels. However, the PL-share in Norway has not reached the level that we can observe in many other European markets. The private label value share in Norway increased from slightly below 5 per cent in 1995 to 8 per cent in 1999, and after a sharp growth lately the present value share is close to 10 %. This is in sharp contrast to Switzerland who has 45 % private labels, and the only European markets that have lower PL-share than Norway are Ireland, Greece and Israel. In Figure 6.2 below we have plotted the correlation between PL-penetration and the market share of the top five retailers in a number of countries.

Figure 6.2: PL-penetration (value) and aggregate market share of the five largest retailers in 2005 in selected European markets. Source: ACNielsen (2005)
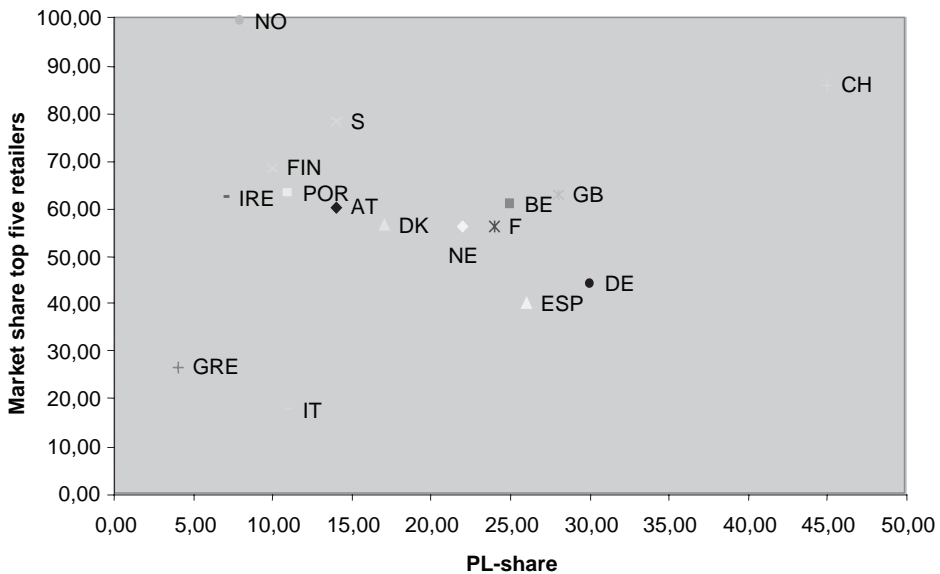
From Figure 6.2 it is interesting to note that it seems be the case that the higher the concentration in the market the lower the PL-share. There are however some exceptions to this rule. Greece and Italy both combine low penetration of private labels with low market concentration whereas Switzerland both has high concentration and a high PL-share. Norway, and to some degree Switzerland and Sweden, are special by having an extremely high market concentration compared to other countries.

Worldwide the share of private labels in retailing is growing, particularly in Europe. The main reason for this is the growing presence of hard discounters that mainly sells private labels.[2] In Norway the German group Lidl is the only hard discounter, but currently Lidl has gained only a minor market share in Norway. It is also noteworthy that PL-growth is very pronounced in Norway's neighbouring countries Sweden and Finland (an annual growth of 10 and 16 %, respectively), but until quite recently the growth in Norway has been fairly modest.

Private label shares are not equally distributed among product categories. Refrigerated food currently has the highest overall share of Private Labels (ACNielsen, 2005). The table below shows worldwide PL-shares and PL growth for a number of product areas.

2 Hard discounters sell a very limited selection of products (primarily shelf stable food) at a very low price.

Table 6.2: Private label share by product area (value sales), first quarter 2005. ACNielsen (2005).

	Product area	PL share		Product area	PL share
1	Refrigerated food	32	8	Non-alcoholic beverages	12
2	Paper, plastic and wraps	31	9	Home care	10
3	Frozen food	25	10	Snacks & confectionary	9
4	Pet food	21	11	Alcoholic beverages	6
5	Shelf-stable food	19	12	Personal care	5
6	Diapers & feminine hygiene	14	13	Cosmetics	2
7	Health care	14	14	Baby food	2

Globally, private labels were on average priced 31 % lower than their manufacturer counterparts (national brands). The biggest price differential between private and national brands are found in the category 'personal care'. The figure below shows the PL price differential versus the corresponding national brand.

Figure 6.3: Private label price differential versus national brands. ACNielsen (2005)
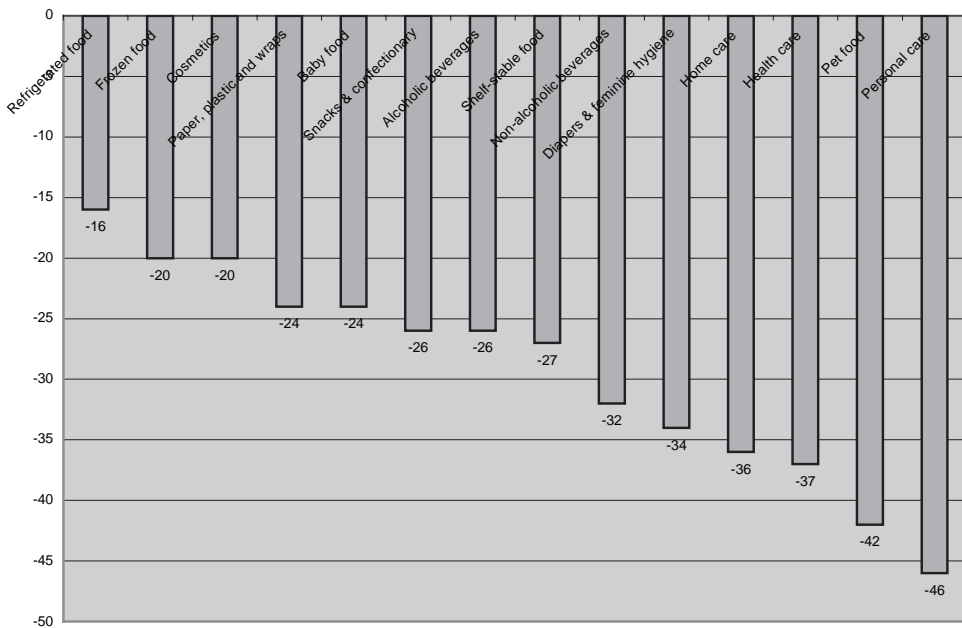
It is interesting to note that the product category with the highest PL share (refrigerated food) has the lowest price differential, while personal care items has the largest price differential and a very low PL share (5 %). In Norway the average price differential between national brands and private labels is 34 %, which is slightly above the world average, but the PL value share in Norway still remains among the lowest in Europe.

According to ACNielsen it is also a fact that the private label spending as a share of total spending on groceries is highest among low-income groups and in large households. In Europe the middle-aged consumers tend to purchase more private labels than the young and old, but this varies greatly in other parts of the world.

6.2.3 The manufacturing sector

The manufacturing sector in Norway is also relatively concentrated. In most product categories there are only one or a few dominant firms. Suppliers to the Norwegian markets are international conglomerates (Nestlé, Proctor and Gamble, Unilever), Norwegian corporations selling mainly to the Norwegian market (Rieber & Søn, Orkla, Oluf Lorentzen), producers of international branded products (Coca Cola, Kellogs, Santa Maria), domestic producers of national brands (Tine, Gilde, Mills, Sætre) and small local niche producers (mainly agricultural products). In addition there are producers of private labels. These are mainly producers without any strong national brands. Lately there is a tendency that also national producers of strong national brands produce and supply private labels to the retail groups.

6.3. Consumer prices and product variety

Grocery products consist of both non-food and food items. It is a well-established fact that Nordic prices on grocery products are higher than in many other countries, and it is also relatively well documented that Norwegian prices are especially high. Also product variety, as measured by the number of product lines in each category for a typical outlet, tends to be lower on the Norwegian market than elsewhere. A recent report produced by a working group established by the Nordic competition authorities (NCA) (2005) investigates these issues, and we will briefly review the main results from this investigation. NCA (2005) concludes that prices on food products tend to be significantly higher in the Nordic countries than in other European countries, and Norway and Iceland stand out as the countries with the highest prices. Part of the price differences can be explained by differences in VAT, taxes and campaigns. However, even after correcting for this the Norwegian prices on food items (excluding alcoholic beverages) are around 40% higher than the EU15 average.

Table 6.3: Net food and beverages prices (excl. taxes) 2004. NCA (2005).
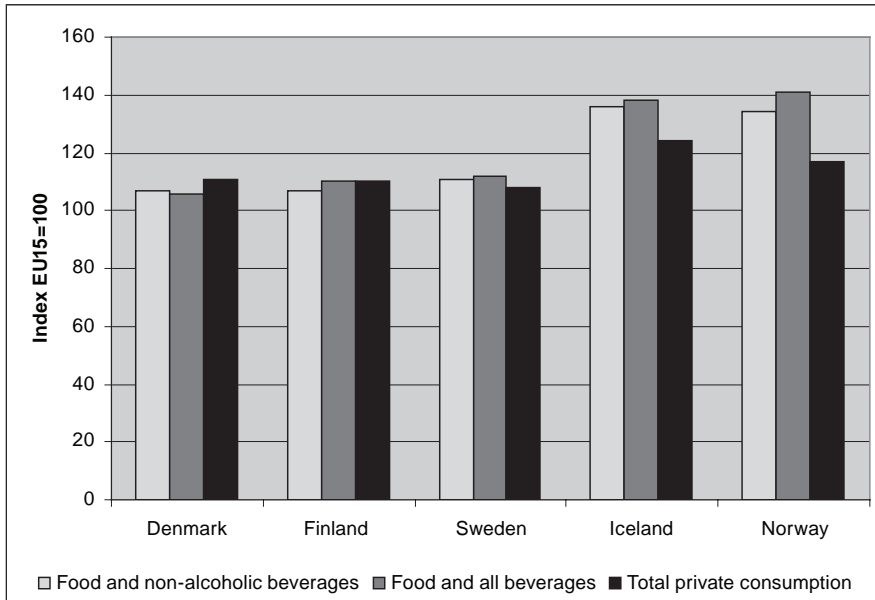
The gap between the Nordic countries and EU15 has been decreasing somewhat the last few years but still remains considerable. Also within European markets the price differences on grocery product are significant, but the difference seems to be decreasing. In a recent survey published by ACNielsen (2005)[3] the prices of international branded grocery goods in 15 markets are compared. In 2002 the span from the cheapest to the most expensive country was 71 %, but lately the gap is reduced to around 50%. According to this survey Norway is Europe's most expensive country in which to buy international branded grocery products. On average Norwegian prices on these products are 42.5 % higher than Germany, Europe's cheapest country for the same products.

Figure 4 below shows the development in CPI, food, non-alcoholic beverages and personal care items in Norway the last 20 years.

---

3 ACNielsen produces a biannual 'Euro Price Barometer' that covers prices and price differences of 160 international branded products for fifteen European markets.

Figure 6.4: CPI and the development in the prices of food, non-alcoholic beverages and personal care in Norway, 1985-2006. Source: Statistics Norway.
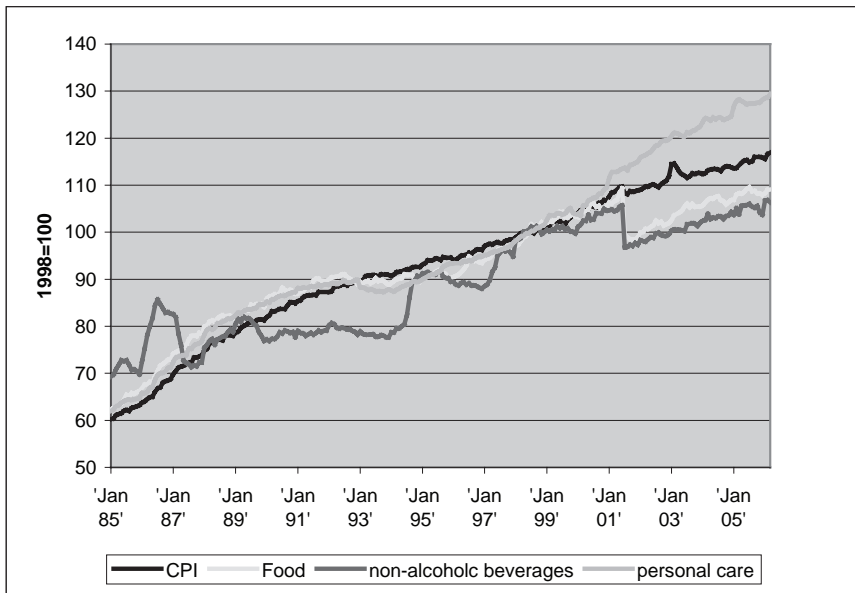
The sharp drop in the prices of food and non-alcoholic beverages in mid 2001 is due to a reduction in VAT for these products. Up till 1995 there was a tendency that prices on grocery products in Norway grew less than other products. The last decade this has changed somewhat. From Figure 4 above we see that both the prices of food items and especially prices on personal care items seem to have increased more than the general CPI the last 5 years.

There exist few studies that document differences in product variety in European markets. However, NCA (2005) claims that the selection of products is especially narrow in the Norwegian market. The study finds that the number of food items in an average (artificial) supermarket was lowest in Norway (a little less than 300 items) whereas the average supermarket in France exhibits a little over 700 food items. The difference is probably less due to a lower number of categories in Norway in comparison with France, but rather that the number of lines within each category is more limited in Norway. Table 6.4 reports the number of lines in super- and hypermarkets in different countries for the broad categories of beverages, dairy products and meat.

Table 6.4: Relative number of product lines in broad categories for different markets (Average = 100.) Source: NCA (2005).

	Beverages		Dairy		Meat	
	Superm.	Hyperm.	Superm.	Hyperm.	Superm.	Hyperm.
Denmark	120	97	99	89	85	122
Finland	111	118	125	151	75	81
France	142	143	114	127	152	131
Iceland	57	44	95	66	117	103
Norway	71	95	49	62	52	69
Sweden	101	102	116	104	123	92
Average	100	100	100	100	100	100

We see that Norway scores below the average in every category. Thus, there seems to be ample evidence that Norwegian prices are relatively high in a European context and even in a Nordic context. Moreover Norwegian consumers seem to be exposed to lower product variety than consumers in other European countries.

In the following section we will briefly discuss some potential explanations for why prices on the Norwegian grocery market seem to be considerably higher than elsewhere and also why consumers' selection of products is low.

6.4 Explaining why prices are high and variety low

The observed high prices and relatively low product variety on the Norwegian grocery market can be explained in a number of ways. In this context we will focus on two different explanations; a structural explanation and a strategic explanation. The structural explanation mainly says that the observed outcome is a consequence of specific structural features and characteristics of the Norwegian market as well as specific Norwegian institutions. The alternative, or maybe complementary, explanation takes the view that the outcome, i.e. market structure, product selection and prices, on the Norwegian market comes as a result of strategic behaviour from the agents in the grocery industry. Which of these explanations is the right one will potentially have a big impact on how public policy should be designed to cope with the challenges in the industry. For instance, if product selection is low due to the fact that consumers' preferences for variety are low in Norway, there is not much scope for public intervention. If, on the other hand, product variety is lower than consumers will prefer due to a strategic battle over the rents in the industry, public policy may have an important role to play.

6.4.1 The structural explanation

The central ingredients in this proposal are that high prices and low product variety are a consequence of demography, consumer preferences and inherent costs in the Norwegian

economy. Norway is a scarcely populated country, and with the exception of the area around the capital of Oslo, the population density is also relatively low. Casual empiricism suggests that product variety in shops is increasing with population size and density; larger cities provide a larger range of shops and more product variety within each shop than smaller places, and larger cities more so than smaller cities. An obvious and intuitive reason for this is that a larger population may exhibit more taste variance. With more taste variance more products are demanded and each taste may get sufficient mass to sustain sufficient sales.

Low population density, long distances and relatively poor infrastructure in many places in Norway also make distribution of grocery products more costly than in densely populated areas with better infrastructure. Thin markets and high transportation and distribution costs can contribute to higher prices and lower product variety as one moves further away from the most densely populated areas in the south-eastern part of Norway.

Norwegian regional policy has traditionally aimed at making it possible to inhabit most parts of Norway, and especially the northern part of Norway has been of special concern. An important instrument in regional policy has been agricultural policy. Norwegian agricultural policy protects Norwegian agricultural products to a large extent from international competition with tariffs and quotas, and contributes to inefficient and expensive production of many agricultural products in Norway. Naturally these policies have induced higher prices on many food products than would have resulted if Norwegian wholesalers and retailers could freely import agricultural products from international markets.

Another potential important factor is the wage distribution and specific labour market regulations in Norway. The Scandinavian countries in general, and Norway in particular, are renowned for their relatively high level of wage compression. In addition the labour market is regulated by a relatively high (in comparison with other markets) minimum wage.[4] Wage compression and minimum wages probably means that labour costs in the Norwegian grocery industry are relatively high compared to other international markets, which in turn puts yet another upward pressure on prices. Another feature often attributed to Norwegian consumers is uniformity of preferences. Clearly, with wage compression and homogeneous consumer preferences there is pressure towards demand uniformity as well, in the sense that the consumers' typical shopping baskets are fairly similar. It may also be the case that prices and product variety in Norway are due to specific Norwegian preferences, and that these differ in a systematic way from other countries. There is some evidence that consumers in different national markets have a special preference for domestic products to imported ones. Under this view Norwegian consumers prefer Norwegian agricultural products and are willing to pay the price needed to sustain their production.

---

4 Actually, Norway does not have a minimum wage, but the union negotiated wages in different industries de facto act as minimum wages.

The structural explanation also highlights high costs as one of the main reasons for high prices. Product variety is low simply because there is insufficient demand for product variety due to a small market size or that most consumers prefer the same products within each category. Many interpret the relative strong influx and popularity of discount stores in Norway as an evidence of the latter.

As we have seen above, private label penetration in Norway is lower than in many other markets. When we know that the average price differential between national brands and private labels is huge (31 % globally), clearly this may contribute to the high average prices in Norway. However, this observation begs the question of why there are so few private labels on the Norwegian market. A plausible explanation for this can be that Norwegian consumers have a high loyalty to and preference for national brands and are willing to pay the premium these brands charge.

In sum, under this view the consumers get the product variety and to some extent the prices they demand and deserve. If Norwegian consumers had preferences for more variety, the retailers would have provided it. In other words, the retailers merely act as agents for the consumers. The alternative strategic explanation takes a completely different view. Under this view retailers and producers use their market power to limit product the product range and charge high prices to increase their profits at the consumers' expense.

6.4.2 The strategic explanation

An important characteristic of the Norwegian market is high concentration at the retail level and a high degree of vertical integration. Increased retailer concentration and vertical integration may affect retail prices and product variety in two fundamental ways. First, higher retail concentration tends to increase retailer bargaining power. Secondly, high concentration may in many circumstances reduce the competitive pressure and increase the risk of collusive behaviour.

Increased retailer bargaining power may make it possible to obtain lower retail costs and better wholesale terms from the producers. The question then is how reduced retailer costs will influence retail prices and product variety. This seems to depend crucially how retailer power is exploited towards the manufacturing sector. In this context there are several instruments frequently used in this industry that seems to be of particular interest: slotting allowances, exclusivity requirements and private label introduction.

Slotting allowances are fixed fees paid by the producers to retailers for access to the retail shelves. In the literature there is an interesting debate of whether such fees can contribute to higher prices and reduced product variety, or if slotting fees simply is an efficient instrument to transfer rent from the upstream to the downstream sector[5]. Exclusivity clauses reduce product variety by definition. The question then is if consumers are compensated for exclusivity with lower retail prices. The idea is that by

---

5 See Rao and Sudhir (2004) and Shaffer (1991) for contributions to this debate.

agreeing on exclusivity retailers can obtain better wholesale terms and more efficient distribution, which in turn benefit consumers with lower prices. The alternative view is that exclusivity clauses are instruments to dampen or eliminate competition to the detriment of consumers. Finally, retailer-owned brands (private labels) are often seen as instruments for retailers to elicit further price concessions from the producers of strong national brands. In addition, private labels are often though of as instruments that enable retailers to segment the consumer market and price discriminate between different consumer groups.

As noted above high retailer concentration may also influence consumers' welfare through a more direct channel, namely by reducing competition and increasing the danger of collusion. In this section we will we review these arguments in turn, and we will start with the bargaining power channel for retailer influence. We will take the view that retail costs consists of fixed and variable distribution costs plus payments for goods from the suppliers and services rendered by either party. Different payments for goods and services may be fixed or dependent on the quantity traded between suppliers and retailers.

The focus in this section will be on the effects of retail concentration and high retailer bargaining power on retail prices and product selection and variety. Of course, low prices are always good for consumers, but the effects from changes in variety are less straightforward. The society's valuation for product variety stems from the consumers' preferences. It is normal to assume that consumer preferences are heterogeneous, but the degree and structure of heterogeneity may vary both between markets and even within segments of each national market. In order to evaluate the eventual loss for society from low product variety, we first need to discuss the structure and distribution of consumers' preferences in the Norwegian market. Recall that the structural explanation saw low variety as a result of low heterogeneity in consumers' preferences. In the following we will allow for heterogeneous preferences, and we will discuss whether in spite of consumer preference for variety there may be strategic reasons for the players in this industry to limit product selection possibilities.

6.4.2.1 Consumer preferences

With heterogeneous preferences reduced variety and high prices affect consumers' utility. High prices are always negative, but the effect from reduced variety need not necessarily be negative. For instance, reduced variety in an outlet has no impact on the utility of a consumer who still can get his most preferred product in that outlet. Only those consumers who find their most preferred product removed from the shelves will suffer from reduced variety. For the latter consumers, the loss can be fully or partially compensated by lower prices, or if it is relatively easy to find their most preferred product in another shop close by.

It is also feasible that some individual consumers value diversity per se. For instance, it may be the case that a consumer would like to see different product lines within each

category from which he can choose what he wants. However, this is probably more relevant for goods where the consumers need to see the goods before they know what the most preferred good is.[6] For grocery products it is probably more natural to assume that a typical consumer knows his most preferred brands in advance and the question is where to find a shop closest to him that carries his most preferred product in as many product categories as possible and at the lowest possible prices. With this interpretation it is the consumers as a group that values diversity and not the individual consumer, and this is probably the most reasonable assumption in the grocery market.

In Norway it is clear that there are retailers with a very limited product range. If consumers with heterogeneous preferences are uniformly distributed in the geographical space, reduced variety reduces utility only for those consumers that will be unable to find their most preferred products or that will have to travel further to find it. On the other hand, if discount chains with low variety typically are located in areas where the population of consumers is such that the majority of the consumers have strong preferences for the variety chosen by the discount chains, the loss from reduced variety will be minimal. Exactly how preferences are distributed within the Norwegian population is of course an empirical question. However, clear segregation in the housing market based on income (and preference) differences seems to be the exception rather than the rule in Norway, and if it exists it is probably most pronounced in the capital. This means that it is most likely that a reduction in product variety will hurt consumer utility, and that a store that limits its product range for strategic reasons will suffer a loss in demand if consumers have alternative places to do their shopping. Can it still be profitable for a chain to limit its product range?

6.4.2.2 Retailer bargaining and wholesale terms

Increased retailer bargaining power naturally enables retailers to extract more of the rent generated in the distribution channel. As illustrated above, the Norwegian market is characterized with a high degree of vertical integration where each of the four retail groups to a large extent use separate and exclusive distribution. When a retailer with a market share of, say, 25% or more bargains with the suppliers it is clear that in many product categories this retailer has a strong bargaining power. A big retailer can easily pit the producers up against each other to obtain better wholesale terms. In some categories the upstream sector has countervailing power as an upstream firm may control the dominant brand in the category or controls a bundle of brands that constitutes the most preferred brands in the industry. In these cases even a retailer with more than 25% of the market may be relatively weak when bargaining with a dominant producer controlling 50% or more of the sales within a given category.

The division of bargaining power between producers and retailers affects wholesale terms, but exactly how depends on how a typical wholesale contract looks like. The sim-

---

6 For instance, if you want to buy a painting you may wish to explore different paintings before deciding on which you like best.

plest wholesale contract we can imagine is a simple uniform wholesale price. A more sophisticated contract is a two-part tariff. A two-part tariff consists of a fixed fee and a marginal (per unit) wholesale price. The fixed fee can be positive, and is then called a franchise fee, but can also be negative, a slotting fee (allowance). A two-part tariff has the advantage that efficiency can be separated from the division of rents between the contracting parties. An optimal wholesale price can be chosen to give the retailer the right incentive to maximize the joint surplus and then the maximal surplus can be divided between the parties by an appropriate fixed fee.

Two-part tariffs alleviate problems with double margins. Double margins tend to render retail prices that are too high as seen from the integrated vertical structure's perspective. Since these contracts are so simple and have superior efficiency properties it seems natural to assume that they are frequently used in practice. If so, the upstream and downstream firms have no conflicting interest when it comes to the optimal wholesale and retail prices. Both parties would like to see them set at the level that maximizes joint surplus. Bargaining power then is manifested in the size (and sign) of the fixed fees in the wholesale contract. If producers have strong bargaining power we would expect to see large franchise fees, and less so if the retailers have the bargaining power. Strong retail bargaining power can also be an explanation for negative fixed fees, slotting fees. According to this view slotting allowances are not only symptoms of the existence of retail bargaining power, but at the same time also instruments to exploit this power.

In general fixed fees can be used to transfer rent between the contracting parties, but can also be payments covering costs for services. Whether fixed fees are used in practice is an empirical question. A recent report by SIFO (2005) revealed that fixed fees flourish in the Norwegian grocery market. The fees carry many different names, some times indicating what kind of costs and expenses they are meant to cover. However, from an empirical point of view it is extremely difficult to tell whether the fees exactly cover costs for the services rendered by a party or if the fees also include an element of rent shifting. Clearly, it is possible to transfer rent with the different fees. If this is correct, then a shift in bargaining power from producers to retailers is unlikely to affect wholesale prices at all. Efficient bargaining would involve that wholesale prices are set such that the optimal retail prices as seen from the vertically integrated structure will be set. Fixed fees can then distribute rents based on bargaining power.

If shelf space is scarce, slotting allowances may be an efficient instrument to allocate shelf space to products most valued by the consumers.[7] The likelihood of observing slotting allowances is highest in cases where there are upstream margins and retailer bargaining power is high. Upstream margins can arise for several reasons and here will explore two candidates; i) upstream margins are due to decreasing returns to scale, and ii) upstream margins controls for strong intrabrand competition.

---

7 In the literature there are several efficiency arguments for the use of slotting fees. For instance, Lariviere and Padmanabhan (1997) show that slotting fees may be use as signalling devices when producers are better informed than the retailers.

Gabrielsen (2005) shows that if production costs exhibits decreasing returns to scale, wholesale prices equal to producer marginal costs will leave rent at the producers. Retailers can effectively appropriate this rent if they have the ability to dictate their terms to the producers, or if the producers must compete for exclusive retail supply rights. The rent extraction effect is illustrated in the figure below. Here we assume that a producer that contracts with a retailer who sets the price p to consumers. The wholesale contract is a two-part tariff with a constant marginal wholesale price w and a fixed fee A. A positive fixed fee is a franchise fee, and when the fixed fee is negative we will denote this as a slotting allowance (SA). Further, suppose producer marginal cost is increasing and also that the retailer has all the bargaining power and can decide on the level of the fixed fee. In this situation it is easy to demonstrate that in equilibrium the joint maximizing wholesale price is set, inducing the retailer to charge the monopoly price to consumers. Because of decreasing returns to scale the producer will earn some rent, illustrated by the shaded area in the figure below. However, if the retailer has all the bargaining power it can capture this rent by charging a slotting allowance (SA) of exactly this size.

Figure 6.5: Decreasing returns to scale and retailer bargaining power.
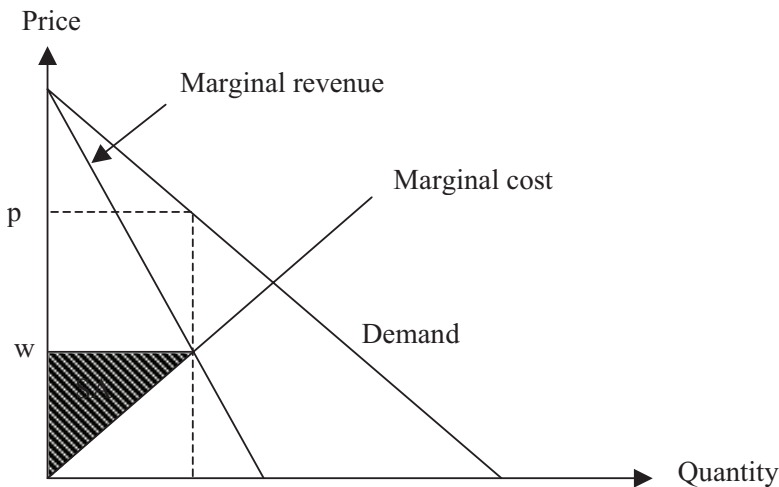
When a slotting allowance of this size is charged, the producer earns exactly zero, and the retailer is able to capture the entire monopoly profit, partly by a margin on the wholesale price and partly by the slotting fee. Here a prohibition on the use of slotting fees could be circumvented by a contract where the producer agreed to supply the same quantity as in the figure above but at a lower wholesale price. To achieve full rent extraction the retailer should demand the wholesale price that exactly covers the average production costs. If so, retail prices are unaffected by the use of slotting allowances.

A second motive for the use of slotting fees is the following. Think of a situation where a producer with some degree of market power sells his product through several retailers that are close substitutes in the sense that they cover more or less the same market. When selling to all retailers, intrabrand competition at the downstream level will be hard. To

avoid competing with itself the producer will then wish to charge high wholesale prices to give the retailers an incentive to charge high prices to the consumers, hence margins are left at the upstream level. Irrespective of the division of the bargaining power, the retailers will be equally interested in the high prices generated by the high wholesale prices as this maximizes the aggregate channel surplus. If the retailers have a strong bargaining position they can appropriate the upstream rent by charging slotting allowances from the producer.

6.4.2.3 Exclusivity

In the Norwegian grocery market we have seen that several retail chains hold a limited product range in selected categories. This means that some products are excluded. In general, the use of exclusivity contracts may induce full or partial foreclosure of competitors. If a product is excluded in all chains in all retail groups we have total foreclosure, and if a product is excluded from some chains but not all we have partial foreclosure.

The traditional motives for exclusivity discussed in the literature are that exclusivity may protect producers from downstream opportunism but also that exclusivity may be an instrument to dampen or eliminate competition.[8] The school of thought known as the Chicago school heavily criticizes the anticompetitive argument. The Chicago school argument is that a producer could not profitably insist on an exclusive arrangement with a buyer. In competition with a producer of an equally profitable alternative product the compensation to the buyer would have to be the entire profit that could be generated from the sale of its product. As long as each product has a positive contribution to the aggregate profit in the industry, the implication is that the three parties would always find a set of contracts and a division of the profit than involves an outcome where socially inefficient exclusion never occurs.

However, from a dominant buyer's perspective the incentives for foreclosure may be very different (Gabrielsen and Sørgard, 1999). If the buyer can commit to exclusivity he may sometimes want to do that. If a dominant buyer is able to commit to exclusivity this will induce tough competition for access to the retail asset. When wholesale contracts are two-part tariffs, the competition reduces fixed fees but not wholesale prices (that are set to maximize the joint surplus). Hence, a commitment to exclusivity is likely to increase profits for the retailers while reducing variety. With heterogeneous preferences this may be to the detriment for consumers.

Several of the Norwegian retail groups have discount chains were the basic business idea is that the number of product lines in each category is limited. In addition to its effects on wholesale terms, the commitment to fewer product lines may have direct cost effects as well. Intuitively, if lower product variety saves variable costs, all else equal low product variety should go hand in hand with lower retail prices. If, on the other hand, low

---

8 For an example, see for instance Segal and Whinston (2000).

product variety saves on fixed retail costs we should expect the effect on retail prices to be minimal. In the latter case low product variety should instead increase retail profit.

Exactly how fewer product lines affect retail costs is of course an empirical question, but in theory it is possible to list some plausible alternatives. First, when carrying many product lines retailers may fail to exploit economies of scale. The economies of scale may stem from cost reductions from handling fewer product lines within each store, but also from rebates obtained from suppliers because of larger quantities (drop-size effects). Second, having many product lines may decrease the turnover within each line that in turn may increase losses related to products with limited durability. Third, as noted above a commitment to few product lines within each category may induce tough competition between producers for access to the retail shelves, and this may be beneficial for the retailers. It seems clear that there might be fixed costs related to each product line within a category, and better wholesale terms may involve lower fixed payments as well as lower marginal wholesale prices. Hence if the major costs savings involve savings on fixed payments, the effects on retail prices are expected to be minor but the effect on retail profits may be significant. However, it is also perceivable that exclusivity may yield efficiency effects by reducing marginal production and distribution costs. If that were the case we would expect consumer prices to fall in response to this.

With retail competition the direct effect from reduced product variety on wholesale terms, retail costs and retail prices will be complemented by a strategic effect. Under price competition rival retailers will reduce their prices in response to the price decrease from a retailer that faces lower marginal costs due to reduced variety. This strategic response is stronger the more differentiated products are and will evaporate completely when products becomes perfect substitutes. There is however another effect as well, and this is easier to illustrate if we assume that a reduction in product variety has no effect on wholesale costs. Suppose two retail stores that each has two differentiated products within a category. If one product is removed from the shelves in one of the stores intrabrand competition for this product will disappear. This means that the store that still has this product in its store will be able to increase its price. Under normal assumptions of retail competition this may induce a price increase of the competing brand (that is carried by both retailers) as well. Hence, exclusivity may increase retail profit by reducing retail costs and increasing prices and at the same time be detrimental to consumers because of increased prices and reduced variety.

Clearly, under strong intra- and interbrand competition retailers and strong producers have incentives to foreclose weaker rivals. To illustrate these incentives, consider a situation where two retailers are located very close to one another and that each store carries two products that are close substitutes. If so, intra- and interbrand competition would result in both wholesale and retail prices close to marginal costs and the profit in the industry would be extremely low. Shaffer (2005) shows that in this situation the producer of the most profitable product and the two retailers jointly have an incentive to foreclose the weaker product from the market by signing exclusive contracts. Furthermore, under exclusion the dominant firm will be able to realize the monopoly profit that in turn is divided between the

contracting parties by means of fixed payments. In this case there is relatively small loss for consumers due to reduced variety (because products are close substitutes) but a huge loss due to the change from competitive pricing to the monopoly price.

A final issue is that retailers can end up with exclusive agreements with different producers in any given category. Clearly, under strong retail competition retailers have an incentive to choose different products. The reason is that this will differentiate stores and dampen competition. This use of exclusivity agreements will inherently lead to lower product variety within each store, but since the alternative can be obtained in other stores, product variety is not reduced per se. The major loss for the consumers in this case may be that retail stores do not compete that much because consumers cannot directly compare the products in different stores. This type of differentiation can come about by one chain choosing the number one national brand in a category and where a second chain chooses the number two national brand. Alternatively, the second chain may choose to introduce a private label instead of the number one national brand. However, such a strategy calls for some type of coordination between different retailers. The reason is that one type of strategy is probably more profitable than the other, and both retailers will want to choose the most profitable strategy.

Hence, the lessons from economic theory is that there is relatively little support for the claim that retailers in general will act as agents for consumers, at least not along every dimension that is important for the consumers. There is some evidence that to some extent retailers will internalize consumers' preferences with respect to product variety, but when it comes to prices this is certainly not the case. In a general set-up retailers and producers do have joint interests when deciding on wholesale and retail prices, but these interests will more often be in conflict with the interests of the consumers.

6.4.2.4 Private labels

As demonstrated above private label penetration is still modest in the Norwegian grocery market. However, the private label share is expected to grow over the next few years. For the retailers private label introduction offers several advantages. First, it may be an effective instrument to elicit price concessions from producers with few competitors. By introducing, or threaten to introduce, a private label better wholesale terms for national brands can be obtained. Second, with heterogeneous preferences and differences in the willingness to pay among the consumers, private label may offer the retailer an instrument to price discriminate between different consumer groups. Finally, since private labels by definition are sold exclusively in the store of one retail group, the retailers become more differentiated which in turn may dampen competition.

Several authors formalize the price concessions argument.[9] The basic idea is that private label introduction induces a price cut by the national brand. Depending on the specific wholesale contract, retail prices are either reduced (if no fixed fees are used) or there is

---

9 Scott-Morton and Zettelmeyer (2004) , Mills (1995) and Narasimham and Wilcox (1998).

no effect at all (if the concession implies a reduction in a fixed fee). With different consumer groups there may be several effects. When a retailer introduces a private label aimed for the low price segment, the national brand has two possible strategies. The national brand can either decide to compete by offering the retailer such beneficial terms that the retailer either decides not to introduce the label or decides not to promote its sale. Alternatively, the national brand may give up the low demand segment and instead concentrate on the loyal segment. In the latter case, it is feasible that private label introduction may increase the retail price of the national brand and in sum induces a welfare loss.[10] Finally, the chain differentiation effect may reduce retail competition and increased retail prices.

Hence it is not at all clear from a theoretical point of view that more private labels always will benefit consumers by lowering retail prices. By a revealed preference argument it must be beneficial for the retailers, as they could have chosen not to introduce private labels. The ambiguous effect on retail prices is well documented in the empirical literature. There are many studies that find that the price response from private label introduction in many categories may be a price increase rather than a decrease in prices.[11]

As illustrated above the private label share in Norway is relatively small when compared to other markets. Also, it is noteworthy that the private label share seems to be lower in markets with high retail concentration. One explanation for this is that if private labels may serve as a bargaining tool versus the producers, and that retailers with high market shares are in less need for such a tool compared to smaller retailers.

6.4.2.5 Buyer power and the incentives to innovate

Retail concentration, vertical integration, slotting fees, exclusivity contracts and private label introduction are all practices that potentially are aimed at both generating and exploiting retailer buyer power. In this section we will briefly discuss the implications from this for the producers' incentives to innovate in the grocery market.

In the economic literature the incentives to innovate is largely driven by producer profits. Two main effects are discussed; the efficiency effect and the replacement effect. The latter comes into play when an unthreatened monopolist considers whether to innovate or not. This situation provides the monopolist rather weak incentives to innovate because the innovation will replace the monopolist's old product and therefore only may create a small increase in profit. If on the other hand, the monopolist is threatened by entry of an innovator of a close substitute, the incumbent firm will – absent innovation – lose its monopoly profit. Hence, the monopolist will in this case have a large incentive to be innovative. The main conclusion from this literature is that competition fosters innovation more than protected monopoly.

---

10 See Gabrielsen and Sørgard (2006).
11 See for instance Parker and Kim 1995 and Gabrielsen et al., 2001.

A potential problem with this literature is that it does not take into account the vertical structure of the industry. From our discussion above we know that may be little correlation between the market structure on the producer level and the profits earned by the producers. The reason is that even with monopoly at the upstream level, the monopoly rent may be appropriated by a retail sector with high bargaining power. If the Norwegian retailers do indeed have strong bargaining power – which there is reason to believe – this may potentially be an important explanation for lack of new product launches on the Norwegian grocery market.

However, the picture is not clear-cut. There also exist arguments in economic literature claiming that for instance slotting allowances may be beneficial for product innovation[12]. The argument is that innovators may be able to signal the market potential of new products by paying slotting fees up front to the retailers. If there is asymmetric information with respect to the product's market potential, a slotting fee may sometimes be an efficient instrument to signal profitability.

In sum however, it seems obvious that if the retail sector is able to appropriate a too large share of the industry profit this may have important negative effects of the producers' incentives to innovate.

The public policy implications from this are far from easy. If the problem is that there is too strong buyer power, the appropriate policy would be to reduce this power. This would shift some of the profits in the industry from the retail level to the producer level. However as we have seen, retailer bargaining power potentially stems from a series of sources, both structural and strategic, and it is not an easy task to pinpoint exactly what should be the optimal policy to reduce retailer bargaining power.

6.4.2.6 Retail competition

The last point we will discuss with respect to the competitiveness of the Norwegian grocery market is facilitating practices with the potential to sustain collusive behaviour. As noted above the structure on the Norwegian retail market is one where four large retail groups controls almost 100% of the market. This basic structure was more or less in place by the mid 90s. Even if not perfectly symmetric, the market shares of the groups have converged in the sense that the smaller group (Rema) have been taking market shares from the second and third groups. Also, all groups set their prices at the central level that naturally leaves no potential for local competition. The retail groups therefore compete at the national level.

Low number of players, market share convergence and national as opposed to local competition are all features that potentially facilitate collusion. It is a well-established fact in economic theory that the fewer the number of players, the easier it is to sustain a collusive agreement. On the other hand asymmetry in market shares will tend to destabilize collusive

---

12 Sullivan (1997) and Rao and Mahi (2003) look at the innovation incentives and slotting allowances.

agreements. The reason is that the firm with the smallest market share will have stronger incentives to deviate than larger firms. The reason is of course that the smallest firm have less to gain by colluding and more to win by deviating than a bigger firm. It will therefore be beneficial for collusion to bring the smallest firm in the industry closer to the bigger ones in terms of market shares, and this is exactly what has happened in Norway. The smallest group, Rema, has gained market shares over the last years.

Central price fixing is also a practice that potentially facilitates collusion. This practice works in a very similar way as reducing the number of firms in the industry. What matters for the incentives to collude is how many players are able to change prices and therefore potentially deviate from a collusive agreement. Central price fixing removes the incentives for local retailers to deviate by cutting prices to gain market shares at the local level.

6.5 Concluding remarks

The Norwegian grocery market has changed a lot over the last decades. From a situation with dominant producers and wholesalers were able to dictate the terms to a fragmented retail sector, retail mergers and vertical integration have turned the market 'upside down'. Today four large retail groups jointly control almost 100% of the retail and wholesale markets. Initially the Norwegian consumers undoubtedly benefited greatly from the restructuring of the Norwegian grocery industry. The consolidation on the retail level and vertical integration with the former independent wholesale sector has without doubt created considerable efficiency gains that early in the restructuring phase gave Norwegian consumers and the Norwegian society considerable gains.

An intriguing question is if the pendulum swung too far? Even if the efficiency gains that could be realized in the short run in the early 80s were obvious, the question is whether the Norwegian market today is stuck with a structure that dampens price competition and blocks innovative activities. Relatively high prices on Norwegian grocery products and a low level of product selection are indications that this may be the case. As noted above, the restructuring of the industry was largely in place by the late 80s, and since then little has happened. The most surprising fact about the last 10 years in the industry is the lack of penetration from international competitors from abroad. Neither has any international retail group gained significant terrain on the Norwegian market, nor has there been any significant improvement in the market shares and product range for the large international conglomerates producing internationally branded products.

In this chapter I have discussed some plausible reasons for why Norwegian grocery products are expensive and why the product range is so limited compared to many other markets. In doing so, I have focused on two main explanations. The first explanation is a structural explanation. The essence of this explanation is that prices are high because of high costs and variety is low because of homogeneous preferences among Norwegian consumers. This can be a plausible explanation for some products (e.g. agricultural products), but cannot explain the price differences in other categories. Hence, while the

structural explanation seems to have some merit for some product groups, it does not seem to offer the complete story.

The strategic explanation relies on the argument that retail competition is low due to high concentration. Moreover, high concentration gives retailers significant bargaining power that is complemented by the strategic use of exclusivity clauses and actual or potential private label introduction. Retail bargaining power and the strategic use of exclusivity hurts consumers because product variety is reduced without offering consumers a full compensation by a significant lowering of retail prices. Furthermore, strong retail bargaining power seems to be a major obstacle for product innovations among the suppliers. Last, there are several characteristics and practices in the market that facilitates collusion between the major retail chains. This is a potential threat to retail competition that should be taken serious by antitrust authorities.

References

ACNielsen (2005): The Power of Private Labels.
http://www2.acnielsen.com/reports/documents/2005_privatelabel.pdf

Dobson et al. (2003): The Patterns and Implications of Increasing Concentration in European Food Retailing, Journal of Agricultural Economics, 54 (1), 111-125.

Gabrielsen, T.S. (2005): Slotting allowances, bargaining power and rent extraction, preliminary manuscript.

Gabrielsen, T. S. and L. Sørgard (1999): Discount Chains and Brand Policy, Scandinavian Journal of Economics, 101: 127-42.

Gabrielsen, T.S. and L, Sørgard (2006): Private Labels, Price Rivalry and Public Policy, European Economic Review, forthcoming.

Gabrielsen, T. S., F. Steen and L. Sørgard (2001): Private Label Entry as a Competitive Force? An analysis of price responses in Norwegian food sector, mimeo.

Lariviere, M.A. and V. Padmanabhan (1997): Slotting Allowances and New Product Introductions, Marketing Science, 16: 112-128.

Mills, D. E. (1995): Why Retailers Sell Private Labels, Journal of Economics and Management Strategy, 4, 509-528.

Narasimhan, C. and R. T. Wilcox (1998): Private Labels and the Channel Relationship: A Cross-Category Analysis, Journal of Business, 71, 573-600.

NCA (2005): Nordic competition authorities: Nordic Food Competition – a taste for competition. Report from the Nordic competition authorities, No. 1 2005.

Parker, P. M. and N. Kim (1995): National Brands versus Private-Labels: An empirical Study of Competition, Advertising, and Collusion", mimeo.

Rao, A. and H. Mahi (2003), The Price of Launching a New Product: Empirical Evidence on Factors Affecting the Relative Magnitude of Slotting Allowances, Marketing Science, 22(2), 246-68.

Rao, A. and Sudhir K. (2004) Are Slotting Allowances Efficiency-Enhancing or Anti-Competitive?

Scott Morton, F. and F. Zettelmeyer (2004): The Strategic Positioning of Store Brands in Retailer-Manufacturer Negotiations, Review of Industrial Organization, 24: 161-194.

Segal, I. R. and M. D. Whinston, 2000, Naked Exclusion: Comment, American Economic Review, 90: 296-309

SIFO (2005): Dulsrud A. and J. R. Beckstrøm: Å sette pris på hylleplassen, Statens institutt for forbruksforskning, fagrapport nr. 2-2005.

Shaffer, G. (1991), Slotting Allowances and Retail Price Maintenance: A Comparison of Facilitating Practices, RAND Journal of Economics, 22(1), 120-35.

Shaffer, G. (2005): Slotting Allowances and Optimal Product Variety, B.E. Journal in Economic Analysis & Policy, 5, no. 1, article 3.

Sullivan, M. W. (1997), Slotting Allowances and the Market for New Products, Journal of Law and Economics, 40(2), 461-93.

Frode Steen and Lars Sørgard

Inefficiencies of cartels: The Norwegian cement cartel 1923 to 1968

7.1 Introduction

The Norwegian cement cartel lasted for 45 years before the members merged into a monopoly in 1968. Here we will show that not only was the cartel very inefficient, but inefficient enough that the merger to monopoly in 1968 actually was welfare enhancing. The cartel had chosen a setup that gave them a strong incentive to undertake investments in excess capacity that was used for production and non-profitable sales abroad. In particular the cartel's sharing rule in the domestic cartelized market will be shown to be crucial for this investment behaviour. In addition to detailed industry and trade data, knowledge from annual reports and historical studies about the functioning of this cartel, this chapter will build heavily on two former studies by Steen and Sørgard (1999) and Röller and Steen (2006).[1]

The Norwegian cement cartel and the way in which it functioned illustrate several things. First that cartel functioning is complicated and is often undermined by incentives to defect on some or several aspects, leading to often unanticipated competition effects. Second, the cement case shows that the presence of competition regulation and competition authorities can be crucial to ensure consumer welfare. Even though we will show that the merger that took place in 1968 was welfare enhancing, the merger was still a merger from cartel to monopoly. There was another alternative to an outright merger, namely competition (a la Cournot). While the merger yielded positive welfare gains after 1968, Röller and Steen (2006) show that competition would have resulted in considerably higher welfare gains. In this sense, the merger which took place in 1968 was only second best. This underlines the importance of well functioning competition authorities.

There are relatively few empirical studies on the workings of a cartel. The most obvious reason for this is that cartels are often illegal and therefore data are difficult to obtain. Even though antitrust agencies sometimes achieve detailed information on cartels, strict confidentiality rules often keep data from academic research. Some notable exceptions is studies based on historical data as e.g., the seminal work by Porter (1983), which investigates price wars in a railroad cartel operating in U.S. in the late 19th century and Genesove and Mullin (1998) that use data from 1892 to 1914 of the American sugar industry, where the American Sugar Refining Company controlled (through acquisition) 95% of the US sugar market by 1895. More recent examples include the so-called Lysine

---

1 In this chapter we will present and discuss the results from these studies in great detail, but all the technicalities and the modelling framework will not be discussed.

cartel, an industry producing feed additive used to ensure the proper growth of livestock, has provided more information on the workings of cartels international settings (see Griffin, 2001). The cartel was in place for the period 1992 to 1995 and was fined in the order of $100 million plus personal fees and prison sentences for some of the employees. To expose the cartel the FBI used covert cameras to tape cartel meetings, providing us with detailed information on the workings of the Lysine cartel.

As opposed to most empirical studies on cartels that focus on markets where a known cartel exists and investigate the cartel's efforts to prevent individual members from cheating on the agreement, we will later show that the Norwegian cement cartel was tightly organized and these issues where not at stake.[2] Coordination of the Norwegian cartel's activities was achieved through the common sales office A/S Portland cementkontor and various other cross industry information sharing and coordination institutions as Norwegian Cementforening.

By contrast, the Norwegian cartel had a market sharing agreement that towards the end of the cartel period led to excessive over production. In particular they divided the domestic market according to capacity shares resulting in incentives to over invest in capacity in order to gain a larger share of the domestic market.[3] We will therefore through this Norwegian case show that the way in which one determines how the monopoly rents are divided up amongst the cartel members will influence cartel efficiency and welfare.[4]

The Norwegian cement cartel setup is also a prominent example of semicollusion, i.e., coordination along some dimensions and competition along others. In our case the cartel colluded on prices and quantities but competed on market shares through investment in capacity used for non-profitable exports. In Steen and Sørgard (1999), this is the main focus where one tailor makes a three stage game of the cartel behaviour, performs a reduced form empirical test of the model predictions and shows that semicollusion took place. There are a number of related studies that have studied the semicollusion set-up that is present in the Norwegian cement industry. Davidson and Deneckere (1990) - look at a game where firms tacitly collude on price, but compete in capacity. Building on

---

2 See for instance Levenstein (1997) for a historical study of the stability of cartels looking at the pre World War I Bromine industry. Genesove an Mullins (2001) discuss how rules and frequent meetings prevented unnecessary retaliations among the sugar cartel members and how they maintained a collusive price level. See also McCutcheon (1997) for a discussion on the importance of information sharing in cartels. Suslow (1988) provides a comprehensive list of different cartels that were active in the inter-war period.

3 As long as prices are higher than short run marginal cost (which is the case, see below) it is optimal to use all the installed capacity such that capacity equals production. Storage is not an option either, due to limited storage capacity for cement. In the annual report from NORCEM in 1968 we could read (p.9): "Because the capacity for storing finished cement is so small, production has to conform quite closely to sales." The sharing rule can thus also be considered a production sharing rule.

4 Osborne (1976) in a seminal contribution refers to these two problems as "the sharing problem" and the "locate the contract surface" problem. See also Eswaran (1996) for a study on cartel unity in the face of business cycle effects.

work by Benoit and Krishna (1987) they show that equilibriums exist where firms will carry excess capacity in order to support collusive outcomes (see also Osborne and Pitchik (1987)). They do not explain as to why firms can not collude in capacity, but rather cite a number of examples of where firms are in such a situation of "semi-collusion" (or as it is also called "mixed games", see Brander and Harris (1983)). They also state that "it is well-known that even in cases of overt collusion (such as the German Cement cartel in the 1920s and 1930s, or the Texas oil industry in the 1930s) firms find it exceedingly difficult to collude in capacities - emphasis added" (see Davidson and Deneckere p.523). Scherer (1980, pp. 370-71) writes that "In Germany during the 1920s and 1930s, shares were allocated on the basis of production capacity. Cartel members therefore raced to increase their sales quotas by building more capacity".

Röller and Steen (2006) have a somewhat different focus in their analysis of the Norwegian cement cartel. They assume an equilibrium approach where they are able to provide also a complete welfare analysis of the Norwegian cement cartel. They consider the decisions of the members of the cartel in a simple two-stage analysis where each cartel member first decides on how much capacity to install, taking the sharing rule into account, then collectively decides on how much of total production to allocate to the domestic market. The sharing rule creates an incentive to "overproduce" and export (even when marginal costs are above the world market price), since each member of the cartel increases their share of the domestic rent. This overproduction reduces the cartel's effectiveness in the sense of lowering profits to the cartel. They also show that the effectiveness crucially depends on the world market price. Since the world market price represents the opportunity costs of not exporting, the common sales office maximizes the cartel's profits by equating marginal domestic revenue with the world market price. As a result, a lower world market price implies that the cartel allocates more production to the domestic market, which reduces the cartel's domestic rents (to the benefit of domestic consumers). One of the main contributions is the way in which marginal costs is identified since this allows a complete welfare analysis. They find that the ineffectiveness of the sharing rule was increasing over time, implying that consumers benefited more (relative to monopoly), while producers were losing both domestically as well as in the export market. In this sense it was consumers, not firms, that benefited from the sharing rule. In particular they find that the ineffectiveness of the cartel was becoming so large that domestic welfare of a merger to monopoly was in fact positive at around 1968, which is exactly when the merger actually took place.

The chapter is organized as follows. We begin by presenting the Norwegian cement industry and the cement cartel. We then discuss sharing rules more generally before we discuss the results from Steen and Sørgard (1999) and Röller and Steen (2006) in more detail. Finally policy implications and some conclusions are provided.

7.2 The Norwegian Cement Industry[5]

The first Norwegian cement plant, A/S Christiania Portland Cementfabrikk (CPC) was established in 1892. At the end of World War I, three new plants were established in Norway: A/S Dalen Portland-Cementfabrikk (DPC) in 1916, CE-NO Portland Cement A/S in 1917 and a firm in Northern Norway, Nordland Portland Cementfabrikk A/S (NPC) in 1918. The capacity expansion, combined with the recession in Norway from 1920, led in the early 20s to a domestic capacity amounting to almost twice the domestic demand [see Gartmann (1990; 114)]. The mismatch between capacity and demand triggered a price war and later the establishment of A/S Norsk Portland Cementkontor in 1923, a joint sales office for the three firms in Southern Norway (CPC, DPC and CE-NO). Five years later, NPC became a member of the common sales office as well. CE-NO was acquired by DPC in 1927 (see Gartmann, 1990), which increased DPCs market share to the level of CPC. The Norwegian cement industry has been cartelized through the common sales office since 1923. The reasons for the creation of the sales office were clearly to remove competition: "both companies (in the south) had to sacrifice something on the alter of collaboration. The sales office primary task was to organize the sale in a better way, to prevent cross-transportation and unprofitable competition." (Gartmann, 1990 p. 46). The reorganization implied that from 1928 all cement was sold trough one agency, and no cement was sold directly from the factories. In particular, the common sales office determined the total domestic sale and sets domestic quotas according to each firm's total capacity (domestic production plus exports, see also footnote 3)

After establishing the common sales office, more institutional ties were developed. In 1927, "Norsk Cementforening (NC)" was founded. NC was an institution (funded by the industry) that coordinated standards, lobbied government committees, and took part in the education of engineers and cement workers. Gartmann (1990 p.47) claims that "the sales office and NC, were forerunners to the full merger in 1968. Norcem came to a finished table arrangement with coordinated sales and information already established over a long period". In the beginning only the two big producers in the south joined NC, but later also the northern firm entered.[6]

The production technology in this industry was gradually improving over the period we are looking at. A cement kiln is built as a tube-like oven, and the kiln's production capacity is primarily determined by the length of the "tube". In the beginning a kiln would be

---

5 The sections on market description and the next section on market sharing rules are primarily based on Röller and Steen (2006).

6 The cement producers started several other institutions as they moved into downstream activities. For instance, the production of cement products as tubes and panels had their own body, called "The Cement producers price co-ordination body" that was founded in 1928, whose task was to "collaborate on prices and rebates to prevent non serious producers to enter the market " (Gartmann, 1990, p. 62). In the same fashion NC controlled that the local concrete mixers, which were small firms often organized in local oligoplolies produced according to quality standards. In 1964 these firms founded a collaboration body: "The local concrete-mixers institute". The institute however had its secretariat in the same offices as NC, suggesting that NC played a rather influential role in co-ordination of the cement industry during the cartel period.

in the order of 20-30 meters long, whereas the newest kilns installed after 1965 was several hundred meters long. In 1920 an efficient rotary kiln produced 50 000 tonnes annually. After the World War II the corresponding amount was 150 000 tonnes, whereas in 1966 and 1967 the largest kilns at Dalen and Slemmestad produced 500 000 tonnes each. The technology also changed from "wet process" to "dry process" over this period, where the newer "dry process" was more efficient and required less energy. The enormous new kilns that are in use today have a capacity of more than 1 million tonnes per kiln, but none of these where installed in our sample period (1927-1982). Gradually expansion of kiln size together with the fact that older kilns only gradually where phased out as they got non profitable made therefore the technology improvement relatively smooth in Norway.

As we mentioned above, the common sales office and the sharing rule will create an incentive to export. The three firms' exports fell gradually during the 30s, from more than 50% of total domestic production to approximately 10% of domestic production at the beginning of World War II. In the mid 50s exports grew rapidly, and in the late 60s over 40% of the domestic production was exported. In 1968 the three firms merged and established the firm Norcem, and during the 70s Norcem closed down the excess capacity. The development in the export share is shown in Figure 7.1.

Figure 7.1:
The development in the Norwegian Export Share the period 1927 to 1987
(Source Norwegian historical industry and trade statistics)
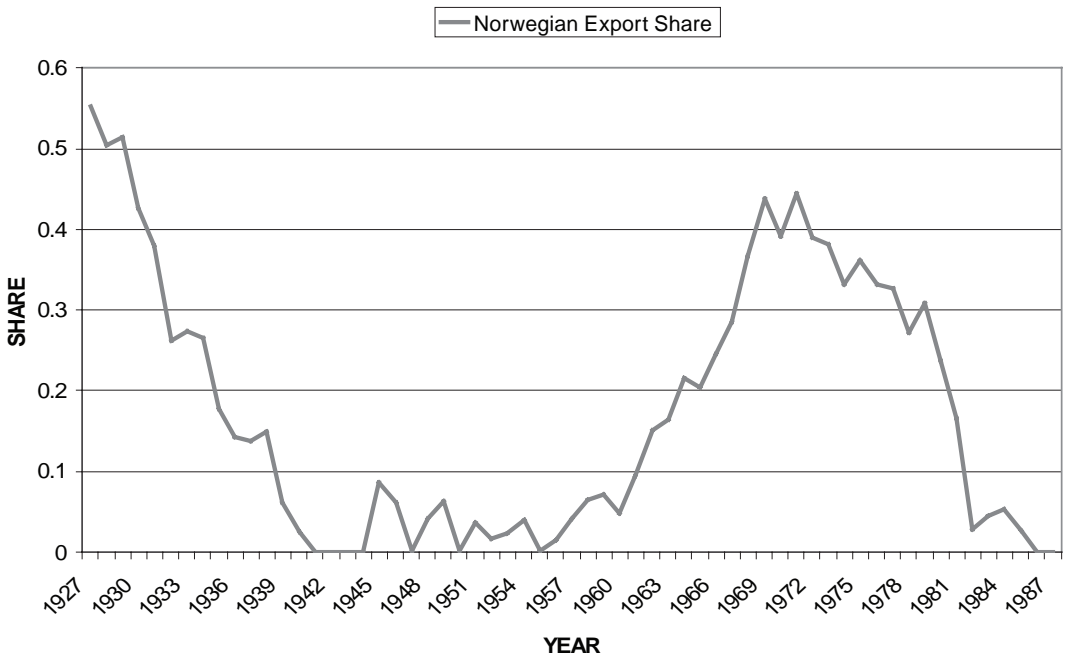
Norwegian exports predominantly went to non-European markets, such as South- and North-America, as well as Africa. The reason why little cement was exported to other European countries has been explained through a retaliation game. Essentially, competition is a multi-market game where credible threats to enter each others markets prevent firms form entering other countries (see for example Röller and Friederiszick (2002)). Aiginger and Pfaffermayr (1997) undertake a study of the competition in the cement and paper industries. On page 252 they state that: "the cement industry is faced with limited geographical competition". Later they say that: "we are confident that the EU is the relevant geographic market for the paper industry, for cement this is clearly not the case" (p.263). As a result, Norwegian exports went to non-European markets in order to prevent possible retaliation from neighbouring European countries.

What about imports? A further implication of the European stale-mate was that there were little imports (at least from other European countries) into Norway. In addition, there were also few imports from other parts of the world. The reason for this was that the domestic market was protected both by high tolls and by relatively high transport costs. For instance, in 1959 the toll was 8 NOK per ton. This was approximately 9% of the factory price. CPC considered this a significant toll barrier. (CPC annual report 1959 p. 4). It is therefore no surprise that imports were low.[7]

In addition, there was a relative high transport cost to Norway, primarily due to the trade pattern at this time period. Norwegian boats had excess capacity for bulk transport leaving Norway, depressing transport prices out of Norway (some minimum ballast is in fact needed for oversea journeys). By contrast, for coming back to Norway there was plenty of cargo from ports in the US, Latin America. Thus, there has been little import of cement into Norway and most of the exports have been to non-European countries.

The common sales office existed since 1923. The industry had a sub-optimal capacity level in the early cartel phase after the entry of the three new firms. Due to the large investments in the early post World War I years the firms had a large capacity and did already export considerable quantities of cement when they established the sales office. In the long run, however, the cartel could close down (not renew) capacity. This is also what we saw happened towards World War II. Hence, it is first when we approach the late cartel phase after World War II that the incentives for over production came into play (see Steen and Sørgard 1999, for more thorough discussion on this). However, the post war period just after 1945 and in to the 1950's, firms had to ask the government for permission to undertake capacity investments. The reason for this was that imports of technology to undertake capacity expansions were rationed, due to shortages after World

---

[7] To the extent that we saw imports it was small and typically seasonal. In some periods more cement then what was produced was needed. "To cover the max-consumption during the fall season there was an import of 32 000 tonnes of cement … and as usual in addition to this some minor quantities of special cement that is not produced in Norway was imported" (CPC-annual report 1959 p.3) In 1959 total production was 1103 000 tonnes, suggesting an import less then 3%. The export in this year was 80 000 tonnes, so Norway was a net exporter also in 1959. Note that the overproduction in 1959 was small compared to what we saw develop during the 60s.

War II. The firms had to ask for permission to undertake capacity investments, and this regulation lasted until the mid 50s.[8] The 1953 annual report for CPC, the largest domestic cement producer, states (p.14): *"...the negotiations about increased production capacity have still not succeeded..."*. However, the 1956 annual report states: *"As is well known, the Norwegian producers have maintained an agreement on market sharing, but two of the producers have in the post-war period not been able to exploit their quotas. These circumstances have now changed..."* We would thus expect the "overproduction" due to semicollusion and competition on market shares to emerge after the rationing disappeared in the mid 1950's.

Figure 7.2 shows domestic production and domestic consumption of cement for the period 1927 to 1987. If we focus on the period from 1955 to 1968, production increased by 150%, whereas the Norwegian consumption only increased by 50%. By 1968 this lead to an export of some 828 000 tonnes, almost as much as Norway's total production in 1955. There is thus rather striking evidence that overproduction took place in the latter period of the cartel.

Figure 7.2:
Norwegian Domestic Production and Consumption of Cement the period 1927 to 1987
(Source Norwegian historical industry and trade statistics)


---
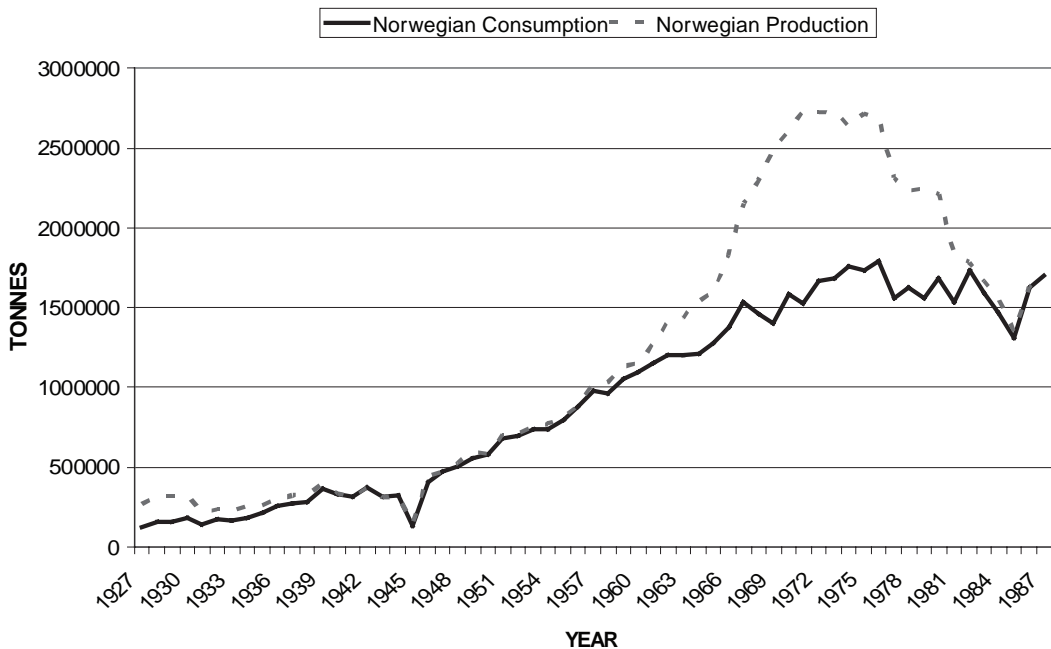
8 This suggests that even if the semicollusion effect was triggered in the post-war period, the firms' export opportunities were constrained.

The analyses of the Norwegian cement cartel focus on the sharing rule and its incentives to explain the above phenomenon. What about alternative explanations? In principle, there may be two other reasons for the observed capacity increase (see also Steen and Sørgard (1999)): the Norwegian producers built up such high capacity levels due to unrealistically high anticipation of increased future consumption, or to deter entry. Let us take these alternative explanations in turn.

Regarding the unanticipated consumption slowdown, the CPC undertook a very comprehensive and detailed ten year forecast of Norwegian cement consumption in 1957 (annual CPC-report 1958 pp.14-28), including a number of different economic and demographic trends (such as fertility, household size, average number of rooms per house, building and construction trends, GNP, population growth). Comparing the 1957 forecast with actual realized consumption, one finds that the forecast was rather accurate with a margin of error below 5% (except for 1959). The prediction and the actual production is shown in Figure 7.3.

Figure 7.3:
Christiania Portland Cementfabrikk's (CPC) 1957 ten-year prediction of Norwegian domestic consumption and actual consumption for the period 1958 to 1968.
(Source: CPC annual report 1957)
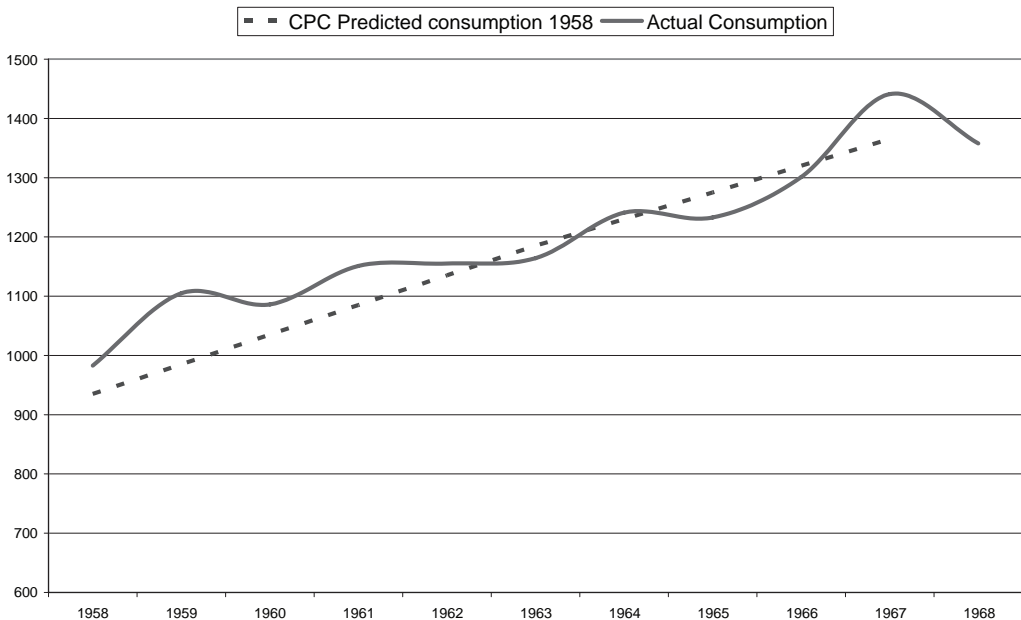
The forecast for 1967 (made in 1957) predicted a Norwegian consumption of 1.35 million tonnes, while the actual consumption in 1967 was 1.358 million tonnes! It appears that the industry's ability to predict future domestic consumption was exceedingly good, making an argument for a dramatic capacity built-up based on optimistic consumption expectations implausible.

If entry deterrence was the motive for over production, we should have expected other European countries to have a similar capacity expansion at that time, as they would have had the same strategic incentives to deter entry. As can be seen in Figure 7.4 the built-up in other European countries was much later.

Figure 7.4:
The expansion in European and Norwegian export the period 1955 to 1989 (Source CEMBUREAU)
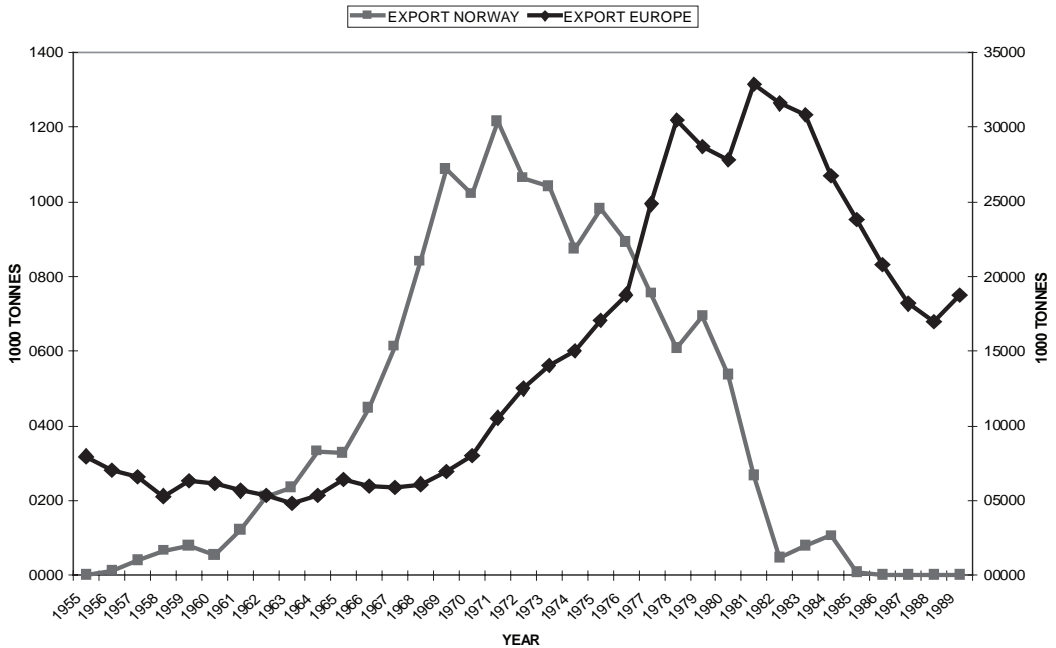
We therefore conclude that the incentives created by the cartel's sharing rule is the most plausible explanation for the large capacity investments in Norway.

7.3 Sharing rules in practice

The Norwegian cartel was subject to a very formal agreement where the market sharing rule was implemented with rigor. This can be illustrated by the fact that CPC and DPC implemented side payments to adjust for sales that were in excess of the firms' domestic market share. The background for this was that DPC had better export facilities (such as port loading technology). This need for market division and organization of the exports was clearly stated by the industry: "CPC's deliveries to its ordinary, domestic market increased from 464.000 tonnes in 1963 to 484.000 tonnes in 1964. In addition, it delivered 54.000 tonnes to DPC's customers, which implied that DPC's export increased with an identical amount. For this indirect export, CPC compensated DPC according to the ordinary export prices." (p. 13). There are several similar statements in other annual reports.

There are other examples of similar sharing rules that have been used by other cartels. We have already mentioned the German cement cartel of the 1920s and 1930s. Another example is the domestic cartels in Japan, which allocated quotas according to relative capacity, led to excess capacity in many Japanese industries during the 50s and 60s [see Matsui (1989)].

Another prominent case of a cartel that divided the market according to production capacity is the so-called Lysine cartel that operated in the period 1992-95. According to Griffin (2001) the cartel members typically met late in the year in order to determine how much each producer had sold in the preceding year. The members then proceeded by estimating the market growth for the upcoming year and allocated the growth among themselves. The volume allocation agreement then became the basis for an annual "budget" for the cartel, a reporting and auditing function and a compensation scheme. The international lysine cartel did not face the same incentive problem as the domestic Norwegian cartel, both because of its international nature and because they did not use a common distribution system. However, the lysine cartel still faced the common cartel problem of how to limit cheating.

The most recent examples of production sharing rules are found in the agricultural cooperatives (Bergman, 1997). The US had 5800 farm marketing and supply cooperatives in 1986 (Sexton, 1986). According to Bergman there were 4536 primary cooperatives just in Germany in 1997. Similar arrangements are found in many other European countries. Typically cooperatives purchase whatever their members have been able to produce, and then decide how much to sell at home. The rest is sold (often at much lower prices) on world markets. Since the cooperatives usually cannot restrict their members production, the incentive structure is analogous to our set-up. In several countries (as e.g., Denmark, Finland, Sweden, France, Germany and Netherlands) agricultural marketing cooperatives are explicitly exempt from prohibitions that regulate other firms (Bergman, 1997).

There are, of course, other sharing rules, most notably geographic market segmentation. An example of this is the so-called marine construction and transportation cartel, where the conspirators reached an agreement to allocate customers and agree on pricing heavy-lift derrick barge and related marine construction services in the major oil and gas production regions of the world. Heavy-lift derrick barges are floating crane vessels with a capacity to lift heavy structures, such as the decks of offshore oil platforms, in a marine environment. The conspiracy originally targeted contracts in the North Sea, but grew to include projects in the Gulf of Mexico and the Far East.[9] Two firms owned all (six) heavy-lifter derricks in the world. In 1997 the two firms (and one of the firm's subsidiaries) were accused of regional market sharing and price fixing.

Geographic sharing rules have other incentive problems than the one studied in this article, such as when economic growth varies considerably across regions. Since cement

---

[9] Information on this cartel can be found for instance in the Department of Justice's press release Monday December 22th 1997, (see http://www.usdoj.gov/atr/public/press_releases/1997/1325.htm.)

production is observable, it can be measured and the market can easily be divided. Using production as a sharing rule will ensure that regional differences in consumption patterns will effect the individual cartel members' profitability in a symmetric way.

7.4 The Norwegian cement cartel – A story of overinvestment and inefficiencies?

We will discuss two studies. The first study that was done by Steen and Sørgard (1999) analyse both the whole period of the cartel and the preceding monopoly period. Here the incentives to overinvest due to the market sharing rule is established theoretically and empirically. In particular it is shown that it is the later period of the cartel that this overinvestment in capacity took place, and this way was ended up being an important determinant for the 1968 merger. The next study by Röller and Steen (2006) therefore concentrate on the late cartel period and develop a structural model that allows them to calculate the magnitude of both the inefficiencies and the total welfare effects. We start out by presenting some of the main findings in the first study before we present Röller and Steen's inefficiency costs and welfare results.

7.4.1 The Norwegian cement cartel - an illustration of an industry with overinvestments due to semicollusion

In Steen and Sørgard (1999) both the early and late cartel period and the monopoly period are analysed. By tailor making a three stage game they show that the semicollusion effect will be triggered as the domestic market gets above a certain level. When the "domestic monopoly cake" gets large enough, the incentives to invest in capacity increases and the semicollusion effect is triggered. This effect can be measured through changes in the export level. In particular they formulate hypotheses for how export will change with changes in export prices (R), costs (C) and domestic market size (A). We will here discuss these hypotheses and present some of the empirical tests and results from Steen and Sørgard. The hypotheses are summarized in Table 7.1.

Table 7.1: Industry Hypotheses
(Source: Steen and Sørgard, 1999, Table 1)

		Cartel (1923-67)			Monopoly (1968-)	
		Early		Late		
$\frac{\partial E}{\partial R}$		+		+	+	
$\frac{\partial E}{\partial C}$		0(-)		-	0(-)	
	(i)	-	and		-	-
$\frac{\partial E}{\partial A}$	(ii)	-	and	+	-	
	(iii)	+	and	+	-	

The relationship between export price and exports is straightforward. The higher the export price, the higher the exported quantity, all else equal. This effect will be the same regardless of which cartel phase one is looking at. Hence, the three plus signs. Second, we consider the relationship between exports and costs. The effect is determined by the level of capacity. With an optimal capacity level this relationship will be negative - an increase in costs is followed by a reduction in exports. However, with a capacity above the optimal level this is not necessarily true. In an excess capacity situation changes in costs will not influence the exported quantity. As discussed above there are reasons to believe that the industry had a sub-optimal capacity level in the early cartel phase. The large investments in the early post World War I years had led to a large capacity and the firms did already export considerable quantities of cement when they established the sales office, therefore, the zero with the minus sign in the parenthesis. In the long run, however, the cartel could close down (not renew) capacity. Hence, as we approach the late cartel phase we therefore expect the ordinary relationship between costs and export to dominate: the higher the costs are, the lower is the exported quantity. In the monopoly phase, after the merger in 1968, the capacity was very high, and export was not profitable.[10] It is thus reasonable to assume that the capacity was above the optimal level for the monopoly, and that costs had no effect on the exported quantity. Hence, the zero with the minus in the parenthesis.

Let us now consider the relationship between exports and domestic market size. By distinguishing between three time periods (phases) we can pin down the semicollusion effect. The model predicts three possible regimes, (i) - (iii). In the monopoly phase, an increase in domestic demand is expected to reduce the exported quantity. Thus, the last minus sign in all three outcome paths. In the cartel period, the situation is more complicated. If each cartel member has a capacity which is larger than the optimal capacity initially, the model predicts that an increase in domestic demand will reduce the exported quantity. This is what we intuitively expect, since an increase in domestic demand should result in a reshuffling of sales from the export market to the domestic market. However, the result is reversed if the domestic demand is sufficiently large, the costs sufficiently low or the export price sufficiently high. In that case each of the cartel members invests in capacity in order to increase its quota in the domestic market. Then we have the semicollusion effect: an increase in domestic demand will increase the investment in capacity and thereby increase the exported quantity. We know that the semicollusion effect is triggered if the domestic demand is sufficiently large, the costs sufficiently low, or the world market price sufficiently high. We know from Figure 1 that the domestic consumption increased substantially in the cartel period, especially after World War II. Technically speaking, we therefore expect A to become large enough to trigger the semicollusion effect in this period.

Hence, our hypothesis is that it is more likely that each firm overinvests in capacity in the late than in the early cartel. In line with this reasoning, there are three possible

---

10 'With no (profitable) export possibilities Norcem had to reduce their capacity with more than one third.' [Gartmann (1990; 246)].

empirical results which all are consistent with our model. They are all shown in Table 7.1. We see that a result where $\partial E / \partial A > 0$ in the early period of the cartel and $\partial E / \partial A < 0$ in the late period of the cartel, is not consistent with the predictions we can draw from our model. *Early* and *late* in is defined as the period 1923 to 1955 and 1956 to 1968 respectively.

The hypotheses in Table 1 can all be captured in a Norwegian export-supply equation, where export is determined by variables as export price, costs and a variable measuring the domestic demand level; $E = f(R, C(\cdot), A)$, where E is Norwegian exports, R is the export price, $C(\cdot)$ is the marginal costs of production and A measures the size of the domestic market. The hypotheses summarised in Table 1 are captured in the signs of the derivatives with respect to W, $C(\cdot)$ and A.

The export equation is assumed to have a log-linear form, and annual data for the period 1927-82 is used to estimate it. The left hand side variable, E, is measured as total yearly Norwegian cement export in tonnes. The export price R is calculated as the average year-ly unit export price per ton cement. To represent the size of the domestic market, A, the yearly Norwegian gross national product (GNP) is used. The data set allows one to cal-culate a very precise short run marginal cost figure as a calculated figure based on the yearly expenses of labour (excluded permanent clerical staff), materials, fuel and elec-tricity per ton produced cement. All values are in real terms. To implement the hypothe-ses in Table 1 dummy variables techniques are used. To test the cost hypotheses, two C variables are included. One cost variable runs for the entire data period, and one is multi-plied by a dummy that takes the value *one* in the period from 1956 to 1967, and *zero* otherwise. To test the semicollusion hypothesis three A variables are included, all multi-plied by different dummy variables that correspond to one of the phases. Hence, one obtains different parameter estimates for each phase. Furthermore, a dummy variable is included to account for the effect of World War II. The econometric model can now be written as[11];

$$\ln E_t = \alpha + \beta_W \ln R_t + \beta_C \ln C_t + \beta_{C56-67} D_{t,56-67} \ln C_t$$

(1) $$+ \beta_{A27-55} D_{t,27-55} \ln A_t + \beta_{A56-67} D_{t,56-67} \ln A_t + \beta_{A68-82} D_{t,68-82} \ln A_t$$

$$+ \lambda D_{t,40-44}^{war} + \varepsilon_t ,$$

---

11 In Steen and Sørgard (1999) we also estimate other versions of the export supply equation where we change on the cost specification and the market size variable.

Table 7.2: Empirical results supply equation (1)
(Source: Steen and Sørgard, 1999, Model A – Table 3)

α	15.511*
	(2.340)
β_W	0.870*
	(0.153)
β_C	-1.225**
	(0.527)
β_{C56-67}	-5.005*
	(1.195)
β_{A27-55}	-2.198**
	(1.174)
β_{A56-67}	4.183*
	(1.096)
β_{A68-82}	-0.244
	(0.929)
λ	-4.503*
	(0.783)
R2	0.88
DW	2.27
$Q(1)_{Ljung-Box}$	1.63
$Q(2)_{Ljung-Box}$	1.66
$Q(3)_{Ljung-Box}$	4.00
F-test	
$H_0 : \beta_{C56-67} = \beta_C$	6.07@
F-test	
$H_0 : \beta_{A27-55} = \beta_{A56-67} = \beta_{A68-82}$	22.22@

significant at a .025 level, ** significant at a .05 level, *** significant at a .10 level. The numbers in parentheses are standard errors.
@ significant at a 2.5% level.

The D_S are dummy variables that take the value one according to the period defined by the subscript, *e.g.,* $D_{t,\,56-67}$ takes the value one in the period from 1956 to 1967, and *zero* otherwise. The World War II effect is captured through the $D_{t,\,40-44}^{war}$ dummy. The error term ϵ, has by assumption the standard properties. The hypotheses from Table 7.1 is implemented as single and joint restrictions. The results and model statistics is presented in Table 7.2.

The World War II dummy is significant at a .025 level, suggesting a negative effect on export. This is reasonable since the cement industry, as well as most other Norwegian industries, experienced a negative shift in production during the war (see also Figure 2).

The remaining parameters can be used to test our predictions. The export price effect (β_R) is positive and significant at a .025 level. As regards costs in early cartel and monopoly, the null hypothesis of $\beta_C > 0$ is rejected at a 0.05 significance level. However, the results are different when we look at the parameter measuring the cost effect in the late cartel phase (β_{C56-67}). This parameter is still negative, but now it is larger in magnitude and it is significant on all interesting significance levels. Thus, the results suggest that the cost is of larger importance in the late cartel phase than in the other two phases.[12]

The last predictions to be tested are the effect of changes in domestic demand on exports, i.e., the semicollusion hypothesis. Dependent on the signs of the three parameters, β_{A27-55}, β_{A56-67} and β_{A68-82}, we will be in situation (i), (ii) or (iii). At first glance, the results suggest that we are in a situation (ii), where the semicollusion effect is triggered in the late cartel phase. When undertaking statistical tests the null hypotheses of β_{A27-55} and β_{A56-67} are rejected at a .05 significance level. The model predicts a negative sign on β_{A68-82}, but the parameters are not significant. The semicollusion effect as measured by the positive sign on β_{A56-67} is strongly significant. Thus, the most reasonable interpretation is that it is in this phase that the semicollusion effect is triggered. It is consistent with the cost parameter results, where the late cartel phase was found to be significantly different from the other two phases.[13] Hence, the domestic market size has a strong and statistical significant positive effect on exports in the period 1956 to 1967 - the late cartel phase. This suggests that there is a semicollusion effect triggered in this period, i.e., higher domestic demand triggers investments in capacity and thereby an increase in export.

In Steen and Sørgard (1999) the semicollusion effect became significant in the Norwegian cement industry first in the late cartel phase - each member of the price cartel increased its investment in capacity to capture a larger share of a growing domestic market. The driving force was the market sharing rule: each firm was given a quota in the domestic market according to its share of total domestic capacity. The study thus suggests that in this particular domestic market the capacity expansion that could have been a virtue - for example an instrument to facilitate collusion - instead became a *problem*. Each of the cartel members should have anticipated this problem and solved it by, for example, exiting from the price cartel in due time. However, the apparent success of the price cartel for many years may explain why they underestimated or simply neglected the potential problem which the growth in the domestic market triggered in the mid 50s. In the 60s - a decade too late - they finally became fully aware of the problem, as the following description from cartel negotiations in the 60s also indicates: *'During the opening round of negotiations for a new cartel agreement there appeared to be much dissatisfaction with the apparent lack of industry coordination when it came to capacity expansion .. . It consequently did not take the parties long to see large potential benefits from a full*

---

12 The hypothesis that there are no differences between the three periods/phases ($H_0: \beta_{C56-67} = \beta_C$), where also tested and rejected.

13 The joint hypothesis that all three market-size parameters are required to be equal; $H_0: \beta_{A27-55} = \beta_{A56-67} = \beta_{A68-82}$ is also rejected.

merger' [Lorange (1973; 33)]. At that time, with excess capacity in the industry, a disso-lution of the price cartel would probably have led to fierce price competition. No sur-prise, then, that the three firms merged in 1968.

Steen and Sørgard established through their results that the cartel was inefficient due to the underlying incentives that led to semicollusion and overinvestments. An interesting question is now what the magnitude of these inefficiencies had, and in particular what the exact welfare implications were. Röller and Steen (2006) therefore go a step further, by looking more particularly on the late cartel period modelling a structural model they are able to quantify both the inefficiency and the welfare effects of the merger.

7.4.2 The Norwegian cement cartel – efficiency costs and welfare implications

Röller and Steen (2006) formulate a two stage structural decision problem that allows them two calculate an estimate of long run marginal costs using observed market data (see their equation (3)), which they also qualify through accounting data on costs.[14] The predicted costs show that the cartel earn a profit in the domestic market, but loose money on exports – the export price is always below marginal costs, whereas the domestic price is always higher than marginal costs (i.e., $R<C<P$). The development in prices and pre-dicted cost are shown in Figure 7.5.

Figure 7.5:
Comparison of domestic price, export price and predicted marginal costs from the model (Source: Röller and Steen, 2006, Figure 4)
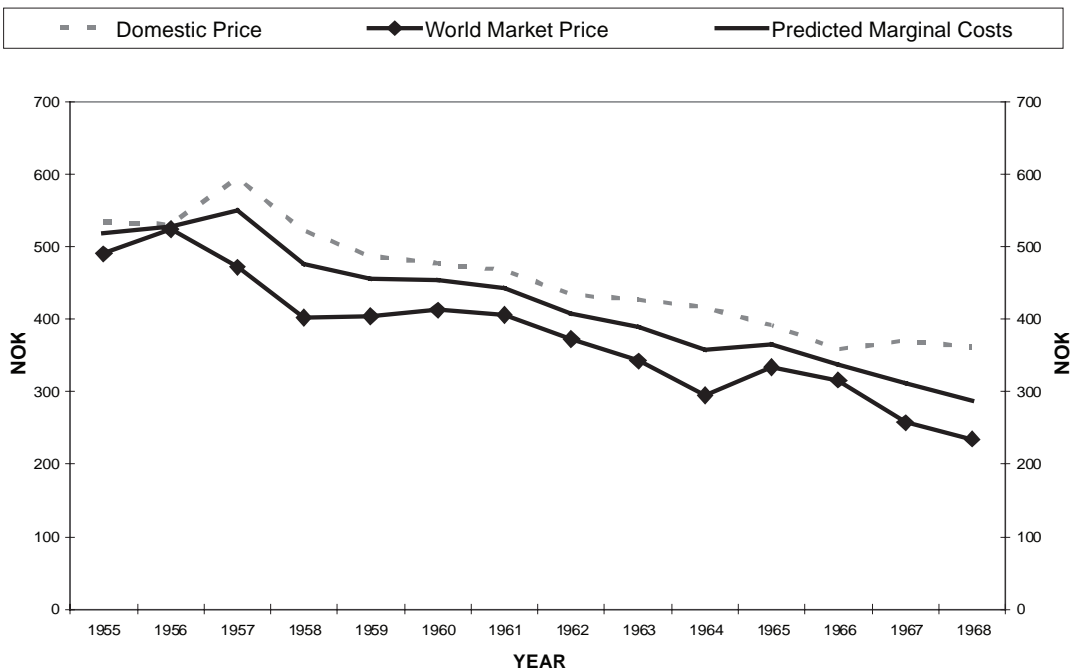

14 Note that long run implies that the marginal cost measure also accounts for the capacity cost.

One of the main assumptions in Steen and Sørgard (1999) was that $R<C<P$. This is also why they conclude that overcapacity used for export was bad for the cartel. This picture is now confirmed from Röller and Steen's calculated marginal cost measure. Whereas changes in the export price will obviously have an effect on the export loss, Röller and Steen however, also show that it affects on total welfare through the consumer surplus. In Röller and Steen's model the cartel allocates domestic output by equating marginal revenue in the domestic market to the world market price. The marginal cost of capacity does in fact not enter the first-order condition for the domestic market equilibrium. By contrast the world market price R enters the first order condition determining domestic sales as it is the opportunity cost of not exporting. As a result, the world market price plays the usual role of marginal costs. Accordingly, the lower R, the lower the domestic price. In particular, when R is below C, the cartels price is below the monopoly price defined by the usual monopoly condition. In this case, the ineffectiveness of the cartel leads to lower domestic prices and profits. This way the sharing rule induces positive domestic welfare effects. However, these gains have to be traded-off against losses in the export markets. The size of this inefficiency will depend on the total amount of capacity that is installed, which is a function of the incentives to gain a bigger share of domestic profits through the semicollusion effect of competing in capacity built up. Overall, they find that a cartel using a capacity-based sharing rule leads to higher domestic consumer surplus, but the impact on profitability is ambiguous.

This trade-off can also be represented graphically. The cartel and monopoly equilibrium are illustrated in Figure 7.6. The monopoly outcome is the usual solution where marginal revenue meets marginal cost, yielding a price P^m and quantity Q^m. No export will take place in monopoly equilibrium, since the world price R is below marginal cost of capacity. The cartel solution is the price-quantity combination (P^{cart}, Q^{cart}) where marginal revenue equals R. Exports are given by the difference between total domestic production, Q, and the domestic quantity sold Q^{cart}. The change in consumer surplus by moving from cartel to monopoly is therefore given by the sum of the areas "A" and "C". The impact on producer surplus is given by "A" minus "B" plus the saved export loss, "D". Finally, the change in welfare is "D-B-C", which is ambiguous.

Figure 7.6:
Welfare Analysis- Cartel and Monopoly
(Source: Röller and Steen, 2006, Figure 3)
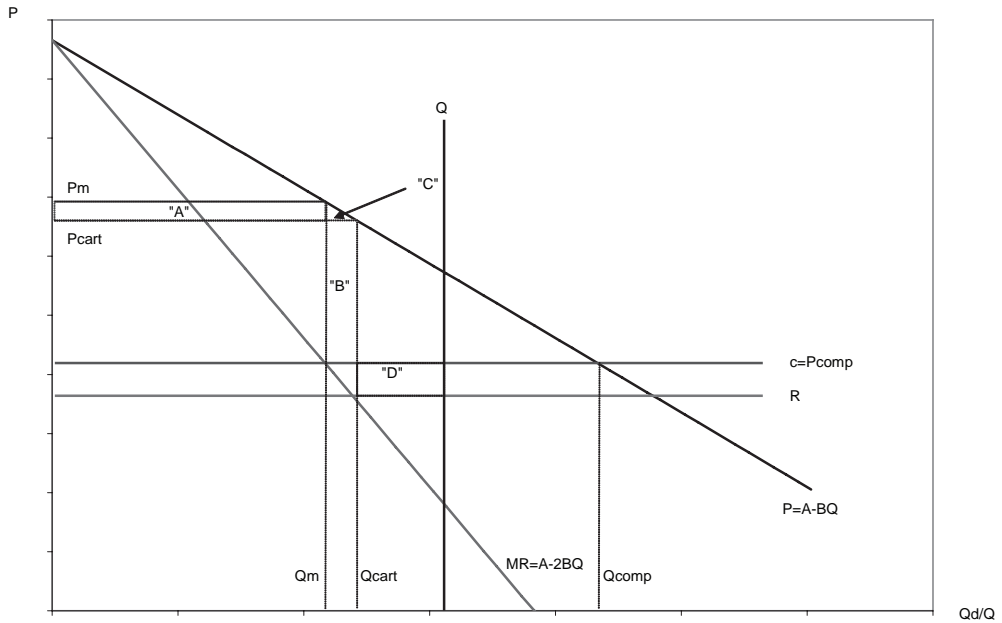
Given that the effectiveness (in terms of cartel profitability) of the Norwegian cement cartel in theory is ambiguous, Röller and Steen now turn to the data in an attempt to evaluate the trade-off empirically. To do so they estimate a domestic demand function, and from this they can find marginal revenue. Knowing the marginal revenue curve, together with the model-predicted marginal cost allows them to calculate producer- and consumer surplus, account for export loss and finally calculate total welfare effects. In order to analyze the impact of the cartel they first compare the cartel situation to that of a monopoly. Figure 7.7 compares moving from the cartel to a monopoly equilibrium for each of the years 1955 to 1968.

Figure 7.7:
Impact of moving from a cartel to monopoly
(Source: Röller and Steen, 2006, Figure 7)
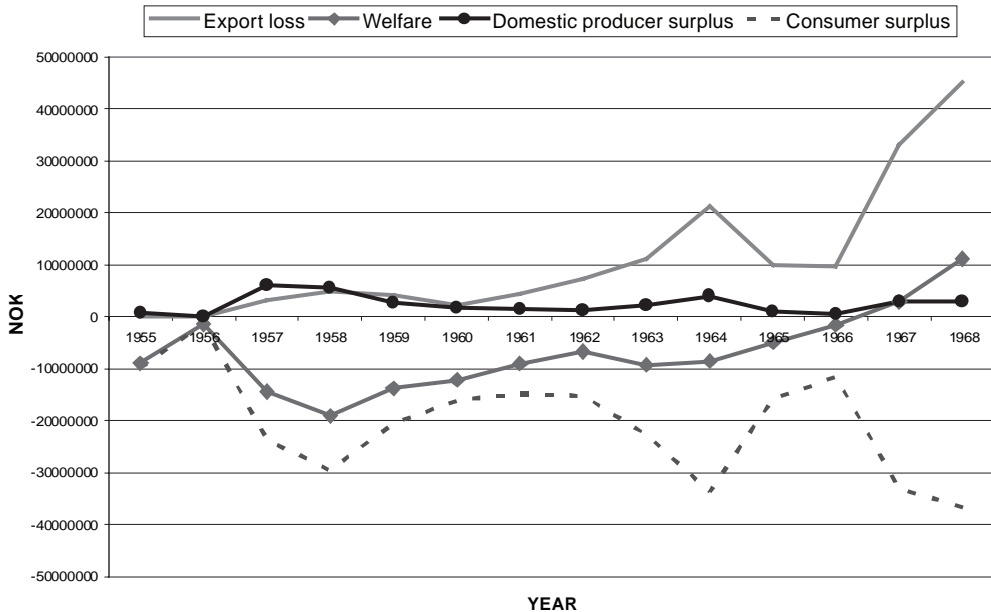
As can be seen in Figure 7.7, the cartel is not effective at all. In particular, losses from exporting are very large. Apparently, the sharing rule creates a considerable incentive problem, leading to significant overproduction and exporting below marginal costs. By contrast, the losses in the domestic market are substantially lower, indicating that the common sales office is rather effective in keeping domestic prices close to monopoly levels. As a consequence of the sharing rule, domestic consumers are better off under the cartel relative to a monopoly. The cartel's ineffectiveness is to the benefit of consumers. Figure 7.7 also shows that the effectiveness of the cartel is declining dramatically over time, as the incentive problem is becoming more and more of a problem for the cartel. Interestingly, the cartel was operating so inefficiently around 1967 that a merger to monopoly actually had a positive effect on welfare. The loss from exporting is so large that the gains to consumers are outweighed, resulting in positive domestic welfare from a merger to monopoly.

The results of Röller and Steen (2006) suggests that the timing of the merger took place exactly at the right time, i.e. when a benevolent domestic dictator would have chosen to merge. Given the likely absence of benevolent dictators in Norway, one may wonder why the merger took place in exactly 1968, i.e. exactly when the net benefit of consumers and firms becomes positive. As already discussed in Section 7.2, a reasonable explanation are the existence of other institutional agreements that have been agreed on a long-term basis. As losses from exporting where mounting up and other agreements were running out in 1968, a merger to monopoly was ultimately implemented. Another factor

allowing a merger to monopoly in 1968 was that antitrust concerns vis-à-vis the merger were unlikely to be significant in Norway at that time.

The previous findings suggest that the merger took place exactly at the optimal time for welfare. However, this conclusion is premature, as it ignores the possibility of competition. Table 7.3 presents the comparison with Cournot competition for the year 1968, as well as the accumulated rents over the sample period 1955-1968.

Table 7.3:
Impact on Producer Surplus, Consumer Surplus, and Welfare (1000 NOK).
(Source Table 3 Röller and Steen, 2006)

	1968	*Accumulated 1955 to 1968*
Cartel to Cournot Competition		
Producer surplus	-106 797	-668 521
Consumer surplus	237 350	1 467 765
Net Welfare Effect	130 553	799 244
Cartel to Monopoly		
Producer surplus	47 891	189 032
Consumer surplus	-36 760	-285 157
Net Welfare Effect	11 131	-96 125

Domestic consumers would have benefited from Cournot competition, i.e. the cartel is not as ineffective that it drives domestic prices down to non-cooperative levels. On the other hand, competition would have lowered producer surplus. In light of this, the wisdom of the merger to monopoly in 1968 has to be reassessed. The merger may have come at the right time, but only if the alternative is to do nothing (i.e. keep the cartel in place). If the alternative is to move to competition, neither cartel nor merger to monopoly have been to the benefit of Norwegian welfare. A well functioning competition policy authority would have broken up the cartel and not allowed the merger to monopoly, something we will return to below. In 1968 alone the welfare gain from breaking up the cartel in favor of competition is some 131 million NOK, while the merger to monopoly increases welfare by only 11 million NOK. In this sense the merger to monopoly was a distant second best solution. The picture is even more dramatic with regard to consumers. While domestic consumers lose from the merger (some 37 million NOK), our model suggests that they would benefit 237 million NOK from competition in 1968 alone.

7.5 Policy implications and some conclusions

Given the incentives created by the sharing rule in the mid 1950's, one may wonder why the firms did not merge earlier than 1968. One explanation is as we said above the first 30 years of cartel success that they had had when the capacity built up started. Another explanation is the existence of other institutional agreements that have been agreed on a long-term basis. The firms entered into two long-term agreements in 1957 and 1962. In these contracts the firms were tied together even stronger. "When the industrial firms start a market collaboration is it natural that this lead to increased contact and exchange of views also within other fields of the firms activities. In the cement industry this lead to an extension of the collaboration, both with regards to particularities and more general issues. Common purchases, standards of cement types, common packaging was agreed upon. This was particular formalized in the agreement of 1957, and even more so with the revision of the agreement in 1962." (Gartmann, 1990 p. 115). CPC themselves described the agreement in their annual report (1962 p. 7) as "An agreement that has as main object to govern a good collaboration between the cement factories to obtain a rational solution of the industry's production and distribution tasks". Interestingly enough the 1962 agreement was denoted "the 7 year agreement" lasted until December 31, 1968. Hence, the merger in 1968 came at a time where either a new market agreement had to be negotiated, or an alternative industry structure. As losses from exporting where mounting up and other agreements were running out in 1968, a merger to monopoly was ultimately implemented.

A third factor allowing a merger to monopoly in 1968 was that antitrust concerns vis-à-vis the merger were not significant in Norway at that time, as an effective merger control did not exist and consumers did not play much of a role in competition concerns. In fact, Norway had no real merger control in 1968. The first formal law dealing with competition policy in Norway was the "trustlaw" approved in 1926. In 1932 Norway passed an extension to the "trustlaw" that allowed authorities to cartelize industries by law. In addition, the 1932 extension outlawed excessively low prices in order to "prevent excessively low profitability in the industry". Consumer interests were practically irrelevant and this cartel-friendly practice continued up to world war I (Nordvik, 1995). In 1953 Norway issued a new law on competition – the so-called price law. The law stated very general objectives on competition issues, but once again the authorities practice was quite cartel friendly. Due to lack of resources the authorities who were responsible at the time ("Prisdirektoratet") did not really focus on the analysis of markets (Halvorsen and Undrum, 1995). Interestingly enough the authorities themselves concluded as late as in 1982 that the "price law from 1960 did not warrant cartel control". However during the 80's the political views changed and merger control was introduced in Norway by 1988. Thus, the general view at the time was that all mergers were good. As a result, there was no visible opposition against the NORCEM merger in 1968.

Röller and Steen (2006) conclude that relative to keeping the cartel in place, the merger to monopoly in 1968 was exactly what a benevolent dictator would have done. However, the picture is rather different, if the alternative is competition. In this case, the Norwegian cement industry is subject to a considerable public policy failure.

In sum, the Norwegian cement cartel and the way in which it functioned illustrate several things. First that cartel functioning is complicated and is often undermined by incentives to defect on some or several aspects, leading to often unanticipated competition effects. Second, the cement case shows that the presence of competition regulation and competition authorities can be crucial to ensure consumer welfare. An imposed deregulation of the cartel to achieve competition rather then monopoly would have been tenfold better in terms of welfare, underlining the importance of well functioning competition authorities.

References

Aiginger K. and M. Pfaffermayr, 1997, "Looking at the cost side of "monopoly"", The Journal of Industrial Economics, Vol 45, No. 3, 245-267

Benoit J.P. and V. Krishna, 1987, "Dynamic Duopoly: Prices and Quantities", Review of Economic Studies, 54, January.

Bergman M.A., 1997, "Antitrust, Marketing Cooperatives, and Market Power", European Journal of Law and Economics, Vol. 4, No. 1, 73-92.

Christiania Portland Cementfabrikk, annual reports 1955-1967.

Brander J.A. and R.G. Harris, (1983) "Anticipated collusion and excess capacity" Institute for Economic Research, Queen's University, Discussion Paper, 530 (April, 1983).

Davidson and Deneckere, 1990, "Excess Capacity and Collusion", International Economic Review, Vol. 31, No.3, August.

Eswaran, M. 1997, "Cartel Unity over the Business Cycle", Canadian Journal of Economics, No.3., August.

Gartmann, F., 1990, "Sement i Norge 100 år" (Cement in Norway 100 Years), Norcem A.S., Oslo.

Genesove, D. and W.P. Mullin, 1998, " Testing static oligopoly models: conduct and cost in the sugar industry, 1890-1914," Rand Journal of Economics, Vol. 29, No. 2, 355-377.

Genesove, D. and W.P. Mullin, 2001, "Rules, Communication, and Collusion: Narrative evidence from the sugar institute case" American Economic Review, Vol. 91, No. 3, 379-398.

Griffin J.M. (2001) "An inside look at a cartel at work: Common characteristics of international cartels" In the proceedings of the 3rd Nordic Competition Policy Conference on "Fighting cartels – why and how?", held in Stockholm on 11-12 September 2000, organized by the Swedish Competition Authorites, published by Elander Graphic Systems, Gøteborg, 2001, pp 29-55.

Green, E., J., and R. H. Porter 1984, "Non-cooperative collusion under imperfect price information", Econometrica, Vol. 66, No. 3, 87-100.

Halvorsen J.A. and S. Undrum, 1995, "From price regulation to competition policy", In "Market, competition and policy: Festschrift for Egil Bakke", Edited by E. Hope, P. Munthe A.C.S. Ryssdal and S. Undrum (In Norwegian), Editor: Fagbokforlaget, 51-78.

Levenstein, M. C., 1997 "Price Wars and the Stability of Collusion: A study of the Pre World War I Bromine Industry" Journal of Industrial Economics, Vol. 45, No. 2, 117-37.

Lorange, P., 1973, Anatomy of a Complex Merger: A Case Study and Analysis, Journal of Business Finance, 5, 32-38.

Matsui, A., 1989, "Consumer-Benefited Cartels under Strategic Capital Investment Competition", International Journal of Industrial Organization, 7, 451-470.

McCutcheon, B., 1997 "Do meetings in smoke filled rooms facilitate collusion?" Journal of Political Economy, Vol. 105, No. 2, 330-50.

NORCEM, annual reports, 1968-82

Nordvik H. W., 1995, "Between the market and the cartel: American and German models in Norwegian competition policy before 1940", In "Market, competition and policy: Festschrift for Egil Bakke", Edited by E. Hope, P. Munthe A.C.S. Ryssdal and S. Undrum (In Norwegian), Editor: Fagbokforlaget, 35-30.

Osborne, D. K., 1976, "Cartel Problems", American Economic Review, 66(5) pp. 835-44 December.

Osborne, D. K. and C. Pitchik, 1987, "Cartels, Profits and Excess Capacity", International Economic Review, Vol 28, No. 2, June.

Porter, R.H., 1983, "A study of cartel stability: The Joint Executive Committee, 1880-1886", Bell Journal of Economics 14, 301-314.

Rotemberg, J.,J., and G. Saloner 1986, "A supergame-theoretic Model of Business Cycles and Price Wars During Booms" American Economic Review, Vol. 76, 390-407.

Röller, L.H. and F. Steen, 2006, "On the Workings of a Cartel: Evidence from the Norwegian Cement Industry from 1955-1968", Vol 96, No. 1, March 2006, pp. 321-338 American Economic Review.

Röller, L.H. and H.W. Friederiszick, "Lokale Märkte unter Globalisierungsdruck? Eine industrieökonomische Studie zur deutschen Zementindustrie" (Local Markets and Globalization: a Study of the German Cement Industry), study commissioned by the German Ministry of Labor and Economics, Berlin: RACR economics, January 2002.

Scherer F.M. 1980, "Industrial market structure and economic performance", Houghton Mifflin Company, second edition.

Sexton, R.J., 1986, "Cooperatives and the Forces Shaping Agricultural Marketing" American Journal of Agricultural Economics, Vol. 19, 1167-72.

Steen, F. and L. Sørgard, "Semi-Collusion in The Norwegian Cement Market", European Economic Review, 43 (1999) 1775-96.

Suslow, V. Y. (1988): 'Stability in international cartels: An empirical survey', Hoover Institutions Working Paper no. E-88-7, Stanford University.

Frode Steen and Lars Sørgard

From failure to success in the Norwegian airline industry

8.1 Introduction

The Norwegian airline industry was deregulated in 1994. A strict regulatory regime was replaced by free entry for domestic airlines. Two large airlines – SAS and Braathens SAFE – started to compete head to head on numerous domestic routes. Unfortunately, the first eight years after deregulation they did not compete the way we could hope for according to the most optimistic predictions from text books in economics. Rather, we experienced some text book examples of market failure. Deregulation led to many empty seats, high prices on flexible tickets, clustering of time scheduling of flights as well as entry, exit and failing firm. In 2001 we returned to monopoly, which implied that a regulated monopoly prior to 1994 was now replaced by an unregulated monopoly. However, in 2002 entry led to price competition. This time it was more in line with the most optimistic predictions from text books in economics.

As the short description indicates, deregulation led to turbulence in the Norwegian airline industry. Some years of market failure was followed by success. In the following we describe the development in the Norwegian airline industry following deregulation, and the lessons we learn from this experience. In Section 2 we describe the initial phase 1994-97 and how competition emerged between SAS and Braathens SAFE. Then we describe in Section 3 the period 1998-99, where a new main airport was opened near Oslo and the airline Color Air entered and 14 months later exited. In Section 4 we describe the period 1999-2001, a period with intense rivalry for large customer contracts, and where Braathens SAFE became a failing firm and the monopoly was re-established. In section 5 we describe the period after 2001, with the abolishment of the frequent flyer program on domestic routes and the entry of Norwegian. In section 6 we conclude our discussion, and discuss what we have learned from the turbulence in the Norwegian airline industry.

8.2 From regulation to semicollusion

The Norwegian airline industry has many of the features observed in other European countries. The largest routes in Norway are of almost equal size as the routes between many specific airports inside Europe as well as the United States.[1] Before 1987 one sin-

---

[1] Not surprisingly, the number of flights between city pairs as, for example, San Francisco-Los Angeles and London-Amsterdam, are much higher than between city pairs in Norway. However, when we take into account the fact that there are several airports in each of these large cities, then the number of flights between specific airports are at the same level as the number of flights on the largest routes in Norway [see Strandenes (1990)].

gle firm was given the exclusive right to have flights on each route. Both prices, the number of flights and time location were regulated. However, there are indications that the regulation had only a minor or no impact on the firm's price setting.[2] In October 1987, a second airline was permitted to have a limited number of flights on some particular routes - four flights at a maximum on each route.

In April 1994, all routes, except those between the smallest airports ('kortbanenettet'), were further deregulated.[3] All domestic firms were free to enter, and they were free to set prices and to determine the time location of their flights as well as the number of flights on each route. Two Norwegian airlines, SAS and Braathens, were the active firms in the Norwegian airline industry before deregulation. They continued to be the only active airlines also after deregulation. On 24 out of 32 routes, the legal monopolist from the era of regulation continued to be a monopolist.

Prior to deregulation, both firms threatened to cut prices following deregulation. However, a study indicates no price reduction on the full fare tickets in the business travellers' segment following deregulation, and only a minor increase in the share of discounted tickets.[4] Although we do not know with certainty whether the airlines colluded on prices (in particular on flexible tickets), there are several reasons for why they could succeed in achieving collusive prices also after the deregulation.

First, there was a potential for collusion in this particular industry. There were only two active firms and no threat of entry of foreign firms until they were permitted to enter in April 1997. Price changes are either announced in the press or through the Amadeus booking system, which in both cases will quickly be observed by the rival. Hence, both firms can quickly respond to the rival's price changes.

Second, the two firms had initially almost equal market shares in the domestic market. Then it was natural to continue with the initial market sharing in the deregulated system. In fact, there were only rather minor changes in the market shares on each route as well

---

2 The regulation dates back to the 40s. Each firm had to apply to the civil aviation authorities concerning price changes, typically once every year. Then each firm could argue that they have had cost increases, an argument that the authorities would find difficult to disprove. Norman and Strandenes (1994) have calibrated the market equilibrium on the route Stockholm-Oslo prior to deregulation in 1993, and they conclude that '[i]nsofar our calibrated coefficients seem "reasonable", the regulatory constraint cannot be severe'. (p. 96) Hence, their study gives support to our conjecture that the regulation had no substantial impact on the price setting.

3 In terms of passengers these smaller airports ("Kortbanenettet") represented less than 8% of the total traffic in 1997 and 1998.

4 This is shown in Lian (1996). He found that the share of the discounted tickets increased with 2.5 %-point from 1992 to 1994-95. According to Lian (1996) this is no dramatic change: *'a 2-3 %-point increase in discount tickets in two-three years is in line with a long term trend and imply no sudden change in this trend'* [our translation] (p. 15). The increase in the share of discounted tickets was larger in the 'leisure' segment than in the business segment [see Lian (1996), table 4.4].

as in the total market shares after deregulation.[5] At 24 out of the 32 city-pair routes, the initial monopoly carrier continued to be a monopolist. For the remaining eight routes, the pre-deregulation dominant firm continued to have a dominant position. On average, the dominant firm had a 13 %-points reduction in market share on these eight routes, and it had no less than 60% market share on any of the routes in the deregulated regime.[6]

Third, for those routes where both firms did have flights, there existed a system for co-ordinating prices. The firms were permitted to consult each other concerning price set-ting. To allow for late changes of flight schedules for normal (no rebate) tickets, from one airline to another, the airlines claimed that they had to have «transferable» prices. To implement such a policy, the firms were permitted to meet regularly to inform each other concerning future prices on non-rebated tickets - labeled interline tickets. Hence, there existed an institutional pre-play communication system where each firm informed its rival about its future prices on normal tickets.

Fourth, the firms signaled an aggressive response to any move by its rival. In particular, each firm matched the rival's offer. For example, prior to deregulation Braathens SAFE introduced a rebate ticket named *Billy* to match SAS' rebate ticket *Jackpot* and set a price NOK 5 below the *Jackpot* price. SAS responded immediately by reducing its *Jackpot* price by NOK 5. A statement by a representative for Braathens SAFE suggested that this was a deliberate policy for the firms in question:

'We will match any offer by SAS within an hour, and we can not accept that SAS has cheaper rebate tickets than what we have' (our translation) [C. Fougli to Dagens Næringsliv, 20/1/94]

Such an apparent aggressive behaviour might have the opposite effect, in line with the effect of a meeting competition clause. When the rival observes that you will match any price cuts, he might find it optimal not to cut prices.

Casual observations suggest that it had been some price competition in the leisure seg-ment, where the firms offered discounted tickets. Both Billy and Jackpot were examples of this kind of tickets. These were discounted tickets with restrictions which made them unattractive for business travelers. Casual observations from the press also suggested that there were no fierce price competition in the business segment. A representative for Braathens, the public relation manager Audun Tjomsland, wrote the following in a news-paper article:

'The two Norwegian firms on Norwegian routes, Braathens and SAS, are of equal size and can follow each other during a price war. A firm starting a price war

---

5 Each firm's market share changed only modestly following deregulation; Braathens SAFE increased its market share from approximately 50% in 1993 to 52% in 1995 [see Lorentzen *et al.* (1996)].

6 The exception is the route Bodø-Tromsø, where each had two non-stop flights both before and after April 1994.

will quickly be followed by the rival firm, so the firm that starts a war will have an advantage only a day or two. Accordingly, the firms are reluctant to trigger a price war' (our translation) [Bergens Tidende, 31/7/95]

Moreover, other statements suggest that the two firms did compete aggressively along other dimensions, among others capacity. For example, Braathens explained its poor result in the first quarter of 1996 in the following way:

'Braathens explains this [poor result] with an increased competition. The firm has increased its capacity, but it has not helped much. The growth results in an increase in employment and other costs of production (our translation) [Dagens Næringsliv, 10/5/96]

A few months earlier, SAS had announced several new initiatives:
'Among the initiatives are recruitment on the ground and in the cabin, adjustment of time-scheduling of flights, an increase in capacity amounting to 400.000 seats annually, better food on business class between Norway and other countries, .. (our translation) [Bergens Tidende, 9/3/96].

Note that none of them mention price cuts. Could it be that the lack of price competition might trigger more competition along other dimensions such as capacity?

Figure 8.1. Capacity and number of passengers 1985-96 on four routes
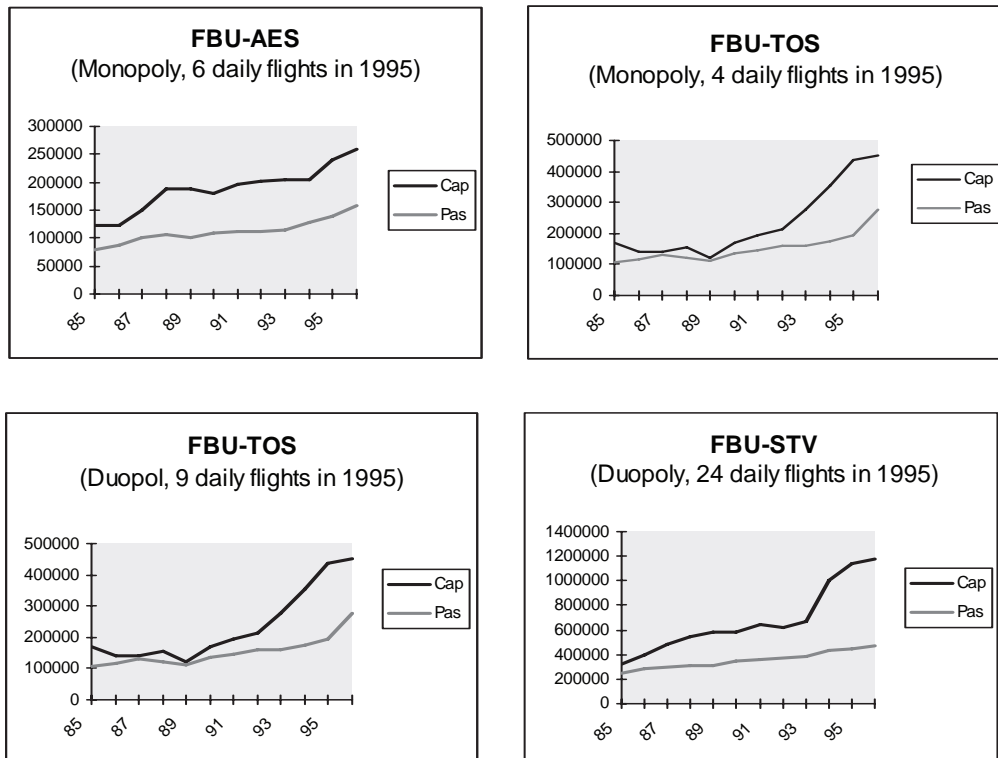
It has been shown in theory that price collusion might lead to tough competition on, for example, capacity (see Fershtman and Gandal, 1994). A high price-cost margin implies that each firm has strong incentives to capture market shares. In the airline industry an increase in capacity – for example an increase in the number of flights on a route – will make an airlines' product more attractive for the customers. See the Appendix, where we present a simple model. This suggests that there is a potential for semicollusion in the Norwegian airline industry: collusion on prices and competition on capacities.[7]

Salvanes, Steen and Sørgard (2003) did an empirical study of the change in capacity following deregulation in the Norwegian airline industry. They conclude that the observations are consistent with semicollusion. To illustrate their results, we have in Figure 1 reported the change in capacity and the number of passengers on four different routes following deregulation. On the two monopoly routes, there are apparently no structural shift in the relationship between capacity and the number of passengers following the deregulation in 1994. In contrast, there seems to be a larger increase in capacity than in the number of passengers on the duopoly routes following the deregulation. As the figure indicates, the load factor (number of non-empty seats) changed rather dramatically following deregulation. On duopoly routes it changed from 61% prior to deregulation to 49% after deregulation. As far as we know, there are no other examples of such a low cabin factor in any airline market.

Collusion on prices may also influence the time scheduling of flights. According to theory of location, it is ambiguous whether a firm should locate close to its rival or not. On the one hand, it should locate close to its rival to capture market shares. This is the classical result from the original Hotelling model. On the other hand, it should locate far away from its rival to dampen price competition. Typically, then, there are incentives for a firm to locate close to its rival - clustering - if there is competition on location, but not on prices.[8]

Salvanes, Steen and Sørgard (2005) did an econometric study of twelve domestic routes in Norway, six of which remained monopolies after the deregulation. There are three main findings we can report from the empirical study, all of which can be illustrated with

---

7 Semicollusion has been present in numerous industries. Price collusion led to intense rivalry on advertising in the American cigarette industry [see Scherer (1980), p. 388-389], the installing of excess capacity in the German [see Scherer (1980), p. 370] as well as the US cement industry [see Scherer and Ross (1990), p. 674], and to excess capacity in ocean shipping [see Scherer and Ross (1990), p. 674]. The existence of cartels in the domestic Japanese market, where quotas were allocated according to relative capacity, led to excess capacity in many Japanese industries during the 50s and 60s [see Matsui (1989)]. The price cartel in the Norwegian cement market led to the instalment of excess capacity in the Norwegian cement industry in the 50s and 60s, which showed up as a large increase in exports [see Steen and Sørgard (1999)].

8 Friedman and Thisse (1993) show that collusion on prices in a duopoly after location is chosen non-collusively results in clustering. This restores the classical result found in Hotelling (1929). However, the fact that each firm has several products - or more than one flight on each route - makes it troublesome to make clear-cut predictions from theory even when there is no competition on prices. There are examples, though, where theory predicts local clustering also in such cases. See Eaton and Lipsey (1975) and Gabszewicz and Thisse (1986).

the time locations on the two routes Oslo-Bodø and Oslo-Stavanger, shown in Figures 8.2 and 8.3.

First, deregulation seems to have no or only a limited effect on the time-scheduling pattern within each airline. In particular, the first carrier - the one with the largest number of flights - seems to spread its flights throughout the day both before and after deregulation.

Figure 8.2: Flight departures Oslo-Bodø before and after deregulation

Before deregulation (winter 93):
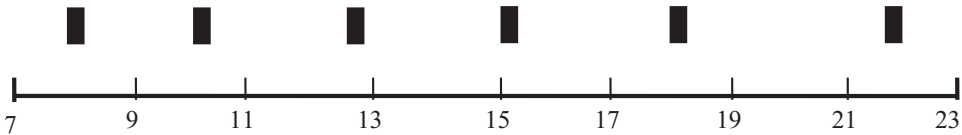
*After deregulation (winter 96):*
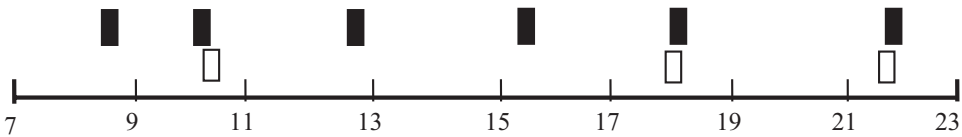
∎ = SAS   ▯ = Braathens

Second, the second carrier on each route has a tendency to locate its departures close to those of the incumbent firm. This is particularly clear on the route Oslo-Bodø, the smallest one of those two routes. On the larger route, the picture is not as clear cut. The econometric results confirm this picture.

Figure 8.3: Flight departures Oslo-Stavanger before and after deregulation

Before deregulation (winter 93):
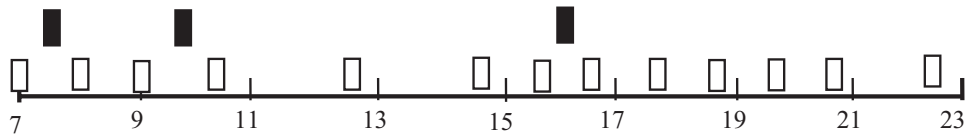
After deregulation (winter 96):
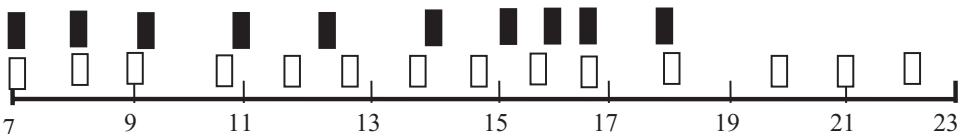
∎ = SAS   ▯ = Braathens

Third, the tendency towards pairwise flights seems to be more prevalent in the morning segment as well as in the afternoon segment than in general. This is especially the case if we look at the route Oslo-Stavanger. In these segments, 7:00-10:00 and 15:30-18:00, the typical passenger is a business traveller.

Interestingly, the econometric evidence shows a tendency toward clustering only in the segment where we expect collusion on prices (business segment). This is consistent with a general conclusion from theory, which says that a relaxation of price competition will give the firms incentives to locate close to each other in order to capture market shares.[9]

8.3 Entry and exit of a third airline

In 1998 several important events changed the competitive environment. A new entrant arrived, and a new main airport in Oslo, Gardermoen, was inaugurated. The slot capacity at Gardermoen allowed both expansion and new entry. Color Air entered in the summer of 1998, but exited 14 month later.

In Figure 8.4 the load factor for SAS and Braathens in the period January 1996 to May 2000 is shown for a route where Color Air competed against both SAS and Braathens. As can be seen from the figure, there was a drop in the load factor in 1998 for both SAS and Braathens when we had a new airport and a new entrant. After the exit of Color Air, the load factor returned to the 1996 level during the spring of 2000. This pattern was parallel for both carriers. Actually, the load factor increased even more during the next 12 months. Why did we observe these rather large changes in the load factor?

Even though Color Air was a low-cost-no-frills concept, Braathens and SAS did not primarily respond to entry with lower prices. Instead they continued to increase their capacity.[10] Hence, the competitive picture did not change, only escalate. We had more capacity, more empty seats and somewhat more price competition in the leisure segment. In total, ten new airplanes entered the Norwegian market after the opening of Gardermoen, only three of them were operated by Color Air.[11]

---

9   See, for example, Tirole (1988) who concludes that one important insight from spatial models is that firms want to differentiate their products from their rivals' products to soften price competition (p. 286-287). In Borenstein and Netz (1997), a study very much in the same spirit as ours, concerning flight departures, time schedules of flights in the US before and after deregulation in 1978 are tested empirically. They found that price competition typically resulted in less clustering of flight departures.

10 This is shown in Steen and Sørgard (2001), but also in the study of Aasheim and Askim (2000) using a shorter dataset.

11 One airplane can be used for a maximum of 16 hours a day in the Norwegian network. Hence, ten new airplanes into the market were a considerable increase of capacity.

Figure 8.4: Load factor Jan 1996 - May 2000 on a duopoly route.
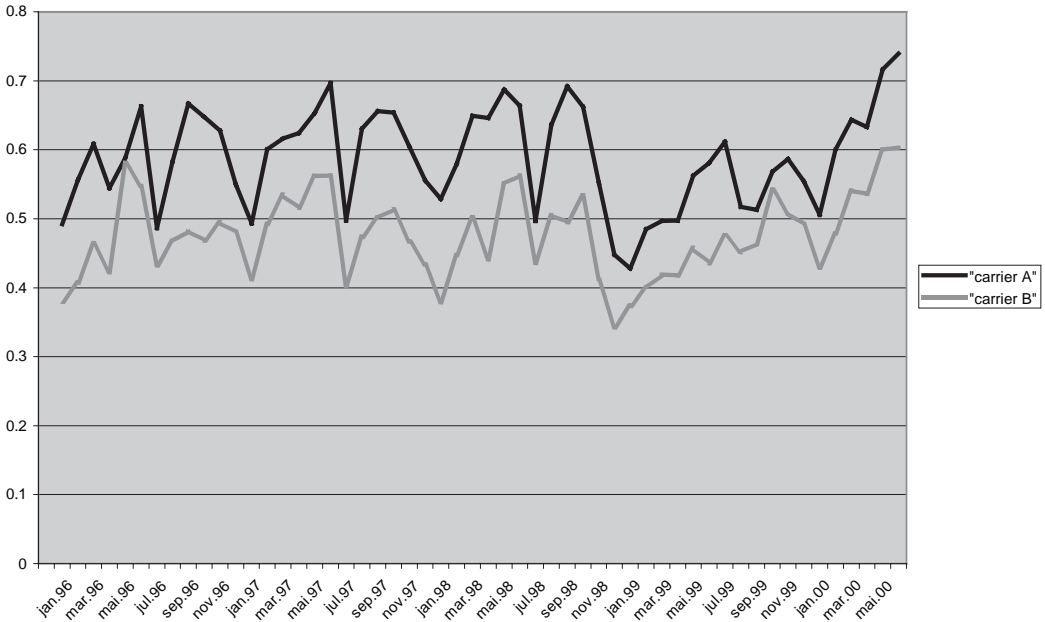
It is interesting to note that SAS was the most aggressive one in its capacity expansion. Several statements from the company can indicate that this was part of a strategy to squeeze Color Air out of the market. The managing director of SAS, Jan Stenberg, said in May 1999: *'SAS has no intention to reduce the excess capacity in the domestic market. The plan is to aim for more aggressive price advertising campaigns in the Norwegian market. ... I think it is a question about only a few months before Color Air will exit the market [our translation] (NTB-press release May 7. 1999).'* The very same day the deputy managing director of SAS, Vagn Sørensen, states: *'This is a question of who is going to give in first, and SAS is very persevering. Our aim is to gain market shares in the Norwegian domestic market – which we will do. [our translation] (Dagbladet, May 7. 1999).* This indicates that SAS was willing to suffer financial losses for a period to have less competitive constraints in the future.

Although we have not analysed the behaviour in detail, these observations are consistent with predatory behaviour. However, there are other possible reasons for the failure of Color Air. First, Color Air did not offer an attractive network. It had rather few flights on each route, which implied that it was not very attractive for business travelers. Second, SAS and Braathens both had frequent flyer programs while Color Air did not have such a program. Third, the competition for large customer contracts between SAS and Braathens made it even more difficult for Color Air to attract business travelers.

In September 1999, four months after SAS' statement to the press (see above) – Color Air had an accumulated loss of 400 MNOK and exited. From then on it seemed as

Braathens and SAS used media to undertake a coordinated reduction in capacity. Just after the exit of Color Air we could read: *'SAS has on several occasions announced that it will reduce their capacity if Braathens start reducing theirs. [our translation] (Bergens Tidende, 12.11. 1999).* Through signaling in the press it was made clear that each firm was willing to reduce its capacity, given that the rival did the same. From then on we saw a sequential game taking place, where both firms took turns in reducing their capacity.

Already in the spring of 2000, seven out of the ten new airplanes that entered in October 1998 had been removed from the market. Only on the route between Oslo and Bergen alone 600,000 seats had been removed, something that correspond to 46% of the passengers traveling this route in 1998. This indicates that the capacity utilization by May 2000 had returned to the level we had before the Gardermoen opening. This has also been verified in systematic econometric tests of several Norwegian routes (Steen and Sørgard, 2001). We still had competition on capacities, but not as aggressive as when Color Air was in the market.

8.4 Large customer contracts and failing firm

When Gardermoen opened, both SAS and Braathens had the possibility to increase their capacity on all routes, also the smaller ones, and both could thereby offer a complete domestic network. This led to more competition on large customer contracts since all large customers were now potential large customers for both carriers. From 1998 and onwards these contracts became more important, both in terms of number of contracts, and in terms of discount size.

A large customer contract is a contract between a large customer (example: Statoil and Telenor) and one of the carriers, where all employees from this firm can travel with this carrier at a contracted price. The contract will specify a percentage reduction in the full (C) price ticket. The typical contract will be a combination of discounts on different routes and a discount according to the customer's total travel volume in the domestic network.

This kind of contract can in theory lead to very intense rivalry on prices. This is an 'all-or-nothing' competition. If the carrier loses one large customer to the other carrier it loses a lot of passenger volume, and since the price cost margin is positive, a lot of profits. Hence, each carrier faces a very high own price elasticity of demand in this large customer market. This gives strong incentives to undercut your rival's price, and might lead to price very close to marginal costs (Bertrand-like competition).

In fact, we did observe intense rivalry on large customer contracts. In 2000, the carriers had around 300 contracts, more than a doubling from 1998, and the discounts had become substantial: *'According to information obtained by Dagens Næringsliv, some of the large customer discounts are in the order of 50% on certain routes. Normally the dis-*

counts are in the range of 5-50% relative to C-price. [our translation] (Dagens Næringsliv 12.09.2001). The SAS' large customer contracts' responsible, Stein Bemer, stated: ' *We hope to reduce the discount level.... it is evident that a possible merger [with Braathens] would make it possible to achieve this goal, ... When the competitive picture changes some of our large customers will not have the same bargaining power to obtain as large discounts as they used to have.'. [our translation] (Dagens Næringsliv 12.09.2001).*

The discounts also differed considerably according to competitive situation on the different routes. The largest discounts could be observed on routes where there was a large asymmetry between the carriers' market shares. The smallest carrier was typically willing to give very large discounts to gain market shares on these routes. Accordingly, on monopoly routes the discounts were smaller (Steen and Sørgard, 2001)

At the same time as we did observe intense rivalry on large customers contracts, the prices for other customers did increase substantially. This is illustrated in Figure 5.

Figure 8.5
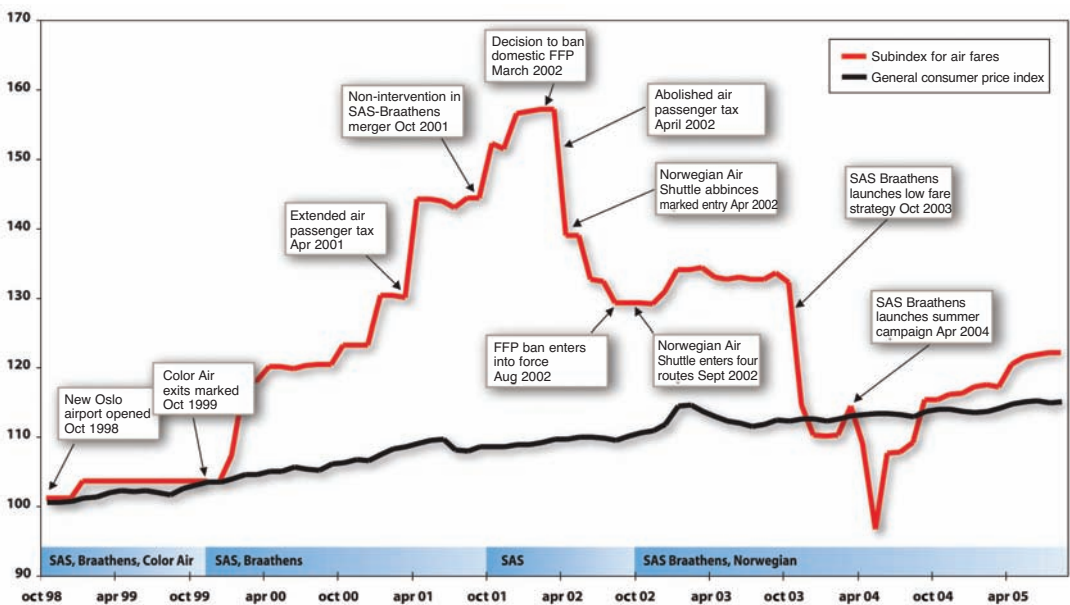
We see from Figure 8.5 that the prices on ordinary tickets, i.e. tickets sold to passengers without any large customer contract, did increase quite substantially from 1998 and onwards. In the period from 1998 to May 2001, when Braathens and SAS announced their merger (see below), the prices increased by an average of more than 27%.

To explain this price increase, let us introduce a simple model. We assume that the carriers via the interline prices were able to collude on prices and therefore act as a de facto

price cartel.[12] To simplify further, let us assume that the large customers only buy C-price tickets. Demand is given by X = A – P, where P is the price before any discounts and A measures the customers' maximum willingness to pay. Furthermore, let s be the share of the consumers that has a large customer contract, and let r denote the average discount (in absolute terms) in the large customer contract. The two carriers will have the following profit function:[13]

$$\Pi = (P - c)(A - P)[1 - s] + (P - c - r)(A - P + r)s$$

The optimal gross (non-discounted) C-price will then be:

$$P^* = \frac{A + c + 2rs}{2}$$

We see that the higher the number of large customer contracts (s) the larger will the C-price become, and the higher the discounts are (r), the higher is the optimal price.[14] The reason for this is quite straightforward. Through an increase in the ordinary C-price the companies can regain some of the discount given in the large customer contracts. The larger the discounts are and the more large customer contracts the carriers sign, the stronger the incentive to increase prices. The problem with this strategy is that customers *without* large customer contracts will face too high prices. Actually, the price will be even higher than the ordinary monopoly price for C-class tickets. Steen and Sørgard (2001) found that the large customer contracts did partly explain this extraordinary price increase in the airline industry.[15]

The study indicates that the discounts in these contracts are the outcome of competition rather than deliberate price discrimination. As is well known from theory of third degree price discrimination, a firm would find it profitable to set a high price in a segment with price inelastic demand, and a low price in a segment with price elastic demand. However, we observed the opposite in the Norwegian airline industry. Large discounts

---

12 This assumption is supported in several of the earlier studies, e.g., Salvanes, Steen and Sørgard (2003; 2005) and Steen and Sørgard (2005). For instance in Salvanes Steen and Sørgard (2000b) their empirical results are consistent with the regime that assumes collusion on prices.

13 An alternative would be to assume that the discount was a percentage discount rather than an absolute discount. It is straightforward to show that a percentage discount has the same qualitative effects as an absolute discount. It can also be shown that if the large customers only act according to the full non-discounted price rather then the net-discounted price the large customer contracts will have less effect on the level of the C-price.

14 In this setup r is exogenously given. This is of course a simplification, since r is decided upon through negotiation between the large customer and the airline company. However, these negotiations are undertaken ex ante. The decision about how much to travel is undertaken ex post. Hence, it is reasonable to assume r to be exogenously determined here.

15 Using monthly data for ten Norwegian routes over the period January 1996 to May 2001, it was found that when we include variables such as the number of large customer contracts or the total revenue that accrues to these contracts in a dynamic price model we find that these variables have a positive and significant effect on the C-price level.

were given to firms who typically buy flexible tickets. Such a firm's demand is typically price inelastic.[16] Then the firms were forced to have a rather perverse price structure seem from their own perspective: a price above monopoly price in the ordinary market, and low prices in the most price inelastic segment of the market.

It was not only the customers that did not gain much from the competition triggered by deregulation. The carriers did not gain either. They operated costly excess capacity, they did not price discriminate correctly according to the demand elasticities, and they lost a lot of revenue from fierce competition for the large customers. After years of competition and several strategic mistakes Braathens was close to bankruptcy in 2001. It had tried to fight against SAS in Sweden and had lost, it struggled to sustain its market shares in Norway, and its new "*back-best*" concept had been a failure.[17] SAS had shown both economic strength and willingness to "bleed" in the battle against Color Air. Braathens might have felt that it would be the next victim in this "war". At the same time a merger would solve all the described problems for the carriers: they could reduce capacity and divide the market, and they could eliminate the competition for the large customers.

In May 21st 2001 it was publicly announced that SAS had made an offer to acquire Braathens. The Norwegian Competition Authority used several months to analyse the situation, and at October 23rd they announced that they would allow the merger. The Norwegian Competition Authority was sceptical to the merger since a monopolisation of the Norwegian market clearly was bad for competition. However, they found that Braathens became a failing firm during the fall of 2001. They permitted the acquisition, since a bankruptcy was considered as a worse alternative. The decision was after this evaluated by the Norwegian ministry, but also they found that an acquisition would be the best alternative. We now returned to a monopoly in the Norwegian airline industry once again, but this time an unregulated monopoly.

---

16 An alternative interpretation is of course that the largest customers receive a quantity discount and the pricing is described by second degree price discrimination. However, this cannot still explain the large discounts given to the large customers. In particular, such a quantity discount should lead to a clear pattern with lower prices the larger the travel volume a firm has. Actually, when we look at the figures this is not the case, several large customers had very favourable contracts with large discounts, but the travel volume of these were exceeded by other firms with significantly less favourable large customer contracts.

17 Braathens introduced the "curtain" also on domestic flights. The full price passengers were given better service and were seated in the front of the plane. The M-class passengers were not served any food and had to sit behind the curtain. This concept was not well received by the Norwegian passengers, and in 2001 Braathens removed the curtain again. In 2002 SAS removed the curtain on inter-Scandinavian flights (Scandinavian Direct concept). The losses in Sweden were primarily a result of the purchase of Malmö Aviation and the losses from the route between Oslo and Stockholm.

8.5 From monopoly to duopoly and price competition

The monopoly position of SAS did not last for long. In September 2002 the airline company Norwegian entered on four domestic routes. From then on we experienced price competition in the Norwegian airline industry, in contrast to what we had in the first eight years after deregulation.

To understand the development from 2002 and onwards, it is important to consider the decision the Norwegian Competition Authority made in April 2002.[18] They banned SAS' frequent flyer program on all domestic routes, which implied that SAS no longer could offer frequent flyer points on domestic routes. A few month later The Ministry turned down a complaint from SAS and confirmed the decision made by the Norwegian Competition Authority. From July 2002 SAS could no longer offer frequent flyer points on domestic flights. According to Norwegian themselves, the ban of frequent flyer points on domestic routes was decisive for their entry into the Norwegian market.

In the literature, it is pointed out that frequent flyer programs are loyalty programs.[19] The consumers become loyal to one firm, in order to accumulate frequent flyer points from this particular firm. One might say that ex ante homogenous goods (an airline flight from A to B) become differentiated ex post. This leads to consumer lock-in. On the other hand, firms compete more aggressively to attract new consumers that can become loyal. Although the net effect is ambiguous in theory, in his survey Klemperer (1995) concludes that loyalty programs typically are detrimental to welfare:

'While there are exceptions to these conclusions, they suggest a presumption that public policy should discourage activities that increase consumer switching costs (such as airlines' frequent flyer programs), and encourage activities that reduce them' (p. 536).

According to his conclusion, frequent flyer programs are expected to have anticompetitive effects. In a situation with a monopoly firm and a potential entrant, it is obvious that this is true. The existing customers are locked in with the established firm. Norwegian's entry in the Norwegian market illustrates this point. In text books for managers, frequent flyer programs are presented as measures that may dampening competition also between existing firms.[20] Both firms have loyal customers, and the net effect may be higher prices.

There were also other reasons that explain the apparent successful entry. They were a lot more cautious about the ticket sale channel than Color Air was when they entered. Norwegian had a well functioning web portal and calling centre from the very start, and they were fully integrated with the travel agents in Norway. They focused on both business- and leisure class customers in their design of prices and frequencies. They had a

---

18 See decision V2002-28.
19 See, for example, Klemperer (1984, 1995) and Carns and Galbraith (1990).
20 See Brandenburger and Nalebuff (1996), chapter 6.

"golden opportunity" with low leasing prices on aircrafts due to September 11th, 2001 and excess supply of well qualified personell from the former Braathens, giving them a head start in the market. Finally, they had already run a successful airline operation for years with small airplanes, providing them with know how and insights in the market.

After entry by Norwegian we have experienced lower prices in the Norwegian airline market. This is illustrated in Figure 8.5. However, as shown in Figure 8.5 the toughness of price competition has changed a lot during the period 2002-2005.

At the outset Norwegian did not act very aggressively. It entered on only four out of 32 domestic routes, and on each of those four routes it had approximately one third of the capacity of SAS. This can be interpreted as what has been labelled a 'puppy dog' strategy. It enters on a limited scale, which implies that the incumbent does not suffer large losses from accommodating entry. If, on the other hand, the incumbent had been responding by cutting prices substantially it would face large losses. Since a puppy dog strategy implies that the incumbent has less incentives to cut prices, such a strategy might give the entrant a friendly welcome in the market. As we see from Figure 8.5, we do not observe dramatic price cuts following entry by Norwegian. It shows that SAS did not respond very aggressively to entry. Actually, SAS did cut back on capacity on some routes following entry.

However, Norwegian did not continue to be a puppy dog. Although it did not increase the frequency on each route the first year after entry, it entered seven new routes during the summer of 2003. In the autumn of 2003 it announced that it planned to have a market share of 18-20 % in 2004/05 in the domestic market in Norway, and that it would be introduced on the Oslo stock exchange. The actual growth and the prospects for future growth of Norwegian forced SAS to respond more aggressively. In October 2003 it launched a low price campaign, apparently to stop Norwegian from capturing larger market shares. The cut in prices from October 2003 was quite substantial, as can be seen in Figure 8.5. In May 2004 SAS continued by launching a Summer campaign with low prices, and as we see from Figure 8.5 that this led to a further and substantial price cut.

Norwegian faced a difficult time following SAS' response in October 2003. Although it did increase the number of passengers in the domestic market after October 2003, it did not succeed in making profits. During 2004 it exited several domestic routes, among others Oslo-Molde and Oslo-Ålesund. Its financial situation worsened during 2004, and some people argued that it might be forced to exit the market. At the same time Coast Air faced financial difficulties. It entered the route Oslo-Haugesund in June 2003, a route where SAS had a monopoly position. It did not succeed in earning profits on this route, and exited the route in the mid of June 2004. The Norwegian Competition Authority investigated the case, and concluded that SAS had abused its dominant position through predatory behaviour. It imposed a fine of MNOK 20 in June 2005. SAS brought the case to court, and Oslo Tingrett decided against the Norwegian Competition Authority's decision in the summer of 2006. The case is now pending for the next court level.

As indicated by Figure 8.5, we observed substantial price cuts in October 2003 followed by further price cuts in May 2004. If the toughness of price competition had continued after May 2004, it is highly unlikely that Norwegian had been able to earn profits. In worst case, this would have resulted in a return to a monopoly situation in the Norwegian airline market. However, SAS partly reversed its price policy. From the late summer of 2004 it raised its prices in the domestic market, as is seen from Figure 8.5. One possible reason for such a shift in strategy was the new Competition Law that came into force in May 2004, and the active role played by the Norwegian Competition Authority. They expressed in public their concern for the competitive situation in the Norwegian airline market. In addition, they made a dawn raid at SAS in June 2004 to investigate whether SAS had abused its dominant position. The dawn raid triggered the investigation of the route Oslo-Haugesund, which led to the predation case mentioned above.

The change in SAS' price policy made it possible for Norwegian to expand further and go from red figures to profits. Although it in late 2004 exited from Oslo-Ålesund, it gradually increased its capacity on domestic routes. For example, it expanded on the large domestic routes in late 2004 from 6 to 9 flights each day and expanded to 12 flights per day in early 2006. It gradually increased its prices. In the second quarter of 2005 it had for the first time a positive profits, and from September 2004 to January 2006 Norwegian's stock market value increased from less than 300 MNOK to more than 1700 MNOK.

Norwegian's initial strategy in 2002 was to serve the domestic market. However, between morning and afternoon rush hours some of the aircrafts were idle. In 2003 it started a few direct routes from Oslo to Europe, to increase the capacity utilization of its aircrafts. This happened at the same time as we saw other carriers like Sterling and Widerøe initiated such routes out of Norway. It turned out that there was a large potential for direct flights to Europe, from Oslo as well as several other domestic airports. In the autumn of 2003 it introduced several new direct routes to the continent. In the summer of 2004 it had approximately 20 direct flights to Europe, and in 2006 the number of direct flights to Europe has increased to almost 50. In Figure 8.6 we have shown Norwegian's direct flights in 2005/06.

Figure 8.6: Norwegian's network in 2005/06

Norwegian's expansion on international routes forced SAS to respond. The traditional system for a network operator is to have a rather limited number of direct flights, and instead send passengers through its hub. However, to respond to Norwegian's direct flights SAS had to offer some alternatives to flights through the hub Kastrup. It introduced several direct flights from different Norwegian cities. For example, in the first half year of 2006 it opened five new direct routes to Europe from other cities in Norway than Oslo. The expansion on direct flights to Europe did trigger a large increase the number of air passengers to and from Norway. For example, in the first nine months in 2005 the number of passengers traveling to and from Norway on traditional airlines increased with almost 11 % while the number of passengers with charter flights increased with approximately 5 %.

Price cuts and the expansion of the number of routes benefited the air passengers. To illustrate the possible effects, let us assume that the prices have decreased with 10-15 % on domestic routes as well as routes between Norway and the rest of Europe. If so, the air passengers have saved approximately 1,5-2,0 BNOK annually.

8.6 Some concluding remarks

The Norwegian airline industry is a story of market failure followed by success. Following deregulation, the carriers competed along the wrong dimensions seen from the customers' point of view. The lack of price competition in the business segment led to competition on location and capacity. The customers did not benefit from competition on location of flight schedules, since it led to clustering. Those who benefited were the passengers in the leisure segment since the supply of low-price tickets increased. The most important segment, however, the business travellers, could only use these tickets to a very small degree since discounted tickets are restricted in use and do not comply with a business traveller's needs. The business travellers within the large companies started to benefit from the large customer discounts towards the end of the period we are looking at, enjoying large discounts after 1998. Those without any large customer contracts, however, received no discounts, and even had to pay an extra premium due to the unfortunate relationship between the large customer contracts and the prices on the flexible tickets.

The way they competed led to destructive competition. Color Air had exited the market in the fall of 1999, 14 months after entry. Braathens became a failing firm in the fall of 2001, and was acquired by SAS. Then the regulated monopoly prior to 1994 was replaced by an unregulated monopoly in 2002.

One year after, though, we experienced entry and competition on prices rather than capacities. Instead of a large number of empty seats, we observed lower prices. Deregulation finally became beneficial for the consumers, and the firms no longer competed in a way that led to higher costs.

There should be many lessons to be learned from this experience.

First, it illustrates that deregulation as such can lead to market failure if the market structure does not promote a pro-competitive outcome. SAS and Braathens split the domestic market in two equal parts, and had even been allowed to have some flights on each other's routes before deregulation. This led to a smooth transition to a deregulated regime, and none of them had incentives to compete on prices. Their main concern became to defend their customer base. In contrast, when Norwegian entered in 2002 it started out with no customer base. The new company had to compete for market shares and the situation was no longer stable. At a certain point in time SAS decided to respond quite aggressively in order to prevent Norwegian from capturing an even larger market share.

Second, no competition on prices may lead to tougher competition along other dimensions. However, non-price competition may not be beneficial neither for customers nor for the airlines. In Norway we observed intense rivalry on capacities. This led to a record-low load factor, below 50 % on some routes at a certain point time. The welfare loss from such an outcome would then not only be the dead weight loss associated with high prices, but also the costs associated with excess capacity.

It is of interest to compare Norway with the US. While deregulation led to intense rivalry on capacities, lack of price competition and a lower load factor in Norway, the opposite was to a large extent true in the US. In our opinion, this contrast highlights the role of the initial regulatory regime and institutional setting. While in Norway the airlines could neither compete on prices nor on capacities, the airlines in the US could compete on capacities. They did so, and the result was a low load factor in the US during the period of regulation. There were thus potentials for cost reductions due to better capacity utilization in the US airline industry, while in Norway a potential for efficiency gains from capacity reduction was clearly more limited. Moreover, as explained above, in Norway the institutional setting supported the two active firm's effort to continue its market sharing and avoid any rivalry on prices. Unfortunately, this triggered intense rivalry on capacities. The contrast between Norway and the US thus illustrates the importance of price competition following deregulation. It might dampen non-price competition and thereby reduce costs, which can add to the traditional welfare gain from a reduction in the dead weight loss associated with lower prices.[21]

Third, the emergence of competition in the Norwegian airline market (and in Scandinavia more generally) illustrates how the prospects for competition might influence the behaviour of firms that earlier have been in a protected market situation with a *de facto* legal monopoly position. SAS was in the late 90s a high cost airline company compared to other network operators. From then and onwards it was forced to respond to the prospects for low cost airlines coming into domestic markets in Scandinavia. In 2002 it launched "SAS Turnaround", a program for implementing cost reductions for the SAS group. In 2006 it had succeeded with its cost reduction plan, and annual costs were reduced with 14 billions. It implied an almost 30 % reductions in costs per unit. Such a dramatic cost reduction illustrates that the prospect for a more competitive market might force firms to save costs.

Fourth, the last years experience in Norway may illustrate the important role of competition policy in deregulated industries. The ban of the frequent flyer program on domestic routes in 2002 made it possible for Norwegian to enter the domestic market. Such an entry, as well as the absence of loyalty inducing programs, triggered price competition. With lower prices the incentives to invest in capacity was rather limited, and we observed a much higher load factor than in the mid 90s when SAS and Braathens set high prices and competed on capacity. However, price competition is beneficial only if low prices are sustainable. A dominant firm might set low prices to force rivals to exit. In the Norwegian airline market we have seen that Color Air exited in September 2001, Braathens became a failing firm in October 2001, and Coast Air exited from Oslo-Haugesund in June 2004. Norwegian faced financial difficulties in 2004, after SAS launched low price campaigns. But a new Competition Act came into force in May 2004,

---

[21]This said, a welfare gain might have taken place as a consequence in Norway since increased capacity also can be viewed as increased "service quality" such as reductions in passengers' schedule delay etc. In fact, two-thirds of the welfare gained increase in the US was from increased flight frequency and not fare reductions (Morrison and Winston, 1986).

where abuse of a dominant position was prohibited. The Norwegian competition authorities expressed its concern publicly, and made a dawn raid at SAS in June 2004. SAS partly reversed its very aggressive price policy, and this allowed Norwegian to expand further and through the increased scale of the network costs have gone down accordingly. Finally in 2005 Norwegian started to earn positive profits.

Appendix: A semicollusion model

Let us present the model introduced in Salvanes, Steen and Sørgard (2003). We consider a duopoly where firms choose both prices and capacities. Since prices are typically more flexible than capacities, we assume that both firms set capacities at stage 1 and set prices at stage 2.

If the firms behave non-cooperatively on both stage 1 and 2, we have a game which is analysed in Kreps and Scheinkman (1983). They show that, when certain assumptions are met, the equilibrium is identical to the Cournot equilibrium. We label this the *competitive regime*. If the firms behave cooperatively on both stages, the firms behave as a cartel and thereby they attain the monopoly equilibrium concerning both price and capacity setting. We label this the *collusive regime*. A third alternative is that the firms behave cooperatively for one choice variable, and non-cooperatively for the second choice variable. As we argued, price is typically easier to change than capacity. As is well known from theory of repeated games, it is easier to collude on a choice variable that can be changed very rapidly. Hence, we find it natural to assume that the firms can collude on prices and compete on capacities.[22] We label this the *semicollusive regime*.[23]

Let us assume the following inverse demand function:

$$P = A - Q_1 - Q_2 \qquad (1)$$

where P is price, Q_i quantity supplied by firm i, $i=1,2$, and A a parameter measuring the demand potential. Furthermore, let C_S denote short run marginal cost and C_L cost per unit of installing capacity. K_i denotes capacity for firm i, where $i=1,2$, and $K = K_1 + K_2$. Let us consider each of the three cases.

Collusive regime (price- and capacity cartel)
Obviously, the firms have no incentives to build idle capacity. Therefore, we have that $Q_i=K_i$ for firm i. The following capacity is installed:

$$K_1^M + K_2^M \equiv K^M = \frac{A - C_S - C_L}{2} \qquad (2)$$

Then we have the following effect of a change in, interpreted as a change in the demand:

$$\frac{\partial K^M}{\partial A} = \frac{1}{2} \qquad (3)$$

---

[22] The fourth alternative would be capacity collusion and price competition. Then the firms could achieve the collusive outcome concerning both prices and capacities simply by setting the monopoly capacity. Hence, the outcome of this fourth alternative would be identical to the outcome in what we labelled the collusive regime.

[23] The semicollusion game we analyse here was first introduced in Fershtman and Gandal (1994).

Competitive regime (price- and capacity competition)
As for collusion, there is no reason for the firms to install idle capacity. The following capacity is installed in equilibrium:

$$K_1^C + K_2^C \equiv K^C = \frac{2(A - C_S - C_L)}{3} \qquad (4)$$

Then we have the following effect of a change in the demand:

$$\frac{\partial K^C}{\partial A} = \frac{2}{3} \qquad (5)$$

Semicollusive regime (price collusion and capacity competition)
The firms succeed in coordinating their price setting. At stage 2, the collusive price is found by solving the following problem:

$$\sum_{i=1}^{2} \pi_i = \max_P (P - C_S)Q - K \cdot C_L \qquad (6)$$

If $K < (A - C_S)/2$, the marginal revenue exceeds the short run marginal cost when all capacity is used for production. Hence, the firms set the price so that the entire capacity is used for production. Then, the market price is $P = A - K$.

If K $(A - C_S)/2$, it is optimal to set $P = (A + C_S)/2$. If so, the firms install excess capacity. Then it remains to determine the sharing rule - each firm's quota in the market. In that case we assume that:

$$Q_i^S = \frac{K_i}{K} D(P) \qquad (7)$$

Each firm's market share is thus identical to its share of total capacity. There are, at least, two reasons for a positive relationship between its own share of total capacity and its own share of total sale. First, the larger the capacity the larger the probability that there is a vacant seat at the airline firm in question. Second, the larger the capacity the larger the number of flights and thereby the service frequency for the airline firm in question. More generally, when products and prices are identical it is reasonable to assume that the demand is distributed so that each firm's sale is related to its share of total supply in the market.

At stage 1, the firms set capacity non-cooperatively. Firm i has the following maximization problem:

$$\pi_i = \max_{K_i} (P - C_S)Q_i - C_L K_i \qquad (8)$$

s.t. (i) $\quad$ if $K \leq \dfrac{A - C_S}{2}$, then $Q_i = K_i$ and $P = A - K$

(ii) $\quad$ if $K > \dfrac{A - C_S}{2}$, then $Q_i = Q_i^S$ and $P = \dfrac{A + C_S}{2}$

Given that K $(A - C_S)/2$, we are back to the case where all capacity is used for production. Then each firm determines its sale by determining its capacity, and price is set to clear the market. Hence, the firms compete for capacity and we have an outcome analogous to the competition regime we specified previously. If $K > (A - C_S)/2$, then the firms installs more capacity than what is demanded in the market at the collusive price. From the first order conditions, we have the following total capacity in equilibrium:

$$K^S = K_1^S + K_2^S = \frac{2(A - C_S)^2}{16C_L} \tag{9}$$

Then we have that the firms install more capacity than what is used for production if:

$$\frac{A - C_S}{2} < \frac{2(A - C_S)^2}{16C_L} \tag{10}$$

Rearranging, we find that the firms install excess capacity if $A > 4C_L + C_S$. This implies that semicollusion may lead to excess capacity. The intuition is that the firms compete for market shares by installing large capacity. When both firms install large capacity, we expect excess capacity in the equilibrium outcome. If we observe (1) no price difference between monopoly and duopoly routes and (2) larger excess capacity in duopoly than in monopoly, this would be consistent with a semicollusive regime.

References

Aasheim, K. B. Askim, 2000, "Turbulens i norsk luftfart – en studie av konkurransemessige effekter av dereguleringen og åpning av ny hovedflyplass" (Turbulence in the Norwegian airline industry), Master thesis, Nowegian School of Economics and Business Administration.

Borenstein, S. and J. Netz, 1999, 'Why do All Flights Leave at 8am? Competition and Departure-Time Differentiation in Airline Markets, *International Journal of Industrial Organization,* 17, 611-640.

Brandenburger, A. M. and B. J. Nalebuff, 1996, *"Co-opetition",* Currency Doubleday, New York.

Carns, R. D. og J. W. Galbraith (1990): 'Artificial Compatibility, Barriers to Entry, and Frequent Flyer Programs, *Canadian Journal of Economics,* 23, 807-16.

Eaton, B. C. and R. C. Lipsey, 1975, 'The Principle of Minimum Differentiation Reconsidered: Some New Developments in the Theory of Spatial Competition', *Review of Economic Studies,* 42, 27-49.

Fershtman, C. and N. Gandal, 1994, 'Disadvantageous Semicollusion', *International Journal of Industrial Organization,* 12, 141-154.

Friedman, J. M. and J.-F. Thisse, 1993, 'Partial Collusion Fosters Minimum Product Differentiation', *Rand Journal of Economics,* 24, 631-645.

Gabszewicz, J. and J.-F. Thisse, 1986, 'Spatial Competition and the Location of Firms', in J. Gabszewicz, J.-F. Thisse, M. Fujiata and U. Schweizer (eds.), *Location Theory,* London, Harwood Academic Press.

Hotelling, H., 1929, 'Stability in Competition', *Economic Journal,* 34, 41-57.

Klemperer, P. (1984): Collusion via switching costs: How frequent flyer programs, trading stamps, and technology choices aid collusion', Research paper 786, Stanford Graduate School of Business.

Klemperer, P. (1995): 'Competition when consumers have switching costs: An overview with applications to industrial organization, macroeconomics, and international trade', *Review of Economic Studies,* 62, 515-539.

Kreps, D. and J. Scheinkman, 1983, 'Quantity Precommitment and Bertrand Competition yield Cournot outcome', *Bell Journal of Economics,* 14, 326-337.

Lian, J. I., 1996, 'Økt luftfartskonkurranse?'(An increase in airline competition?), TØI-rapport nr. 322/1996, Transportøkonomisk Institutt, Oslo.

Matsui, A., 1989, 'Consumer-Benefited Cartels under Strategic Capital Investment Competition', *International Journal of Industrial Organization,* 7, 451-470.

Morrison, S. A. and C. Winston (1990): 'The Dynamics of Airline Pricing and Competition', *American Economic Review,* 80: 389-393.

Norman, V. D. and S. P. Strandenes, 1994, 'Deregulation of Scandinavian Airlines: A Case Study of the Oslo-Stockholm Route', in P. Krugman and A. Smith (eds.), *Empirical Studies of Strategic Trade Policy,* The University of Chicago Press, Chicago.

Salvanes, K. G., F. Steen and L. Sørgard, 2003,'Compete, Collude, or Both? Deregulation in the Norwegian Airline Industry', *Journal of Transport Economics and Policy,* 37(3), 383-416.

Salvanes, K. G., F. Steen and L. Sørgard, 2005, 'Hotelling in the Air? Flight Departures in Norway', *Regional Science and Urban Economics,* 35, 193-213.

Scherer, F., 1980, *Industrial Market Structure and Economic Performance,* Houghton-Mifflin, Boston, MA.

Scherer, F. and D. Ross, 1990, *Industrial Market Structure and Economic Performance,* Houghton-Mifflin, Boston, MA.

Steen, F. and L. Sørgard, 1999, 'Semicollusion in the Norwegian Cement Market', *European Economic Review,* 43, 1775-1796.

Steen, F. and L. Sørgard, 2001, 'Konkurranse versus monopol i norsk luftfart' (Competition versus Monopoly in the Norwegian Airline Industry), confidential report submitted to the Norwegian Competition Authority, October 2001.

Strandenes, S. P., 1990, 'Regulering og konkurranse i skandinavisk luftfart' (Regulation and Competition in Scandinavian Airline Industry), chapter 4 in L. Sørgard (ed.), *Næringsøkonomi - 13 norske bransjestudier (Industrial Economics - 13 Norwegian Industry Studies),* Bedriftsøkonomens Forlag.

Tirole, J., 1988, *The Theory of Industrial Organization,* MIT Press, Massachusetts.

This book describes the impact of competition in seven Norwegian industries: electricity, pharmacy, telecommunication, airlines, cement, groceries and radio. From this experience, we can draw some lessons. It is found that price competition tends to reduce costs and that competition typically generates benefits for the consumers. Moreover, it is found that the existence of a competition law and market design are crucial for having a positive impact of competition.